Petter Mostad
Applied Mathematics and Statistics
Chalmers

Suggested solutions for
MVE550 Stochastic Processes and Bayesian Inference
Re-exam April 8 2020

1. (a) The uniform prior on p is the same as a Beta(1, 1) distribution, and given p, each y_i has a Bernoulli(p) distribution, or in other words a Binomial(1, p) distribution. The Beta distribution is conjugate to the Binomial. Using the formula for the conjugacy, we get the posterior

$$p \mid \text{data} \sim \text{Beta}(1+2, 1+3) = \text{Beta}(3, 4).$$

More directly you may think as follows: Using Bayes formula, we get that

$$\pi(p \mid \text{data}) \propto_p p^2(1-p)^3$$

Comparing with the Beta distribution, we get that $\pi(p \mid \text{data}) = \text{Beta}(p; 3, 4)$.

(b) The predictive distribution for this conjugacy is Beta-Binomial, and we get

$$y_6 \mid \text{data} \sim \text{Beta-binomial}(1, 3, 4).$$

Thus

$$\Pr[y_6 = 1 \mid \text{data}] = \frac{\Gamma(1+3)\Gamma(1-1+4)\Gamma(7)}{\Gamma(3)\Gamma(4)\Gamma(1+3+4)} = \frac{3}{7}.$$

Alternatively you may compute

$$\begin{aligned}\pi(y_6 = 1 \mid \text{data}) &= \int_0^1 \pi(y_6 = 1 \mid p)\pi(p \mid \text{data})\, dp = \int_0^1 p \cdot \frac{\Gamma(3)\Gamma(4)}{\Gamma(7)} p^2(1-p)^3\, dp \\ &= \frac{\Gamma(3)\Gamma(4)}{\Gamma(7)} \int_0^1 p^3(1-p)^3\, dp = \frac{\Gamma(3)\Gamma(4)}{\Gamma(7)} \cdot \frac{\Gamma(8)}{\Gamma(4)\Gamma(4)} = \frac{3}{7} = 0.4286.\end{aligned}$$

In the last line, we use the formula for the Beta(3, 3) density to compute the integral.

2. First, we rearrange the rows and columns so that the states are listed in the order B, C, A, D. We then get the transition matrix

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

The point with this rearrangement is that the matrix now has the standard form used in Dobrow section 3.8, with

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

and

$$R = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{4} \end{bmatrix}.$$

We get

$$F = (I - Q)^{-1} = \begin{bmatrix} 1 & -\frac{1}{3} \\ -\frac{1}{4} & 1 \end{bmatrix}^{-1} = \frac{12}{11}\begin{bmatrix} 1 & \frac{1}{3} \\ \frac{1}{4} & 1 \end{bmatrix}$$

and

$$FR = \frac{12}{11}\begin{bmatrix} 1 & \frac{1}{3} \\ \frac{1}{4} & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{4} \end{bmatrix} = \frac{12}{11}\begin{bmatrix} \frac{1}{2} & \frac{5}{12} \\ \frac{7}{12} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{6}{11} & \frac{5}{11} \\ \frac{7}{11} & \frac{4}{11} \end{bmatrix}.$$

Thus the probability that a chain starting in C is absorbed in A is $\frac{7}{11} = 0.6364$.

3. (a) One may compute that

$$G'(s) = \frac{9+s}{12} \cdot \frac{3+s}{4} \cdot e^{(s-1)/3}$$

and thus that $G'(1) = \frac{5}{6} = 0.8333$. As $G'(1)$ is the expectation of the offspring process, this expectation is less than 1, and thus the branching process is subcritical.

(b) One may further compute that

$$G''(s) = \frac{63 + 18s + s^2}{144} e^{(s-1)/3}$$

and thus that $G''(1) = \frac{41}{72}$. If Z is a variable with the offspring distribution we can then compute

$$\text{Var}(Z) = G''(1) + G'(1) - G'(1)^2 = \frac{41}{72} + \frac{5}{6} - \left(\frac{5}{6}\right)^2 = \frac{17}{24} = 0.7083.$$

(c) As the branching process is subcritical, the probability for extinction is 1.

An alternative solution method to (a) and (b) is to observe that the probability generting function $G(s)$ is the product of $\left(\frac{3}{4} + \frac{1}{4}s\right)^2$ and $e^{\frac{1}{3}(s-1)}$. The first factor is the probability generating function of a random variable $X \sim \text{Binomial}\left(2; \frac{1}{4}\right)$, while the second factor is the probability generating funcion of a random variable $Y \sim \text{Poisson}\left(\frac{1}{3}\right)$. Thus the offspring process is the sum of independent variables X and Y with these distributions. This means that the expectation of the offspring process is $2 \cdot \frac{1}{4} + \frac{1}{3} = \frac{5}{6}$ and the variance is $2 \cdot \frac{1}{4} \cdot \frac{3}{4} + \frac{1}{3} = \frac{17}{24}$.

4. (a) In the stationary distribution for a random walk on an undirected graph, the probability of being at a node is proportional to the degree of the node, i.e., the number of neighbours it has. So if we can show that different sequences has different number of neighbours, we have shown that the stationary distribution is *not* uniform.

   Consider the sequence $(0, 0, \ldots, 0)$ of only zeros. By inspection, it should be clear that it has $2 \cdot 48 = 96$ neighbours. On the other hand, consider the sequence $0, 1, 2, \ldots, 24, 24, \ldots, 1, 0$. A neighbour can be created only by changing either of the two middle values of 24 to 23. Thus one sees this sequence has only 2 neighbours.

   (b) The idea would be to use a Metropolis Hastings algorithm, where the target distribution is the uniform distribution on $\mathcal{A}$. R code could look like

```
S <- rep(0, 50)
N <- 1000000
result <- rep(0, N)
for (i in 2:N) {
  prop = G(S)
  if (F(S)/F(prop)>runif(1))
    S = prop
  result[i] <- max(S)
}
```

   For a current sequence S and a proposed sequence S', the acceptance function is computed as

$$a(S, S') = \frac{\pi(S)q(S \mid S')}{\pi(S')q(S' \mid S)} = \frac{q(S \mid S')}{q(S' \mid S)} = \frac{1/\deg(S')}{1/\deg(S)} = \frac{\deg(S)}{\deg(S')} = \frac{F(S)}{F(S')}$$

   where $q(S' \mid S)$ and $q(S \mid S')$ are the probabilities of selecting S' when the current state is S, and vice versa, respectively.

   NOTE: Example 5.4 (Darwin's finches) in our textbook Dobrow is quite similar to this question. However, there is an error in that example; the acceptance function should be $\deg(i)/\deg(j)$, and not $\deg(j)/\deg(i)$ as written. Students who have been confused by this error will not be deducted for this confusion.

   (c) Using the Metropolis Hastings code above, one can generate a Markov chain of sequences whose stationary distribution is the uniform distribution on $\mathcal{A}$. By the Law of Large Numbers for Markov Chains, the average of $L(S)$ applied to the sequences in the chain will have as limit the true average we seek. In the code above, the average of the `result` vector will approximate the average we seek.

5. (a) The five different states can be described as follows:

   (A) No customers

   (B) Anders preparing hamburger, no customer waiting

   (C) Anders preparing hotdog, no customer waiting

(D) Anders preparing hamburger, one customer waiting

(E) Anders preparing hotdog, one customer waiting

The generator matrix becomes:

$$Q = \begin{bmatrix} -\frac{1}{4} & \frac{1}{4}\cdot 0.6 & \frac{1}{4}\cdot 0.4 & 0 & 0 \\ \frac{1}{3} & -\frac{7}{12} & 0 & \frac{1}{4} & 0 \\ \frac{1}{2} & 0 & -\frac{3}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{3}\cdot 0.6 & \frac{1}{3}\cdot 0.4 & -\frac{1}{3} & 0 \\ 0 & \frac{1}{2}\cdot 0.6 & \frac{1}{2}\cdot 0.4 & 0 & -\frac{1}{2} \end{bmatrix}$$

(b) We need to compute the stationary distribution $v = (v_1, v_2, v_3, v_4, v_5)$. We know that it fulfills $vQ = 0$ and $v_1 + v_2 + v_3 + v_4 + v_5 = 1$. So we may use 4 of the 5 equations from the matrix equation $vQ = 0$ together with $v_1 + v_2 + v_3 + v_4 + v_5 = 1$ to find the solution. In R, one may write

```
Q = matrix(c(-1/4, 0.6/4, 0.4/4, 0, 1,
             1/3, -7/12, 0, 1/4, 1,
             1/2, 0, -3/4, 0, 1,
             0, 0.6/3, 0.4/3, -1/3, 1,
             0, 0.6/2, 0.4/2, 0, 1),
           5, 5, byrow=T)

v = c(0, 0, 0, 0, 1)%*%solve(Q)
print(1 - v[1])
```

which yields 0.516.

(c) When a customer appears at the food truck, it is in state A, B, ..., E with probability $v_1, v_2, \ldots, v_5$, respectively. The probability that the customer does not turn away is $v_1 + v_2 + v_3$. With probability v_1 there is no waiting time. With probability v_2 the expected waiting time is 3 minutes (as we assume the waiting time is Exponentially distributed, so that it does not matter that the hamburger may already have been started on). With probaility v_3 the expected waiting time is 2 minutes. Thus, the expected waiting time given that the customer does not turn away is

$$\frac{3\cdot v_2 + 2\cdot v_3}{v_1 + v_2 + v_3} = 1.038$$

minutes.

6. (a) The mentioned prior corresponds to one where each row is modelled with a Dirichlet(1, 1, 1, 1) distribution. Because of the Multivariate-Dirichlet conjugacy, the posterior for, e.g., the first row, given the observation vector $(3, 4, 0, 0)$, is Dirichlet$(1 + 3, 1 + 4, 1 + 0, 1 + 0) =$ Dirichlet$(4, 5, 1, 1)$. The expectation for this distribution is the vector

$(4/11, 5/11, 1/11, 1/11)$. The posterior expectation for P thus becomes

$$\mathrm{E}[P \mid \text{data}] = \begin{bmatrix} 4/11 & 5/11 & 1/11 & 1/11 \\ 5/14 & 3/14 & 5/14 & 1/14 \\ 1/11 & 5/11 & 2/11 & 3/11 \\ 1/10 & 1/10 & 3/10 & 5/10 \end{bmatrix}.$$

(b) The number of values that may be non-zero in each of the four rows of P are 2, 3, 3, and 2, respectively. Thus the priors used for these rows are Dirichlet(1, 1) Dirichlet(1, 1, 1), Dirichlet(1, 1, 1), and Dirichlet(1, 1). The posteriors become Dirichlet(4, 5), Dirichlet(5, 3, 5), Dirichlet(5, 2, 3), and Dirchlet(3, 5), respectively, so that the expectation of the posterior becomes

$$\mathrm{E}[P \mid \text{data}] = \begin{bmatrix} 4/9 & 5/9 & 0 & 0 \\ 5/13 & 3/13 & 5/13 & 0 \\ 0 & 5/10 & 2/10 & 3/10 \\ 0 & 0 & 3/8 & 5/8 \end{bmatrix}.$$

7. According to the Chapman-Kolmogorov Relationship and using the standard notation from Chapter 3 of Dobrow, we may write

$$P^8_{10} = P^3_{11} P^5_{10} + P^3_{10} P^5_{00}.$$

We know that $P^3_{10} > 0.3$ and that $P^5_{10} > 0.5$. However, in order to find a lower bound on P^8_{10}, we would need to have lower bounds on P^3_{11} and P^5_{00}, and there is no reason why these should not be small. In fact, one can easily find an example where these values are zero: If

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

the chain will simply jump between 1 and 0 at every step. We get $P^3_{11} = 0$ and $P^5_{00} = 0$, while at the same time $P^3_{10} = 1$ and $P^5_{10} = 1$. In this case, $P^8_{10} = 0$, so the answer is that $p = 0$, and the P above exemplifies a Markov chain with the required properties.