Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE550 Stochastic Processes and Bayesian Inference

Exam 19/8 2019, 14:00 - 18:00
Allowed aids: Chalmers-approved calculator.
Total number of points: 30. To pass, at least 12 points are needed.
There is an appendix containing information about some probability distributions.
Unless explicitly allowed, an answer is not complete without a supporting computation or argument.

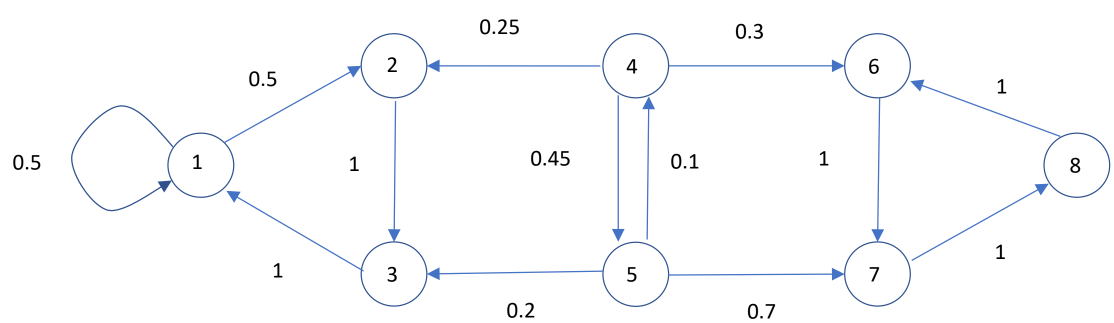

Figure 1: The transition graph for question 1.

1. (7 points) Consider the transition graph of Figure 1 for a discrete time Markov chain.

   (a) List the communication classes. Which of these classes are closed? (No explanation needed).

   (b) List the recurrent states. List the transient states. (No explanations are needed).

   (c) List all proper subsets of the eight nodes such that, if you consider the nodes in the subset and the transition probabilities between these nodes as indicated in the figure, you have an *ergodic* Markov chain. (Explain your conclusion.)

(d) Assume a chain starts at node 1. Compute the limit, as the number of steps goes to infinity, of the probability of being at node 2.

(e) Assume a chain starts at node 8. Compute $\lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} P^m_{8,7}$, where P is the transition matrix of the Markov chain.

2. (4 points) Assume Y has a negative Binomial distribution with parameters p and r (see the Appendix). Assume p has a uniform prior on the interval (0,1) and that r is fixed.

   (a) Find the name of and the parameter or parameters of the posterior density for p given an observation y.

   (b) Before any observations of y have been made, compute the expression for the marginal probability mass function for y, i.e., the probability mass function taking into account the uncertainty in p expressed in the prior.

3. (2 points) Assume a discrete-time Markov chain on the state space consisting of A, B, and C has been observed for 17 steps, with the following values:

$$B, A, C, C, B, B, C, A, A, C, B, B, A, B, C, B, A$$

   (a) Write down an estimate for the transition matrix P based on observed frequencies.

   (b) Assume we use a prior for the transition matrix consisting of a product of Dirichlet distributions, with all pseudo-counts equal to 1. What is the form of the posterior distribution given the data above? What is the expectation of this posterior?

4. (5 points) Assume a discrete probability distribution is specified with a probability vector $p = (p_1, p_2, \dots,)$. Let T be a transition matrix for this state space. The goal of the Metropolis-Hastings algorithm is to define a Markov chain $X_0, X_1, X_2, \dots$ with stationary distribution p.

   (a) Write down the Metropolis-Hastings algorithm where T is used for proposals.

   (b) Prove that the resulting chain $X_0, X_1, \dots$ is time-reversible.

   (c) Are there extra conditions needed to ensure that $X_0, X_1, \dots$ has p as a limiting distribution? If so, what is this condition or what are these conditions?

5. (5 points)

   (a) What is the definition of a Branching process?

   (b) How do you define a critical, supercritical, and a subcritical Branching process?

   (c) Assume the offspring distribution is Poisson with parameter λ. Find the probability generating function for the offspring distribution.

   (d) Assume $\lambda > 1$ and let s be the probability of extinction of the Branching process. Find and simplify an equation that s must satisfy, i.e., one which may be used to compute s for a given λ.

6. (3 points)

   (a) Explain briefly what Gibbs sampling is.

   (b) Explain briefly what Perfect sampling is.

   (c) For the matrix exponential, prove that $e^{(s+t)A} = e^{sA}e^{tA}$.

7. (4 points) A machine has three states: It works OK, it works in a stressed state, or it is broken. If it is OK, it wil stay OK for an exponentially distributed amount of time, with expectation 1000 hours. It will then go into the stressed state. If it is in a stressed state, it will break, according to a Poisson process with rate 0.1 per hour, or it will return to the OK state, according to an independent Poisson process with rate 0.5 per hour.

   (a) Write down the generator matrix Q.

   (b) If it is in the stressed state, what is the expected length of time it will stay in this state before it moves to another state[1]?

   (c) If it starts out OK, what is the expected time until it breaks?

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

---

[1] In the original exam, the last part of the sentence ("before it moves to another state") was missing, making the question somewhat less clear.

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 Negative Binomial distribution

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

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

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

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 $x \in \{0, 1, 2, \dots\}$ 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
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.