Reading guide:
Stochastic Processes and Simulation II

Course literature: S.M. Ross, Introduction to Probability Models, 11th edition, Academic Press, 2014.

Session 1

Repetition of Poisson processes and continuous-time Markov chains

Section 5.1. Read as introduction.

Section 5.2. Most of the main text should be understood. The examples are good illustrations of whether you understand the material. Section 5.2.4 can be skipped.

Section 5.3. Important section. Examples 5.15 and 5.16 are illustrative, but not very important. Section 5.3.5 is probably about the most important and useful properties of a Poisson process, i.e., the order statistics property (Theorem 5.2 and following remark). When solving problems involving Poisson processes it is good to spend first some time on "How can I apply this property?" Only after that, start thinking about other methods. Example 5.21 is very illustrative. Skip Section 5.3.6.

Sections 6.1-6.5. They are all important, but mainly as supporting material for the remainder of the course, e.g., in the part on queueing theory. Get familiar with Kolmogorov's forward and backward equation and learn about computing limiting probabilities. Examples 6.1 and 6.15 are very illustrative.

Session 2

Generalizations of Poisson processes

Section 5.4. Examples 5.24, 5.27 and 5.30 are very illustrative.

In addition to Section 5.4.1: Assume that $\{N(t), t \geq 0\}$ is a homogeneous Poisson process with rate 1, $\{\hat{N}(t), t \geq 0\}$ is a nonhomogeneous Poisson process with intensity function $\lambda(t)$, and define $m(t) = \int_0^t \lambda(s)\, ds$ for all $t \geq 0$. Then $\{\hat{N}(t), t \geq 0\}$ is distributed as $\{N(m(t)), t \geq 0\}$ and it follows that, conditioned on $\hat{N}(t) = n$, the points of $\{\hat{N}(s), s \in (0,t)\}$ are independent and identical distributed on $(0,t)$ with distribution function $m(s)/m(t)$.

In addition to Section 5.4.3: In Exercise 5.95c you can prove that the order statistics property (Theorem 5.2, p. 310) applies to mixed Poisson processes.

Session 3

Renewal theory

Section 7.1. Read as introduction and understand the theory.

Section 7.2. Important to understand this section. At the bottom of page 411 $[t]$ for the index of the largest summand is the floor of t, that is the integer part of t. Further, note that Example 7.2 is about a homogeneous Poisson process.

Section 7.3. Relevant as a whole. Make sure that you also understand the concept of stopping times. You do not need to know the central limit theorem for renewal processes, but it is important that you know that such a theorem exists and gives you an idea of the order of magnitude of the deviation of $N(t)$ from its asymptotic mean. Examples 7.4-7.7, 7.10 and 7.12 are very illustrative.

Session 4

Renewal reward theory

Section 7.4. The entire section, including the examples, is relevant. Examples 7.14-7.15, 7.18 and 7.19 are very illustrative.

Session 5

Regenerative and semi-Markov processes and the inspection paradox

Section 7.5. The entire section is relevant. Note that a renewal process is not regenerative (as opposed to the first example on page 436). This is because $N(t)$ is strictly increasing and therefore at the time of a renewal the process does not regenerate. The age of a renewal process $A(t)$ (see Example 7.18) is regenerative and regenerates at renewal times.

Section 7.6. Read the section. Example 7.30 is very illustrative.

In addition: Recall that a continuous-time Markov chain on $\mathcal{S}$ can be characterized by the departure rates $\{v_i, i \in \mathcal{S}\}$ and the transition probabilities $\{P_{ij}, (i, j) \in \mathcal{S}^2\}$. If the Markov chain is irreducible (all states communicate, i.e., you can go from a state to another in a finite number of steps), and positive recurrent (the mean return time to each state is finite), then a unique stationary distribution ρ exists, satisfying $\rho_i v_i = \sum_{j\in\mathcal{S}} v_j P_{ji} \rho_j$, for all $i \in \mathcal{S}$, and $\sum_{i\in\mathcal{S}} \rho_i = 1$. We also know that $\pi_i = \sum_{j\in\mathcal{S}} \pi_j P_{ji}$ and $\sum_{i\in\mathcal{S}} \pi_i = 1$ give the stationary distribution π of the embedded discrete time model. We have that the long-run proportion of time P_i spent in state i is given by $P_i = \rho_i$ and is proportional to π_i/v_i.

If in the above process the holding times are not exponential, but just random with means $\{\mu_i, i \in \mathcal{S}\}$ then the above process is called a semi-Markov process. Under mild conditions, the asymptotic fraction of time a semi-Markov process is in state i converges a.s. to

$$\frac{\pi_i \mu_i}{\sum_{j\in\mathcal{S}} \pi_j \mu_j}$$

.

Section 7.7. Very important section. The inspection paradox is an example of "size biasing", which is important in many areas of statistics and probability.

Session 6

Queueing theory: exponential models

Section 8.1. Read as an introduction. Note that throughout the entire chapter it is assumed that the queue does not grow out of bounds and that if there are many customers in the system (i.e., more than a given number depending on the parameters of the model), then the departure rate of customers must be higher than the arrival rate.

Section 8.2. Study Section 8.2.1. Further study the rest of the section in Session 7.

Section 8.3. Sections 8.3.1-3 are all important. Example 8.4 is very illustrative. The part of Section 8.3.2 after the solution of Example 8.5 (page 497-499) can be skipped. Sections 8.3.4 and 8.3.5 may be skipped too.

Section 8.9. Study Sections 8.9.1 and 8.9.2. Skip 8.9.3 and 8.9.4.

Session 7

Queueing theory: PASTA and general workloads

Section 8.2. Study Section 8.2.2 and get familiar with working with the PASTA principle. Examples 8.1-8.2 are very illustrative.

Section 8.5. The section goes through examples and computations which are very illustrative.

Session 8

Simulation.

Section 11.1. Read as an introduction. You do not have to know how random number generators usually work (page 646, first two complete paragraphs). However, it is fun to read online reviews of "A Million Random Digits with 100,000 Normal Deviates". Example 11.1 is illustrative.

Section 11.2. Important section. Note that Section 11.2.1 is a bit imprecise, since if F is flat on some interval (i.e., the density is 0 on an interval), then F does not have an inverse, in the sense, that the value of x solving $F(x) = u$ is not unique. To be more formal, we may define, for $u \in [0, 1]$,

$$F^{-1}(u) = \inf\{x \mid F(x) \geq u\}.$$

So, $F^{-1}(u)$ is the "minimum" x for which $F(x)$ is at least u. Examples 11.3-11.5 are very illustrative.

Section 11.3. Not that important. Just read through it and understand Sections 11.3.2 and 11.3.4.

Section 11.4. The beginning of the section is important. Example 11.9 is very illustrative. Section 11.4.1 is not that important since this method gives insight into stochastic processes, but is irrelevant for practical simulations.

Variance reduction techniques

Section 11.6. The part before Section 11.6.1 gives an introduction and motivates the section. Section 11.6.1 shows a theoretical way to obtain a variance reduction, although usually not very efficient. Section 11.6.2 is more relevant and it is important that you understand what is going on here. Examples 11.16 and 11.18 are very illustrative. Section 11.6.3 may be skipped, but might still be illustrative for the interested reader. Section 11.6.4 is relevant and is about an important technique to estimate small probabilities. Examples 11.22-11.23 are very illustrative.

Session 9

Simulation of stochastic processes and MCMC methods

Section 11.5. Introduction and Section 11.5.1 should be well understood. Example 11.13 is illustrative. Section 11.5.2 can be skipped.

Section 4.9. It is important as a whole.

Some additional ideas (partly covered by Section 4.9)

Let $\{X_n, n \geq 0\}$ be a homogeneous discrete time Markov process with (countable) state space $\mathcal{S}$ and transition matrix Q. So, for all $i, j \in \mathcal{S}$ and $n = 0, 1, \ldots,$ we have

$$\mathbb{P}(X_{n+1} = j \mid X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n = i) = \mathbb{P}(X_{n+1} = j | X_n = i) = Q_{ij}.$$

Let $\{\pi_i, i \in \mathcal{S}\}$ be the stationary distribution of this Markov chain (assuming that it exists). That is, $\{\pi_i, i \in \mathcal{S}\}$ constitute a non-negative row vector with $\sum_{i \in \mathcal{S}} \pi_i = 1$, which satisfies $\pi Q = \pi$.

If (not only if) the Markov chain is irreducible, not periodic and has finite state space, then

$$\lim_{n \to \infty} \mathbb{P}(X_n = j \mid X_0 = i) = \pi_j \qquad \text{for all } i, j \in \mathcal{S}$$

and

$$n^{-1} \sum_{k=1}^{n} f(X_i) \to \mathbb{E}[f(X)],$$

where X has density function π and f is a (sufficiently nice) function from $\mathcal{S}$ to $\mathbb{R}$.

Suppose that we want to simulate from some distribution π on $\mathcal{S}$ or approximate $\mathbb{E}[f(X)]$ for X with distribution π. The idea of the Markov chain Monte Carlo (MCMC) method is to create a Markov chain $\{X_n, n = 0, 1, \dots\}$ of which π is a stationary distribution and then use the results of the previous paragraph.

The Ising model

One model for which MCMC methods are particularly useful is the Ising model. The Ising model is a model for ferromagnetism by which one can explain how many "micro-magnets" that interact with each other (in a way specified below) can form a macro-magnet at low temperatures, while this magnetism is lost at high temperatures.

The Ising model is defined as follows. Consider a state space $\{-1, 1\}^V$, where V are the vertices of a large part of a lattice (e.g. a large cube in $\mathbb{Z}^d$, or the integer set $\{1, 2, \cdots, n\}$). The -1's and 1's are called spins. If one of the signs is significantly more present than the other, then one can observe magnetism on the macro-level.

To specify the interaction of the states of the elements of V, we define the Hamiltonian

$$H(\sigma) = \sum_{v \sim w} \mathbb{1}(\sigma(v) \neq \sigma(w)),$$

where $v, w \in V$ and $v \sim w$ if v and w are next to each other (e.g., have Eucledian distance 1 in $\mathbb{Z}^d$). So the Hamiltonian counts the number of "neighbors" that have opposite spins. The probability of state $\sigma \in \{-1, 1\}^V$ is

$$\pi_\sigma = C_\beta e^{-\beta H(\sigma)},$$

where $\beta > 0$ is a constant (inverse temperature) and C_β is a normalizing constant. So, the larger the Hamiltonian of a state, the less the probability of the state. The normalizing constant is hard to compute and direct sampling from the distribution π is therefore very hard.

However, we can apply so-called Gibbs sampling (a special case of MCMC) to this model. In one dimension this is applied as follows. Say we consider state space $\{-1, 1\}^V$, where V are are the points $1, \dots, n$ and interactions are only possible between points at distance one of each other. We consider the Markov chain

$$\{X^{(k)}, k = 0, 1, \dots\} = \{(X_1^{(k)}, \dots, X_n^{(k)}), k = 0, 1, \dots\}$$

on this state space. Start with some initial state, e.g., $X^{(0)} = (1, \dots 1)$ or $X^{(0)} = (-1, \dots, -1)$. When in state $X^{(k)}$, choose one element uniformly at random, say the i-th element $X_i^{(k)}$. The next state $X^{(k+1)}$ is such that $X_j^{(k+1)} = X_j^{(k)}$ for all $j \neq i$ and $X_i^{(k+1)}$ is a simulated random variable with mass

$$\mathbb{P}\left(X_i^{(k+1)} = 1 \,|\, X_{i-1}^{(k)} + X_{i+1}^{(k)} = 0\right) = \frac{1}{2},$$
$$\mathbb{P}\left(X_i^{(k+1)} = 1 \,|\, X_{i-1}^{(k)} = X_{i+1}^{(k)} = 1\right) = \frac{1}{1 + e^{-2\beta}},$$
$$\mathbb{P}\left(X_i^{(k+1)} = 1 \,|\, X_{i-1}^{(k)} = X_{i+1}^{(k)} = -1\right) = \frac{e^{-2\beta}}{1 + e^{-2\beta}}.$$

Note that, with these probabilities, deviating from your neighbors is punished.

Using the theory from Section 4.9 it can be shown that this algorithm has the right stationary distribution.

Session 10

Discussion of the exercises on Chapter 11, in preparation for the computer assignment.

Session 11

Brownian motion

Section 10.1. Very important to gain understanding about Brownian motion. Example 10.1 is illustrative.

In addition: Note that in the construction of the Brownian motion the properties of $X_1, X_2, \ldots$ in equation (10.1) that are used are that (i) they are independent and identically distributed, (ii) have expectation 0 and (iii) variance 1. So, the Brownian motion should also be the "scaling limit" if the X_i's are replaced by other i.i.d. random variables with expectation 0 and variance 1.

Note that $\{X(t), t \geq 0\}$ has the same distribution as $\{-X(t), t \geq 0\}$. So the distribution of a Brownian motion is symmetric with respect to the x-axis.

Note further that if we replace Δx by $\sqrt{c}\Delta x$ and Δt by $c\Delta t$, then we obtain the same distribution of the limiting process (a scaling property). This implies that if you "zoom in" on a Brownian motion (with the right scaling), you will still see a Brownian motion.

Sections 10.2-10.3. Should be understood.

Session 12

Variations on Brownian motion, Gaussian processes and stationary processes

Section 10.5. It is a very illustrative example of how to do computations on Brownian motions and therefore recommended.

Section 10.7. It should be understood entirely.

Section 10.8. It should be understood up to and including Example 10.6 (which is about a very important process). Examples 10.5-10.6 are very illustrative. Examples 10.7-10.9 can help you assess whether you understand the topic.