Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE550 Stochastic Processes and Bayesian Inference

Exam April 24 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.

1. (6 points) In the context of discrete time discrete state space time-homogeneous Markov chains:
   (a) What is a regular transition matrix?
   (b) What is a communication class, and what does it mean that a communcation class is closed?
   (c) What does it mean that a state j is transient?
   (d) What does it mean that a state j is positive recurrent?
   (e) If the state space is finite, what does it mean for the Markov chain to be ergodic?
   (f) If π is a stationary distribution for the Markov chain, what does it mean that it is time reversible?

2. (4 points) Assume $x \mid \lambda \sim \text{Exponential}(\lambda)$, so that x has an Exponential distribution with rate λ
   (a) Assume the prior is $\lambda \mid \alpha, \beta \sim \text{Gamma}(\alpha, \beta)$ for some parameters $\alpha > 0$ and $\beta > 0$. Compute the posterior distribution $\lambda \mid x$ and find its name and parameters.
   (b) Consider a Poisson process with parameter λ, and use as an improper prior for λ the function $\pi(\lambda) \propto 1/\lambda$. Assuming that the three first waiting times for observations in the Poisson process were 1.2, 1.7, and 0.9, find the posterior distribution for λ.

3. (5 points) Consider the discrete time Markov chain whose transition graph is illustrated in Figure 1.
   (a) Write down the transition matrix.
   (b) Compute the fundamental matrix F.
   (c) Given that the chain starts in state 1, what is the probability that it will be absorbed in state 4?

4. (2 points) Formulate the strong law of large numbers for Markov chains.

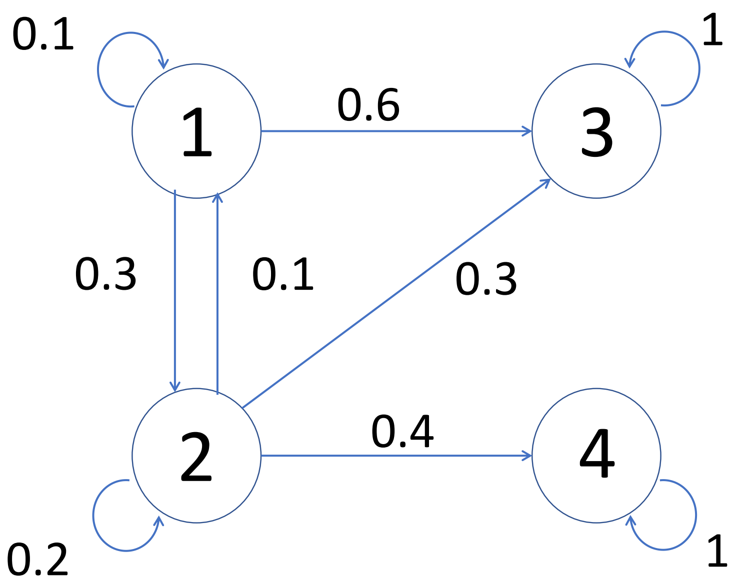

Figure 1: The graph for question 3.

5. (4 points) A machine component can have one of three states: A, B, or C. It stays in each state for an exponentially distributed time length, with expectation 1/2, 1/3, and 1/4 minutes for the states A, B, and C, respectively. When it changes from state A, it goes into state B with 60% probability or state C with 40% probability. When it changes from state B, it goes to state A with 90% probability; otherwise it goes to state C. When it changes from state C, it always goes to state A. Compute the long-term proportion of time that the component spends in state A.

6. (3 points) Explain what a Hidden Markov Model is, in particular describe what are the hidden variables and what are the observed variables in such a model. *Outline* a computational algorithm for finding the marginal posterior distribution for one of the hidden variables given all the observed variables.

7. (4 points) Assume two types of requests arrive at a computer server: Requests of type A arrive as a Poisson process with parameter λ_A and requests of type B arrive as an independent Poisson process with parameter λ_B.

   (a) If $\lambda_A = 3$ and $\lambda_B = 2$, what is the probability that within the first time unit, exactly three requests of type A and exactly 4 requests of type B will arrive?

   (b) For general λ_A and λ_B, what is the formula for the probability that the sequence of the first 7 requests will be A, B, B, A, B, B, A?

8. (2 points) Let B_t and W_t denote independent Brownian motions, and define, for $t \geq 0$ and real a and b,

$$X_t = a + b(B_t + W_{3+t}).$$

Find all pairs (a, b) such that $\{X_t\}_{t\geq 0}$ is Brownian motion; the answer may depend on W_3.

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 $x = (x_1, x_2, \dots, x_n)$ has a Dirichlet distribution, with $x_i \geq 0$ and $\sum_{i=1}^n x_i = 1$ and with parameters $\alpha = (\alpha_1, \dots, \alpha_n)$ with $\alpha_1 > 0, \dots, \alpha_n > 0$, 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 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 April 24 2019

1. (a) A regular transition matrix P is a transition matrix such that there is an $n > 0$ such that P^n is a positive matrix: A positive matrix is one where all the elements are positive.
   (b) A communication class is a subset S of states such that, for all $i, j \in S$, there are $n > 0$ and $m > 0$ such that $P^m_{ij} > 0$ and $P^n_{ij} > 0$, while for any pair $i \in S$ and $j \notin S$, this is not the case. A closed communication class is a communication class with a zero probability of ever leaving the class.
   (c) A state j is transient if the probability that a chain starting at j will ever return to j is less than 1.
   (d) A state j is positive recurrent if the expected number of steps for a chain to return to j if it starts at j is finite.
   (e) A finite state space Markov chain is ergodic if it is irreducible and aperiodic.
   (f) Time reversibility means that, for all states i and j, $\pi_i P_{ij} = \pi_j P_{ji}$.

2. (a) Using Bayes theorem we get

$$\begin{aligned}
\pi(\lambda \mid x) &\propto_\lambda \pi(x \mid \lambda)\pi(\lambda) \\
&\propto_\lambda \text{Exponential}(x; \lambda)\,\text{Gamma}(\lambda; \alpha, \beta) \\
&\propto_\lambda \lambda \cdot \exp(-\lambda x) \cdot \lambda^{\alpha-1} \cdot \exp(-\beta\lambda) \\
&\propto_\lambda \lambda^{\alpha} \cdot \exp(-(\beta + x)\lambda) \\
&\propto_\lambda \text{Gamma}(\lambda; \alpha + 1, \beta + x)
\end{aligned}$$

   In other words, the posterior distribution is a Gamma distribution with parameters $\alpha + 1$ and $\beta + x$.
   (b) The prior corresponds to a Gamma(0, 0) distribution. The posterior is obtained by updating the Gamma distribution as in (a) with the data given, resulting in the posterior

$$\text{Gamma}(3, 1.2 + 1.7 + 0.9) = \text{Gamma}(3, 3.8).$$

3. (a) We get the transition matrix

$$P = \begin{bmatrix} 0.1 & 0.3 & 0.6 & 0 \\ 0.1 & 0.2 & 0.3 & 0.4 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

(b) We get

$$Q = \begin{bmatrix} 0.1 & 0.3 \\ 0.1 & 0.2 \end{bmatrix}$$

and so

$$\begin{aligned} F &= (I-Q)^{-1} = \begin{bmatrix} 0.9 & -0.3 \\ -0.1 & 0.8 \end{bmatrix}^{-1} = \frac{1}{0.9\cdot 0.8 - 0.3\cdot 0.1}\begin{bmatrix} 0.8 & 0.3 \\ 0.1 & 0.9 \end{bmatrix} \\ &= \frac{1}{0.69}\begin{bmatrix} 0.8 & 0.3 \\ 0.1 & 0.9 \end{bmatrix} = \begin{bmatrix} 1.1594 & 0.4348 \\ 0.1449 & 1.3043 \end{bmatrix}. \end{aligned}$$

(c) We have

$$R = \begin{bmatrix} 0.6 & 0 \\ 0.3 & 0.4 \end{bmatrix}$$

and thus

$$FR = \frac{1}{0.69}\begin{bmatrix} 0.8 & 0.3 \\ 0.1 & 0.9 \end{bmatrix}\begin{bmatrix} 0.6 & 0 \\ 0.3 & 0.4 \end{bmatrix} = \frac{1}{0.69}\begin{bmatrix} 0.57 & 0.12 \\ 0.33 & 0.36 \end{bmatrix}.$$

Thus the probability for a process that starts in state 1 to be absorbed in state 4 is $\frac{0.12}{0.69} = 0.1739$.

4. Assume $X_0, X_1, \dots, X_n, \dots$ is an ergodic Markov chain with stationary distribution π. Let r be a bounded real-valued function. Then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} r(X_i) = \mathrm{E}(r(X))$$

where X is a random variable with distribution π.

5. The holding time parameters are $q = (2, 3, 4)$. The embedded chain transition matrix is

$$\tilde{P} = \begin{bmatrix} 0 & 0.6 & 0.4 \\ 0.9 & 0 & 0.1 \\ 1 & 0 & 0 \end{bmatrix}.$$

Thus the generator matrix becomes

$$Q = \begin{bmatrix} -2 & 1.2 & 0.8 \\ 2.7 & -3 & 0.3 \\ 4 & 0 & -4 \end{bmatrix}.$$

The linear system $\pi Q = 0$ gives

$$\begin{aligned} -2\pi_1 + 2.7\pi_2 + 4\pi_3 &= 0 \\ 1.2\pi_1 - 3\pi_2 &= 0 \\ 0.8\pi_1 + 0.3\pi_2 - 4\pi_3 &= 0 \end{aligned}$$

with solution $\pi = \frac{1}{163}(100, 40, 23)$. Thus, the long-term proportion of time that the component spends in state A is $100/163 = 0.6135$.

6. A hidded Markov model consists of a Markov chain $X_1, X_2, \dots, X_n$ of "hidden" random variables, and another sequence $Y_1, \dots, Y_n$ of variables such that the distribution of Y_i only depends on X_i, and possibly on Y_{i-1}. Thes latter variables are the "observed" variables. If the values of the variables Y_i are indeed observed, the posterior distribution for one of the hidden variables, say X_i can be found as follows: In a "Forward" part of the algorithm, for $j = 1, \dots, i$, the posterior for X_j given $Y_1, \dots, Y_j$ is found in a recursive algorithm. In a "Backward" part of the algorithm, for $j = n, \dots, i$, the likelihoods for X_j given the data $Y_j, \dots, Y_n$ are found in a recursive algorithm. Then the two are put together to find the marginal posterior for X_i.

7. (a) The events that three requests of type A arrive during the time unit and that four requests of type B arrive during the time unit are independent, and the probability of both can be computed using the Poisson probability mass function. Thus the answer is
$$e^{-\lambda_A}\frac{\lambda_A^3}{3!}e^{-\lambda_B}\frac{\lambda_B^4}{4!} = e^{-3}\frac{3^3}{3!}e^{-2}\frac{2^4}{4!} = 3e^{-5} = 0.02021384.$$

   (b) The probability of the first event being a request of type A or B is $\frac{\lambda_A}{\lambda_A+\lambda_B}$, respectively $\frac{\lambda_B}{\lambda_A+\lambda_B}$. As the successive events are independent, the probability asked for is
$$\left(\frac{\lambda_A}{\lambda_A+\lambda_B}\right)^3\left(\frac{\lambda_B}{\lambda_A+\lambda_B}\right)^4 = \frac{\lambda_A^3\lambda_B^4}{(\lambda_A+\lambda_B)^7}.$$

8. We have
$$\mathrm{E}(X_t) = a + b(\mathrm{E}(B_t) + \mathrm{E}(W_{3+t})) = a + b(0 + W_3).$$
Setting this to zero gives $a + bW_3 = 0$. Further,
$$\mathrm{Var}(X_t) = b^2\left(\mathrm{Var}(B_t) + \mathrm{Var}(W_{3+t})\right) = b^2(t+t)$$
and setting this to t gives $2b^2 = 1$. Thus we must have $b = \frac{1}{\sqrt{2}}$ and $a = -\frac{1}{\sqrt{2}}W_3$. On the other hand, with these values,
$$a + b(B_t + W_{3+t}) = \frac{1}{\sqrt{2}}(B_t + (W_{3+t} - W_3))$$
fulfills all criteria for a Brownian motion.