Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE550 Stochastic Processes and Bayesian Inference

Re-exam April 5, 2023, 8:30 - 12:30
Examiner: Petter Mostad will be available by phone 031-772-3579
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

1. (6 points) Ari would like to learn about and document the probability p that a randomly chosen car going through his street exceeds the speed limit by more than 10 km/h. From his window, he can clock the cars in such a way that he can detect this, and every morning he counts cars until and including the first one speeding. He assumes his count X is Geometrically distributed.

   (a) If he observes the counts 9, 5, 7, 14, what is the likelihood function for p given this data?

   (b) What kind of assumption does he need to make about p to do a Bayesian analysis? Make a reasonable assumption.

   (c) Write down an expression, using mathematics or R code, that computes the probability that $p < 0.05$ given the observations Ari has made.

   (d) Let Y be the predicted count for tomorrow, using the information from the data observed so far in the prediction. Write down a mathematical expression equal to $\Pr(Y = 10)$.

2. (4 points) Consider a Branching process with offspring distribution X, where

$$\Pr(X = k) = \begin{cases} \alpha^k(1-\alpha^2) & \text{for } k \text{ even} \\ 0 & \text{for } k \text{ odd} \end{cases}$$

   for some parameter α with $0 < \alpha < 1$.

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

   (b) Compute the probability that the branching process goes extinct as a function of α, or show how to compute it.

3. (3 points) Assume a model has been established where observations $y_1, \dots, y_n, y_{n+1}$ are independent given some underlying parameter θ, with given likelihood function f:

$$\pi(y_i \mid \theta) = f(y_i; \theta).$$

Assume also given a prior density function g for θ:

$$\pi(\theta) = g(\theta).$$

(a) Outline the steps in an MCMC algorithm that obtains a sample $\theta_1, \theta_2, \dots, \theta_N$ from a Markov chain whose limiting distribution is the posterior $\pi(\theta \mid y_1, \dots, y_n)$.

(b) Assume a sequence $\theta_1, \dots, \theta_N$ like above has been produced. How can one use this sequence to obtain an approximate sample from the distribution of y_{n+1} conditional on $y_1, \dots, y_n$?

4. (8 points) Elisabeth is studying a system which can have 5 different states: A, B, C, D, and E. She assumes they change according to a continuous-time Markov model. The expected holding times for states A, B, C, D, E are 1/3, 1, 1/5, 1/6, and 1/2 years, respectively.

   When leaving a state, the probabilities for going to other states are specified as

   - From state A one always goes to state B.
   - From state B one goes to state A, C, or D with probabilities 1/4, 1/2, 1/4, respectively.
   - From state C one always goes to state B.
   - From state D one goes to state B or E with equal probability.
   - From state E one goes to state C or D with equal probability.

   Your answers to the questions below (except (b)) should be expressions consisting of matrices and vectors of numbers, and may contain operations such as addition, subtraction, multiplcation, and inversion. You do not need to evaluate the expressions.

   (a) Write down the generator matrix Q. (List states in the order A,B,C,D,E).

   (b) Is the Markov chain time reversible? Prove or disprove.

   (c) What is the long term probability that the system is in state A?

   (d) If the system starts in state B, what is the expected time until it enters state C?

   (e) If the system starts in state A, what is the *number of steps* until it enters state C?

5. (5 points) After a shop opens, customers enter it according to a Poisson process N_t with rate λ per minute. Independently of each other, customers spend Z minutes in the shop, where Z is uniformly distributed on the interval $[0, 10]$. Let X_t be the number of customers in the shop at time t.

   (a) Is X_t a Markov process? Prove or argue yes or no.

   (b) For a time $t \leq 10$, select uniformly at random a customer among those that entered the shop before time t. Compute the probability that the customer is still in the shop at time t.

(c) At a time $t \geq 10$, select uniformly at random a customer among those that entered the shop during the last 10 minutes. Compute the probability that the customer is still in the shop at time t.

(d) Compute the expected number of customers $\mathrm{E}[X_t]$ at time t. You may separate into the cases $t < 10$ and $t \geq 10$.

6. (4 points)

(a) Write down the definition of geometric Brownian motion.

(b) If G_t denotes geometric Brownian motion, compute its expectation $E(G_t)$.

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 $0 \leq p \leq 1$, 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 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 April 5 2023

1. (a) The likelihood function becomes

$$\pi(\text{data} \mid p) = p(1-p)^{9-1}p(1-p)^{4-1}p(1-p)^{7-1}p(1-p)^{14-1} = p^4(1-p)^{31}.$$

(b) He would need to choose a prior density $\pi(p)$ for p. A possibility would be to set $p \sim \text{Uniform}(0,1)$, but there are other possibilities, such as setting $\pi(p) \propto_p \frac{1}{p(1-p)}$.

(c) Using the uniform prior, we get the posterior

$$\pi(p \mid \text{data}) \propto_p \pi(\text{data} \mid p)\pi(p) = p^4(1-p)^{31} \propto_p \text{Beta}(p; 5, 32).$$

Thus

$$\Pr(p < 0.05 \mid \text{data}) = \int_0^{0.05} \text{Beta}(p; 5, 32)\, dp = \frac{\Gamma(5+32)}{\Gamma(5)\Gamma(32)} \int_0^{0.05} p^4(1-p)^{31}\, dp.$$

In R this can be computed with

```
pbeta(0.05, 5, 32)
```

which gives 0.03236271.

(d) Writing $\pi(p)$ for the posterior density Beta(5, 32) found above, the posterior predictive mass function in this case can be computed as

$$\begin{aligned}\pi(y) &= \frac{\pi(y \mid p)\pi(p)}{\pi(p \mid y)} = \frac{p(1-p)^y\,\text{Beta}(p; 5, 32)}{\text{Beta}(p; 5+1, 32+y)} \\ &= \frac{p(1-p)^y \frac{\Gamma(5+32)}{\Gamma(5)\Gamma(32)} p^4(1-p)^{31}}{\frac{\Gamma(6+32+y)}{\Gamma(6)\Gamma(32+y)} p^5(1-p)^{31+y}} = \frac{\Gamma(5+32)\Gamma(6)\Gamma(32+y)}{\Gamma(5)\Gamma(32)\Gamma(32+6+y)}.\end{aligned}$$

Thus

$$\Pr(Y = 10) = \frac{\Gamma(37)\Gamma(6)\Gamma(42)}{\Gamma(5)\Gamma(32)\Gamma(48)}.$$

2. (a)

$$G_X(s) = \mathrm{E}\left[s^X\right] = \sum_{k=0}^{\infty} s^{2k}\alpha^{2k}(1-\alpha^2) = (1-\alpha^2)\sum_{k=0}^{\infty}(\alpha^2 s^2)^k = \frac{1-\alpha^2}{1-\alpha^2 s^2}.$$

(b) The probability of extinction is the smallest positive root of the equation

$$\frac{1-\alpha^2}{1-\alpha^2 s^2} = s$$

i.e., of

$$\alpha^2 s^3 - s + 1 - \alpha^2 = 0.$$

Using that $s = 1$ is a root we can factor this as

$$(s-1)(\alpha^2 s^2 + \alpha^2 s + \alpha^2 - 1) = 0.$$

Solving $s^2 + s + 1 - 1/\alpha^2 = 0$ by completing the square gives the smallest positive root as

$$s = -\frac{1}{2} + \sqrt{\frac{1}{\alpha^2} - \frac{3}{4}}.$$

Alternatively, one might explain how to compute this result numerically.

3. (a) To define an MCMC chain, one first has to select a proposal distribution $g(\theta^* \mid \theta)$ from which one can readily simulate, and an initial distribution for θ. The algorithm then goes as follows:

- Simulate θ_1 from the initial distribution.
- For $i = 1, \ldots, N-1$:
  - Simulate θ^* using $g(\theta^* \mid \theta_i)$.
  - Compute
$$\rho = \min\left(1, \frac{g(\theta^*)\prod_{i=1} f(y_i;\theta^*)q(\theta \mid \theta^*)}{g(\theta_i)\prod_{i=1} f(y_i;\theta_i)q(\theta^* \mid \theta)}\right)$$
  - Set $\theta_{i+1} = \theta^*$ with probability ρ; otherwise set $\theta_{i+1} = \theta_i$.

(b) For example, one can for $i = 1, \ldots, N$ simulate z_i from the distribution with likelihood $\pi(z_i \mid \theta_i) = f(z_i;\theta_i)$. As $\theta_1, \ldots, \theta_N$ is an approximate sample from the posterior, $(z_1, \theta_1), \ldots, (z_N, \theta_N)$ will be an approximate sample from $\pi(y_{n+1}, \theta \mid y_1, \ldots, y_n)$, so $z_1, \ldots, z_N$ will be an approximate sample from $\pi(y_{n+1} \mid y_1, \ldots, y_n)$.

4. (a) We get

$$Q = \begin{bmatrix} -3 & 3 & 0 & 0 & 0 \\ \frac{1}{4} & -1 & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 5 & -5 & 0 & 0 \\ 0 & 3 & 0 & -6 & 3 \\ 0 & 0 & 1 & 1 & -2 \end{bmatrix}.$$

(b) The rate of going from E to C is 1, while the rate of going the other way is zero. A stationary distribution will have positive probabilities for all states, and so it cannot possibly fulfil the detailed balance condition between C and E. so the chain is NOT time reversible.

(c) The chain is ergodic, and its unique stationary distribution is given by the probability vector v satisfying $vQ = 0$. We can find v by replacing the first column of Q by 1s to obtain Q', solving the matrix equation $vQ' = [1, 0, 0, 0, 0]$, and finding the value v_1. In other words, one may compute

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 3 & 0 & 0 & 0 \\ 1 & -1 & \frac{1}{2} & \frac{1}{4} & 0 \\ 1 & 5 & -5 & 0 & 0 \\ 1 & 3 & 0 & -6 & 3 \\ 1 & 0 & 1 & 1 & -2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

(d) We make state C into an absorbing state. Thus we take out from Q the row and column corresponding to C, and then take the negative of the inverse to get the fundamental matrix. Then we take the sum of the *second* row to get the answer to the question. We get that we should compute

$$-\begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} -3 & 3 & 0 & 0 \\ \frac{1}{4} & -1 & \frac{1}{4} & 0 \\ 0 & 3 & -6 & 3 \\ 0 & 0 & 1 & -2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

(e) The transition matrix of the embedded chain is:

$$\tilde{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}.$$

Making C into an absorbing state, we obtain the matrix V by removing the row and column corresponding to C. The fundamental matrix in the case of a discrete-time Markov chain is then $F = (I - V)^{-1}$, and we get that we should now compute

$$\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ -\frac{1}{4} & 1 & -\frac{1}{4} & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

5. (a) X_t is not a Markov process, as the future number of customers in the shop will be influenced not only by the current number of customers there, but also by how long these customers have stayed, i.e., the values of X_s for $s < t$. For example, if we can see from the values of X_s for $s < t$ that the current customers have all been there more than 9 minutes, the number of customers at $t + 1$ will be simply those who arrived between t and $t + 1$, whereas if they had been there less than 1 minute, most will by a high probability still be there at $t + 1$.

(b) Let Y be the arrival time of the customer. Then the density $\pi(y)$ for Y is uniform on the interval $[0, t]$. If we condition on Y, we get

$$\begin{aligned}
\Pr(\text{still in shop}) &= \int_0^t \Pr(\text{still in shop} \mid y)\pi(y)\,dy \\
&= \int_0^t \left(1 - \frac{t-y}{10}\right) \cdot \frac{1}{t}\,dy \\
&= \frac{1}{t}\left[y - \frac{t}{10}y + \frac{y^2}{20}\right]_0^t \\
&= 1 - \frac{t}{10} + \frac{t}{20} \\
&= 1 - \frac{t}{20}.
\end{aligned}$$

(c) This situation is exactly the same as the situation when $t = 10$, as the Poisson process of arriving customers is stationary. Thus the probability becomes $1 - \frac{10}{20} = \frac{1}{2}$.

(d) If $t < 10$,

$$\mathrm{E}[X_t] = \mathrm{E}[\mathrm{E}[X_t \mid N_t]] = \mathrm{E}\left[N_t\left(1 - \frac{t}{20}\right)\right] = \left(1 - \frac{t}{20}\right)\mathrm{E}[N_t] = \left(1 - \frac{t}{20}\right)\lambda t.$$

If $t \geq 10$, as all customers who entered more than 10 minutes ago have left, we get

$$\mathrm{E}[X_t] = \mathrm{E}[\mathrm{E}[X_t \mid N_t - N_{t-10}]] = \mathrm{E}\left[(N_t - N_{t-10})\frac{1}{2}\right] = \frac{1}{2}\,\mathrm{E}[N_{10}] = \frac{10\lambda}{2} = 5\lambda.$$

6. (a) A process G_t is geometric Brownian motion if there are parameters G_0, μ, and σ so that

$$G_t = G_0 e^{t\mu + \sigma B_t}$$

where B_t is Brownian motion.

(b) We get

$$\begin{aligned}
E(G_t) &= E(G_0 e^{t\mu+\sigma B_t}) \\
&= G_0 e^{t\mu} E(e^{\sigma B_t}) \\
&= G_0 e^{t\mu} \int_{-\infty}^{\infty} e^{\sigma s} \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{1}{2t}s^2\right) ds \\
&= G_0 e^{t\mu} \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2t}(s^2 - 2\sigma t s)\right) ds \\
&= G_0 e^{t\mu} \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2t}(s - \sigma t)^2 + t\frac{\sigma^2}{2}\right) ds \\
&= G_0 e^{t\mu} e^{t\sigma^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{1}{2t}(s - \sigma t)^2\right) ds \\
&= G_0 e^{t(\mu+\sigma^2/2)}
\end{aligned}$$