Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE550 Stochastic Processes and Bayesian Inference

Re-exam August 17, 2020, 8:30 - 12:30
Examiner: Petter Mostad, phone 031-772-3579
Allowed aids: All aids are allowed.
For example you may access teaching material on any format and you may use R for computation.
However, you are **not** allowed to communicate with any person other than the examiner and the exam guard.
Total number of points: 30. To pass, at least 12 points are needed.
You need to explain how you derive your answers, i.e., show the steps in computations, unless explicitly stated otherwise.
There is an appendix containing relevant information about some probability distributions.

1. (5 points) Assume the real-valued random variable X has distribution $X \mid \theta \sim \text{Normal}(42, 1/\theta)$, where $\theta > 0$ is a parameter. Assume we use an improper prior $\pi(\theta) \propto_\theta 1/\theta$ for θ.
   (a) Find the name and parameters of the posterior for θ given an observation x.
   (b) Find the name and parameters of the posterior for θ after observing $x = 41.1$, $x = 42.1$, and $x = 41.7$, in that order.
2. (6 points) A discrete-time Markov chain has states A, B, C, D, E. It starts at A and the transition matrix is
$$P = \begin{bmatrix} 0 & 0.1 & 0 & 0.1 & 0.8 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0.7 & 0 & 0.1 & 0.2 & 0 \end{bmatrix}.$$
   (a) Draw the transition graph.
   (b) For each state, write down (without explanation) whether the state is transient, recurrent, and/or absorbing. Also, write down the period of each state.
   (c) Write down two different stationary distributions for the chain.
   (d) Is the chain ergodic? Why/why not?
   (e) For each of its five states, compute the expected number of visits to the state.
3. (5 points) Consider a branching process where the offspring process is a Poisson process with intensity λ truncated at 3: In other words, to generate an offspring number, one generates a value from a Poisson process with intensity λ, and if that value is greater than 3 it is set to 3.
   (a) Find an expression for the probability generating function $G(s)$ in terms of λ.

(b) Let λ_c be the value for λ such that the branching process is critical. Find and simplify an equation that λ_c must satisfy, and explain briefly how one may compute λ_c numerically.

(c) Find and express in terms of λ_c the variance of the offspring process when the branching process is critical.

(d) Compute the extinction probability when the branching process is critical.

4. (4 points) The two real variables x and y, with $y > 0$, have a joint probability density defined by

$$\pi(x, y) \propto_{x,y} y^3 \exp\left(-x^2 y - 2xy - 2y\right)$$

Describe in detail how to set up and implement a Markov chain so that, in the limit when the number of steps increases to infinity, the distribution of the chain will follow the density above. Your algorithm should use Gibbs sampling.

5. (5 points) A continuous-time Markov chain has 5 states: 0, 1, 2, 3, 4. If it is at state 0, it waits a time that is distributed as Exponential(0.5), then X is drawn according a Binomial$(4, 0.8)$ distribution and the process moves to state X. When the process is in a state $X > 0$, it waits an Exponential$(X/5)$ time before moving to state $X - 1$.

(a) Draw a transition graph for the process, and write down its generator matrix Q.

(b) Is the process time reversible? Why / why not?

(c) Compute the proportion of time the process is in state 2.

(d) Another process works as follows: In any state, it waits a time that is Exponential(1) distributed. Then it produces the next state according to a transition matrix P. Find the matrix P so that this process is identical to the one described above.

6. (5 points) Assume B_t is Brownian motion.

(a) Compute the distribution of $B_{at} + B_{a^2t}$, where $a > 0$ is a constant.

(b) Assume $b > 0$. For what combinations of a and b is $B_{abt} + B_{a^2t}$ Brownian motion?

Appendix: Some probability distributions

The Bernoulli distribution

If $x \in \{0, 1\}$ has a Bernoulli distribution with parameter $0 \le p \le 1$, 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 $0 \le p \le 1$, 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 $x_i \ge 0$ 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
Re-exam August 17 2020

1. (a) We get

$$\begin{aligned}
\pi(\theta \mid x) &\propto_\theta \pi(x \mid \theta)\pi(\theta) \\
&\propto_\theta \frac{1}{\sqrt{2\pi/\theta}} \cdot \exp\left(-\frac{1}{2/\theta}(x-42)^2\right) \cdot \frac{1}{\theta} \\
&\propto_\theta \theta^{1/2-1} \cdot \exp\left(-\frac{1}{2}(x-42)^2 \cdot \theta\right).
\end{aligned}$$

This is proportional to a Gamma density with parameters $\alpha = 1/2$ and $\beta = \frac{1}{2}(x-42)^2$, so

$$\theta \mid x \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}(x-42)^2\right)$$

(b) More generally, we get that if the prior for θ is Gamma(α,β), then

$$\begin{aligned}
\pi(\theta \mid x) &\propto_\theta \pi(x \mid \theta)\pi(\theta) \\
&\propto_\theta \frac{1}{\sqrt{2\pi/\theta}} \cdot \exp\left(-\frac{1}{2/\theta}(x-42)^2\right) \cdot \theta^{\alpha-1}\exp(-\beta\theta) \\
&\propto_\theta \theta^{\alpha+1/2-1}\exp\left(-\left(\frac{1}{2}(x-42)^2+\beta\right)\theta\right),
\end{aligned}$$

so that

$$\theta \mid x \sim \text{Gamma}\left(\alpha+\frac{1}{2}, \beta+\frac{1}{2}(x-42)^2\right).$$

Applying this updating rule for the three data observations, we get that

$$\begin{aligned}
\theta \mid \text{data} &\sim \text{Gamma}\left(3\cdot\frac{1}{2}, \frac{1}{2}(41.1-42)^2+\frac{1}{2}(42.1-42)^2+\frac{1}{2}(41.7-42)^2\right) \\
&= \text{Gamma}(1.5, 0.455).
\end{aligned}$$

2. (a)

(b)
- A is transient, not recurrent, not absorbing, and has period 2.
- B is not transient, recurrent, not absorbing, and has period 2.
- C is not transient, recurrent, not absorbing, and has period 2.

- D is not transient, recurrent, absorbing, and has period 1.
- E is transient, not recurrent, not absorbing, and has period 2.

(c) For example

$$\begin{aligned} v_1 &= (0, 0.5, 0.5, 0, 0) \\ v_2 &= (0, 0, 0, 1, 0) \end{aligned}$$

(d) The chain is not ergodic: An ergodic chain needs to be irreducible, i.e., have only one communication class. This chain has 3 communication classes: A,E; B,C; and D.

(e) Using the ordering A, E, B, C, D of the states, we can re-write the transition matrix as

$$P' = \begin{bmatrix} 0 & 0.8 & 0.1 & 0 & 0.1 \\ 0.7 & 0 & 0 & 0.1 & 0.2 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} Q & R \\ 0 & E \end{bmatrix}$$

where $Q = \begin{bmatrix} 0 & 0.8 \\ 0.7 & 0 \end{bmatrix}$. For the purposes of computing the expected number of visits to A and E we can regard the three states C, D, E as a single absorbing state. We then get

$$F = (I - Q)^{-1} = \begin{bmatrix} 1 & -0.8 \\ -0.7 & 1 \end{bmatrix}^{-1} = \frac{1}{1 - 0.56} \begin{bmatrix} 1 & 0.8 \\ 0.7 & 1 \end{bmatrix} = \begin{bmatrix} 2.2727 & 1.8182 \\ 1.5909 & 2.2727 \end{bmatrix}.$$

As the chain starts in A, the expected number of visits to A is 2.2727 and the expected number of visits to E is 1.8182. For the remaining states, we see that there is a positive probability of reaching them, and that, when the chan has reached them, it will revisit them an infinite number of times. Thus the expected number of visits is infinite.

3. (a) We get

$$\begin{aligned} G(s) &= \mathrm{E}(s^X) = \sum_{n=0}^{3} s^n \Pr(X = n) \\ &= e^{-\lambda} + e^{-\lambda}\lambda s + e^{-\lambda}\frac{\lambda^2}{2}s^2 + (1 - e^{-\lambda}(1 + \lambda + \frac{\lambda^2}{2}))s^3 \end{aligned}$$

(b) The process is critical when $\mathrm{E}(X) = 1$ where X is the offspring process. This gives

$$\begin{aligned}
e^{-\lambda_c}\lambda_c + 2 \cdot e^{-\lambda_c}\frac{\lambda_c^2}{2} + 3 \cdot (1 - e^{-\lambda_c}(1 + \lambda_c + \frac{\lambda_c^2}{2})) &= 1 \\
\lambda_c + \lambda_c^2 + 3e^{\lambda_c} - 3 - e\lambda_c - \frac{3}{2}\lambda_c^2 &= e^{\lambda_c} \\
-\frac{1}{2}\lambda_c^2 - 2\lambda_c - 3 &= -2e^{\lambda_c} \\
\frac{1}{4}\lambda_c^2 + \lambda_c + \frac{3}{2} &= e^{\lambda_c}
\end{aligned}$$

The positive solution of this equation can be found numerically, for example by various optimization algorithm.

(c) We may compute the variance for example via the probability generating function. We have

$$\begin{aligned}
G'(s) &= e^{-\lambda}\lambda + e^{-\lambda}\lambda^2 s + (1 - e^{-\lambda}(1 + \lambda + \frac{1}{2}\lambda^2))3s^2 \\
G''(s) &= e^{-\lambda}\lambda^2 + (1 - e^{-\lambda}(1 + \lambda + \frac{1}{2}\lambda^2)6s \\
G''(1) &= e^{-\lambda}\lambda^2 + 6 - 6e^{-\lambda}(1 + \lambda + \frac{1}{2}\lambda^2) = 6 - 6e^{-\lambda} - 6e^{-\lambda}\lambda - 2e^{-\lambda}\lambda^2.
\end{aligned}$$

As we have

$$\mathrm{Var}(X) = G''(1) + G'(1) - G'(1)^2$$

and $G'(1) = \mathrm{E}(X) = 1$ when the process is critical, we get, in this case,

$$\mathrm{Var}(X) = 6 - 6e^{-\lambda_c} - 6e^{-\lambda_c}\lambda_c + 2e^{-\lambda_c}\lambda_c^2.$$

(d) The extinction probability is 1, as this is a critical process.

4. The chain will alternate between simulating from two conditional distributions:

$$\begin{aligned}
\pi(x \mid y) &\propto_x \exp\left(-x^2y - 2xy\right) \\
&\propto_x \exp\left(-y(x^2 + 2x)\right) \\
&\propto_x \exp\left(-y(x^2 + 2x + 1)\right) \\
&\propto_x \exp\left(-\frac{2y}{2}(x + 1)^2\right)
\end{aligned}$$

Comparing with the normal density we get

$$x \mid y \sim \mathrm{Normal}\left(-1, \frac{1}{2y}\right).$$

Similarly,

$$\pi(y \mid x) \propto_y y^3 \exp(-y(x^2 + 2x + 2)).$$

Comparing with the Gamma density we get

$$y \mid x \sim \text{Gamma}(4, x^2 + 2x + 1)$$

A Gibbs sampling algorithm will start at some reasonable value, for example $(x, y) = (1, 1)$, and then alternate between simulating x and y from the distributions above.

5. (a) The transition rates out of state 0 can be computed as

$$\begin{aligned} q_{04} &= 0.5 \cdot 0.8^4 = 0.2048 \\ q_{03} &= 0.5 \cdot 4 \cdot 0.8^3 \cdot 0.2 = 0.2048 \\ q_{02} &= 0.5 \cdot 6 \cdot 0.8^2 \cdot 0.2^2 = 0.0768 \\ q_{01} &= 0.5 \cdot 4 \cdot 0.8 \cdot 0.2^3 = 0.0128 \end{aligned}$$

and thus we also get

$$q_0 = q_{01} + q_{02} + q_{03} + q_{04} = 0.4992$$

For the remaining non-zero transition rates we can compute

$$\begin{aligned} q_{43} &= \frac{4}{5} = 0.8 \\ q_{32} &= \frac{3}{5} = 0.6 \\ q_{21} &= \frac{2}{5} = 0.4 \\ q_{10} &= \frac{1}{5} = 0.2 \end{aligned}$$

so that we get

$$Q = \begin{bmatrix} -0.4992 & 0.0128 & 0.0768 & 0.2048 & 0.2048 \\ 0.2 & -0.2 & 0 & 0 & 0 \\ 0 & 0.4 & -0.4 & 0 & 0 \\ 0 & 0 & 0.6 & -0.6 & 0 \\ 0 & 0 & 0 & 0.8 & -0.8 \end{bmatrix}.$$

(b) The process is not time-reversible. The chain is irreducible and has a unique stationary distribution with positive probabilities for each state. Using this stationary distribution, then, for example, the probability to be in state 3 and move to state 2 is positive, while the probability to be in state 2 and then move to state 3 is zero, contradicting a condition of time-reversibility.

(c) If $v = (v_0, v_1, \ldots, v_4)$ is the unique stationary distribution, we know that $vQ = 0$. One can solve this on computer, but it is also easy to solve manually the equations

$$\begin{aligned} 0 &= -0.4992v_0 + 0.2v_1 \\ 0 &= 0.0128v_0 - 0.2v_1 + 0.4v_2 \\ 0 &= 0.0768v_0 - 0.4v_2 + 0.6v_3 \\ 0 &= 0.2048v_0 - 0.6v_3 + 0.8v_4 \\ 0 &= 0.2048v_0 - 0.8v_4 \end{aligned}$$

together with the equation $v_0+v_1+v_2+v_3+v_4 = 1$. Specifically, successive substitution of v_4 and v_3 yields

$$\begin{aligned} 0 &= -0.4992v_0 + 0.2v_1 \\ 0 &= 0.4864v_0 - 0.4v_2 \\ 0 &= 0.4096v_0 - 0.6v_3 \\ 0 &= 0.2048v_0 - 0.8v_4 \end{aligned}$$

which togethre with $v_0 + v_1 + v_2 + v_3 + v_4 = 1$ yields

$$v = (0.1769, 0.4416, 0.2153, 0.1208, 0.0453)$$

The answer is that the proportion of the time the chain is in state 2 is 0.2153.

(d) The new process will be a Poisson subortination of the orignial process. Note that we can use a Poisson rate of $\lambda = 1$ as this rate is higher than any of the transition rates q_i found above. According to the theory, we can write

$$P = \frac{1}{\lambda}Q + I = Q + I = \begin{bmatrix} 0.5008 & 0.0128 & 0.0768 & 0.2048 & 0.2048 \\ 0.2 & 0.8 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.6 & 0.4 & 0 \\ 0 & 0 & 0 & 0.8 & 0.2 \end{bmatrix}.$$

6. (a) For any $a > 0$ we have $\mathrm{E}(B_{at} + B_{a^2t}) = \mathrm{E}(B_{at}) + \mathrm{E}(B_{a^2t}) = 0$. If $a \geq 1$ then $a^2t \geq at$ and we can write

$$B_{at} + B_{a^2t} = B_{a^2t} - B_{at} + 2B_{at}$$

where $B_{a^2t} - B_{at}$ and B_{at} are independently normally distributed. The variance is

$$\begin{aligned} \mathrm{Var}(B_{a^2t} - B_{at} + 2B_{at}) &= \mathrm{Var}(B_{a^2t} - B_{at}) + \mathrm{Var}(2B_{at}) \\ &= \mathrm{Var}(B_{a^2t-at}) + 4\,\mathrm{Var}(B_{at}) \\ &= a^2t - at + 4at = a^2t + 3at \end{aligned}$$

Thus when $a \geq 1$,

$$B_{at} + B_{a^2t} \sim \text{Normal}(0, a^2t + 3at).$$

If $a \leq 1$ then $a^2t \leq at$ and we can write

$$B_{at} + B_{a^2t} = B_{at} - B_{a^2t} + 2B_{a^2t}.$$

The variance now becomes

$$\begin{aligned} \text{Var}(B_{at} - B_{a^2t} + 2B_{a^2t}) &= \text{Var}(B_{at-a^2t}) + 4\,\text{Var}(B_{a^2t}) \\ &= at - a^2t + 4a^2t = at + 3a^2t \end{aligned}$$

and we get

$$B_{at} + B_{a^2t} \sim \text{Normal}(0, at + 3a^2t).$$

(b) Similar to (a), we divide up into the cases $b \geq a$ and $b < a$. When $b \geq a$, similar computations as above gives

$$\text{Var}(B_{abt} + B_{a^2t}) = at(b + 3a)$$

so in order for this to be Brownian motion, we need that $a(b + 3a) = 1$, i.e., $b = 1/a - 3a$. Together with the conditions $0 < a \leq b$, we find that

$$\begin{aligned} 0 < \quad a &\leq \frac{1}{2} \\ b &= \frac{1}{a} - 3a \end{aligned}$$

are combinations of a and b fulfilling the criteria. On the other hand when $b < a$ we get

$$\text{Var}(B_{abt} + B_{a^2t}) = at(a + 3b)$$

and we need $a(a + 3b) = 1$, i.e., $b = \frac{1}{3}(\frac{1}{a} - a)$. Together with the conditions $0 < b < a$ we find that

$$\begin{aligned} \frac{1}{2} < \quad a &< 1 \\ b &= \frac{1}{3}\left(\frac{1}{a} - a\right) \end{aligned}$$

are combinations of a and b fulfilling the criteria.