Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE550 Stochastic Processes and Bayesian Inference

Re-exam April 8, 2020, 8:30 - 12:30
Examiner: Petter Mostad, phone 031-772-3579
Allowed aids: All aids are allowed.
For example you may access teaching material on any format and you may use R for computation.
However, you are **not** allowed to communicate with any person other than the examiner and the exam guard.
Total number of points: 30. To pass, at least 12 points are needed.
There is an appendix containing relevant information about some probability distributions.

1. (4 points) Assume the variables $y_1, y_2, \ldots, y_k, \ldots$ each can have possible values 0 or 1. Assume the parameter p is uniformly distributed on the interval from 0 to 1. Assume that given p, the y_i are independent, with the probablity p of being 1. Assume you have made the observations $y_1 = 1, y_2 = 0, y_3 = 0, y_4 = 1, y_5 = 0$.

   (a) What is the posterior for p given the data?

   (b) Given the information above, what is the probability that $y_6 = 1$?

2. (3 points) A discrete-time Markov chain has states A, B, C, D, and transition matrix

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

   If the chain starts in state C, what is the probability that it is absorbed in state A? (Show the steps in the computation).

3. (4 points) Let $Z_0, Z_1, \ldots$ be a branching process where the offspring distribution has probability generating function

$$G(s) = \left(\frac{3+s}{4}\right)^2 e^{(s-1)/3}.$$

   (a) Is the process critical, subcritical, or supercritical? Prove your answer.

   (b) Find the variance of the offspring process.

   (c) Find the extinction probability of the branching process.

4. (6 points) We will consider sequences $S = (x_1, \ldots, x_{50})$ of 50 integers so that $x_1 = 0$, $x_{50} = 0$, and $|x_i - x_{i+1}| \leq 1$ for all $i = 1, \ldots, 49$. In other words, the sequences change with at most 1 at each step. An example is illustrated with

```
0  0  0 -1 -2 -2 -2 -3 -3 -2 -2 -1 -1  0  0  0  1  2  1  2  3  3  2  3  2
2  3  3  4  3  4  4  5  4  4  5  6  5  6  5  5  4  3  3  2  3  2  1  0  0
```

Let $\mathcal{A}$ be the set of all sequences of the type above. For each $S \in \mathcal{A}$ there is a largest value $L(S) = \max_{i=1,\ldots,50} x_i$ that the integers reach; for the sequence above, $L(S) = 6$. We would like to estimate the average of $L(S)$ over all sequences S in the set $\mathcal{A}$.

(a) Two such sequences are called *neighbours* if they are identical except at a single position in the sequence. Define a graph where the nodes are the possible sequences and an edge connects two sequences if they are neighbours. If you make a random walk on this graph, will the stationary distribution be uniform? Why or why not?

(b) Assume you have computer code which for any valid sequence S computes $F(S)$, the number of neighbour sequences, and another function $G(S)$ which randomly selects a neighbour sequence uniformly among the neighbours. Write down R code (or pseudo code) which generates a Markov chain of sequences whose stationary distribution is the uniform distribution. Also write down any computation you need to make in order to derive any formulas you use in your code.

(c) Describe how you can use the code above to estimate[1] the average of $L(S)$ over all sequences S in the set $\mathcal{A}$.

5. (6 points) Anders runs a food truck selling hotdogs and hamburgers. A hotdog meal takes on average 2 minutes to prepare while a hamburger meal on average takes 3 minutes to prepare. 60% of customers choose the hamburger meal. We assume that the preparation times are exponentially distributed. Customers arrive according to a Poisson process with an intensity of 1 person every four minutes. If Anders is already working on an order when a customer arrives, the new customer waits. However, if there is already another customer waiting, the new customer goes away.

(a) Describe the five states that Ander's food truck can be in. Write down the generator matrix for the corresponding Markov chain.

(b) Compute the expected proportion of time Anders will be spending tending to customers. You may do this using R. An alternative is to just describe in detail how one can do such a computation.

(c) Among the customers that do not turn away, what is the average waiting time?

[1] This would not be the most computationally efficient way to get an estimate

6. (4 points) Consider a discrete time Markov chain with the possible states 1, 2, 3, 4. Assume it has been observed for 30 transitions, and the resulting counts of transitions between states is given in the following table (with the rows listing the state before the transition and the columns listing the state after the transition):

	1	2	3	4
1	3	4	0	0
2	4	2	4	0
3	0	4	1	2
4	0	0	2	4

   (a) Let P be the unknown transition matrix for the chain. Use as a prior for P the uniform distribution on the set of matrices P where all entries are non-negative and all rows sum to 1. Compute the expectation of the posterior of P given the data above.

   (b) Assume now we get the new information that the Markov chain can only change its value with +1, 0, or -1 in each step. Change the prior distribution to reflect this, so that it is a product of Dirichlet distributions with pseudo counts 1, but restricted to the entries in the rows of P that could be non-zero. Compute the expectation of the posterior of P using this prior and using the data above.

7. (3 points) A discrete-time Markov chain has states 0 and 1 and starts at 1. The probability that it will be 0 after exactly 3 steps is *at least* 0.3, and the probability that it will be 0 after exactly 5 steps will be *at least* 0.5. Find the probability p so that no Markov chain fulfilling the above has a probability below p of being 0 after exactly 8 steps, and give the transition matrix of a Markov chain fulfilling the above and having probability p of being 0 after exactly 8 steps.

Appendix: Some probability distributions

The Bernoulli distribution

If $x \in \{0, 1\}$ has a Bernoulli distribution with parameter $0 \leq p \leq 1$, then the probability mass function is

$$\pi(x) = p^x(1 - p)^{1-x}.$$

We write $x \mid p \sim \text{Bernoulli}(p)$ and $\pi(x \mid p) = \text{Bernoulli}(x; p)$.

The Beta distribution

If $x \in [0, 1]$ has a Beta distribution with parameters with $\alpha > 0$ and $\beta > 0$ then the density is

$$\pi(x \mid \alpha,\beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1}.$$

We write $x \mid \alpha,\beta \sim \text{Beta}(\alpha,\beta)$ and $\pi(x \mid \alpha,\beta) = \text{Beta}(x; \alpha,\beta)$.

The Beta-Binomial distribution

If $x \in \{0, 1, 2, \ldots, n\}$ has a Beta-Binomial distribution, with n a positive integer and parameters $\alpha > 0$ and $\beta > 0$, then the probability mass function is

$$\pi(x \mid n, \alpha,\beta) = \binom{n}{x} \frac{\Gamma(x + \alpha)\Gamma(n - x + \beta)\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n + \alpha + \beta)}.$$

We write $x \mid n, \alpha,\beta \sim \text{Beta-Binomial}(n, \alpha,\beta)$ and $\pi(x \mid n, \alpha,\beta) = \text{Beta-Binomial}(x; n, \alpha,\beta)$.

The Binomial distribution

If $x \in \{0, 1, 2, \ldots, n\}$ has a Binomial distribution, with n a positive integer and $0 \leq p \leq 1$, then the probability mass function is

$$\pi(x \mid n, p) = \binom{n}{x} p^x(1 - p)^{n-x}.$$

We write $x \mid n, p \sim \text{Binomial}(n, p)$ and $\pi(x \mid n, p) = \text{Binomial}(x; n, p)$.

The Dirichlet distribution

If $x = (x_1, x_2, \ldots, x_n)$ has a Dirichlet distribution, with $x_i \geq 0$ and $\sum_{i=1}^{n} x_i = 1$ and with parameters $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $\alpha_1 > 0, \ldots, \alpha_n > 0$, then the density function is

$$\pi(x \mid \alpha) = \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_n)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_n)} p_1^{\alpha_1-1} p_2^{\alpha_2-1} \cdots p_n^{\alpha_n-1}.$$

We write $x \mid \alpha \sim \text{Dirichlet}(\alpha)$ and $\pi(x \mid \alpha) = \text{Dirichlet}(x; \alpha)$.

The Exponential distribution

If $x \geq 0$ has an Exponential distribution with parameter $\lambda > 0$, then the density is

$$\pi(x \mid \lambda) = \lambda \exp(-\lambda x)$$

We write $x \mid \lambda \sim \text{Exponential}(\lambda)$ and $\pi(x \mid \lambda) = \text{Exponential}(x; \lambda)$. The expectation is $1/\lambda$ and the variance is $1/\lambda^2$.

The Gamma distribution

If $x > 0$ has a Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ then the density is

$$\pi(x \mid \alpha\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x).$$

We write $x \mid \alpha, \beta \sim \text{Gamma}(\alpha, \beta)$ and $\pi(x \mid \alpha, \beta) = \text{Gamma}(x; \alpha, \beta)$.

The Geometric distribution

If $x \in \{1, 2, 3, \ldots\}$ has a Geometric distribution with parameter $p \in (0, 1)$, the probability mass function is

$$\pi(x \mid p) = p(1-p)^{x-1}$$

We write $x \mid p \sim \text{Geometric}(p)$ and $\pi(x \mid p) = \text{Geometric}(x; p)$. The expectation is $1/p$ and the variance $(1-p)/p^2$.

The Normal distribution

If the real x has a Normal distribution with parameters μ and σ^2, its density is given by

$$\pi(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right).$$

We write $x \mid \mu, \sigma^2 \sim \text{Normal}(\mu, \sigma^2)$ and $\pi(x \mid \mu, \sigma^2) = \text{Normal}(x; \mu, \sigma^2)$.

The Poisson distribution

If $x \in \{0, 1, 2, \ldots\}$ has Poisson distribution with parameter $\lambda > 0$ then the probability mass function is

$$e^{-\lambda} \frac{\lambda^x}{x!}.$$

We write $x \mid \lambda \sim \text{Poisson}(\lambda)$ and $\pi(x \mid \lambda) = \text{Poisson}(x; \lambda)$.

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

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

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 $v_1, v_2, \dots, v_5$, 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

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

(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

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

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.