Petter Mostad
Applied Mathematics and Statistics
Chalmers and GU

MVE550 Stochastic Processes and Bayesian Inference

Re-exam August 25, 2025, 8:30 - 12:30
Examiner: Petter Mostad will be available by phone 0707163235,
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.

1. (6 points) Anja is a geologist measuring low-level earthquakes. After arriving at a new location, she detects 12 the first week, 7 the next week, and 8 the third week.

   (a) She assumes that the number of earthquakes per week is Poisson distributed with a rate parameter λ. If she uses the improper prior $\lambda \propto_\lambda 1/\lambda$, what is her posterior distribution for λ after 3 weeks of observations?

   (b) Using her prior, her assumption, and her data, Anja computes a probability distribution for the number of earthquakes she will observe in the fourth week. Find this probability distribution (or its density function).

   (c) Anja's boss has more experience, and believes a better prediction can be obtained than the one obtained in (b). Mention two separate ideas for how the predictions might be improved.

2. (4 points) Consider the transition graph for a discrete-time Markov chain given in Figure 1:

   (a) Write down the communication classes of the chain.

   (b) For each state, find whether it is transient or recurrent, and find its periodicity.

   (c) Is the chain ergodic? Prove or disprove.

   (d) Is the chain absorbing? Prove or disprove.

3. (5 points) While most DNA is inherited from either parent to their child, mitochondrial DNA is inherited only from the mother.

   (a) Assume, in a population where each woman gets on average 1.4 daughters, a mutation happens in some mitochondrial DNA. Assume the mutation does not affect the reproduction rate of the mother. After 10 generations, what is the expected number of women carrying this mutation?

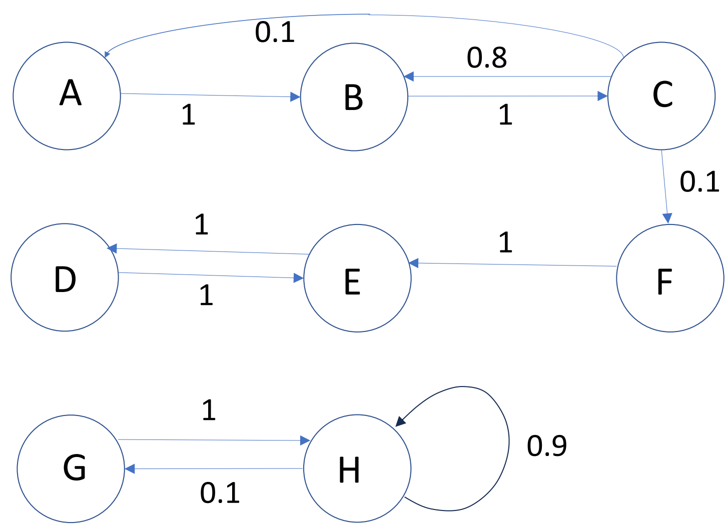

Figure 1: The Figure for question 2.

(b) Assume the number of daughters a woman gets is Poisson distributed. What is the probability that the mutation will eventually disappear from the population? (You may not be able to complete the calculations on your calculator, but please describe how to complete calculations numerically).

(c) Assume that, in fact, the mutation somehow reduces fertility, so that the probabilities of having 0, 1, 2, 3, or more than 3 daughters become 0.35, 0.35, 0.2, 0.1, and 0, respectively. In this situation, is it certain that the mutation would eventually disappear from the population, or is there a possibility that it would persist indefinitely?

4. (2 points) Explain what the Strong Law of Large Numbers for Markov Chains is.

5. (7 points) A machine can be operating, or in a cool-down state, or off. It changes from off to operating according to a Poisson process with rate 0.2 per hour. It then moves from operating to cool-down according to a Poisson process with a rate of 2 per hour. Finally, it moves from cool-down to off according to a Poisson process with rate 4 per hour.

   (a) The situation can be represented by a continuous-time Markov chain. Make a transition rate graph for such a chain, and write down its generator matrix. Compute the expected proportion of time that the machine is operating.

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

   (c) In fact, while in its cool-down phase, the machine can accumulate errors. These errors accumulate over the cycles of the machine, so errors persist even if the machine is turned off. When the machine reaches three such errors, it breaks. While the machine is in a cool-down state, errors occur according to a Poisson process with a rate of 0.1 per hour.

Now, this situation can be represented by a continuous-time Markov chain. Make a transition rate graph for such a chain, and write down its generator matrix.

(d) If the machine is currently off and it has no errors, describe how to compute the expected time until it breaks. (You do not need to obtain the numerical answer; just describe how it can be computed).

6. (5 points) A Brownian motion can be defined as stochastic process $\{X_t\}_{t\geq 0}$ fulfilling

- $\{X_t\}_{t\geq 0}$ is a Gaussian process.
- $X_0 = 0$.
- $\mathrm{Cov}[X_s, X_t] = \min(s, t)$.
- $\mathrm{E}[X_t] = 0$.
- The function $t \mapsto X_t$ is a continuous function with probability 1.

Use this to either prove or disprove that the following processes are Brownian motion, where B_t denotes Brownian motion:

(a) $X_t = \frac{1}{2}B_{2t}$.

(b) $X_t = B_{t+2} - B_2$.

(c) $X_t = B_t + t(B_1 - B_t)$.

(d) $X_t = \sqrt{|B_t|}$.

7. (1 point) In the context of Markov chain Monte Carlo sampling, what exactly separates "perfect sampling" from other types of Markov chain Monte Carlo sampling?

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 $x \in \{0, 1, 2, \dots, n\}$ 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 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 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. We write $x \sim \text{Negative-Binomial}(r, p)$ and $\pi(x \mid r, p) = \text{Negative-Binomial}(x; r, p)$.

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 $\pi(x \mid \lambda) = \text{Poisson}(x; \lambda)$. 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 August 25, 2025

1. (a) For the posterior we get

$$\begin{aligned}
\pi(\lambda \mid \text{data}) &\propto_\lambda \pi(\text{data} \mid \lambda)\pi(\lambda) \\
&\propto_\lambda \text{Poisson}(12;\lambda)\,\text{Poisson}(7;\lambda)\,\text{Poisson}(8;\lambda)\frac{1}{\lambda} \\
&\propto_\lambda e^{-3\lambda}\lambda^{12}\lambda^{7}\lambda^{8}\lambda^{-1} \\
&\propto_\lambda \text{Gamma}(\lambda; 27, 3)
\end{aligned}$$

(b) What we want is the posterior predictive for y_4, the number of earthquakes in the fourth week. We may compute it as follows:

$$\begin{aligned}
\pi(y_4 \mid \text{data}) &= \frac{\pi(y_4 \mid \lambda, \text{data})\pi(\lambda \mid \text{data})}{\pi(\lambda \mid y_4, \text{data})} \\
&= \frac{\text{Poisson}(y_4;\lambda)\,\text{Gamma}(\lambda;27,3)}{\text{Gamma}(\lambda; 27+y_4, 4)} \\
&= \frac{e^{-\lambda}\frac{\lambda^{y_4}}{y_4!}\cdot\frac{3^{27}}{\Gamma(27)}\lambda^{27-1}e^{-3\lambda}}{\frac{4^{27+y_4}}{\Gamma(27+y_4)}\lambda^{27+y_4-1}e^{-4\lambda}} \\
&= \frac{\Gamma(27+y_4)}{\Gamma(27)\cdot y_4!}\cdot\frac{3^{27}}{4^{27+y_4}}.
\end{aligned}$$

You may express the result as above, or you may re-write it as

$$\frac{\Gamma(27+y_4)}{\Gamma(27)\Gamma(y_4+1)}\cdot\left(\frac{3}{4}\right)^{27}\cdot\left(\frac{1}{4}\right)^{y_4}.$$

Recognizing that this is in fact the same as a Negative-Binomial distribution, we may also write

$$\pi(y_4 \mid \text{data}) = \text{Negative-Binomial}(y_4; 27, 3/4).$$

(c) One possible thing to do to improve predictions is to use an informative prior. One would then use data from similar areas, or previous data from the same area, to construct a prior $\pi(\lambda)$ containing some more information.

Another important issue is the assumption that earthquakes are Poisson distributed. Implicitly, this means that one assumes they are independent of each other, which is not necessarily a reasonable assumption for a geologist. A more realistic model than the Poisson distribution could be used to make a more realistic prediction. Either, one could use a different distribution for the counts of these kinds of measurements per week (based on scientific experience) or more fundamentally one could use a continuous-time model for the times that is not a Poisson process.

2. (a) The communication classes are $\{A, B, C\}, \{D, E\}, \{F\}, \{G, H\}$.

   (b) A: Transient, period 1.
   B: Transient, period 1.
   C: Transient, period 1.
   D: Recurrent, period 2.
   E: Recurrent, period 2.
   F: Transient, period ∞.
   G: Recurrent, period 1.
   H: Recurrent, period 1.

   (c) The chain is not ergodic. One of the conditions for ergodicity is irreducibility. But the chain is not irreducible as it has more than one communication class.

   (d) The chain is not absorbing. To be absorbing it would have to have at least one state that is absorbing, and it does not.

3. (a) We model this as a Branching process, where the offspring variable has expectation $\mu = 1.4$. The expected size of the population (with the mutation) is then $\mu^{10} = 1.4^{10} = 28.93$ after 10 generations.

   (b) We need to compute the extinction probability in a branching process where the offspring distribution X has a Poisson distribution with expectation $\lambda = 1.4$. For the probability generating function we get

$$\begin{aligned} G_X(s) &= \mathrm{E}(s^X) \\ &= \sum_{k=0}^{\infty} \text{Poisson}(k; \lambda) s^k \\ &= \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} s^k \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!} \\ &= e^{-\lambda} e^{s\lambda} \\ &= e^{(s-1)\lambda}. \end{aligned}$$

   The answer is given by finding the smallest positive root of the equation $G_X(s) = s$, i.e., $e^{(s-1)\lambda} = s$. This may be done numerically by some optimization algorithm.

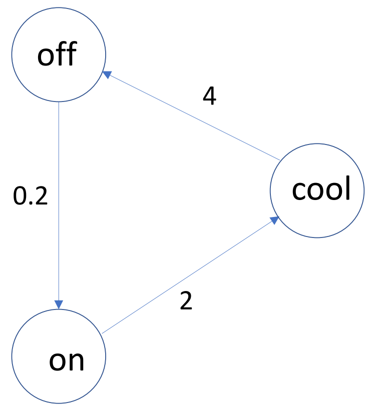

Figure 1: Solution for 5a.

(c) With the new numbers, the expected number of offspring is

$$\mu' = 0.35 \cdot 1 + 0.2 \cdot 2 + 0.1 \cdot 3 = 1.05.$$

As $\mu' > 1$ the process is supercritical, and there will be a non-zero probability that it will not go extinct. So it is *not* certain that the mutation would eventually disappear.

4. The strong law of large numbers says that, if $x_1, x_2, \ldots, x_n$ is a sample from a random variable X with a finite variance, then (with probability 1)

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = \mathrm{E}(X).$$

Assume instead that $x_1, x_2, \ldots, x_n$ is a realization from a Markov chain whose limiting distribution is that of X. Then the above still holds. This is the strong law of large numbers for Markov chains.

5. (a) The transition rate graph is seen in Figure 1.
The generator matrix is

$$Q = \begin{bmatrix} -0.2 & 0.2 & 0 \\ 0 & -2 & 2 \\ 4 & 0 & -4 \end{bmatrix}$$

The chain is ergodic, and the stationary distribution v can be found solving the equation $vQ = 0$ subject to the sum of the terms of v being 1. We get for example the equations

$$\begin{aligned} 0.2v_1 &= 2v_2 \\ 2v_2 &= 4v_3 \\ v_1 + v_2 + v_3 &= 1 \end{aligned}$$

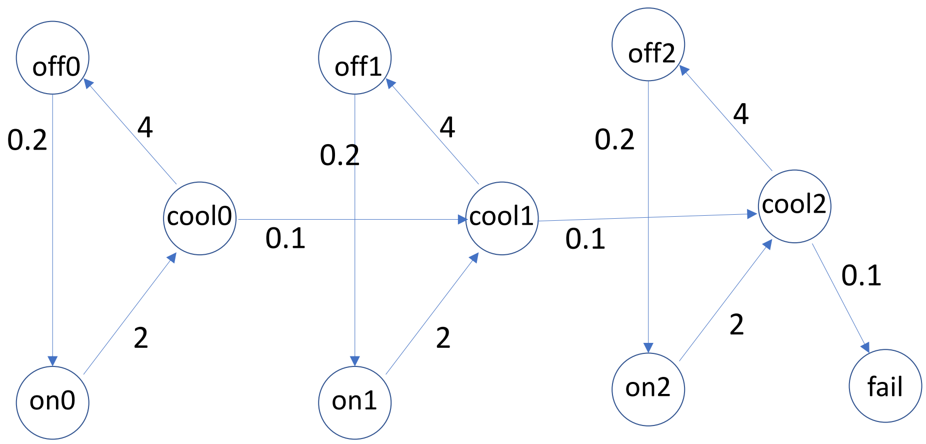

Figure 2: Solution for 5c.

resulting in $v = \frac{1}{23}(20, 2, 1)$. So the proportion of time spent in the off state will be $2/23 = 0.08696$.

(b) The Markov process is not time reversible. For example, $v_1 q_{12} > 0$ while $v_2 q_{21} = 0$, and these would have to be equal if the chain should be time reversible.

(c) For each of the three states used so far, we need to use three versions, one with zero errors having occurred, one with 1 error having occurred, and one with 2 errors having occurred. The transition rate graph is shown in Figure 2 and the generator matrix becomes

$$Q' = \begin{bmatrix} -0.2 & 0.2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & -4.1 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.2 & 0.2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & -4.1 & 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.2 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & -4.1 & 0.1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The states are listed in the order off, on, cool with zero errors; off, on cool with one error; off, on, cool with two errors, and fail.

(d) Let V be the Q' matrix with the last row and column deleted. Then the matrix $F = -V^{-1}$ consists of the expected times staying in a state, before being absorbed (i.e., the machine fails). Thus the answer is obtained by summing the first line of the matrix F.

6. (a) This is is NOT Brownian motion. We see this as

$$\text{Cov}\left[X_s, X_t\right] = \text{Cov}\left[\frac{1}{2}B_{2s}, \frac{1}{2}B_{2t}\right] = \frac{1}{4}\text{Cov}\left[B_{2s}, B_{2t}\right] = \frac{1}{4}\min(2s, 2t) = \frac{1}{2}\min(s, t).$$

This is not the correct covariance.

(b) This is a Brownian motion: To show that it is a Gaussian process, select a sequence $t_0, t_1, \ldots, t_k$ of numbers and real values $a_1, \ldots, a_k$. Then

$$a_1 X_{t_1} + \cdots + a_k X_{t_k} = a_1 B_{t_1+2} + \cdots + a_k B_{t_k+2} - (a_1 + \cdots + a_k)B_2.$$

As Brownian motion is a Gaussian process, the linear combination above is also normal, and X_t is a Gaussian process. Clearly $X_t = 0$, $\text{E}[X_t] = 0$, and the function $t \mapsto X_t$ is continuous with probability 1; all these statements follow from the corresponding property of B_t. Finally, assuming $s < t$,

$$\begin{aligned}\text{Cov}\left[X_s, X_t\right] &= \text{E}((X_s - \text{E}(X_s))(X_t - \text{E}(X_t))) \\ &= \text{E}(X_s X_t) \\ &= \text{E}((B_{s+2} - B_2)(B_{t+2} - B_2)) \\ &= \text{E}(B_{s+2}B_{t+2} - B_2 B_{t+2} - B_{s+2}B_2 + B_2^2) \\ &= \min(s+2, t+2) - \min(2, t+2) - \min(s+2, 2) + 2 = s\end{aligned}$$

so the covariance is also correct.

(c) This is NOT Brownian motion. For example, $X_1 = 0$, which does not have the correct variance.

(d) This is NOT Brownian motion. The simplest way to see this is to notice that $\sqrt{|B_t|}$ is not normally distributed, which it needs to be in order for this to be a Gaussian process.

7. If x_n is the n'th value in a Markov chain Monte Carlo chain with a limiting distribution, it is generally not known how close the distribution of x_n is to the limiting distribution. However, in perfect sampling, you achieve a situation where the distribution of the generated value x is exactly the limiting distribution.