Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE550 Stochastic Processes and Bayesian Inference

Exam April 5, 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.

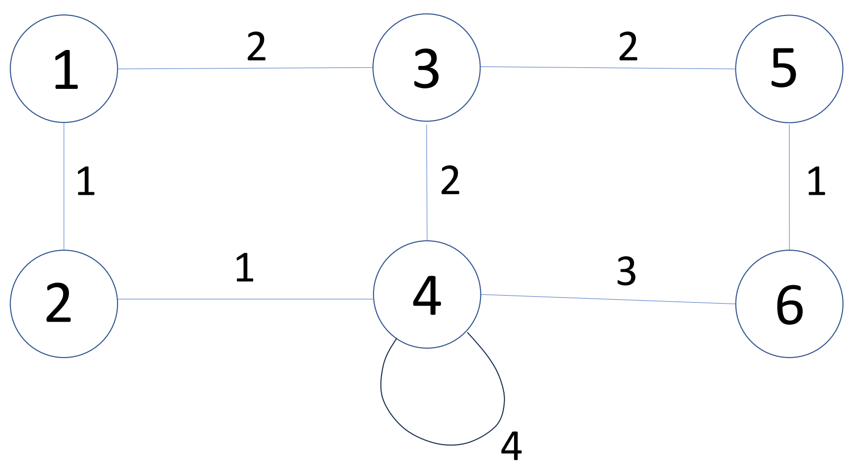

Figure 1: The undirected weighted graph used in question 1.

1. (8 points) Consider the Markov chain defined as the random walk on the weighted undirected graph of Figure 1.
   (a) Draw the transition graph for the Markov chain, and write down its transition matrix P.
   (b) Which of the states are recurrent, and which are transient?
   (c) Does the Markov chain have a limiting distribution? If so, compute it.
   (d) Prove or disprove that the Markov chain is time reversible.
   (e) If the chain starts in node 1, what is the probability that it visits node 5 before it visits node 6? You may express your result as an equation involving numerical matrices and matrix algebra.

2. (9 points) A process with repeating events is modelled with a Poisson process with intensity parameter λ events per minute. The process is monitored over 10 minutes, and the waiting time in minutes for the first event, between the first and the second event, between the second and third event, and between the third and fourth event, is observed as

$$x_1 = 0.7, x_2 = 4.7, x_3 = 3.9, x_4 = 0.5$$

respectively. The sum of these waiting times is $S = 9.8$ minutes. No more events occur during the 10 monitored minutes. Below, we use these observations and Bayesian statistics to learn about λ:

(a) Use the four observed waiting times to obtain a function of λ proportional to a posterior for λ, using $1/\lambda$ as a prior.

(b) Use the single observation that exactly 4 events happened during the 10 monitored minutes to obtain a function of λ proportional to a posterior for λ, using $1/\lambda$ as a prior.

(c) If you have obtained different posteriors in (a) and (b), can you explain this difference? Is there a way to adjust the way computations are made, to make the results identical?

(d) Using the information and the computational methods above, compute the expected arrival time of the fifth event.

+	+	-	+
+	-	+	+
-	+	-	-
-	-	+	+

Figure 2: The configuration of the Ising model used in question 3.

3. (6 points) Recall the Ising model: A configuration σ is a vector consisting of values +1 or -1, one value for every node in a grid. We use the 4×4 grid of Figure 2, where 9 values are +1 and 7 values are -1. For nodes v and w in the grid, we say that they are neighbours,

written $v \sim w$, if the distance between them in the grid is equal to the minimal distance between any nodes in the grid (so nodes have a maximum of 4 neighbours). The energy of a configuration is defined as

$$E(\sigma) = -\sum_{v \sim w} \sigma_v \sigma_w$$

where σ_v and σ_w are the values (+1 or -1) of σ at nodes v and w respectively. For a parameter $\beta > 0$ we define a probability mass function on the set of possible configurations by

$$\pi(\sigma) \propto_\sigma \exp(-\beta E(\sigma)).$$

(a) What is meant by Gibbs sampling in general? In the particular case of the model above, how would Gibbs sampling be performed?

(b) Assume one is using Gibbs sampling on the model above, and that the current configuration is shown in Figure 2. If the next step in the Gibbs sampling may change the state σ_v of the node v in the upper left hand corner of the grid, compute the probability that σ_v is indeed changed in the next step.

(c) What is meant by perfect sampling in general? In the special case of the model above, outline (no details reqired) how it would work.

4. (3 points) Prove the Equation $P(t)P(s) = P(t + s)$ for the transition function matrix for a continuous-time discrete state space Markov chain X. Use the definition of $P(t)$, not the theory for exponential matrices. Mention the properties of continuous-time discrete state space Markov chains that you use in your proof.

5. (2 points) What is the definition of a birth-and-death process? Prove that a birth-and-death process is time reversible.

6. (2 points) If B_t is Brownian motion, compute the distribution of $B_2 + 3B_3 + B_5$.

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 $\text{Gamma}(\alpha,\beta)$ 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 $\text{Inverse-Gamma}(\alpha,\beta)$ 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 April 5 2023

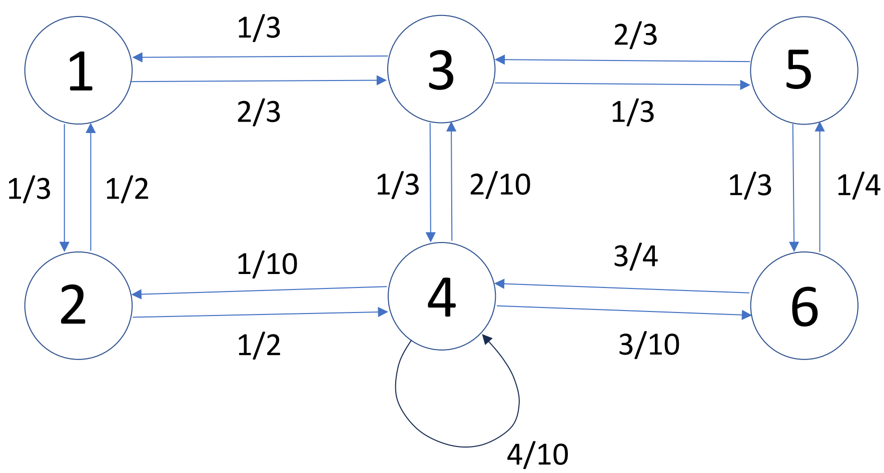

Figure 1: The transition graph answering question 1a.

1. (a) We get Figure 1 above, and

$$P = \begin{bmatrix} 0 & 1/3 & 2/3 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 1/10 & 2/10 & 4/10 & 0 & 3/10 \\ 0 & 0 & 2/3 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 3/4 & 1/4 & 0 \end{bmatrix}.$$

(b) All are recurrent, as can be seen from the transition graph. It is also true that all states are recurrent in a random walk on any weighted undirected connected graph.

(c) Any random walk on a weighted undirected graph has a limiting distribution. By adding the weights of the edges going out of each node, we get that the limiting distribution is

$$\pi = \frac{1}{C}(3, 2, 6, 10, 3, 4)$$

where

$$C = 3 + 2 + 6 + 10 + 3 + 4 = 28.$$

(d) Any Markov chain constructed as a random walk on a weighted undirected graph is time reversible.

(e) To solve this question, we change the Markov chain to make both 5 and 6 absorbing. The transition matrix of the new chain is

$$\begin{bmatrix} Q & R \\ 0 & I \end{bmatrix}$$

where

$$Q = \begin{bmatrix} 0 & 1/3 & 2/3 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/3 \\ 0 & 1/10 & 2/10 & 4/10 \end{bmatrix}$$

and

$$R = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1/3 & 0 \\ 0 & 3/10 \end{bmatrix}.$$

The probability of starting in state number $1, 2, 3$ or 4 and ending in state 5 or 6 can now be computed as the elements of the matrix

$$FR = (I - Q)^{-1}R$$

Thus the probability of starting in state 1 and being absorbed in state 5 is given by

$$\begin{aligned}
& \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} (I - Q)^{-1} R \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
= & \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1/3 & -2/3 & 0 \\ -1/2 & 1 & 0 & -1/2 \\ -1/3 & 0 & 1 & -1/3 \\ 0 & -1/10 & -2/10 & 6/10 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 1/3 \\ 0 \end{bmatrix} \\
= & \; 0.533333
\end{aligned}$$

2. (a) The waiting times are Exponentially distributed with parameter λ. The posterior for λ is proportional to the likelihood times the prior, so

$$\begin{aligned}
\pi(\lambda \mid (x_1, x_2, x_3, x_4) \quad & \propto_\lambda \quad \pi(x_1, x_2, x_3, x_4 \mid \lambda)\pi(\lambda) \\
& = \prod_{i=1}^{4} \text{Exponential}(x_i \mid \lambda) \cdot \frac{1}{\lambda} \\
& = \lambda^4 \prod_{i=1}^{4} \exp(-\lambda x_i) \cdot \frac{1}{\lambda} \\
& = \lambda^3 \exp(-S\lambda)
\end{aligned}$$

It can be noted that this function is proportional to the Gamma density with parameters 4 and S.

(b) The number of events after 10 minutes has a Poisson(10λ) distribution. Thus the posterior for λ becomes

$$
\begin{aligned}
\pi(\lambda \mid 4 \text{ events during } 10 \text{ mins}) &\propto_\lambda \pi(4 \text{ events during } 10 \text{ mins} \mid \lambda)\pi(\lambda) \\
&= \text{Poisson}(4; 10\lambda) \cdot \frac{1}{\lambda} \\
&= \exp(-10\lambda)\frac{(10\lambda)^4}{4!} \cdot \frac{1}{\lambda} \\
&\propto_\lambda \lambda^3 \exp(-10\lambda)
\end{aligned}
$$

It can be noted that this function is proportional to the Gamma density with parameters 4 and 10.

(c) The reason for the difference is that the first computation does not take into accout the extra information that there were no events during the last 0.2 monitored minutes. In fact, given λ, the probability that there are no events during 0.2 minutes can be computed using either the cumulative distribution for the Exponential(λ), yielding $\exp(-0.2\lambda)$, or that we have zero observations for a Poisson(0.2λ) distribution, also yielding $\exp(-0.2\lambda)$. In either case, the extra factor in the likelihood of solution (a) makes the posterior identical to the posterior of solution (b).

(d) We can compute the waiting time for a new observation using the law of total expectation, and using that the expected value of an exponential distribution with parameter λ is $1/\lambda$:

$$\text{E(waiting time)} = \text{E}\left(\text{E}\left(\text{waiting time} \mid \lambda\right)\right) = \text{E}(1/\lambda).$$

We have found above that λ has posterior distribution Gamma(4, 10), so $1/\lambda$ has posterior distribution Inverse-Gamma(4, 10), which according to the Appendix has expectation $10/3 = 3.333$.

Note that we know that no new events occur between 9.8 and 10 minutes. The expected waiting time after 10 minutes is 3.333 minutes. Thus the expected time for the next observation to occur is at 13.333 minutes.

3. (a) Gibbs sampling can be described as a version of the Metropolis Hastings algorithm where the proposal function proposes a change in only one variable at a time. The proposal distribution is then the conditional distribution give the current value of all other variables, and the acceptance probability is 1.

In the Ising model example, one would cycle through each of the nodes in the grid and simulate either +1 or -1 as a new value at that node using the conditional distribution given the values at all the other nodes.

(b) As currently $\sigma_v = +1$, we need to compute the probability of σ_v being -1 in the conditional distribution given σ_{-v}, i.e., the current configuration of all other nodes.

We get

$$\begin{aligned}
\pi(\sigma_v = -1 \mid \sigma_{-v}) &= \frac{\pi(\sigma_v = -1, \sigma_{-v})}{\pi(\sigma_{-v})} \\
&= \frac{\pi(\sigma_v = -1, \sigma_{-v})}{\pi(\sigma_v = -1, \sigma_{-v}) + \pi(\sigma_v = +1, \sigma_{-v})} \\
&= \frac{\exp(-\beta E(\sigma_v = -1, \sigma_{-v}))}{\exp(-\beta E(\sigma_v = -1, \sigma_{-v})) + \exp(-\beta E(\sigma_v = +1, \sigma_{-v}))} \\
&= \frac{\exp(-\beta((-1)(+1) + (-1)(+1)))}{\exp(-\beta((-1)(+1) + (-1)(+1))) + \exp(-\beta((+1)(+1) + (+1)(+1)))} \\
&= \frac{\exp(2\beta)}{\exp(2\beta) + \exp(-2\beta)} \\
&= \frac{1}{1 + \exp(-4\beta)}.
\end{aligned}$$

(c) In general, perfect sampling is an algorithm which guarantees that the value at the last step of the Markov chain of a pre-defined length is indeed sampled from the limiting distribution. In the Ising model example, one would simulate a chain of maximal configurations, starting with all +1 values, and a chain of minimal configurations, starting with all -1 values, using coupled Gibbs simulations, so that steps where nodes have identical neighbours would get identical outcomes. If the two simulations have converged after a predetermined number of steps, the sample is perfect.

4. We get for the ij term of the matrix $P(t+s)$

$$\begin{aligned}
P(t+s)_{ij} &= \Pr[X_{t+s} = j \mid X_0 = i] \\
&= \sum_k \Pr[X_{t+s} = j, X_t = k \mid X_0 = i] \\
&= \sum_k \Pr[X_t = k \mid X_0 = i] \Pr[X_{t+s} = j \mid X_t = k, X_0 = i] \\
&= \sum_k \Pr[X_t = k \mid X_0 = i] \Pr[X_{t+s} = j \mid X_t = k] \\
&= \sum_k P(t)_{ik} \Pr[X_s = j \mid X_0 = k] \\
&= \sum_k P(t)_{ik} P(s)_{kj} \\
&= (P(t)P(s))_{ij}
\end{aligned}$$

which completes the proof. In the 4th line we use the Markov property, and in the 5th line we use the stationary increments property.

5. A birth-and-death process is a continuous time Markov process whose state space is the non-negative integers and where the only transitions occur between adjacent integers. As the transition graph is then a line, it is also a tree, and thus the process is time reversible.

6. We may write

$$B_2 + 3B_3 + B_5 = B_2 + 4B_3 + B_5 - B_3 = 5B_2 + 4(B_3 - B_2) + B_5 - B_3.$$

As

$$\begin{aligned} B_2 &\sim \text{Normal}(0, 2) \\ B_3 - B_2 &\sim \text{Normal}(0, 1) \\ B_5 - B_3 &\sim \text{Normal}(0, 2) \end{aligned}$$

and these random variables are independent, we can compute the variance of their linear combination as

$$\text{Var}(5B_2 + 4(B_3 - B_2) + B_5 - B_3) = 5^2 \cdot 2 + 4^2 \cdot 1 + 2 = 68.$$

As the sum of normal variables is normal, and as the expectation is linear, we get

$$B_2 + 3B_3 + B_5 \sim \text{Normal}(0, 68).$$