Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE550 Stochastic Processes and Bayesian Inference

Exam January 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 relevant information about some probability distributions.

1. (4 points) Assume a variable $x > 0$ has density

$$\pi(x \mid \theta) = \frac{\theta^2 e^{-\theta/x}}{x^3}$$

where $\theta > 0$ is a parameter.

   (a) Write down a proof that the Gamma family of densities is a conjugate family to the likelihood above.

   (b) Assuming $\theta \sim \text{Gamma}(\alpha, \beta)$ and that $x \mid \theta$ has the distribution above, compute and simplify the marginal density for x.

2. (7 points) A Markov chain is defined as a random walk on the weighted undirected graph displayed in Figure 1. Note that the nodes are called A, B, C and the weights are $w_1, \dots, w_6$ where these are positive numbers.

   (a) Given specific values for $w_1, \dots, w_6$, what is the limiting distribution for the Markov chain?

   (b) Assume the chain has been observed for 28 steps, and that the table below lists counts of observed transitions from the node given on the left column to the node given on the top row.

	A	B	C
A	2	5	2
B	3	1	5
C	3	4	2

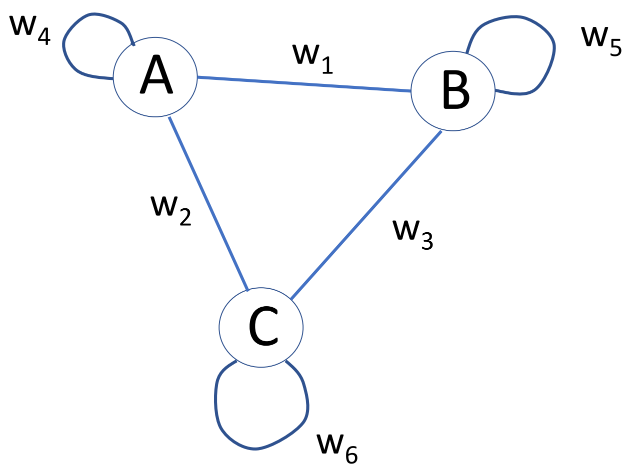


Figure 1: The graph for question 2.

Assume we use a prior for the weights with density $\pi(w_1, \ldots, w_6) = \exp(-w_1 - \cdots - w_6)$. Write down and simplify a function proportional to the posterior density for the weights $w_1, \ldots, w_6$.

(c) Describe in detail an algorithm that computes the (approximate) expected posterior limiting probability for the chain to be in state A. You may use R code or pseudo-code to give a precise description of your algorithm. You don't need to run the algorithm.

(d) In the situation above, we could have assumed that the Markov chain was represented by a stochastic matrix P and used Dirichlet priors for the rows of P. What, if any, would be the difference for the interpretation of the result? [1]

3. (6 points) Consider the Markov chain with states space $\{1, 2, \ldots, n\}$ and transition graph given in Figure 2, where p is a parameter satisfying $0 \leq p \leq 1$.

   (a) For each possible value of p determine the number of communication classes.

   (b) For each possible value of p and each state, determine its period.

   (c) For each possible value of p compute all possible stationary distributions for the chain, if any exist.

   (d) For each possible value of p compute all possible limiting distributions for the chain, if any exist.

[1] A better formulation of this question, unfortunately not used in the actual exam, would have been "What, if any, would be the difference in the posterior model if the amount of data approached infinity?"

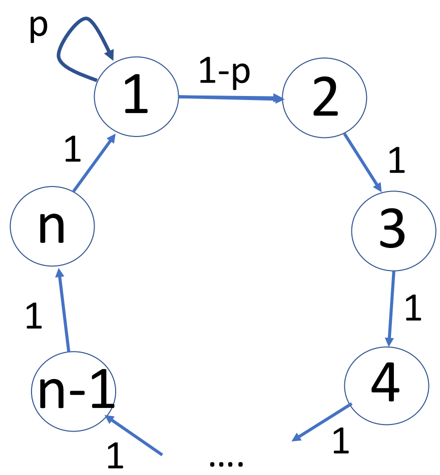


Figure 2: The transition graph for question 3

4. (4 points)

   (a) Assume the offspring process in a Branching process has probability 1/4 for zero offspring, probabiliy 1/2 for 1 offspring, and probability 1/4 for 3 offspring. Calculate the probability that the process will eventually go extinct.

   (b) Assume another Branching process uses as offspring process in the first generation a Poisson distribution with parameter λ. After this, the offspring distribution of (a) is used. Compute the probability that this branching process will go extinct.

5. (5 points) Adam is the main salesperson in a store that sells candy. Customers arrive according to independent Poisson processes. Adult customers arrive on average with one customer every 3 minutes while on average one child arrives every minute. The time it takes to service a customer is exponentially distributed. For adult customers it takes on average 2 minutes, while for child customers it takes on average 1 minute. If Adam is busy with a customer when another customer arrives, that customer moves on to another salesperson.

   (a) Compute the long time average proportion of time Adam serves adult customers.

   (b) Write down the transition rate graph for the process above, and also the graph with transition probabilities for a Poisson subordinated process to the process above.

   (c) Based on the above, give a short proof that the continuous-time Markov process you derived above is time reversible[2].

[2] A better formulation of the question would have been "give a proof that is as short as possible"

6. (2 points) Assume N_t is a Poisson process with parameter $\lambda = 2$. Prove that $N_t - 2t$ is a martingale with respect to N_t.

7. (2 points) Prove that the stochastic process $(X_t)_{0 \leq t \leq 1}$ defined by conditioning Brownian motion on $B_1 = a$ for some real a is identical to the process $Y_t = B_t - tB_1 + ta$ for $0 \leq t \leq 1$.

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
Exam January 9 2021

1. (a) Assuming that $\theta \sim \text{Gamma}(\alpha, \beta)$, we get

$$\begin{aligned}
\pi(\theta \mid x) &\propto_\theta \pi(x \mid \theta)\pi(\theta) \\
&= \frac{\theta^2 e^{-\theta/x}}{x^3} \cdot \text{Gamma}(\theta; \alpha, \beta) \\
&\propto_\theta \theta^2 e^{-\theta/x} \theta^{\alpha-1} e^{-\beta\theta} \\
&= \theta^{\alpha+2-1} e^{-(\beta+1/x)\theta} \\
&\propto_\theta \text{Gamma}\left(\theta; \alpha+2, \beta+\frac{1}{x}\right)
\end{aligned}$$

So if the prior is any Gamma density then the posterior is also a Gamma density. This proves conjugacy for the Gamma family.

(b) We may compute

$$\begin{aligned}
\pi(x) &= \frac{\pi(x \mid \theta)\pi(\theta)}{\pi(\theta \mid x)} \\
&= \frac{\frac{\theta^2 e^{-\theta/x}}{x^3} \cdot \text{Gamma}(\theta; \alpha, \beta)}{\text{Gamma}\left(\theta; \alpha+2, \beta+\frac{1}{x}\right)} \\
&= \frac{\frac{\theta^2 e^{-\theta/x}}{x^3} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\beta\theta)}{\frac{(\beta+1/x)^{\alpha+2}}{\Gamma(\alpha+2)} \theta^{\alpha+2-1} \exp(-(\beta+1/x)\theta)} \\
&= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \cdot \frac{\beta^\alpha}{(\beta+1/x)^{\alpha+2}} \cdot \frac{1}{x^3} \\
&= \frac{\alpha(\alpha+1)\beta^\alpha}{(\beta+1/x)^{\alpha+2} x^3}
\end{aligned}$$

2. (a) Using the theory for undirected weighted graphs, the limiting distribution for the states A, B, C is

$$\left(\frac{w_1+w_2+w_4}{W}, \frac{w_1+w_3+w_5}{W}, \frac{w_2+w_3+w_6}{W}\right)$$

where

$$W = 2(w_1+w_2+w_3) + w_4 + w_5 + w_6.$$

(b) We get

$$\begin{aligned}
& \pi(w_1,\dots,w_6 \mid \text{data}) \\
\propto_{w_1,\dots,w_6} \quad & \pi(\text{data} \mid w_1,\dots,w_6)\pi(w_1,\dots,w_6) \\
\propto_{w_1,\dots,w_6} \quad & \frac{w_4^2 w_1^5 w_2^2}{(w_1+w_2+w_4)^9}\cdot\frac{w_1^3 w_5^1 w_3^5}{(w_1+w_5+w_3)^9}\cdot\frac{w_2^3 w_3^4 w_6^2}{(w_2+w_3+w_6)^9}\cdot\exp\left(-w_1-\cdots-w_6\right) \\
= \quad & \frac{w_1^8 w_2^5 w_3^9 w_4^2 w_5^1 w_6^2}{(w_1+w_2+w_4)^9(w_1+w_5+w_3)^9(w_2+w_3+w_6)^9}\exp(-w_1-\cdots-w_6)
\end{aligned}$$

(c) The idea would be to simulate a sample from the posterior for the vector of weights $(w_1, w_2, \dots, w_6)$, and then take the average of $\frac{w_1+w_2+w_4}{W}$ over this sample. There are many ways to generate such a sample. Below is a basic example:

```
post <- function(w) { w[1]^8*w[2]^5*w[3]^9*w[4]^2*w[5]*w[6]^2/
    (w[1]+w[2]+w[4])^9/(w[1]+w[5]+w[3])^9/(w[2]+w[3]+w[6])^9*
    exp(-sum(w))
}
N <- 10000
result <- rep(0, N-1)
w <- wprop <- rep(1, 6)
for (i in 2:N) {
  wprop <- abs(w + rnorm(6, 0, 0.1))
  if (runif(1) < post(wprop)/post(w)) w <- wprop
  result[i-1] <- (w[1]+w[2]+w[4])/(sum(w)+w[1]+w[2]+w[3])
}
print(mean(result))
```

Many improvements could be made to the algorithm above to improve its accuracy. For example, one should transform so that one simulated the variables $u_i = \log(w_i)$ instead of the variables w_i, and one should compute the logarithm of the posterior density instead of the density itself. One should also remove burn-in.

The most important point is that the formula from (a), for the long-term probability for state A, should be computed for the simulated vector of weights in each step, and the average should be computed afterwards.

(d) The assumption that the Markov chain is represented as a random walk on a weighted graph is equivalent to the assumption that the Markov chain is time reversible. In the alternative model, no such assumption would be made. The difference between the priors would make also make a difference, but this difference would diminish as the amount of data increased. The remaining difference would be that the Markov chain using the model of this task would be time-reversible, while in the alternative model it would not.

3. (a) When $p = 1$ there are n communication classes, one for each state. When $p < 1$, there is a single communication class.

(b) When $p = 1$, the chain does not return to states $i > 1$, so these have period ∞, while state 1 has period 1. When $0 < p < 1$ the states are all aperiodic. When $p = 0$ all states have period n.

(c) When $p = 1$ the chain is absorbed in the state 1, so the distribution $(1, 0, \dots, 0)$ is a stationary distribution, and there can be no other.

When $p < 1$ the chain is irreducible, so there exists a single stationary distribution. It can be found as the probability vector v satisfying $vP = v$ where

$$P = \begin{bmatrix} p & 1-p & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

We get

$$v_1 p + v_n = v_1$$
$$v_1(1-p) = v_2$$

and

$$v_2 = v_3 = \cdots = v_n.$$

Together with $v_1 + v_2 + \cdots + v_n = 1$ this yields

$$v = \frac{1}{p + n - pn}(1, 1-p, \dots, 1-p)$$

as the unique stationary distribution.

(d) When $p = 1$ the chain is absorbed in the state 1 so the limiting distribution is clearly $(1, 0, \dots, 0)$. When $0 < p < 1$ the Markov chain is irreducible and aperiodic, so it has a unique limiting distribution that is identical to the stationary distribution found above. When $p = 0$ the Markov chain is periodic, and thus does not have a limiting distribution.

4. (a) We need to find the smallest positive root of $G(s) = s$ where $G(s) = \frac{1}{4} + \frac{1}{2}s + \frac{1}{4}s^3$ is the probability generating function. We get the equation

$$4s = 1 + 2s + s^3$$

or $1 - 2s + s^3 = 0$. Using that this equation has a root $s = 1$ (as we know that $G(1) = 1$) we can factorize

$$1-2s+s^3 = (s-1)(s^2+s-1) = (s-1)\left(\left(s + \frac{1}{2}\right)^2 - \frac{5}{4}\right) = (s-1)\left(s + \frac{1}{2} + \frac{\sqrt{5}}{2}\right)\left(s + \frac{1}{2} - \frac{\sqrt{5}}{2}\right).$$

Thus the smallest positive root, and the extinction probability, is $c = -\frac{1}{2} + \frac{\sqrt{5}}{2} = 0.618034$.

(b) Conditioning on the size of the first generation and using the value c computed above, we get

$$\begin{aligned}
\Pr[\text{extinction}] &= \mathrm{E}\left[\mathrm{E}\left[\text{extinction} \mid Z_1\right]\right] = \mathrm{E}\left[c^{Z_1}\right] = \sum_{k=0}^{\infty} c^k e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(c\lambda)^k}{k!} = e^{-\lambda} e^{c\lambda} = \exp(-0.381966\lambda).
\end{aligned}$$

5. (a) We can model the situation with a continuous time Markov chain with three states: O (Adam has no customers), A (Adam has an adult customer), and C (Adam has a child customer). The generating matrix becomes

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

and the equation $vQ = 0$ yields the two equations $v_1/3 - v_2/2 = 0$ and $v_1 - v_3 = 0$. Together with the equation $v_1 + v_2 + v_3 = 1$ we easily get the solution

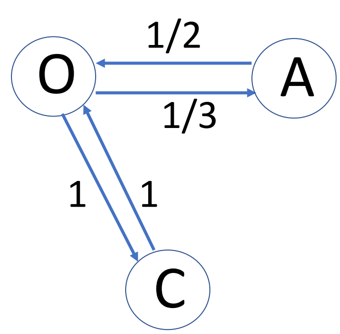
$$v = \left(\frac{3}{8}, \frac{1}{4}, \frac{3}{8}\right)$$

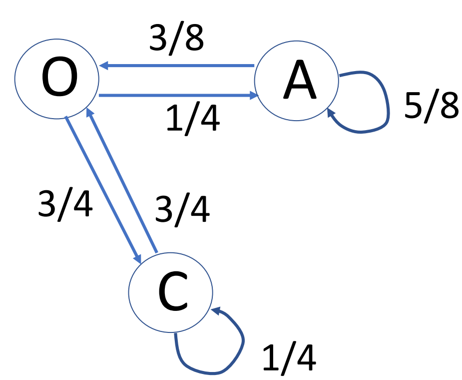
and the answer to the question is a quarter of the time.

(b) To make a Poisson subordination, we choose $\lambda = 4/3$, which yields

$$R = \frac{1}{\lambda} Q + I = \frac{3}{4} \begin{bmatrix} -\frac{4}{3} & \frac{1}{3} & 1 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ 1 & 0 & -1 \end{bmatrix} + I = \begin{bmatrix} -1 & \frac{1}{4} & \frac{3}{4} \\ \frac{3}{8} & -\frac{3}{8} & 0 \\ \frac{3}{4} & 0 & -\frac{3}{4} \end{bmatrix} + I = \begin{bmatrix} 0 & \frac{1}{4} & \frac{3}{4} \\ \frac{3}{8} & \frac{5}{8} & 0 \\ \frac{3}{4} & 0 & \frac{1}{4} \end{bmatrix}.$$

The graphs become





(c) As the transition rate graph is a tree, it automatically follows that the Markov process is time reversible.

6. We have

$$\begin{aligned}
& \mathrm{E}\left[N_t - 2t \mid N_r, 0 \le r \le s\right] \\
= \; & \mathrm{E}\left[N_s + N_t - N_s \mid N_r, 0 \le r \le s\right] - 2t \\
= \; & \mathrm{E}\left[N_s \mid N_r, 0 \le r \le s\right] + \mathrm{E}\left[N_t - N_s\right] - 2t \\
= \; & N_s + \mathrm{E}\left[N_{t-s}\right] - 2t \\
= \; & N_s + 2(t-s) - 2t \\
= \; & N_s - 2s
\end{aligned}$$

Further,

$$\mathrm{E}\left[|N_t - 2t|\right] \le \mathrm{E}\left[|N_t|\right] + 2t = 2t + 2t < \infty.$$

7. We have

$$X_t \sim B_t \mid (B_1 = a) \sim B_t - tB_1 + tB_1 \mid (B_1 = a) \sim B_t - tB_1 + ta \mid (B_1 = a)$$

Now $B_t - tB_1$ is a Brownian bridge, and according to Dobrow it is independent of the value of B_1. Thus

$$B_t - tB_1 + ta \mid (B_1 = a) \sim B_t - tB_1 + ta \sim Y_t.$$

Alterantively, one may observe that the processes X_t and Y_t are Gaussian processes so it is enough to prove that they have the same expectation and covariance functions to prove that they are identical. This can be done with direct computation, using similar computations as those in Dobrow when proving the statement above for $a = 0$. We get

$$\mathrm{E}\left[X_t\right] = at = \mathrm{E}\left[Y_t\right]$$

and when $s \le t$

$$\mathrm{Cov}\left[X_s, X_t\right] = s - st = \mathrm{Cov}\left[Y_s, Y_t\right].$$