Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE550 Stochastic Processes and Bayesian Inference

Re-exam August 26, 2024, 8:30 - 12:30
Examiner: Petter Mostad will be available by phone 031-772-3579
and will visit the 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.
All answers need to be explicitly computed or explicitly argued for.

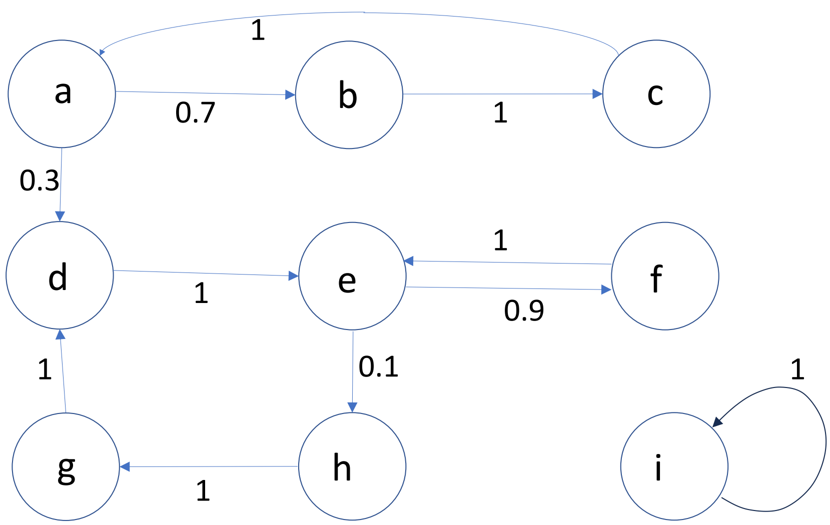

Figure 1: The transition graph used in question 1.

1. (4 points) Consider the transition graph in Figure 1 for a discrete time Markov chain.

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

   (b) In general, for a discrete time Markov chain with a finite state space, what does it take to make it ergodic? Define the words you use.

2. (8 points) Consider a continuous time Markov chain with state space $\{A, B, C, D, E, F\}$. It starts in state A.

   - It stays an average time of 5 seconds in state A before moving on to states B, C, or E with equal probability.
   - It stays an average time of 1 second in state B before moving on to states A or C with equal probability.
   - It stays an average time of 2 seconds in state C before moving on to state B.
   - It stays an average time of 3 seconds in state D before moving on to state F.
   - It stays an average time of 1 second in state E before moving on to states A, B, or C with equal probability.
   - It stays an average time of 2 seconds in state F before moving on to state D.

   (a) Write down the generator matrix Q for the process, and draw its transition rate graph.

   (b) Write down and simplify an equation, consisting of matrices of numbers, matrix operations, and vectors of numbers, that computes the long-term probability that the process is in state A.

   (c) Write down and simplify an equation, consisting of matrices of numbers, matrix operations, and vectors of numbers, that computes the expected time until the process hits state C for the first time.

   (d) Imagine that you make counts of the number of times each state is visited. Write down and simplify an equation, consisting of matrices of numbers, matrix operations, and vectors of numbers, that computes the long-term proportion of visits to state A.

3. (6 points) Tarja is studying data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ where, in each pair, the probability mass function for y_i given x_i and a real-valued parameter θ is modelled by $f(y_i, x_i, \theta)$. She wants to use the data and MCMC to obtain a sequence $\theta_1, \theta_2, \ldots, \theta_{100000}$ where each θ_i is approximately sampled from the posterior of θ given the data. She uses a flat prior on θ.

   (a) Write down a mathematical formula for the likelihood of the parameter θ given the data.

   (b) Write down a mathematical formula for the logarithm of the posterior density for θ given the data.

   (c) Assuming the function in (b) has been implemented in a function called `logpost`, write down pseudo-code (or actual code) for an algorithm generating a sequence $\theta_1, \theta_2, \ldots, \theta_{100000}$ as described above. Explain all items in the code that you need to introduce.

4. (4 points) Consider a Branching process with offspring distribution X, where

$$\Pr(X = k) = \begin{cases} f(k)(1-\beta) & \text{for } k \text{ even (including } k = 0) \\ \beta & \text{for } k = 1 \\ 0 & \text{otherwise} \end{cases}$$

where f is some function and β is some parameter satisfying $0 \leq \beta < 1$.

(a) Find and simplify an expression for the probability generating function for X.

(b) Prove that the probability of extinction for this brancing process does not depend on the parameter β.

5. (6 points) Consider simulations from a Brownian motion.

(a) At what specific time t in the simulation will you be 90% sure that the value is within the interval $[-100, 100]$? (Express the answer either as an (approximate) number or in terms of standard functions (e.g., R functions) of numbers).

(b) Consider a random walk where, at each time step, the walker adds to her position either +1 or -1. The walk starts at zero. Use the Donsker principle to find out how many steps the walker can make to be 90% sure she has stayed within the interval $[-100, 100]$.

(c) Now, assume the random walker, at each time step, will move either +1, -1, or 0, with equal probabilities. How many time steps should the walker now make to be 90% sure to be within the interval $[-100, 100]$?

6. (2 points) What is the definition of a spatial Poisson process? Explain.

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 $0 \le p \le 1$, 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 $x \ge 0$ 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 Gamma$(\alpha,\beta)$ distribution has expectation $\frac{\alpha}{\beta}$ and variance $\frac{\alpha}{\beta^2}$.

The Inverse Gamma distribution

If $x > 0$ has an Inverse 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 $x \mid \alpha,\beta \sim \text{Inverse-Gamma}(\alpha,\beta)$ and $\pi(x \mid \alpha,\beta) = \text{Inverse-Gamma}(x;\alpha,\beta)$. If $x \sim \text{Gamma}(\alpha,\beta)$ then $1/x \sim \text{Inverse-Gamma}(\alpha,\beta)$. The Inverse-Gamma$(\alpha,\beta)$ distribution has expectation $\frac{\beta}{\alpha-1}$ and variance $\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$.

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 $x \in \{0, 1, 2, \ldots\}$ 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 $\pi(x \mid \lambda) = \text{Poisson}(x; \lambda)$. 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
Re-exam August 26 2024

1. (a) The communication classes of the chain are the sets $\{a, b, c\}$, $\{d, e, f, g, h\}$, and $\{i\}$. The states $\{a, b, c\}$ are transient and have period 3. The states $\{d, e, f, g, h\}$ are recurrent and have period 2. The state $\{i\}$ is recurrent and has period 1.

   (b) It needs to be irreducible and aperiodic. Irreducible means that, for any two states, there is an n so that one can go from the first to the second in n steps. The period of a state is the largest common divisor of the lengths of all possible paths that start and end in the state. A state is aperiodic if it has period 1. A Markov chain is aperiodic if all its states are aperiodic.

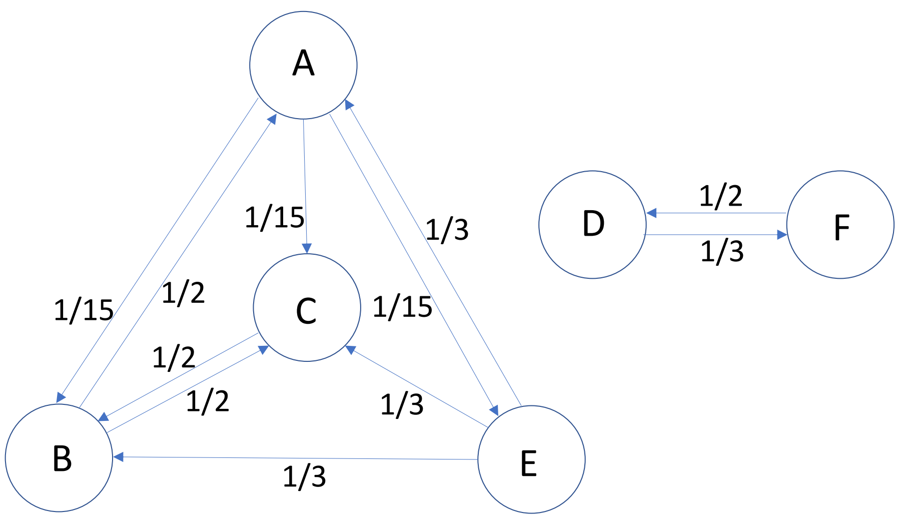

Figure 1: The transition rate graph for question 2.

2. (a) We get Figure 1 and

$$Q = \begin{bmatrix} -\frac{1}{5} & \frac{1}{15} & \frac{1}{15} & 0 & \frac{1}{15} & 0 \\ \frac{1}{2} & -1 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & -1 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}.$$

(b) To get the simplest equation, we first notice that the process as described is reducible, and that it will only visit the states A, B, C, and E, as it starts in A. So we can limit ourselves to looking at the Markov chain for these states, with generator matrix

$$Q' = \begin{bmatrix} -\frac{1}{5} & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} \\ \frac{1}{2} & -1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & -1 \end{bmatrix}.$$

The answer to the question is the first coordinate of the probability vector v satisfying $vQ' = 0$. Replacing for example the first column of Q' with 1's to produce a matrix Q'', we get that we want to solve the equation

$$vQ'' = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}.$$

Multiplying on the right with the inverse of Q'', and extracting the value corresponding to A, the solution can be computed with

$$\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} \\ 1 & -1 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 1 & \frac{1}{3} & \frac{1}{3} & -1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The numerical answer is $45/88 = 0.5113636$.

(c) If we make state C absorbing, the remaining matrix has form

$$V = \begin{bmatrix} -\frac{1}{5} & \frac{1}{15} & \frac{1}{15} \\ \frac{1}{2} & -1 & 0 \\ \frac{1}{3} & \frac{1}{3} & -1 \end{bmatrix}$$

and the fundamental matrix is $F = -V^{-1}$. As the process starts in A, the answer to the question is the sum of the first row of this matrix, which can be computed with

$$-\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{5} & \frac{1}{15} & \frac{1}{15} \\ \frac{1}{2} & -1 & 0 \\ \frac{1}{3} & \frac{1}{3} & -1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

The numerical answer is $26/3 = 8.66666$.

(d) To solve this, we first find the transition matrix for the embedded chain of the reduced process concering states A, B, C, and E:

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

The answer to the question is the first term of the probability vector v that satisfies $vP = v$, i.e., $v(I - P) = 0$. To compute it, we may compute

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

The numerical answer is $9/40 = 0.225$. An alternative solution could be based on the full solution from (b): Finding there that

$$(v_A, v_B, v_C, v_E) = \frac{1}{88}(45, 16, 24, 3),$$

we get that the answer to (d) can be calculated as

$$\frac{v_A/5}{v_A/5 + v_B/1 + v_C/2 + v_E/1} = \frac{45/5}{45/5 + 16 + 24/2 + 3} = \frac{9}{40}.$$

3. (a) The likelihood function $L(\theta)$ becomes the product over the different functions for each i:

$$L(\theta) = \prod_{i=1}^{n} f(y_i, x_i, \theta).$$

(b) As the posterior density is proportional to the likelihood times the prior, we get for the logarithm of the posterior density

$$\ell(\theta) = \log\left(\prod_{i=1}^{n} f(y_i, x_i, \theta)\right) + C = \sum_{i=1}^{n} \log(f(y_i, x_i, \theta)) + C$$

where C is a constant chosen such that

$$\int_\theta \exp(\ell(\theta))\, d\theta = 1.$$

(c) We can write, in R, or similar in pseudo-code, for example

```
N <- 100000
theta <- rep(0, N)
for (i in 2:N) {
  proposal <- prop(theta[i-1])
  diff <- logpost(proposal) - logpost(theta[i-1]) +
    logdiff(proposal, theta[i-1]) - logdiff(theta[i-1], proposal)
  if (runif(1) < exp(diff))
    theta[i] <- proposal
  else
    theta[i] <- theta[i-1]
}
```

We choose to run the MCMC Markov chain for $N = 100000$ steps. We choose the initial value $\theta_1 = 0$ in the code above. To run the MCMC, we need a proposal function `prop`, which randomly simulates a new proposed value given the old value of θ. In addition to the `logpost` function, we also use above a `logdiff` function which computes the probability density for proposing its first argument given that the `prop` function is given the its second argument.

4. (a)

$$G_X(s) = \mathrm{E}\left[s^X\right] = \beta s + \sum_{k=0,k \text{ even}}^{\infty} s^k f(k)(1-\beta) = \beta s + (1-\beta)\sum_{i=0}^{\infty} s^{2k} f(2k).$$

(b) The extinction probability is the smallest positive root of the equation $G_X(s) = s$, i.e., of the equation

$$\beta s + (1-\beta)\sum_{i=0}^{\infty} s^{2k} f(2k) \quad = \quad s$$

which can be rewritten as

$$(1-\beta)\sum_{i=0}^{\infty} s^{2k} f(2k) = (1-\beta)s.$$

As $1-\beta > 0$, the above is equivalent to

$$\sum_{i=0}^{\infty} s^{2k} f(2k) = s$$

which does not depend on β, so the extinction probability is independent of β.

5. (a) The value at time t will be normally distributed with expectation zero and variance t. So we want to find a t so that a Normal distribution with standard deviation $\sqrt{t}$ and zero expectation has a 90% probability of being in the interval $[-100, 100]$. A variable with the standard normal has a 90% probability of being in the interval $[-1.64, 1.64]$, so we get

$$\sqrt{t} \cdot 1.64 = 100,$$

i.e.,

$$t = \left(\frac{100}{1.64}\right)^2 = 3718.025$$

So at time 3718, we will be 90% sure that simulations are in the interval $[-100, 100]$. If you do not remember the value 1.64, you may express the solution for example with the R code

```
(100/qnorm(0.95))^2
```

which gives the slightly more accurate answer 3696.115.

(b) Let X be the random variable for a single time step. Then $\mathrm{E}[X] = 0$ and

$$\mathrm{Var}[X] = \mathrm{E}\left[X^2\right] - EX^2 = 1 - 0 = 1$$

As this is a random walk, we get that the expectation and variance at time step t is 0 and t, respectively, and the Donsker principle tells us that, for large t, we can use Brownian motion as an approximate model. Thus the answer to (b) is the same as the answer to (a).

(c) If Y is the random variable for one step in the new process, we have $\mathrm{E}[Y] = 0$ and

$$\mathrm{Var}[Y] = \mathrm{E}\left[Y^2\right] - \mathrm{E}[Y]^2 = \frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 0 = \frac{2}{3}.$$

For a Brownian motion, we have that $\mathrm{Var}[B_{at}] = at$. Setting $a = 2/3$, we get that the Brownian motion $B_{2t/3}$ can be used to approximate our random walk. To be 90% sure that are within $[-100, 100]$, we need

$$\frac{2t}{3} = \left(\frac{100}{1.64}\right)^2$$

giving $t = 5577.037$, or, more accurately, using R,

```
3/2*(100/qnorm(0.95))^2
```

giving 5544.173

6. Given a set $U \subseteq \mathbb{R}^k$, a spatial Poisson process on U is a set $\{N_A\}_{A \subseteq U}$ of random variables such that

- N_A is Poisson distributed with parameter $\lambda(A)$ where $\lambda(A)$ is the size (e.g., area, volume, etc.) of A.
- If A and B are disjoint, then N_A and N_B are independent.

(The sets A need to be measurable, and $\lambda(A)$ is their Lebesgue measure).