Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE550 Stochastic Processes and Bayesian Inference

Exam January 8, 2022, 8:30 - 12:30
Examiner: Petter Mostad, 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) A Branching process $Z_0, Z_1, \dots$ has an offspring process with expectation μ and variance σ^2.
   (a) Compute $\mathrm{E}(Z_n)$.
   (b) Compute $\mathrm{Var}(Z_n)$ in terms of μ, σ^2, and $\mathrm{Var}(Z_{n-1})$.
   (c) Compute $\mathrm{Var}(Z_n)$ for $n = 0, 1, 2, 3, 4$. Guess at a general formula expressing $\mathrm{Var}(Z_n)$ in terms of μ, σ^2, and n, and prove the formula by induction.

2. (6 points) Alex is trying to model the inflow of customers to his shop, which is open daily 10:00 - 18:00. He has a data file where he has recorded the arrival time of all customers during the last 5 days. During these days, there have been a total of 23 customers arriving before 14:00 and 44 after 14:00.
   (a) Initially, Alex assumes his customers arrive according to a Poisson process with parameter λ customers per hour, using a prior $\pi(\lambda) \propto_\lambda 1/\lambda$. Find the posterior probability that $\lambda > 1.2$. You may express your answer in terms of a integral, or in terms of a suitable R function (but make sure to simplify your answer).
   (b) Alex observes he often has more customers in the afternoon, so he now wants to use an inhomogenenous Poisson process where the rate of customer arrivals is λ before 14:00 and $\lambda\mu$ after 14:00, with μ an extra parameter with prior $\pi(\mu) \propto_\mu 1/\mu$. Find an expression proportional to the posterior density for pairs (μ, λ).
   (c) Alex gets more ambitions and wants to use a model where the rate of customer arrivals is $\lambda, \lambda\mu_1, \lambda\mu_2, \dots, \lambda\mu_7$ for each of his 8 opening hours, respectively. Name and give a brief outline of an algorithm with which Alex can obtain an approximate sample from the posterior for his parameters $\theta = (\lambda, \mu_1, \mu_2, \dots, \mu_7)$.

3. (6 points) Chess games are played on an 8×8 board of black and white squares as shown in Figure 1. A move of the knight piece consists of two steps in some direction and then one step to the side, as illustrated in Figure 1. Some experimentation may convince you that for any two squares there is a sequence of moves that may bring a knight from one

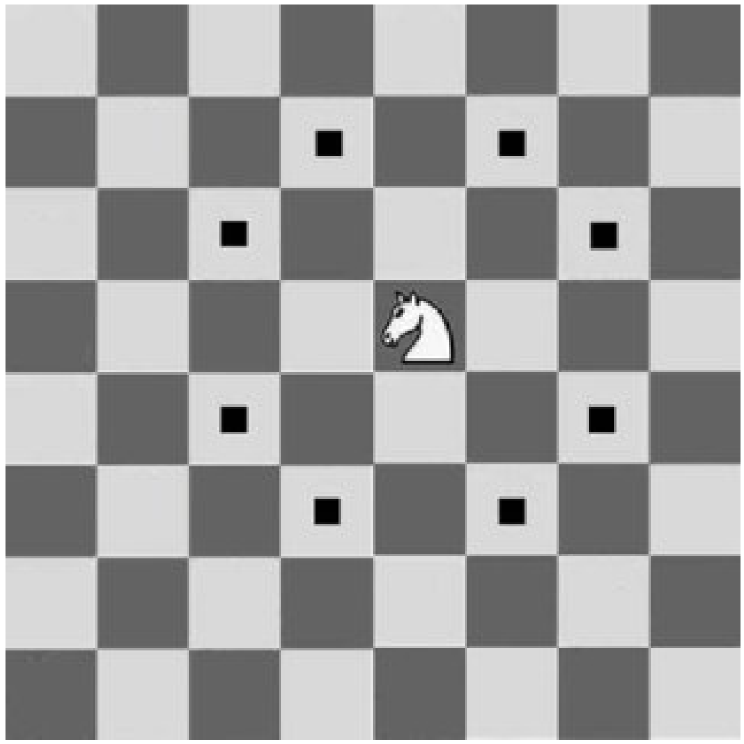

Figure 1: Illustration of the 8×8 chess board consisting of alternating black and white squares, and the possible moves for a knight.

square to the other. Assume that a knight starts in the lower left corner of the board and that at every time step it moves randomly, with equal probability, to each of the squares it may move to.

(a) Is this a Markov chain? If so, is it an ergodic Markov chain, and does it have a limiting distribution? Explain (prove) all parts of your answer.

(b) Assume we change the rules for how the knight can move by looping the board in all directions, so that instead of being limited by the edge of the board the knight can simply jump to the other side of the board. For example, possible moves from the lower left corner are illustraded in Figure 2. With these changed rules, what is the expected number of moves until the knight returns to its starting point at the lower left corner?

(c) Going back to the original rules for moving the knight, compute the expected number of moves until the knight returns to its starting point at the lower left corner. Note: If you (to save time) just describe how to make such a computation, instead of doing the whole computation, you will only lose one point.

4. (8 points) The rate graph for a continuous-time Markov chain is given in Figure 3. Your answers to the questions below should consist of numbers or expressions which might

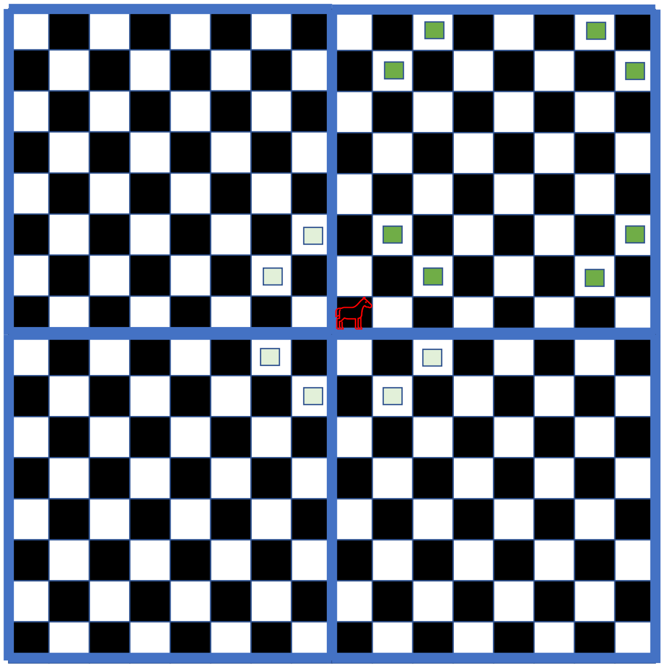

Figure 2: The extended moves for a knight positioned at the bottom left corner of the board: The dark greed squares indicate the 8 possible positions the knight can move to. The light green squares illustrate how these positions appear when the board is "looped" around its edges.

include matrices of numbers, matrix multiplication, matrix inversion, and similar.

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

(b) Prove or disprove that the process is time reversible.

(c) What is the long-term proportion of time that the process will spend in state C?

(d) Assuming that the process is in state C, what is the expected time until the first time it hits either state A or state F?

(e) Assuming that the process is in state C, what is the expected *number of new visits* to states other than A it will make until it hits state A?

5. (4 points)

(a) Define a Brownian bridge.

(b) Describe briefly how one may simulate a realization of a Brownian bridge.

(c) If X_t is a Brownian bridge and $0 < s < r < 1$, compute $\mathrm{cov}(X_s, X_r)$.

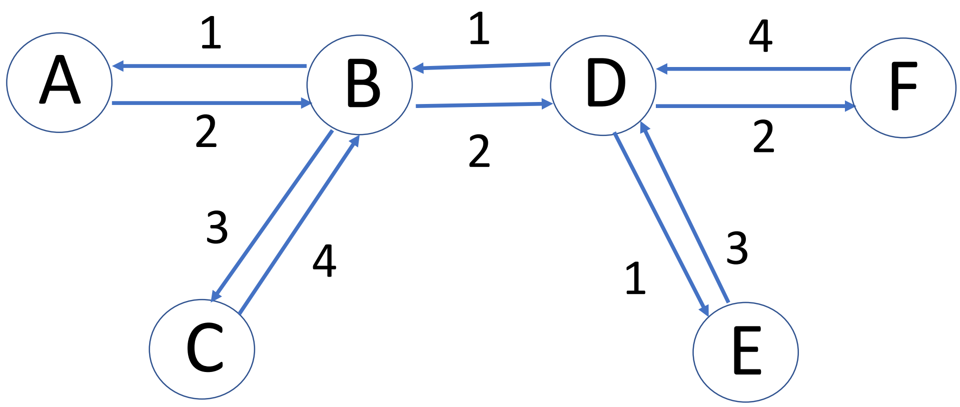

Figure 3: The rate graph for the Markov chain of question 4.

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 $x \in \{0, 1, 2, \dots, n\}$ 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 $x \in \{0, 1, 2, \dots, n\}$ 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 $x = (x_1, x_2, \dots, x_n)$ has a Dirichlet distribution, with $x_i \geq 0$ and $\sum_{i=1}^n x_i = 1$ and with parameters $\alpha = (\alpha_1, \dots, \alpha_n)$ with $\alpha_1 > 0, \dots, \alpha_n > 0$, 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 $x \in \{1, 2, 3, \dots\}$ 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 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 $x \in \{0, 1, 2, \dots\}$ 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 8 2022

1. (a) We get

$$\begin{aligned}
\mathrm{E}[Z_n] &= \mathrm{E}\left[\sum_{i=1}^{Z_{n-1}} X_i\right] = \mathrm{E}\left[\mathrm{E}\left[\sum_{i=1}^{k} X_i \mid Z_{n-1}=k\right]\right] = \mathrm{E}\left[\sum_{i=1}^{k} \mathrm{E}[X_i] \mid Z_{n-1}=k\right] \\
&= \mathrm{E}[k\mu \mid Z_{n-1}=k] = \mathrm{E}[\mu Z_{n-1}] = \mu\,\mathrm{E}[Z_{n-1}].
\end{aligned}$$

Application of recursion and the fact that $\mathrm{E}[Z_0] = \mathrm{E}[1] = 1$ gives

$$\mathrm{E}[Z_n] = \mu^n.$$

(b) We get

$$\begin{aligned}
\mathrm{Var}[Z_n] &= \mathrm{Var}\left[\sum_{i=1}^{Z_{n-1}} X_i\right] \\
&= \mathrm{Var}\left[\mathrm{E}\left[\sum_{i=1}^{k} X_i \mid Z_{i-1}=k\right]\right] + \mathrm{E}\left[\mathrm{Var}\left[\sum_{i=1}^{k} X_i \mid Z_{n-1}=k\right]\right] \\
&= \mathrm{Var}[\mu Z_{n-1}] + \mathrm{E}\left[\sigma^2 Z_{n-1}\right] = \mu^2\,\mathrm{Var}[Z_{n-1}] + \mu^{n-1}\sigma^2.
\end{aligned}$$

(c) We get directly $\mathrm{Var}[Z_0] = 0$ and $\mathrm{Var}[Z_1] = \mathrm{Var}[X_1] = \sigma^2$ Using the result from (b) repeatedly we get

$$\begin{aligned}
\mathrm{Var}[Z_2] &= \mu^2\sigma^2 + \mu\sigma^2 \\
\mathrm{Var}[Z_3] &= \mu^4\sigma^2 + \mu^3\sigma^2 + \mu^2\sigma^2 \\
\mathrm{Var}[Z_4] &= \mu^6\sigma^2 + \mu^5\sigma^2 + \mu^4\sigma^2 + \mu^3\sigma^2
\end{aligned}$$

We hypothesize that for $n \geq 1$

$$\mathrm{Var}[Z_n] = \sigma^2 \sum_{i=n-1}^{2n-2} \mu^i$$

and prove the formula by induction: First, it is true for $n = 1$, and secondly, assuming it is true for $n-1$ we get

$$\begin{aligned}
\text{Var}\,[Z_n] &= \mu^2\,\text{Var}\,[Z_{n-1}] + \mu^{n-1}\sigma^2 \\
&= \mu^2\sigma^2 \sum_{i=n-2}^{2(n-1)-2} \mu^i + \mu^{n-1}\sigma^2 \\
&= \sigma^2 \left(\sum_{i=n}^{2n-2} \mu^i + \mu^2 \right) \\
&= \sigma^2 \sum_{i=n-1}^{2n-2} \mu^i
\end{aligned}$$

so the proof is complete. Note that we can write the result as

$$\text{Var}\,[Z_n] = \sigma^2 \sum_{i=n-1}^{2n-2} \mu^i = \sigma^2\mu^{n-1} \sum_{i=0}^{n-1} \mu^i = \sigma^2\mu^{n-1}\frac{\mu^n - 1}{\mu - 1}.$$

2. (a) There have been a total of $23 + 44 = 67$ customers during the $5 \cdot 8 = 40$ hours of observation. Thus the likelihood function is Poisson$(67; 40\lambda)$. We get

$$\begin{aligned}
\pi(\lambda \mid \text{data}) &\propto_\lambda \pi(\text{data} \mid \lambda)\pi(\lambda) \\
&\propto_\lambda e^{-40\lambda}\frac{(40\lambda)^{67}}{67!} \cdot \frac{1}{\lambda} \\
&\propto_\lambda e^{-40\lambda}\lambda^{67-1}
\end{aligned}$$

so that

$$\pi(\lambda \mid \text{data}) = \text{Gamma}(\lambda; 67, 40).$$

The probability p asked for can be expressed as an integal as

$$p = \int_{1.2}^{\infty} \frac{40^{67}}{67!}\lambda^{67-1}\exp(-40\lambda)\,d\lambda$$

or in R

```
1 - pgamma(1.2, 67, 40)
```

(b) We get for the posterior

$$\begin{aligned}
\pi(\lambda, \mu \mid \text{data}) &\propto_{\lambda,\mu} \pi(\text{data} \mid \lambda, \mu)\pi(\lambda, \mu) \\
&\propto_{\lambda,\mu} \text{Poisson}(23; 20\lambda) \cdot \text{Poisson}(44, 20\lambda\mu) \cdot \frac{1}{\lambda} \cdot \frac{1}{\mu} \\
&\propto_{\lambda,\mu} e^{-20\lambda}(20\lambda)^{23}e^{-20\lambda\mu}(20\lambda\mu)^{44}\frac{1}{\lambda\mu} \\
&= \exp(-20\lambda(1+\mu))\lambda^{67-1}\mu^{44-1}
\end{aligned}$$

(c) Alex might use a Metropolis Hastings algorithm to obtain such an posterior. The algorithm would start with reasonable values for the parameters (for example the value 1) and use a proposal density $q(\theta^* \mid \theta)$ in each iteration. Generally, the algorithm would iterate between making a proposed new density according to $q(\theta^* \mid \theta)$ and accepting it with probability

$$a = \min\left(1, \frac{\pi(\theta^* \mid \text{data})q(\theta \mid \theta^*)}{\pi(\theta \mid \text{data})q(\theta^* \mid \theta)}\right).$$

If θ^* is not accepted, the old value θ would be repeated.

For the posterior $\pi(\theta \mid \text{data})$ we get (writing $\mu_0 = 1$, assuming the counts of the different hours are $c_1, c_2, \dots, c_8$, respectively, and using the priors $\mu_i \propto_{\mu_i} 1/\mu_i$)

$$\begin{aligned}
\pi(\theta \mid \text{data}) &\propto_\theta \prod_{i=1}^{8} \text{Poisson}(c_i; 5\lambda\mu_{i-1}) \cdot \frac{1}{\lambda\mu_1 \dots \mu_7} \\
&\propto_\theta \prod_{i=1}^{8} e^{-5\lambda\mu_{i-1}}(5\lambda\mu_{i-1})^{c_i} \cdot \frac{1}{\lambda\mu_1 \dots, \mu_7} \\
&\propto_\theta \lambda^{67-1}\mu_1^{c_2-1}\mu_2^{c_3-1} \dots \mu_7^{c_8-1} \exp(-5\lambda(1 + \mu_1 + \dots + \mu_7))
\end{aligned}$$

An alternative would be to use Gibbs sampling, in which case one would cycle through simulating from the conditional distribution of each of the parameters given fixed values for the others. From the expression of the posterior above we see that these conditional distributions would all be Gamma distributions.

3. (a) This will be a Markov chain, as the position at each time step only depends on the position at the previous time step. However, this Markov chain is not ergodic: In fact it is periodic, of period 2, as the knight will alternate between black and white squares. Because of the periodicity, there is also no limiting distribution.

(b) The Markov chain may be viewed as a random walk on a graph: The graph would consist of all the 64 squares in the board game, and each square has a degree 8 because of the extended way we allow the knight to move. Given the comment in the question about getting from any square to any other, the Markov chain is irreducible. Thus there is a unique stationary distribution. Because all the 64 states have degree 8, the stationary distribution is uniform. The long-term proportion of steps spent at the starting square is 1/64, and the expected return time to the starting square becomes 64.

(c) It is still possible to look at this as an irreducible random walk on a graph, but now the states do not all have degree 8. To do computations, one needs to find the degree of each of the 64 states. If d denotes the sum of all the degrees, we know that the long-term proportion of steps spent at the start square is $\frac{2}{d}$, as the start square has degree 2. Thus the expected number of steps to return to the start square becomes $d/2$.

In fact, there are 4 squares with degree 2, 3 with degree 3, 20 with degree 4, 16 with degree 6, and 16 with degree 8. This means that $d = 336$ and that the expected number of steps to return to the start square is 168.

4. (a) We get

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

(b) As the rate graph is a tree, the process is time reversible.

(c) It is clear that the process is irreducible and ergodic. To find the limiting distribution we find the v with positive values summing to 1 such that $vQ = 0$. We can do this by replacing the first column of Q with ones, producing Q', and then requiring that vQ' should be the vector $(1, 0, 0, 0, 0, 0)$. Finally, to find the long term proportion of time that the process will spend in C, we take the third element of v. In matrix terms we need to compute

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

(d) To find this expected time we make states A and F absorbing states. Removing rows and columns connected to these states we are left with a matrix

$$Q_0 = \begin{bmatrix} -6 & 3 & 2 & 0 \\ 4 & -4 & 0 & 0 \\ 1 & 0 & -4 & 1 \\ 0 & 0 & 3 & -3 \end{bmatrix}.$$

The matrix F of expected times spent in each state before absorbtion is given by $F = -Q^{-1}$. The answer is given by the sum of the second line (corresponding to state C) of this matrix, so we must compute

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

(e) To answer this question, we first find the transition matrix of the embedded discrete-time Markov chain, which becomes

$$\tilde{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1/6 & 0 & 1/2 & 1/3 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 & 1/4 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

The fundamental matrix when making A into an absorbing state will be $F = (I - \tilde{P}_{-A})^{-1}$, where $\tilde{P}_{-A}$ is $\tilde{P}$ with the row and columnd representing A removed. Our desired answer is the sum of the second row of this matrix, i.e, the answer is

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

5. (a) A Brownian bridge is Brownian motion on the interval $[0, 1]$ conditional on $B_1 = 1$.

(b) One may simulate Brownian motion as usual and then for each t subtract tB_1 from the simulated values.

(c) Using that we can write $X_t = B_t - tB_1$ we get

$$\begin{aligned} \operatorname{Cov}[X_s, X_r] &= \operatorname{Cov}[B_s - xB_1, B_r - rB_1] \\ &= \operatorname{Cov}[B_s, B_r] - s\operatorname{Cov}[B_1, B_r] - r\operatorname{Cov}[B_s, B_1] + sr\operatorname{Cov}[B_1, B_2] \\ &= s - sr - rs + sr\operatorname{Var}[B_1] \\ &= s - sr \end{aligned}$$