Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE550 Stochastic Processes and Bayesian Inference

Trial exam autumn 2018
Allowed aids: Chalmers-approved calculator.
Total number of points: 30. To pass, at least 12 points are needed

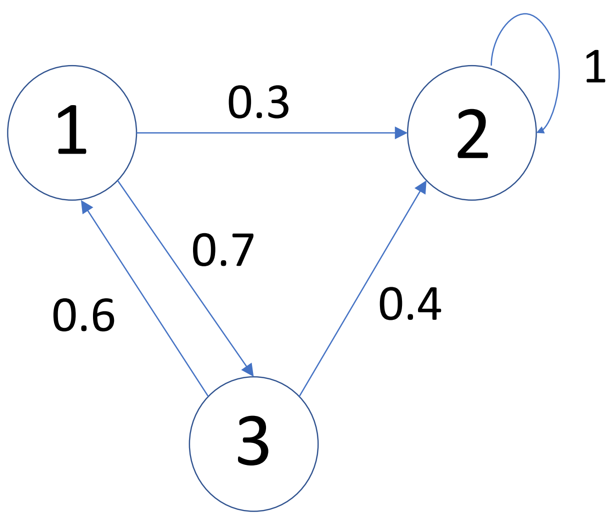

Figure 1: The chain for question 1.

1. (4 points) Consider the discrete-time Markov chain with state space $\{1, 2, 3\}$ and transition probabilities given in Figure 1.
   (a) What is the limiting distribution for this chain?
   (b) Write down its transition matrix, and compute its fundamental matrix F.
   (c) If a chain starts in state 1, what is the expected number of steps it will be in state 3?
2. (4 points) Consider an ergodic discrete-time Markov chain with a finite state space, transition matrix $P = [P_{ij}]$, and stationary distribution v.
   (a) If we simulate k independent Markov chains, with each chain having a different starting value at X_0, what is the probability that all chains are in the same state after n steps, as $n \to \infty$?

(b) Describe a way to set up the simulation so that each chain is still a realization from the Markov chain with the given transition matrix, and each chain starts with a different starting value, but the probability that all chains are in the same state after n steps approaches 1 as $n \to \infty$.

(c) Describe a type of sampling using the simulation method of (b), and explain why this is useful in this sampling type.

3. (3 points) Let X be a random variable with values in the set $\{0, 1, 2, \dots\}$ and let $G(s)$ be its probability generating function. The variance of X can be expressed in terms of derivatives of $G(s)$ evaluated at specific values for s. Derive the correct formula.

4. (4 points) Assume $\{N_t\}_{t \geq 0}$ is a Poisson process with parameter λ. Assume each arrival is independently marked with M (with probability p) or with F (with probability $1 - p$). Let $N_t^{(M)}$ and $N_t^{(F)}$ be the number of events of type M or F, respectively, in the interval $[0, t]$. Prove that $N_t^{(M)}$ and $N_t^{(F)}$ are independent random variables and derive their distributions.

5. (5 points) Define a discrete-time continuous-valued Markov chain $X_0, X_1, X_2, \dots$, by defining $X_0 \sim \text{Normal}(0, 1)$ and for $k = 1, 2, \dots$,

$$X_k \sim \text{Normal}(aX_{k-1}, 1)$$

where a is a real parameter.

(a) Assume we use the prior $a \sim \text{Normal}(1, 1)$ for a. Find the posterior distribution for a given observations x_0 and x_1 of X_0 and X_1, respectively.

(b) Prove that the posterior for a given observations of $X_0, X_1, \dots, X_k$ is normal. You do not need to derive the parameters of the distribution.

6. (5 points) Consider a continuous-time Markov chain on the state space $\{1, 2, 3, 4\}$. The transition rates between the states are indicated in the transition rate graph in Figure 2.

(a) Is this chain ergodic? Why/why not?

(b) Write down the (infinitesimal) generator matrix Q for this chain.

(c) Let $v = (v_1, v_2, v_3, v_4)$ be the probability vector representing the staionary distribution for the continuous-time Markov chain. Write down four linear equations that the numbers v_1, v_2, v_3, v_4 need to satisfy, and which determine these numbers uniquely. (You do not need to solve the equations).

(d) Write down the transition matrix for the embedded Markov chain.

(e) Let $w = (w_1, w_2, w_3, w_4)$ be the probability vector representing the stationary distribution for the discrete embedded Markov chain. Write down four linear equations that the numbers w_1, w_2, w_3, w_4 need to satisfy, and which determine these numbers uniquely. (You do not need to solve the equations).

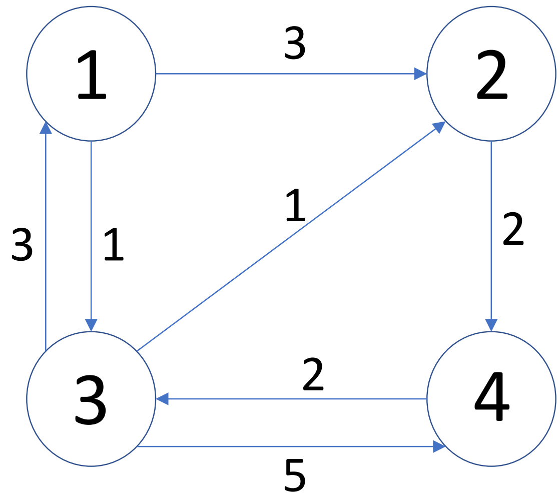

Figure 2: The chain for question 6.

7. (5 points) Assume $\{B_t\}_{t\geq 0}$ is the Brownian motion stochastic process.

   (a) Find the distribution of $B_1 + B_2 + B_3$.

   (b) Give the definition of a Gaussian process (as defined in Dobrow).

   (c) Let $a > 0$. For $n = 1, 2 \ldots,$, let M_n be the event consisting of those $t > 0$ such that there exists $t_1 < t_2 < \cdots < t_n < t$ with $B_{t_i} = a$ for $i = 1, \ldots, n$. Let T_n be the largest t smaller than or equal to all $t \in M_n$. Compute the distribution of $T_{100} - T_1$.

Appendix: Some probability distributions

The Beta distribution

If $x \geq 0$ has a Beta(α, β) distribution with $\alpha > 0$ and $\beta > 0$ then the density is

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

The Binomial distribution

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

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

The Exponential distribution

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

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

and the cumulative distribution function is

$$F(x) = 1 - \exp(-\lambda x).$$

The Gamma distribution

If $x > 0$ has a Gamma(α, β) distribution, with $\alpha > 0$ and $\beta > 0$, then the density is

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

The Normal distribution

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

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

The Poisson distribution

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

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

Petter Mostad
Applied Mathematics and Statistics
Chalmers

Suggested solutions for
MVE550 Stochastic Processes and Bayesian Inference
Trial exam autumn 2018

1. (a) As state 2 is absorbing while the other states are transient, the limiting distribution is $(0, 1, 0)$.

   (b) The transition matrix is $P = \begin{bmatrix} 0 & 0.3 & 0.7 \\ 0 & 1 & 0 \\ 0.6 & 0.4 & 0 \end{bmatrix}$. The part concerning the transient states is $Q = \begin{bmatrix} 0 & 0.7 \\ 0.6 & 0 \end{bmatrix}$ and thus the fundamental matrix is

$$F = (I - Q)^{-1} = \begin{bmatrix} 1 & -0.7 \\ -0.6 & 1 \end{bmatrix}^{-1} = \frac{1}{1 - 0.6 \cdot 0.7} \begin{bmatrix} 1 & 0.7 \\ 0.6 & 1 \end{bmatrix} = \begin{bmatrix} 1.7241 & 1.2069 \\ 1.0345 & 1.7241 \end{bmatrix}.$$

   (c) This can be read of the fundamental matrix: The answer is 1.2069.

2. (a) As $n \to \infty$ the distribution of each chain will approach the stationary distribution v. Let us write s for the number of states in the state space. We can compute the probability p that the k chains have the same state by summing over the possible states and computing the probability that each chain has this state. Letting $n \to \infty$, we get

$$p = \sum_{i=1}^{s} v_i^k.$$

   (b) One may use “coupling”: Pick an ordering of the states in the state space and name the states $1, \ldots, s$. Define a function f sending pairs (i, u), where u is a real number in $[0, 1]$ and i is a state, to a new state, as follows:

$$f(i, u) = \max \left\{ j : \sum_{r=1}^{j-1} P_{ir} \leq u \right\}.$$

   One may see that $\mathrm{P}(f(i, U) = j) = P_{ij}$ when U is a random variable with a uniform distribution on $[0, 1]$. Thus, if a chain is at state i, one may choose its state in the next step according to $f(i, U)$ where $U \sim \text{Uniform}(0, 1)$.

   The values of the k chains are now constructed by using, for each step m, the same random number U_m for all the chains. Each chain will then be a Markov chain with the given transition matrix, while the chains will not be independent, they will be “coupled”.

It should be clear, that, if two chains have the same state at a certain step, they will continue to have the same state at all following steps. Furthermore, although we do not prove this rigorously, it should be clear that unless all rows of the transition matrix are euqal, there exist some pairs of states with a non-zero probability that two chains in these two states at one step will have the same state at the following step. Using ergodicity, it may be shown that as $n \to \infty$, all chains will end up in the same state.

(Note that our coupling will not work if all the rows of the transition matrix are equal. However, in this case, all the rows are equal to v, all chains converge immediately to the limiting distribution, and all questions about convergence are rather uninteresting).

(Note: A fairly short explanation would give students full points for this subquestion).

(c) Perfect sampling uses the simulation method of (b). In Perfect sampling, one decides on a certain number of steps n, and then simulates s chains, all with different starting states, for n steps, using coupling as above. If all the chains are in the same state at the final step, one knows that this step is a sample from the stationary distribution. For this to happen with a reasonable probability, one needs the effect of (b), i.e., that the chains tend to end up in the same state.

3. We have

$$\begin{aligned} G(s) &= \mathrm{E}\left[s^X\right] \\ G'(s) &= \mathrm{E}\left[Xs^{X-1}\right] \\ G''(s) &= \mathrm{E}\left[X(X-1)s^{X-2}\right] \end{aligned}$$

so that $G'(1) = \mathrm{E}[X]$ and $G''(1) = \mathrm{E}[X(X-1)] = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]$. Thus

$$\begin{aligned} \mathrm{Var}[X] &= \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 \\ &= \mathrm{E}\left[X^2\right] - \mathrm{E}[X] + \mathrm{E}[X] - \mathrm{E}[X]^2 \\ &= G''(1) + G'(1) - (G'(1))^2 . \end{aligned}$$

4. The joint probability mass function for $N_t^{(M)}$ and $N_t^{(F)}$ can be derived as

$$\begin{aligned} & P(N_t^{(M)} = i, N_t^{(F)} = j) \\ = \;& P(N_t^{(M)} = i, N_t = i + j) \\ = \;& P(N_t^{(M)} = i \mid N_t = i + j) P(N_t = i + j) \\ = \;& \frac{(i+j)!}{i! \cdot j!} p^i (1-p)^j e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} \\ = \;& e^{-\lambda} \frac{1}{i!} (p\lambda)^i \frac{1}{j} ((1-p)\lambda)^j \\ = \;& \left[e^{-\lambda p} \frac{1}{i!} (p\lambda)^i \right] \left[e^{-\lambda(1-p)} \frac{1}{j!} ((1-p)\lambda)^j \right] \\ = \;& \mathrm{Poisson}(i; \lambda p) \cdot \mathrm{Poisson}(j; \lambda(1-p)) \end{aligned}$$

From this it follows that the marginal distribution for $M_t^{(M)}$ is Poisson(λp) and that the marginal distribution for $M_t^{(F)}$ is Poisson$(\lambda(1-p))$, and that the two random variables are independent.

5. (a) We get

$$\begin{aligned}
\pi(a \mid x_0, x_1) &\propto_a \pi(x_0, x_1 \mid a)\pi(a) \\
&= \pi(x_1 \mid a, x_0)\pi(a) \\
&= \text{Normal}(x_1; ax_0, 1) \cdot \text{Normal}(a; 1, 1) \\
&\propto_a \exp\left(-\frac{1}{2}(x_1 - ax_0)^2\right)\exp\left(-\frac{1}{2}(a-1)^2\right) \\
&= \exp\left(-\frac{1}{2}\left[a^2x_0^2 - 2ax_0x_1 + x_1^2 + a^2 - 2a + 1\right]\right) \\
&\propto_a \exp\left(-\frac{1}{2}\left[(x_0^2+1)a^2 - 2a(x_0x_1+1)\right]\right) \\
&\propto_a \exp\left(-\frac{1}{2}(x_0^2+1)\left(a - \frac{x_0x_1+1}{x_0^2+1}\right)^2\right)
\end{aligned}$$

so $a \sim \text{Normal}\left(\frac{x_0x_1+1}{x_0^2+1}, \frac{1}{x_0^2+1}\right)$.

(b) In the above computation a had the prior Normal$(1, 1)$, but it should be clear that an entirely similar computation would result in a normal posterior also for any other normal prior for a. Furthermore, the posterior for a given a sequence $x_0, x_1, \dots, x_k$ of observations for the variables $X_0, X_1, \dots, X_k$ can be computed by sequentially using each observation as data, and the posterior for a from previous computations as as prior for a. The posterior is then at each stage normal, so, by induction, the final posterior, given all observations, is also normal.

6. (a) The chain is ergodic as there are non-zero transition rates for example from 1 to 2 to 4 to 3 and back to 1, and thus only one communication class.

(b)

$$Q = \begin{bmatrix} -4 & 3 & 1 & 0 \\ 0 & -2 & 0 & 2 \\ 3 & 1 & -9 & 5 \\ 0 & 0 & 2 & -2 \end{bmatrix}.$$

(c) We must have $\sum_{i=1}^4 v_i = 1$ and $vQ = 0$. The four columns of Q give rise to 4 dependent equations. Selecting 3 of these, we get, for example,

$$\begin{aligned}
1 &= v_1 + v_2 + v_3 + v_4 \\
0 &= -4v_1 + 3v_3 \\
0 &= 3v_1 - 2v_2 + v_3 \\
0 &= v_1 - 9v3 + 2v_4
\end{aligned}$$

(d)

$$P = \begin{bmatrix} 0 & \frac{3}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 1 \\ \frac{3}{9} & \frac{1}{9} & 0 & \frac{5}{9} \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

(e) We must have $\sum_{i=1}^{4} w_i = 1$ and $w(I - P) = 0$. The four columns of $I - P$ give rise to 4 dependent equations. Selecting 3 of these, we get, for example,

$$\begin{aligned} 1 &= w_1 + w_2 + w_3 + w_4 \\ 0 &= w_1 - \frac{3}{9} w_3 \\ 0 &= -\frac{3}{4} w_1 + w_2 - \frac{1}{9} w_3 \\ 0 &= -w_2 - \frac{5}{9} w_3 + w_4 \end{aligned}$$

7. (a) We can write

$$B_1 + B_2 + B_3 = B_1 + 2B_2 + B_3 - B_2 = 3B_1 + 2(B_2 - B_1) + B_3 - B_2.$$

where each of B_1, $B_2 - B_1$, and $B_3 - B_2$ are independent and normally distributed with expectation zero. Thus the sum is normally distributed with expectation zero, and variance

$$\begin{aligned} \mathrm{Var}\,[B_1 + B_2 + B_3] &= \mathrm{Var}\,[3B_1 + 2(B_2 - B_1) + B_3 - B_2] \\ &= \mathrm{Var}\,[B_1] + \mathrm{Var}\,[2(B_2 - B_1)] + \mathrm{Var}\,[B_3 - B_2] \\ &= 9\,\mathrm{Var}\,[B_1] + 4\,\mathrm{Var}\,[B_2 - B_1] + \mathrm{Var}\,[B_3 - B_2] \\ &= 9\,\mathrm{Var}\,[B_1] + 4\,\mathrm{Var}\,[B_1] + \mathrm{Var}\,[B_1] \\ &= 9 + 4 + 1 = 14. \end{aligned}$$

Thus

$$B_1 + B_2 + B_3 \sim \mathrm{Normal}(0, 14).$$

(b) A Gaussian process $\{X_t\}_{t\geq 0}$ is a continuous-time stochastic process with the property that for all $n = 1, 2, \dots$ and $0 \leq t_1 < t_2 < \cdots < t_n$, the random variables $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ have a multivariate Normal distribution.

(c) Note that T_1 is the first hitting time for a. This is a stopping time, and thus the stochastic process $\{X_t\}_{t\geq 0}$ defined by

$$X_t = B_{T_1 + t} - B_{T_1}$$

is Brownian motion. This means that it has infinitely many zeroes in the interval $(0, \epsilon)$ for any $\epsilon > 0$. Thus, for any $n > 0$ there is infinitely many sequences $T_1 < t_1 < t_2 < \cdots < t_n < T_1 + \epsilon$ such that $B_{t_i} = a$ for $i = 1, \dots, n$. This shows that $T_{100} = T_1$, i.e., its distribution is concentrated in the value 0.