Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE550 Stochastic Processes and Bayesian Inference

Exam January 11, 2023, 8:30 - 12:30
Examiner: Petter Mostad, phone 031-772-3579, visits exam at 9:30 and 11:30
Allowed aids: Chalmers-approved calculator
Total number of points: 30. At least 12 points are needed to pass.
See appendix for some information about some probability distributions

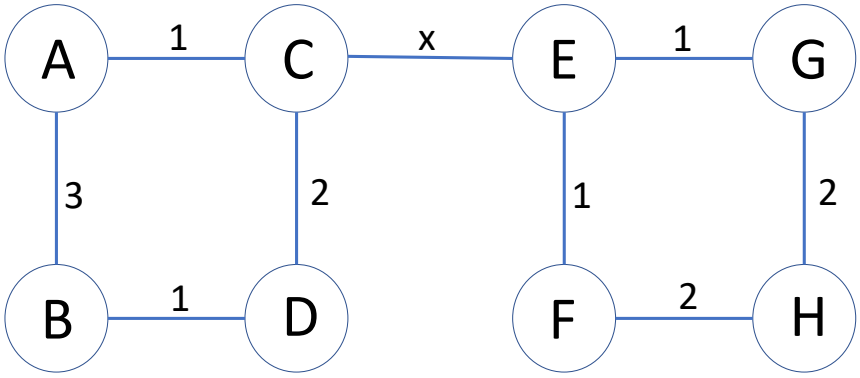

Figure 1: The weighted graph for question 1.

1. (8 points) Consider the weighted undirected graph of Figure 1. The weight x is a real-valued parameter that may be zero or positive. Define a Markov chain as a random walk on this weighted graph.

   (a) List the recurrent states and the transient states of the chain. Argue why the states are recurrent or transient.

   (b) For all possible fixed values of x, present, if possible, a stationary distribution.

   (c) For all possible fixed values of x, present, if possible, a limiting distribution.

   (d) Assume $x = 3.7$. Is the Markov chain time reversible? Argue for your answer.

   (e) For the Markov chain that starts in A, write an expression that computes the expected number of steps before it hits either G or H. Your expression should be written in terms of matrices and vectors of numbers and x, and operations on these such as multiplication, subtraction, addition, and inversion.

2. (6 points) Consider a branching process where the offspring distribution X has the probability generating function

$$G_X(s) = \frac{2}{3} + \frac{1}{3}s^2\frac{1-k}{1-sk}$$

where k is some real-valued parameter satisfying $0 < k < 1$.

   (a) Compute which values of k that make the process subcritical, critical, or supercritical.

   (b) What is the probability that there is exactly one offspring, i.e., what is $\mathrm{P}[X = 1]$?

   (c) Compute the probability of extinction as a function of k.

3. (4 points)

   (a) Describe the Metropolis Hastings algorithm in your own words: What is its purpose? What do you need to choose when using it? What are the steps in the algorithm?

   (b) What are the main ideas in perfect sampling?

4. (8 points) A continuous-time Markov chain has two possible states, A and B, and generator matrix

$$Q = \begin{bmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{bmatrix}$$

for positive λ_1 and λ_2. It starts in state A.

   (a) Let T be the time at which the the Markov chain returns to A the first time. Compute the expectation and variance of T.

   (b) For $t \geq 0$ define M_t as the number of times the chain has *returned* to A before time t. Prove or disprove that this is a Poisson process. Is it a counting process?

   (c) A realization of the process has been observed for a short while, as it jumps between A and B. The times spent in each state have been recorded as

A	B	A	B	A	B	A	B	A	B
0.2	4.0	0.1	3.5	0.3	2.0	0.2	2.5	0.3	3.0

   Assume we use the following prior density on Q:

$$\pi(Q) = \mathrm{Gamma}(\lambda_1; 4, 1) \cdot \mathrm{Gamma}(\lambda_2; 1, 3).$$

   Compute the posterior density for Q given the data above.

5. (4 points) Let $\{B_t\}_{t\geq 0}$ denote Brownian motion.

   (a) Compute the expectation and variance of $B_1 + 2B_4 + 3B_7$.

   (b) For each of the processes below, make an argument about whether or not $\{M_t\}_{t\geq 0}$ is Brownian motion:

i. $M_t = B_{2.3+t} - B_{2.3}$

ii. $M_t = B_{T+t} - B_T$ where T is the last time smaller than time 2.3 that the process is equal to zero.

iii. $M_t = B_{T+t} - B_T$ where T is the first time the process reaches the value 2.3.

Appendix: Some probability distributions

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 $x \in \{1, 2, 3, \dots\}$ 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 Negative Binomial distribution

A stochastic variable x taking on as possible values any nonnegative integer has a Negative Binomial distribution if its probability mass function is given by

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

where $r > 0$ and $p \in (0, 1)$ are parameters.

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 $\pi(x \mid \lambda) = \text{Poisson}(x; \lambda)$. The Poisson distribution has expectation λ and variance λ.

Petter Mostad
Applied Mathematics and Statistics
Chalmers

Suggested solutions for
MVE550 Stochastic Processes and Bayesian Inference
Exam January 11 2023

1. (a) All states are recurrent. A possible way to argue: If $x = 0$ there are two closed communication classes. If $x > 0$ there is one closed communication class. All states in finite closed communication classes are recurrent. (Note that the definition of recurrence is about *returning* to a state, not about reaching it.)

   (b) For any Markov chain defined as a random walk on a weighted graph, we get a stationary distribution by, for each node, computing the sum of the weights on the edges connected to it, and then normalizing so that the probabilities sum to 1. For the nodes $A, \ldots, H$ we get the sums

   $$4, 4, 3 + x, 3, 2 + x, 3, 3, 4$$

   respectively, with the total sum 26+2x. Thus a stationary distribution is

   $$\frac{1}{26 + 2x}(4, 4, 3 + x, 3, 2 + x, 3, 3, 4).$$

   Note that, for $x = 0$, there are also other stationary distributions.

   (c) For any x, the Markov chain is periodic with period 2. Thus it cannot have a limiting distribution, as the probabilities for what the current state is will always depend on the starting state.

   (d) Yes, the chain is time reversible, as all Markov chains defined as random walks on weighted graphs are time reversible.

   (e) When $x = 0$ the chain is reducible and the expected number of steps is infinite. For $x > 0$, let us first write down the transition matrix for the Markov chain without any absorbing states:

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

To answer the question, we set the two last states (G and H) as absorbing, and take the inverse of the identity minus the submatrix of P where the rows and columns corresponding to states G and H are removed. Then we take the sum of the first row of this matrix, as we start in state A. Thus the answer becomes

$$\begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix} 1 & -3/4 & -1/4 & 0 & 0 & 0 \\ -3/4 & 1 & 0 & -1/4 & 0 & 0 \\ -\frac{1}{3+x} & 0 & 1 & -\frac{2}{3+x} & -\frac{x}{3+x} & 0 \\ 0 & -1/3 & -2/3 & 1 & 0 & 0 \\ 0 & 0 & -\frac{x}{x+2} & 0 & 1 & -\frac{1}{x+2} \\ 0 & 0 & 0 & 0 & -1/3 & 1 \end{bmatrix}^{-1}\begin{bmatrix}1\\1\\1\\1\\1\\1\end{bmatrix}.$$

2. (a) As we know that $\mathrm{E}[X] = G'_X(1)$, we first compute the derivative of G_X:

$$G'_X(s) = \frac{1}{3}\left[2s\frac{1-k}{1-sk} + s^2\frac{k(1-k)}{(1-sk)^2}\right].$$

From this we get

$$\mathrm{E}[X] = G'_X(1) = \frac{1}{3}\left[2 + \frac{k}{1-k}\right].$$

Solving $G'_X(1) = 1$ yields $k = \frac{1}{2}$. From this we get that if $0 < k < \frac{1}{2}$ then the process is subcritical. If $k = \frac{1}{2}$ the process is critical. And if $k > \frac{1}{2}$ the process is supercritical.

(b) We get that

$$\mathrm{P}[X = 1] = G'_X(0) = 0$$

Thus the probability is zero. One can also derive this result by expanding the factor $\frac{1-k}{1-sk}$ as a power series; we see then that $G_X(s)$ is expressed as a power series where the first-degree term (containing s^1) is zero.

(c) We solve for s the equation

$$G_X(s) = s,$$

i.e.,

$$\frac{2}{3} + \frac{1}{3}s^2\frac{1-k}{1-sk} = s$$

which may be transformed to

$$s^2(2k+1) - s(2k+3) + 2 = 0.$$

As we know that, as always, $s = 1$ is a root, we may factor this as

$$(s-1)(s(2k+1) - 2) = 0$$

and we see that the other root is given by

$$s(2k+1) - 2 = 0$$

i.e., $s = \frac{2}{2k+1}$. So when $k \leq \frac{1}{2}$ the extinction probability is 1, while when $k > \frac{1}{2}$ the extinction probability is $\frac{2}{2k+1}$.

3. (a) Given a function $f(x)$ proportional to a density, the purpose of the Metropolis Hastings algorithm is to generate an approximate sample from this density. More precisely, one constructs a Markov chain which has the density as its limiting distribution, and generates a realization from this chain. To do so, one needs to select a starting value (or distribution) for the Markov chain, and a *proposal distribution* $q(x^* \mid x)$, a density for all x from which one knows how to generate a sample.

   An outline of the steps in the algorithm is:

   - Set x_0 to the starting state of the Markov chain, or simulate it from the starting distribution.
   - For $i = 1, \ldots, N$:
     - Simulate x^* using the density $q(x^* \mid x_{i-1})$.
     - Compute the acceptance probability

$$\rho = \min\left(1, \frac{f(x^*)q(x \mid x^*)}{f(x)q(x^* \mid x)}\right)$$

     - Set $x_i = x^*$ with probability ρ, otherwise set $x_i = x_{i-1}$.
   - Now x_N is an approximate sample from the density proportional to $f(x)$.

   (b) The purpose of perfect sampling is to find an algorithm so that the simulated value x_N is exactly from the target distribution, without involving an approximation or limiting process. One achieves that by generating x_N in such a way that it is independent of the starting value x_0 of the chain. In principle, one can obtain such an independence by considering S *coupled* chains, where S is the size of the sample space, and where each chain starts at a different state. An important idea is then to couple these chains so that if two chains are at equal states in one time step, they will be equal at the next time step too. Finally, in order to avoid keeping track of S chains, one can, if possible, use a partial ordering so that only a maximal and a minimal chain needs to be produced: If they have converged to the same state at some predefined time step, one knows the value at this time step is a sample from the target distribution. (Note that a shorter answer than this can give full points).

4. (a) Let X be a holding time when the process is in state A and Y a holding time when the process is in state B. Then

$$\begin{aligned} X &\sim \text{Exponential}(\lambda_1) \\ Y &\sim \text{Exponential}(\lambda_2). \end{aligned}$$

   Thus we get

$$\mathrm{E}[T] = \mathrm{E}[X] + \mathrm{E}[Y] = \frac{1}{\lambda_1} + \frac{1}{\lambda_2}$$

   and

$$\mathrm{Var}[T] = \mathrm{Var}[X] + \mathrm{Var}[Y] = \frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2}.$$

(b) M_t is NOT a Poisson process. If it were a Poisson process with parameter λ, we know that the time T between events would be exponentially distributed with parameter λ. Thus we would get

$$\mathrm{E}[T] = \frac{1}{\lambda}$$

and

$$\mathrm{Var}[T] = \frac{1}{\lambda^2}.$$

This would mean that

$$\begin{aligned} 1/\lambda &= 1/\lambda_1 + 1/\lambda_2 \\ 1/\lambda^2 &= 1/\lambda_1^2 + 1/\lambda_2^2. \end{aligned}$$

These two together would yield

$$\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right)^2 = \frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2}$$

which is impossible as $\frac{2}{\lambda_1\lambda_2} > 0$.
However, M_t is clearly a counting process.

(c) Let us write $a_1, \ldots, a_5$ and $b_1, \ldots, b_5$ for the observed holding times for state A and state B, respectively, and note that $a_1 + \cdots + a_5 = 1.1$ and $b_1 + \cdots + b_5 = 15$. We then get for the posterior

$$\begin{aligned} \pi(Q \mid \text{data}) &\propto_Q \pi(\text{data} \mid Q)\pi(Q) \\ &\propto_Q \left[\prod_{i=1}^{5} \text{Exponential}(a_i; \lambda_1) \cdot \text{Exponential}(b_i; \lambda_2)\right] \pi(Q) \\ &\propto_Q \left[\prod_{i=1}^{5} \lambda_1 \exp(-\lambda_1 a_i)\lambda_2 \exp(-\lambda_2 b_i)\right] \lambda_1^{4-1} \exp(-\lambda_1)\lambda_2^{1-1} \exp(-3\lambda_2) \\ &= \lambda_1^8 \exp -2.1\lambda_1\lambda_2^5 \exp(-18\lambda_2) \\ &= \text{Gamma}(\lambda_1; 9, 2.1) \cdot \text{Gamma}(\lambda_2; 6, 18) \end{aligned}$$

As the last function integrates to 1, it is the actual density for Q.

5. (a) We get

$$\mathrm{E}[B_1 + 2B_4 + 3B_7] = \mathrm{E}[B_1] + 2\,\mathrm{E}[B_4] + e\,\mathrm{E}[B_7] = 0$$

and

$$\begin{aligned} \mathrm{Var}[B_1 + 2B_4 + 3B_7] &= \mathrm{Var}[B_1 + 2B_4 + 3(B_7 - B_4) + 3B_4] \\ &= \mathrm{Var}[B_1 + 5(B_4 - B_1) + 5B_1 + 3(B_7 - B_4)] \\ &= \mathrm{Var}[6B_1 + 5(B_4 - B_1) + 3(B_7 - B_4)] \\ &= 36\,\mathrm{Var}[B_1] + 25\,\mathrm{Var}[B_4 - B_1] + 9\,\mathrm{Var}[B_7 - B_4] \\ &= 36 \cdot 1 + 25 \cdot 3 + 9 \cdot 3 = 138. \end{aligned}$$

(b) We have proven in the course that $B_{T+t} - B_T$ is Brownian motion if T is a *stopping time*. The random variable T is a stopping time if the value of T can be determined by knowing the Brownian motion on the interval $[0, T]$. This means that

i. This is Brownian motion, as 2.3 trivially is a stopping time.

ii. This is NOT Brownian motion, as the value of T here cannot be determined by looking at the Brownian motion in the interval $[0, T]$.

iii. This is Brownian motion, as the value of T can be determined by looking at the Brownian motion in the interval $[0, T]$.