Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE550 Stochastic Processes and Bayesian Inference

Exam January 18, 2020, 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 relevant information about some probability distributions.

1. (4 points) Assume a random variable X with values in the positive real numbers has density $\pi(x \mid \theta) = \theta^2 x \exp(-\theta x)$, where $\theta > 0$ is a parameter. Assume we use the improper prior $\pi(\theta) \propto_\theta 1/\theta$.

   (a) Find a function of θ proportional to the posterior for θ given an observation x.

   (b) Find the name and parameters of the posterior for θ given an observation x.

   (c) Find the name and parameters of the posterior for θ after having made all the observations $x = 2.1$, $x = 1.2$, and $x = 3.9$.

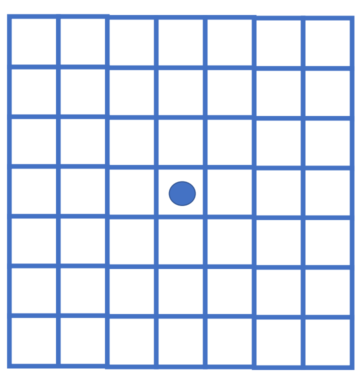

Figure 1: The board for question 2.

2. (4 points) A game is played on a board consisting of 7×7 squares, see Figure 1. The game starts in the square in the middle of the board. In each move, one moves with equal probability to one of the directly adjacent squares: There are 4, 3, or 2 such adjacent squares, with squares on the edges and in the corners having fewer neighbours.

(a) Show that the game can be described as a random walk on a graph.

(b) Compute the expected number of steps until you are back at the central square of the board.

3. (5 points) In the offspring process of a Branching process, the probability to obtain 3 offspring is 4/13 and the probability to obtain 2 offspring is also 4/13. With the remaining probability, there is no offspring.

   (a) Is the process supercritical, subcritical, or critical?

   (b) What is the expected size of Z_n, the size of the n'th generation?

   (c) Find the probability generating function for the offspring distribution.

   (d) Compute the probability that the process goes extinct.

4. (3 points) What is the Ising model? Describe how one can use Gibbs sampling on this model. (For partial credit: Just describe how Gibbs sampling works).

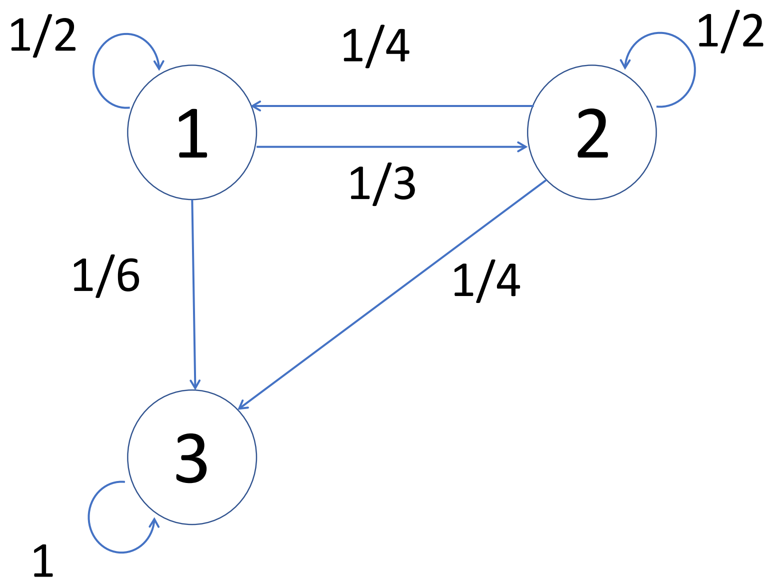

Figure 2: The network for question 5.

5. (4 points) Consider the discrete time Markov chain on the three states 1, 2, 3 illustrated in Figure 2. The chain starts in state 1.

   (a) Describe the communication classes of the chain. For each state, describe whether it is transient or recurrent.

   (b) Compute the expected number of visits to state 2.

6. (6 points) A biological system is described by a continuous time Markov chain with 5 diffeerent possible states: A, B, C, D, and E. We have that

   - If in state A, the system stays there for an expected time of 2 hours, then moves to state B.
   - If in state C, the system stays there for an expected time of 3 hours, then moves to state B.
   - If in state E, the system stays there for an expected time of 4 hours, then moves to state D.
   - If in state B, the system moves to state A with a rate of 0.5 per hour, to state C with a rate of 0.5 per hour, and to state D with a rate of 1 per hour.
   - If in state D, the system moves to state B at the rate of 1 per hour, and to state E with a rate of 0.5 per hour.

   (a) Write down the generator matrix Q.

   (b) Describe why this chain has a limiting distribution $v = (v_1, v_2, \ldots, v_5)$.

   (c) Is this a time-reversible chain? Prove or disprove this.

   (d) Write down 5 equations whose joint solution determine v.

7. (4 points) Let B_t denote Brownian motion, and define, for all $t > 0$ and some constants $a > 0$ and $b > 0$, $X_t = bB_{at}$.

   (a) Define what it means that B_t is a Brownian motion.

   (b) Compute the distribution of $X_1 + X_2$, expressed in terms of a and b.

   (c) Determine a formula a and b need to satisfy in order for X_t to be Brownian motion, and prove that X_t is Brownian motion if a and b satisfy this formula.

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
Exam January 18 2020

1. (a) We get

$$\pi(\theta \mid x) \propto_\theta \pi(x \mid \theta)\pi(\theta) \propto_\theta \theta^2 x \exp(-\theta x)\theta^{-1} = \theta \exp(-\theta x)$$

   (b) By comparing with a Gamma distribution with parameters $\alpha = 2$ and $\beta = \theta$, we get that

$$\theta \mid x \sim \text{Gamma}(2, x)$$

   (c) We get

$$\begin{aligned} & \pi(\theta \mid x_1 = 2.1, x_2 = 1.2, x_3 = 3.9) \\ \propto_\theta \; & \pi(x_1 = 2.1 \mid \theta)\pi(x_2 = 1.2 \mid \theta)\pi(x_3 = 3.9 \mid \theta)\pi(\theta) \\ \propto_\theta \; & \theta^2 \exp(-2.1\theta)\theta^2 \exp(-1.2\theta)\theta^2 \exp(-3.9\theta)\theta^{-1} \\ = \; & \theta^5 \exp(-7.2\theta) \end{aligned}$$

   By comparing with a Gamma distribution with parameters $\alpha = 8$ and $\beta = 7.2$, we get that

$$\theta \mid x_1 = 2.1, x_2 = 1.2, x_3 = 3.9 \sim \text{Gamma}(6, 7.2).$$

   Another way to obtain this result is to observe that the Gamma family of densities is conjugate to the presented likelihood: If the prior is Gamma(α,β) then the posterior becomes Gamma($\alpha + 2$,$\beta + x$). With the prior corresponding to the improper distribution Gamma(0, 0) the posterior becomes

$$\text{Gamma}(0 + 2 + 2 + 2, 0 + 2.1 + 1.2 + 3.9) = \text{Gamma}(6, 7.2).$$

2. (a) Consider a graph where the nodes correspond to the 49 squares of the game. Two nodes are connected with an edge if the corresponding squares are adjacent. As one moves to any of the adjacent squares with equal probability, this move corresponds to a random walk along the graph.

   (b) In the graph, we can count that there are 25 nodes with degree 4, 20 nodes with degree 3, and 4 nodes with degree 2. So the sum of all degrees is $25 \cdot 4 + 20 \cdot 3 + 4 \cdot 2 = 168$. Thus, in the limiting distribution, the probability of being at the central square is $4/168 = 1/42$. This means that the expected return time to the central square is 42.

3. (a) The expected size of the offspring distribution is

$$3\frac{4}{13} + 2\frac{4}{13} = \frac{20}{13} > 1$$

so the process is supercritical.

(b) We get

$$\mathrm{E}(Z_n) = \left(\frac{20}{13}\right)^n.$$

(c) The probability generating function is

$$G(s) = \frac{4}{13}s^3 + \frac{4}{13}s^2 + \frac{5}{13}.$$

(d) The probability of extinction is the smallest positive root of $G(s) = s$, i.e., the smallest positive root of

$$\frac{4}{13}s^3 + \frac{4}{13}s^2 + \frac{5}{13} = s$$

Multiplying by 13 and noting that 1 is necessarily a root of this equation, as we always have $G(1) = 1$, we get

$$\begin{aligned} 4s^3 + 4s^2 - s + 5 &= 0 \\ (s-1)(4s^2 + 8s - 5) &= 0 \\ (s-1)(2s-1)(2s+5) &= 0 \end{aligned}$$

where we go from the first to the second line by polynomial division and from the second to the third for example by finding the roots of the 2nd degree polynomial $4s^2 + 8s - 5$. We see that the roots are 1, 1/2, and $-5/2$, so the smallest positive root, and thus the extinction probabliity, is 1/2.

4. The Ising model consists of a grid of variables σ_i, each of which have the possible values $+1$ and -1. Two variables σ_i and σ_j are *neighbours* if i and j are *neighbours* in the grid, denoted $i \sim j$. A probability density is defined on the set $\sigma = \{\sigma_i\}$ by writing

$$\pi(\sigma) \propto_\sigma \exp\left(\beta \sum_{i \sim j} \sigma_i \sigma_j\right)$$

where the sum runs over all pairs of neighbours i and j.

Gibbs sampling can be used to generate an (approximate) sample from this density in the following way: Starting with some arbitrary value on σ, one goes through all the components of σ, simulating each σ_i using the conditional density $\pi(\sigma_i \mid \sigma_{-i})$, where σ_{-i} denotes the all components of σ except i. This is repeated until reasonable convergence.

As σ_i has only two possible values, the conditional distribution can be found by computing

$$\begin{aligned}
\pi(\sigma_i = 1 \mid \sigma_{-i}) &= \frac{\pi(\sigma_i = 1, \sigma_{-i})}{\pi(\sigma_i = 1, \sigma_{-i}) + \pi(\sigma_i = -1, \sigma_{-i})} \\
&= \frac{\exp\left(\beta \sum_{i\sim j} 1 \cdot \sigma_j\right)}{\exp\left(\beta \sum_{i\sim j} 1 \cdot \sigma_j\right) + \exp\left(\beta \sum_{i\sim j} (-1) \cdot \sigma_j\right)} \\
&= \frac{1}{1 + \exp(-2\beta \sum_{i\sim j} \sigma_j)}.
\end{aligned}$$

5. (a) States 1 and 2 are in one communication class, while state 3 is in a separate communication class. States 1 and 2 are transient, as there is a positive probability of never returning to these states, while state 3 is absorbing, and thus recurrent.

   (b) The transition matrix is

$$P = \begin{bmatrix} 1/2 & 1/3 & 1/6 \\ 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1 \end{bmatrix}$$

   and thus the fundamental matrix is

$$\begin{aligned}
F &= (I - Q)^{-1} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1/2 & 1/3 \\ 1/4 & 1/2 \end{bmatrix} \right)^{-1} \\
&= \begin{bmatrix} 1/2 & -1/3 \\ -1/4 & 1/2 \end{bmatrix} = \frac{1}{\frac{1}{4} - \frac{1}{12}} \begin{bmatrix} 1/2 & 1/3 \\ 1/4 & 1/2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 3/2 & 3 \end{bmatrix}.
\end{aligned}$$

   If the chain starts in 1, the expected number of visits to state 2 until absorbtion is 2.

6. (a) We get

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

   (b) From the description, we see that all states are in the same communication class, and there is no absorbing state. Thus the chain is irreducible, and as such it has a unique stationary distribution which is the limiting distribution.

   (c) If we draw the transition graph for this Markov chain, we can see that it is a tree. This implies that it is time reversible.

   (d) In order for v to be a stationary distribution, we need that $vQ = 0$ and also that the sum of the components of v is 1. The five equations corresponding to $vQ = 0$ will be dependent, so we can remove one of theses. Thus, for example disregarding the fourth column of Q, the following 5 equations need to be satisfied:

$$\begin{aligned}
v_1 + v_2 + v_3 + v_4 + v_5 &= 1 \\
-\frac{1}{2}v_1 + \frac{1}{2}v_2 &= 0 \\
\frac{1}{2}v_1 - 2v_2 + \frac{1}{3}v_3 + v_4 &= 0 \\
\frac{1}{2}v_2 - \frac{1}{3}v_3 &= 0 \\
\frac{1}{2}v_4 - \frac{1}{4}v_5 &= 0
\end{aligned}$$

In fact, substituting in $v_1 = v_2$, $v_3 = \frac{3}{2}v_2$, and $v_5 = 2v_4$ we get two equations in v_2 and v_4:

$$\begin{aligned}
v_2 + v_2 + \frac{3}{2}v_2 + v_4 + 2v_4 &= 1 \\
\frac{1}{2}v_2 - 2v_2 + \frac{1}{2}v_2 + v_4 &= 0
\end{aligned}$$

which gives $v_2 = v_4 = \frac{2}{13}$ and the full solution

$$v = (v_1, v_2, v_3, v_4, v_5) = \frac{1}{13}(2, 2, 3, 2, 4).$$

7. (a) $\{B_t\}_{t\geq 0}$ needs to be a stochastic process satisfying:
   - $B_0 = 0$.
   - $B_t - B_s \sim \text{Normal}(0, t-s)$ for all $s < t$.
   - $B_t - B_s$ is independent from $B_r - B_q$ whenever $s < t \leq q < r$.
   - $t \mapsto B_t$ is a continuous map with probability 1.

   (b) We can for example write

$$X_1 + X_2 = bB_a + bB_{2a} = bB_a + b(B_{2a} - B_a + B_a) = 2bB_a + b(B_{2a} - B_a).$$

   As $B_a \sim \text{Normal}(0, a)$ and, independently, $B_{2a} - B_a \sim \text{Normal}(0, a)$, we get that $X_1 + X_2$ is normally distributed with expectation $\text{E}(bB_a + bB_{2a}) = 0$ and variance

$$\text{Var}(2bB_a + b(B_{2a} - B_a)) = 4b^2\,\text{Var}(B_a) + b^2\,\text{Var}(B_{2a} - B_a) = 4b^2a + b^2a.$$

   Thus,

$$X_1 + X_2 \sim \text{Normal}(0, 5b^2a).$$

   (c) To obtain $X_t \sim \text{Normal}(0, t)$ for all t, we need that

$$t = \text{Var}(X_t) = \text{Var}(bB_{at}) = b^2at$$

   so we must have $b = 1/\sqrt{a}$. But if this holds, X_t is indeed Brownian motion:

- $X_0 = B_{0a}/\sqrt{a} = B_0/\sqrt{a} = 0$.
- $X_t - X_s = \frac{1}{\sqrt{a}}(B_{at} - B_{as})$. But $B_{at} - B_{st} \sim \text{Normal}(0, at - as)$, so

$$\frac{1}{\sqrt{a}}(B_{at} - B_{as}) \sim \text{Normal}\left(0, \frac{at - as}{a}\right) = \text{Normal}(0, t - s)$$

- Assuming $s < t \leq q < r$, we have that

$$X_t - X_s = b(B_{at} - B_{as})$$

  and

$$X_r - X_q = b(B_{ar} - B_{aq})$$

  Using that $as < at \leq aq < ar$ and the fact that the Brownian motion has independent increments, we get that X_t also has independent increments.
- The map $t \mapsto bB_{at}$ is continuous with probability 1, as it is a composition of the map $t \mapsto B_t$, which is continuous with probability 1, and functions multiplying with a and b.