

# Stochastic Processes and Modeling

## MATH 6660–1 – Fall 2015

### Homework 4

Due Friday, December 11 at 5:00 PM

This homework has 350 points plus bonus available but, as always, homeworks are graded out of 100 points. Full credit will generally be awarded for a solution only if it is both correctly and efficiently presented using the techniques covered in the lecture and readings, and if the reasoning is properly explained. If you used software or simulations in solving a problem, be sure to include your code, simulation results, and/or worksheets documenting your work. If you score more than 100 points, the extra points do count toward your homework total. Generally speaking, computations regarding statistical properties of Markov chains should be done using deterministic formulas from class and the texts (when appropriate) and not just through direct Monte Carlo simulation, unless otherwise instructed.

## 1 Calculation and Modeling Problems

### 1.1 Continuous Time Machine Repair Model (30 points)

This problem is a continuous-time version of a model discussed in a previous problem for discrete-time Markov chain models. Suppose  $N$  machines are available for service, of which only  $M$  are needed to be operational at any given time. Each has a rate  $\mu_{\text{on}}$  of failing while operating, and  $\mu_{\text{off}}$  of failing while dormant. Assume that the probability a machine fails in a given interval is independent of its history and of the status of other machines. Broken machines are sent to a service facility, which can handle a maximum of  $R$  machines at a time. The repair time for a machine under service is an exponentially distributed random variable with average  $T_r$ .

- a. **(10 points)** Construct a continuous-time Markov chain model for the state

of the system of machines. Clearly define all parameters and the infinitesimal generator (transition rate matrix).

- b. **(10 points)** Develop expressions for the stationary distribution for the model. Explain whether or not the model should always converge to this stationary distribution at long times.
- c. **(10 points)** Use a computer to help you compute and plot the stationary distribution for some interesting choices of parameters.

## 1.2 Extinction of Population with/without Finite Carrying Capacity (30 points)

Consider an asexual population in an environment with limited resources, so that there is a maximum population size  $K$ . One continuous-time Markov chain model for the dynamics of the population is through a birth-and-death process, with aggregate birth rate  $\lambda i(1 - i/K)$  and aggregate death rate  $\mu i$  when the population has size  $i$ , where  $\lambda$  and  $\mu$  are the natural per capita birth and death rates of the population in the absence of competition for resources.

Suppose the population starts with a single individual. For both the case of finite and infinite  $K$ , calculate:

- a. the probability that the population will eventually become extinct,
- b. the expected time until the population becomes extinct.

Simplify your formulas as much as possible, and provide plots of these values with respect to  $K$  for some suitable values of  $\lambda$  and  $\mu$ .

Does the limit as  $K \rightarrow \infty$  agree with your results for the case with no environmental constraints ( $K = \infty$ )? If not, explain why these results need not agree.

## 1.3 Age Distribution in Pure Birth Process (20 points)

Consider a rapidly growing asexual population with constant per capita birth rate, negligible death rate, and negligible competition for resources. Recall that a per capita birth rate is the average rate at which an individual gives birth to offspring.

Assume that the population starts with the introduction of a single newly created individual.

- a. **(5 points)** Calculate the probability distribution for the size of the population after a time  $t$ .
- b. **(15 points)** Calculate the probability distribution for the size of the population at time  $t$  which is younger than some arbitrary age level  $a$ .

## 1.4 Population Model with Immigration (30 points)

Consider a population which has a per capita birth rate  $\lambda$ , a per capita death rate  $\mu$ , and an aggregate rate of immigration  $\nu$ . Recall that per capita rates refer to individuals, while aggregate rates refer to the population as a whole. Suppose the dynamics of the population size can be modeled by a continuous-time Markov chain with the usual birth-and-death model, modified to account for immigration. Derive criteria on  $\lambda$ ,  $\mu$ , and  $\nu$  for the following to have probability 1 to be true (consider each separately):

- a. The population will always remain bounded by some fixed value (not specified in advance).
- b. The population size is guaranteed to eventually never fall below 1000.
- c. The population will repeatedly be reduced to zero size.
- d. The expected time is finite between the first and the second time the population falls to zero size.

## 1.5 Fluctuations of Particle Number (30 points plus bonus)

Consider a container of gas molecules with volume  $V$  which has a small opening allowing gas molecules to pass in and out. We assume the environment has a constant density  $\rho$  of gas molecules, and that this density is negligibly affected by the number of gas molecules that enter or leave the container. We wish to model  $N(t)$ , the number of gas molecules that happen to be inside the container at time  $t$ . We initialize the system with some fixed number  $N_0$  of gas molecules distributed uniformly in the container. Note that we may have initialized the gas in the container in a highly compressed or near-vacuum state, so the number of molecules in the container  $N(t)$  could change significantly.

- a. **(5 points plus bonus)** Thermal fluctuations will cause gas molecules to occasionally pass through the opening (from either direction) between the container and the environment. Assume both the environment and the container are

at the same temperature, so that the gas molecules have the same thermal speeds both inside and outside of the container. You can also assume the gas is “ideal,” meaning that you can neglect any interference or interaction between the molecules. Formulate a simple continuous-time Markov chain model for the number of gas molecules in the container, with the state space of  $N(t)$  being the nonnegative integers. You do not need to know physics to get a reasonable model (except that there may be a constant you do not know how to choose). For bonus credit, use any knowledge of physics you may have to describe any constants in the model more precisely.

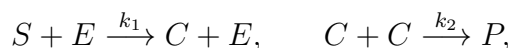
- b. **(25 points)** Use your continuous-time Markov chain model to predict the mean and standard deviation of  $N(t)$ . One way to do this is to define the generating function

$$G_N(s, t) = \mathbb{E}s^{N(t)}$$

and use the Kolmogorov forward equation for the probability distribution of  $N(t)$  to obtain a partial differential equation for  $G_N(s, t)$ . You should be able to solve the resulting partial differential equation for  $G_N(s, t)$  using the method of characteristics.

## 1.6 Stochastic Chemistry (30 points)

Consider an enzyme-catalyzed (irreversible) chemical reaction:



where  $S$  denotes a substrate molecule,  $E$  an enzyme molecule,  $C$  an intermediate complex molecule, and  $P$  the product molecule. The rate constants  $k_1$  and  $k_2$  mean that the rates per volume of the two chemical reactions are, respectively,  $k_1(N_S/V)(N_E/V)$  and  $k_2(N_C/V)((N_C - 1)/V)$ , where the subscripted  $N$  variables refer to the number of that type of molecule in the reaction chamber with volume  $V$ , in which we assume all molecules are well-mixed. The factor  $N_C - 1$  in the second reaction rate expression arises from the fact a complex molecule must find another complex molecule to react with, and there are  $N_C - 1$  complex molecules other than itself to pair with. We will further assume that the substrate is available in such great supply that we can neglect the change in  $N_S$  due to the chemical reaction.

- a. **(10 points)** Formulate a continuous-time Markov chain model to describe the concentration density (number of molecules per volume) of the intermediate complex molecule. Choose a discrete state space which respects the fact that the number of complex molecules in the container must be an integer.

- b. **(20 points)** After a sufficiently long time, the reaction will settle down to a steady-state “burn” and the concentration of intermediate-state complex molecules will approach a stationary limit distribution. Calculate the mean and variance of this limit distribution for the concentration density of complex molecules. One way to do this is to introduce a generating function

$$G_N(s) = \mathbb{E}s^N$$

for the number  $N$  of intermediate complex molecules in the limit distribution. You can derive an ordinary differential equation with boundary conditions for  $G_N(s)$  from the equations for the stationary distribution for  $N$ . You may need to invoke special functions in writing down the solution to this equation.

## 1.7 Time to Infect an Entire Population (30 points)

Consider a stochastic SIR epidemic model as was discussed in class (or Chapter 2 of Andersson and Britton’s *Stochastic Epidemic Models and Their Statistical Analysis*), but suppose that there is no recovery from infection. (This is the so-called SI model, which can also be used to model other irreversible contact processes. One example might be the spread of information; unaware “susceptible” people are “infected” with knowledge when they come in contact with it). In particular, suppose there are initially  $n$  susceptible and  $m$  infected members of the population, and that each infected person attempts to infect each susceptible person according to a Poisson process with (per capita) rate  $\lambda/(n + m)$ .

- a. **(15 points)** Compute the probability that everyone in the population eventually becomes infected, and the expected time until this happens. Discuss in particular what happens when the number  $n$  of initial susceptibles becomes very large.
- b. **(15 points)** Conduct a numerical simulation of this stochastic SI model. Show some example realizations of the infection process, and by performing a reasonably large number of runs, check the answers you computed analytically in the previous part.

## 2 Mathematical Problems

### 2.1 Stationary Distributions for Continuous-Time and Embedded Discrete-Time Markov Chain (10 points)

Suppose an irreducible continuous-time Markov chain with transition rate matrix  $A$  has stationary distribution  $\pi$ . Consider the embedded discrete-time Markov chain defined by the successive states visited by the continuous-time Markov chain. Describe precisely how the stationary distribution of this discrete-time Markov chain is associated to  $A$  and  $\pi$ . Be sure to argue precisely.

### 2.2 Absorption Probabilities for Continuous-Time Markov Chain (10 points)

In class we presented the following formula for the absorption probabilities  $U_{ij} = P(X(\tau) = j | X(0) = i)$  where  $\tau = \inf\{t > 0 : X(t) \notin T\}$  where  $T$  is the set of transient states:

$$A_{ij} + \sum_{k \in T} A_{ik} U_{kj} = 0 \text{ for } i \in T, j \in T^c.$$

Show that this will give the same result for  $U_{ij}$  as would be achieved by computing absorption probabilities using the discrete-time formula for the absorption probability of the embedded Markov chain.

### 2.3 Service in a Queue in Stationary State (20 points)

Consider a simple single-server queueing model where customers arrive at rate  $\lambda$  and are served at rate  $\mu$ . Suppose that the system has evolved to the point where it is described by its stationary distribution.

- a. (5 points) Compute the rate at which customers leave.
- b. (5 points) Why does your answer make sense?
- c. (10 points) Consider the process  $Z(t)$  which describes the number of people who have left the queue by time  $t$ . Is  $Z(t)$  a Poisson counting process? For full credit, support your answer with mathematical calculations. For partial credit, provide an intuitive argument.

## 2.4 Alternative Characterization of Markov Time (20 points)

In class, we defined a Markov time with respect to a filtration  $\{\mathcal{A}_t\}_{t \geq 0}$  as a random time  $\tau$  such that for all  $t > 0$ , the event  $\{\tau \leq t\} \in \mathcal{A}_t$ . Would it be equivalent to define a Markov time as a random time  $\tau$  such that for all  $t > 0$ , the event  $\{\tau \geq t\} \in \mathcal{A}_t$ ? If so, show the definitions are equivalent. If not, find a counterexample and explain carefully why it satisfies one definition but not the other.

## 2.5 Martingale Sums (10 points)

Suppose  $\{X_n\}_{n=0}^\infty$  are a collection of random variables such that the summed random process

$$S_n \equiv \sum_{m=0}^n X_m$$

is a martingale (with respect to the filtration generated by the  $\{X_n\}_{n=0}^\infty$ ). Prove that the random variables in the sums must be uncorrelated:

$$\langle (X_n - \langle X_n \rangle)(X_{n'} - \langle X_{n'} \rangle) \rangle = 0 \text{ for } n \neq n'.$$

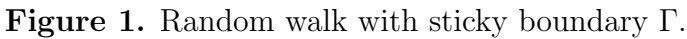
## 2.6 Martingale Argument (10 points)

Consider a discrete-time random walk on the positive quadrant of the two-dimensional lattice of integers. (See Figure 1). Suppose that at each time step, the random walker moves either up one unit or to the right one unit with equal probability. Let  $\Gamma$  be any continuous path through the first quadrant extending from some point on the positive  $y$  axis to the positive  $x$  axis. The path is drawn with only horizontal or vertical segments connecting lattice points. Suppose the random walker starts at the origin, and gets trapped as soon as it reaches the path  $\Gamma$ .

Use a martingale argument to show that on average, such a random walker will have made as many right steps as it will have made up steps before getting stuck on the path  $\Gamma$ .

## 2.7 Random Monogamous Mating (15 points)

You may have noticed that many population models assume the species in question is entirely licentious, mating randomly with any agent of the opposite sex (if sexes are even distinguished in the model!) it happens to encounter. As a slight improvement, consider a model in which the males and females randomly choose a mate (of the opposite sex), but once they have chosen a mate, they marry/bond/cohabitate (sorry,





I forgot the Wiccan version) for life, and produce offspring through this bonded relationship. So, let  $M_n$  denote the number of males and  $F_n$  the number of females surviving to reproductive maturity and seeking mates in generation  $n$ . Assuming they can all find each other, this generation will form  $C_n = \min(M_n, F_n)$  monogamous couples. Assume that the number of offspring produced by each couple is an independent, identically distributed random variable, and that of those offspring which survive to reproductive maturity and seek a mate, an average fraction  $p$  are female and the rest are male. Prove that if the average number of *either* male offspring or female offspring per couple surviving to reproductive maturity and seeking mates is less than 1, then the random process  $\{C_n\}_{n=0}^\infty$  is a supermartingale, meaning that

$$\mathbb{E}[C_{n+1}|C_0, C_1, \dots, C_n] \leq C_n.$$

## 2.8 Dogbert’s Turn to Play in the Maze (20 points)

After observing your experiments with Sam and Ratbert, Dogbert insists you put him in a maze with cheese so he can demonstrate his superior intelligence to the other creatures. He does demand you remove the shock generators, though. Also he finds the maze you gave Sam and Ratbert “trivial” and demands you give him a large sequence of different mazes for him to demonstrate his cheese-seeking supremacy. Because of Dogbert’s force of will, you comply, but you don’t really feel like taking detailed statistics about how quickly Dogbert is solving each maze because you don’t really care. Yet you know he will demand an answer about the average time it takes him to solve a maze and will be really annoying if you don’t give him an answer that will make him happy. So you try to make a good guess though stochastic methods.

By watching Dogbert cleverly navigate a few mazes, you find it difficult to construct a simple stochastic model to describe his motion. To make some progress, you characterize his motion simply in terms of distance from the cheese and discretize your observations into 10 second time intervals. Dogbert’s motion toward the cheese is neither independent, nor identically distributed over different time intervals, because he is smart enough to learn the maze as he goes along. All you can reliably say is that during each 10 second interval:

- Dogbert changes his distance to the cheese by no more than  $b$  (whether an increase or a decrease) just because of his speed limitations.
- Regardless of his past motion, Dogbert will, by smell alone, on average decrease his distance to the cheese by at least a positive amount  $m$ . Note that because of his learning capacity, Dogbert’s decrease in distance to the cheese can depend on his past history, and could in some instances be expected to decrease by more than  $m$ .

You also know that you've constructed the mazes to all have the same level of difficulty, and every time you start Dogbert a distance  $d$  from the cheese. Once Dogbert approaches within a distance  $c$  of the cheese, he declares victory by jumping up onto the cheese and dancing on it.

Prove that with this information, you can deduce that the average amount of time it takes for Dogbert to perform his cheese-dance can be no more than  $10(b + d - c)/m$  seconds. This gives you some guidance on what number to make up so that Dogbert won't be insulted.

### 3 Numerical Simulation

#### 3.1 Stochastic Epidemic Model (35 points)

Write a computer program to simulate the progress of an epidemic according to the stochastic SIR model discussed in class and in Chapter 2 of Andersson and Britton. Make some generic (but reasonable) assumption for what the shape of the probability distribution for the duration of the infectious period of the illness should look like, with a few parameters the user can choose. One of these parameters should be the average duration of the illness, but you can include a few other adjustable parameters as well. You are not allowed to use the exponential distribution to model the infectious period for this exercise.

Your program should allow the user to specify:

- the initial susceptible population
  - the initial infected population
  - the rate at which the infection is communicated to susceptibles.
  - any free parameters in your model for the probability distribution for the duration of an illness.
- a. **(25 points)** Generate a plot that shows one realization of the evolution of the number of susceptible and infected people in the population as a function of time.
  - b. **(10 points)** Perform a reasonably large number of simulations (say 1000) and collect data to produce:
    - (a) A plot of the average number of susceptible and infected people versus time.

- (b) A histogram for the total number of people eventually infected by the disease.