Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE550 Stochastic Processes and Bayesian Inference

Re-exam April 9, 2021, 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 information about some probability distributions.

1. (6 points) Assume the variable x has non-negative integers $\{0, 1, 2, \dots\}$ as possible values and a probability mass function

$$\pi(x \mid \theta) = \theta(1-\theta)^x$$

where θ is a parameter satisfying $0 < \theta < 1$.

   (a) Guess at a family of distributions for θ that might be a conjugate family, and prove that this family is conjugate. (Hint: Consider how we made inference for the parameter of the Binomial distribution).

   (b) Find an expression for the marginal mass function for x when $\theta \sim \text{Uniform}(0, 1)$.

   (c) Assume instead that the prior for θ is a discrete probability distribution on the set $1/n, 2/n, \dots, (n-1)/n$ for some n. Give an outline for how one can compute the posterior distribution for θ given several observations of x with the probability mass function above.

2. (4 points) Consider the discrete-time Markov chain with transition graph given in Figure 1. We assume the chain starts at c.

   (a) What is the expected number of steps before hitting d? (You will get full points if you write down in a precise way how to compute the result, using functions that can be run for example in R, but you are of course also allowed to do the computation).

   (b) Assume you would like to compute the expected number of steps before the chain produces the sequence abc. Construct and draw the transition graph for a Markov chain that can be used to compute this result. (You do not need to to the actual computation).

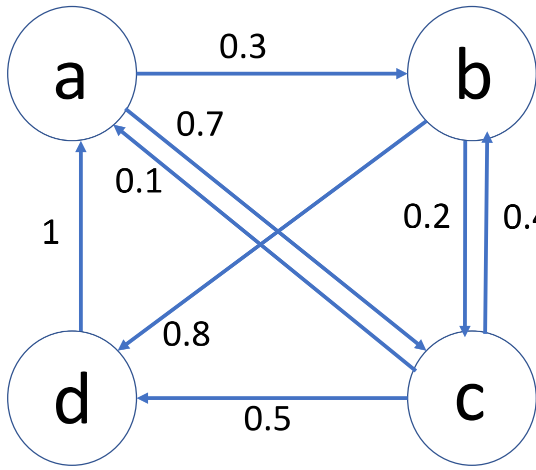

Figure 1: The graph for question 2.

3. (7 points) A Branching process has an offspring distribution X given by

$$P(X = k) = \begin{cases} a_0 & \text{if } k = 0 \\ 0 & \text{if } k = 1 \\ \frac{1-a_0}{2^{k-1}} & \text{if } k = 2, 3, \ldots \end{cases}$$

where a_0 is a parameter satisfying $0 < a_0 < 1$.

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

(b) For which values of a_0 is the branching process supercritical?

(c) Find the extinction probability in terms of a_0.

(d) Assume now that a_0 has a prior that is uniform on the interval $(0, 1)$. Assume the branching process given in Figure 2 has been observed. Find the posterior distribution for a_0.

4. (3 points) Alfons is running an antiques store. He divides his customers into three categories: A, B, and C. His experience is that these customers all arrive according to independent Poisson processes: Customers of type A arrive at a rate of 3 per hour, and customers of type B at a rate of 2 per hour. On average he has 7 customers per hour in total.

(a) What is the probability that at least 3 customers of type A and exactly 2 customers of type B arrive during the first two hours?

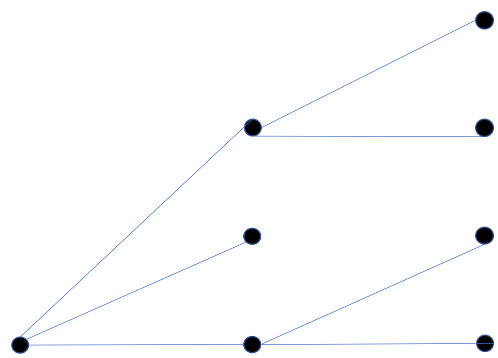

Figure 2: A picture of the three first generations of the branching process in question 4d: There is no information about possible offsprings from the third generation.

(b) Assume that exactly 9 customers arrive during the first 2 hours one day. Select one of these customers uniformly at random, and compute the probability that the customer has arrived within the first 3/4 of the first hour after the opening.

5. (4 points) A computer varies between 3 states, denoted as 1, 2, and 3. We model its satus at time t with a continuous-time Markov chain with these states. With the rate of change from state i to state j denoted by q_{ij}, we have $q_{12} = 2$, $q_{13} = 0.5$, $q_{21} = 0.3$ $q_{23} = 0.1$, $q_{31} = 1.5$, $q_{32} = 0$.

   For each of the questions below, you can get full points if you write down in a precise way how to compute the result, using functions that can be run for example in R, but you are of course also allowed to do the computation and report the result.

   (a) What is the long-term proportion of time the computer is in state 2?

   (b) If we ignore the lengths of stays in various states and only count the visits, what is the long-term proportion of visits to state 2?

6. (6 points) Let B_t denote Brownian motion.

   (a) Find the probability distribution of $aB_t + bB_{2t} + cB_{3t}$ where a, b, c are fixed constants.

   (b) Prove that $-B_t$ is Brownian motion.

   (c) Find the probability that $B_t = 1.4$ for exactly one t in the interval $0 < t < 1$.

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
Re-exam April 9 2021

1. (a) We try with the Beta family: Assume $\theta \sim \text{Beta}(\alpha,\beta)$. Then

$$\begin{aligned}\pi(\theta \mid x) &\propto_\theta \pi(x \mid \theta)\pi(\theta) \\ &\propto_\theta \theta(1-\theta)^x\theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &= \theta^\alpha(1-\theta)^{\beta+x-1}\end{aligned}$$

so $\theta \mid x \sim \text{Beta}(\alpha+1,\beta+x)$ and the Beta family is a conjugate family of priors.

(b) One way to compute is the following

$$\pi(x) = \int_0^1 \theta(1-\theta)^x\,d\theta = \frac{\Gamma(2)\Gamma(x+1)}{\Gamma(3+x)} = \frac{1}{(x+1)(x+2)}$$

where we have used the formula for the density of a $\text{Beta}(2, x+1)$ distribution to compute the integral. Another way is to compute

$$\begin{aligned}\pi(x) &= \frac{\pi(x \mid \theta)\pi(\theta)}{\pi(\theta \mid x)} = \frac{\theta(1-\theta)^x \cdot \text{Beta}(\theta; 1, 1)}{\text{Beta}(\theta; 2, 1+x)} \\ &= \frac{\theta(1-\theta)^x}{\frac{\Gamma(3+x)}{\Gamma(2)\Gamma(1+x)}\theta(1-\theta)^x} = \frac{\Gamma(2)\Gamma(1+x)}{\Gamma(3+x)} = \frac{1}{(x+1)(x+2)}.\end{aligned}$$

(c) Let p be the vector of length $n-1$ containing the prior, so that $p_i = \Pr[\theta = i/n]$. If we have observations $x_1, \dots, x_k$, define vectors $v_1, \dots, v_k$ by

$$v_{ji} = \frac{i}{n}\left(1 - \frac{i}{n}\right)^{x_j}$$

for $j = 1, \dots, k$, $i = 1, \dots, n-1$. Then compute the vector v with

$$v_i = p_i \cdot v_{1i} \cdot \dots \cdot v_{ki}$$

and normalize it so that it sums to 1: This probability vector is then the posterior for θ.

2. (a) The transition matrix, after changing d into an absorbing state, becomes

$$P' = \begin{bmatrix} 0 & 0.3 & 0.7 & 0 \\ 0 & 0 & 0.2 & 0.8 \\ 0.1 & 0.4 & 0 & 0.5 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} Q & R \\ 0 & 1 \end{bmatrix}$$

where

$$Q = \begin{bmatrix} 0 & 0.3 & 0.7 \\ 0 & 0 & 0.2 \\ 0.1 & 0.4 & 0 \end{bmatrix}.$$

The fundamental matrix is then $F = (I - Q)^{-1}$ and the expected number of steps before hitting d is the sum of the entries in the third row of this matrix. In R we can write

```
Q <- matrix(c(0, 0, 0.1, 0.3, 0, 0.4, 0.7, 0.2, 0), 3, 3)
F <- solve(diag(3)-Q)
print(sum(F[3,]))
```

which yields the numeric result 1.812796. It is also possible to use what Dobrow calls "first step analysis" to obtain the same result.

(b) We need a Markov chain which records not only the current state, but also how far we might have come in constructing the sequence *abc*. We can use the transition graph in Figure 1.

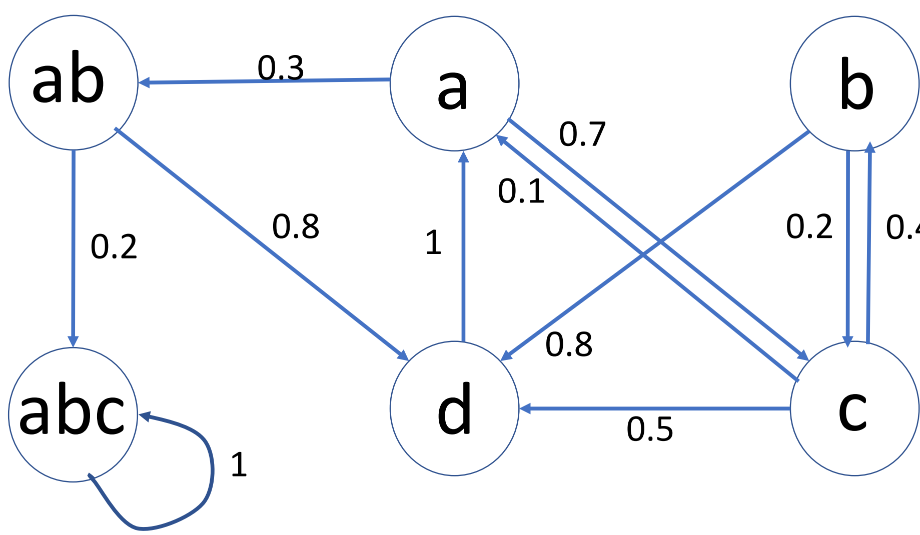

Figure 1: The graph for question 2

3. (a) We get

$$\begin{aligned} G(s) &= \mathrm{E}\left[s^X\right] = a_0 + (1-a_0)\sum_{k=2}^{\infty}\left(\frac{1}{2}\right)^{k-1} s^k \\ &= a_0 + (1-a_0)s\sum_{k=2}^{\infty}\left(\frac{s}{2}\right)^{k-1} \\ &= a_0 + (1-a_0)s\frac{s/2}{1-s/2} = a_0 + (1-a_0)\frac{s^2}{2-s} \end{aligned}$$

(b) To find the expectation we may differentiate $G(s)$:

$$G'(s) = (1-a_0)\frac{(2-s)2s + s^2}{(2-s)^2} = (1-a_0)\frac{4s - s^2}{(2-s)^2}$$

Thus $\mathrm{E}[X] = G'(1) = 3(1-a_0)$. The Branching process is supercritical if and only if $\mathrm{E}[X] > 1$, i.e., if

$$3(1-a_0) > 1$$

which gives $a_0 < \frac{2}{3}$.

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

$$a_0 + (1-a_0)\frac{s^2}{2-s} = s,$$

which yields the 2nd degree equation

$$s^2 - \frac{2+a_0}{2-a_0}s + \frac{2a_0}{2-a_0} = 0.$$

We know that $G(1) = 1$, so 1 is a root of this equation. Using that, we get the factorization

$$(s-1)\left(s - \frac{2a_0}{2-a_0}\right) = 0$$

and the smallest positive root, when $a_0 < \frac{2}{3}$, is $\frac{2a_0}{2-a_0}$. In summary: When $0 < a_0 < \frac{2}{3}$, the excinction probability is $\frac{2a_0}{2-a_0}$, while when $\frac{2}{3} \leq a_0 < 1$, the extinction probability is 1.

(d) There are 4 observations of the offspring distribution in Figure 2. In one of those there is no offspring, while in the other 3 there are 2 or more offspring. The likelihood for the first observation is a_0, while the likelihoods for the other three observations are proportional to $1 - a_0$ as a function of a_0. With a prior that is uniform on $(0, 1)$ we get that the posterior is proportional to

$$a_0^1(1-a_0)^3$$

Comparing with the Beta density, we see that

$$p_0 \mid \text{data} \sim \text{Beta}(2, 4).$$

4. (a) Let X_A and X_B be the number of customers of type A and B, respectively, during the first two hours. We get $X_A \sim \text{Poisson}(2 \cdot 3) = \text{Poisson}(6)$ and $X_B \sim \text{Poisson}(2 \cdot 2) = \text{Poisson}(4)$. The answer to the question becomes

$$\begin{aligned} & \Pr[X_A \geq 3]\Pr[X_B = 2] \\ = & \ (1 - \Pr[X_A = 0] - \Pr[X_A = 1] - \Pr[X_A = 2])\Pr[X_B = 2] \\ = & \ (1 - e^{-6}(1 + 6 + 6^2/2))e^{-4}4^2/2 = 0.1374451 \end{aligned}$$

This can also be computed in R with

```
(1-ppois(2, 6))*dpois(2, 4)
```

(b) Given that a fixed number of customers arrive, the arrival time of a randomly selected customer among these will be uniformly distributed. Thus the probability is $\frac{3/4}{2} = 0.375$.

5. (a) We get

$$Q = \begin{bmatrix} -2.5 & 2 & 0.5 \\ 0.3 & -0.4 & 0.1 \\ 1.5 & 0 & -1.5 \end{bmatrix}$$

for the generator matrix. To find the limiting distribution $v = (v_1, v_2, v_3)$ we need to solve the equations $vQ = 0$ and $v_1 + v_2 + v_3 = 1$. If we let Q' be the matrix Q with the last column replaced by 1's, we get that we need to solve the equation

$$vQ' = (0, 0, 1)$$

Possible R code is

```
Q <- matrix(c(-2.5, 0.3, 1.5, 2, -0.4, 0, 1, 1, 1), 3, 3)
print(c(0, 0, 1)%*%solve(Q))
```

yielding the numerical answer

$$(0.15, 0.75, 0.1)$$

Thus the answer to the original question is 0.75.

(b) We first find the transition matrix for the embedded chain:

$$\tilde{P} = \begin{bmatrix} 0 & 0.8 & 0.2 \\ 0.75 & 0 & 0.25 \\ 1 & 0 & 0 \end{bmatrix}.$$

In order to find the limiting distribution $w = (w_1, w_2, w_3)$ for the discrete-time Markov chain, we need to solve the equations $w_1 + w_2 + w_3 = 1$ and $w\tilde{P} = w$, or equivalently $w(\tilde{P} - I) = 0$. With similar computations as in (a), we get

```
Q <- matrix(c(-1, 0.75, 1, 0.8, -1, 0, 1, 1, 1), 3, 3)
print(c(0, 0, 1)%*%solve(Q))
```

yielding the numerical answer

$$(0.4545455, 0.3636364, 0.1818182)$$

Thus the answer to the original question is 0.3636364. Note that the result can also be found directly from (a) using the relationship between the limiting distributions of a continuous-time Markov chain and its embedded chain:

$$\psi_2 = \frac{\pi_2 q_2}{\pi_1 q_1 + \pi_2 q_2 + \pi_3 q_3} = \frac{0.75 \cdot 0.4}{0.15 \cdot 2.5 + 0.75 \cdot 0.4 + 0.1 \cdot 1.5} = \frac{4}{11} = 0.3636364.$$

6. (a) We get

$$\begin{aligned} & aB_t + bB_{2t} + cB_{3c} \\ = \ & aB_t + b(B_{2t} - B_t) + bB_t + c(B_{3t} - B_{2t}) + c(B_{2t} - B_t) + cB_t \\ = \ & (a+b+c)B_t + (b+c)(B_{2t} - B_t) + c(B_{3t} - B_{2t}) \end{aligned}$$

This is a sum of three independent normally distributed variables, and it has a normal distribution. We see directly that the expectation is zero, and for the variance we get

$$\begin{aligned} & \mathrm{Var}\left[(a+b+c)B_t + (b+c)(B_{2t} - B_t) + c(B_{3t} - B_{2t})\right] \\ = \ & (a+b+c)^2 t + (b+c)^2 t + c^2 t \\ = \ & \left((a+b+c)^2 + (b+c)^2 + c^2\right) t \end{aligned}$$

So

$$aB_t + bB_{2t} + cB_{3c} \sim \mathrm{Normal}\left(0, \left((a+b+c)^2 + (b+c)^2 + c^2\right) t\right)$$

(b) One may prove this directly from the definition: One must then prove each of the 5 defining properties of Brownian motion mentioned in Dobrow. Alternative one may first argue that $-B_t$ is a Gaussian process: As Brownian motion is a Gaussian process, any linear combination of variables from the process has a multivariate normal distribution, so this is also true for any linear combination of variables from the process $-B_t$, so $-B_t$ is a Gaussian process. It also satisfies $-B_0 = 0$, $\mathrm{E}[-B_t] = 0$, and $\mathrm{Cov}[-B_s, -B_t] = \mathrm{Cov}[B_s, B_t] = \min\{s, t\}$. Finally, $t \mapsto -B_t$ is clearly a continuous map. By a theorem in Dobrow, $-B_t$ is Brownian motion.

(c) If there is exactly one such t, that implies that there is at least one such t, which implies that $T_{1.4} < 1$, where $T_{1.4}$ is the first hitting time for 1.4. As the first hitting time is a stopping time, we get that $B_{T_{1.4}+t} - B_{T_{1.4}}$ is brownian motion. We know that the probability that this process has a zero in the interval $(0, \epsilon)$ is 1, for any ϵ. Thus we can find another t, with $t < 1$, where the original Brownian motion will be equal to 1.4. In fact, with probability 1, there will be infinitely many t with $t < 1$ where $B_t = 1.4$, as long as we assume there is at least one such t. But this means that the probability that there is exactly one such t is zero.