ECE440 - Introduction to Random Processes

Final Exam

December 14, 2025

Instructions:
- This is an **individual** take-home exam, **collaborations are not allowed**.
- You will need to use Matlab to answer some of the questions in this exam.
- You are not allowed to use generative AI tools to aid you in solving the questions.
- You should prepare a report in electronic format with your answers including all necessary derivations, discussions, Matlab code, and plots.
- Write your answers concisely and clearly. Show all your work.
- Your solutions should be submitted via Gradescope as a single pdf file.
- **The submission deadline is 10 pm ET, Tuesday December 16, 2025.**
- Late submissions will not be accepted.
- Perfect score: 100 (out of 105, extra points are bonus points).
- This exam has 5 numbered pages.

GOOD LUCK!

Epidemic processes. The term *epidemic* refers to a phenomenon that is prevalent in excess to what might be expected. It is most commonly used in the context of diseases and their dissemination throughout a population – such as with malaria, COVID-19, bubonic plague, and AIDS – but it is also at times used more broadly in other contexts, such as in describing the spread of perceived problems in a society as well as (mis)information cascades in online social media. Epidemic modeling is concerned with three primary issues: (i) understanding the mechanisms by which epidemics spread; (ii) predicting the future course of epidemics; and (iii) achieving an ability to control the spread of epidemics. Random processes and stochastic modeling has an important role to play in this context.

The most commonly used class of continuous-time epidemic models is the class of *susceptible-infected-removed* (SIR) models. Consider a closed population of N elements, of which S of those elements are susceptible to infection (called 'susceptibles'), I elements are infected (called 'infectives'), and R elements recovered and immune (or, alternatively, 'removed' if you prefer sad endings and also want to account here for deaths). Naturally, because of the closed system assumption one must have $N = S + I + R$. The reactions that define the SIR model describe the infection of susceptibles, and the recovery with subsequent development of immunity (or death,

in any case removal) of infectives. As shorthand (symbolic) notation for the model we use

$$R_1 : S + I \xrightarrow{\beta} 2I$$
$$R_2 : I \quad \xrightarrow{\gamma} \emptyset$$
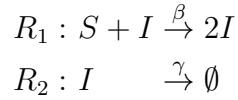
The first reaction R_1 denotes infection of a susceptible element by chance encounter with an infective. The rate of encounters between susceptible and infective individuals is specified to be β, and is referred to as the infection rate. We will adopt the customary assumption of *homogeneous mixing* among members of the population, which asserts that the population is: (i) homogeneous, and (ii) well mixed, in the sense that any pair of members are equally likely to interact with each other. Accordingly, for S susceptibles and I infectives the rate of reaction R_1 is simply βSI. Reaction R_2 denotes recovery and immunity of an infective individual, and the recovery rate is specified to be γ. For a population of I infectives, the rate of reaction R_2 is thus γI.

This exam is roughly divided in two sections. In the first section, which comprises parts *A-E*, you are asked to simulate a *deterministic* model for the population dynamics of the SIR epidemic. In the second section, which comprises parts *F-M*, you are asked to build and analyze a more realistic *stochastic* SIR model based on a continuous-time Markov chain (CTMC), and determine the probability that the epidemic dies out. Note that each part may have multiple questions, **so read carefully to make sure you answer all questions.**

A) Deterministic SIR model (10 points). Towards studying the dynamics of the SIR epidemic process, for given continuous time instant t let $S(t)$ be the number of susceptibles, $I(t)$ be the number of infectives, and $R(t) = N - S(t) - I(t)$ be the number of removed elements in the population. Consistent with the reactions R_1 and R_2, the *deterministic* SIR model is given by the following system of nonlinear differential equations

$$\frac{dS(t)}{dt} = -\beta S(t)I(t)$$
$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t) \tag{1}$$
$$\frac{dR(t)}{dt} = \gamma I(t)$$

with initial conditions $S(0) = N - i_0$, $I(0) = i_0$, and $R(0) = 0$. Write a Matlab script that solves the SIR model differential equations, and plots the evolution of $\{S(t), I(t), R(t)\}$ versus time over the interval $[0, t_{\max}]$. I will help you by sharing a function that solves the deterministic SIR model, which you can download from:

http://www.hajim.rochester.edu/ece/sites/gmateos/ECE440/Final/SIR_deterministic.m

For your simulation adopt a population size of $N = 100$, a single initial infective $i_0 = 1$ (a.k.a. 'patient zero'), and $t_{\max} = 30$. Try out three sets of parameters $\{\beta_1, \gamma_1\} = \{0.05, 0.5\}$, $\{\beta_2, \gamma_2\} = \{0.01, 0.5\}$, as well as $\{\beta_3, \gamma_3\} = \{0.01, 2\}$, and for each one generate a plot which overlays the resulting curves $S(t)$, $I(t)$ and $R(t)$.

B) Disease-free equilibrium (5 points). If $\beta, \gamma > 0$ and regardless of the number of initial infectives, then $\lim_{t \to \infty} I(t) = 0$ meaning the disease eventually disappears from the population. Provide a *qualitative* justification of this asymptotic behavior for the deterministic SIR model.

C) Threshold theorem (8 points). As argued in part *B* the asymptotic behavior of $I(t)$ does not depend on the system parameters and is thus beyond our control. Different is the situation for the

transient dynamics, which are determined by the epidemic's *basic reproduction number* defined as the number of secondary infections caused by one infected individual in an entirely susceptible population. For the SIR model the basic reproduction number is defined as $\mathcal{R}_0 := S(0)(\beta/\gamma)$, and it plays a critical role in the following threshold theorem.

Theorem 1: For the deterministic SIR model, the following holds:
- If $\mathcal{R}_0 = S(0)(\beta/\gamma) > 1$, then there is an initial increase in the number of infectives $I(t)$ and the population exhibits an outbreak (epidemic).
- If $\mathcal{R}_0 = S(0)(\beta/\gamma) \leq 1$, then $I(t)$ decreases monotonically to zero.

Evaluate $\mathcal{R}_0$ for the three parameter settings adopted in part *A*. Do your simulations agree with the transient behavior predicted by the threshold theorem? Explain the initial increase in $I(t)$ when $\mathcal{R}_0 > 1$. [Hint: What is $\lim_{t \to 0^+} \frac{dI(t)}{dt}$ in (1)?]

D) Final size of the epidemic (10 points). As argued in part *B*, under the SIR model the epidemic eventually ends. Still, of interest is to quantify the total number of infectives during the course of the epidemic, i.e., the final size of the epidemic given by $R(\infty) := \lim_{t \to \infty} R(t)$ (since all infectives sooner or later recover). Now, by solving the differential equations of the SIR model one can show that

$$S(t) + I(t) = S(0) + I(0) + \frac{\gamma}{\beta} \ln\left(\frac{S(t)}{S(0)}\right), \quad t \geq 0. \tag{2}$$

Use the identity (2) (which you can accept, no need to prove it), to show that the final size of the epidemic $R(\infty)$ solves the nonlinear equation

$$R(\infty) = \frac{\gamma}{\beta} \ln\left(\frac{N - i_0}{N - R(\infty)}\right). \tag{3}$$

For $N = 100$, $i_0 = 1$, $t_{\max} = 30$ and $\{\beta, \gamma\} = \{0.01, 0.5\}$, solve (3) numerically to determine the final size of the epidemic $R(\infty)$. Corroborate your solution agrees with your plot in part *A*. [Hint: You can solve (3) by finding a zero or 'root' of the function $f(x) := x - \frac{\gamma}{\beta} \ln\left(\frac{N-i_0}{N-x}\right)$, since $f(R(\infty)) = 0$. To that end, Matlab's function `fzero` could become handy.]

E) Deterministic model shortcomings (2 points). Briefly describe limitations of the deterministic SIR model.

F) Stochastic SIR model and infection process (12 points). To specify the stochastic SIR model, we adopt the homogeneous mixing assumption and suppose elements of the population act independently. It is thus reasonable to model the time $T_{R_1}(1,1)$ until the occurrence of a chance encounter between a susceptible individual and an infective individual (i.e., reaction R_1) as exponentially distributed with mean $1/\beta$, i.e.,

$$T_{R_1}(1,1) \sim \exp(\beta). \tag{4}$$

We say that an infection occurs whenever reaction R_1 takes place. Fix a given time t and let S and I be the number of susceptibles and infectives in the population at time t, respectively. Accordingly, there are SI different susceptible-infective element pairs which we index by $j = 1, \ldots, SI$. Let $t + T_{R_1,j}(1,1)$ be the random time at which the j-th pair of elements reacts. According to (4), $T_{R_1,j}(1,1)$ is exponentially distributed with parameter β. Now, denote by $T_{R_1}(S,I)$ the random time until the next infection. The probability distribution of $T_{R_1}(S,I)$ is

exponential with parameter βSI. Write $T_{R_1}(S,I)$ as a function of the random times $T_{R_1,j}(1,1)$, $j = 1, \ldots, SI$, and hence explain why the latter statement is true.

G) Recovery process (3 points). The time $T_{R_2}(1)$ until an infective individual recovers (i.e., reaction R_2) is also random, and modeled as exponentially distributed with rate γ, i.e.,

$$T_{R_2}(1) \sim \exp(\gamma).$$

We say that a recovery occurs whenever reaction R_2 takes place. Fix a given time t and let I be the number of infectives in the population at time t. Denote by $T_{R_2}(I)$ the random time until the next recovery, and use arguments similar to those in part *F* to explain why $T_{R_2}(I) \sim \exp(\gamma I)$.

H) Four simple questions on the recovery process (8 points). Supposing that there are currently $I \geq 2$ infectives including Alice and Bob, what is the probability that Bob will be the first one to recover from the disease. Bob has been infected for 3 days, while Alice has been infected for one week. What is the probability that Bob recovers before Alice? If $1/\gamma = 4$ days and $T_{R_2,A}(1)$ denotes the time until Alice recovers, what is $P\left[T_{R_2,A}(1) > 2 \text{ days}\right]$? What about $P\left[T_{R_2,A}(1) > 5 \text{ days} \,\big|\, T_{R_2,A}(1) > 1 \text{ days}\right]$?

I) Continuous-time Markov chain (CTMC) model (10 points). Since $R(t) = N - S(t) - I(t)$, the number $\{S(t), I(t)\}$ of susceptibles and infectives in the population at time t can be modeled as a *two-dimensional* CTMC with states (s,i), where $0 \leq s, i \leq N$ and $0 \leq s+i \leq N$. Explain why and specify the transition rates $\nu_{(s,i)}$ out of state (s,i) and the transition probabilities $P_{(s,i),(\underline{s},\underline{i})}$. Recall that the $P_{(s,i),(\underline{s},\underline{i})}$ denote the probability of going from state (s,i) to state $(\underline{s},\underline{i})$ given that the CTMC is transitioning out of state (s,i). Notice that most of these transition probabilities are null.

J) Alternative CTMC representation (10 points). Give expressions for the transition rates $q_{(s,i),(\underline{s},\underline{i})}$ from state (s,i) to state $(\underline{s},\underline{i})$. Notice that most of these transition rates are null. Drawing a transition diagram is somewhat cumbersome for a two-dimensional CTMC such as the stochastic SIR model. Instead, you are asked to draw 'local' transition diagrams showing only the direct transitions in and out of state (s,i), where: (i) $s, i > 0$; (ii) $s = 0$, $i > 0$; and (iii) $s > 0$, $i = 0$. Argue that states $(s,0)$ with $s \geq 0$, are absorbing states of the CTMC.

K) System simulation (20 points). Write a Matlab script to simulate the sample paths of the SIR stochastic process $\{S(t), I(t)\}$, with initial distribution $P(S(0) = N - i_0, I(0) = i_0) = 1$ (i.e., the initial state is $(N - i_0, i_0)$ almost surely). Recall that given $\{S(t), I(t)\}$, the corresponding realization for the number of recovered elements in the population can be obtained via $R(t) = N - S(t) - I(t)$. Run your simulation for a population size $N = 100$, a single initial infective $i_0 = 1$, time horizon $t_{\max} = 30$, and rates $\{\beta, \gamma\} = \{0.05, 0.5\}$. Plot one realization of $\{S(t), I(t), R(t)\}$ for $0 \leq t \leq t_{\max}$ (overlay the three curves in the same figure). In a separete figure, plot three realizations of $I(t)$ superimposed to its deterministic counterpart you already obtained in part *A*.

L) Ergodicity of the CMTC (2 points). Is the CTMC irreducible? Is it ergodic? [Hint: There is no need to construct the embedded discrete-time Markov chain to answer these.]

M) Almost sure disease-free equilibrium (5 points). Argue that

$$\lim_{t \to \infty} \sum_{s=0}^{N-i_0} P\left(S(t) = s, I(t) = 0 \,\big|\, S(0) = N - i_0, I(0) = i_0\right) = 1$$

and conclude that the stochastic SIR epidemic model reaches a *disease-free equilibrium almost surely*, i.e. $\lim_{t \to \infty} I(t) = 0$ with probability 1.

Addendum 1

I have used a disease model as a working example, but the problem you just solved and related variants appears in many different contexts dealing with spread of (broadly defined) 'contagions'. Moving on to communication networks this type of analysis is needed to predict and control the spread of computer viruses. A minor variation would tell you about (mis)information cascades in online social networking platforms such as Twitter/X.

Addendum 2

Despite the substantial amount of research involving the general epidemic SIR model and others like it, the underlying assumption of homogeneous mixing is admittedly simple and, for many diseases, too poor of an approximation to reality. As a result, interest has turned increasingly towards so-called *structured population models*, in which assumed contact patterns take into account some structure(s) within the population of interest. Such structure might derive from spatial proximity (e.g., diseases of plants, in which infection occurs through the help of carriers over short distances), social contact (e.g., sexual contact in the transmission of AIDS), or demographics (e.g., households, age brackets, etc.). Often it is convenient to represent this structure as a graph or network, and one can expect the characteristics of the epidemics to be affected to at least some extent by the characteristics of the network. These and other exciting topics related to statistical analysis and learning with graphs will be covered in 'ECE 442 - Network Science Analytics', the class I will teach during Spring '26. If you are interested, I will be happy to see you again in the classroom this coming January. If you want to know more about this class, for the time being you can check the class website here:

https://www.hajim.rochester.edu/ece/sites/gmateos/ECE442.html

Time estimate

To complete this exam I estimate that the total amount of time required is roughly 8 hours. A rough breakdown by parts is the following:

Problem part	Time for execution	Problem part	Time for execution
Problem comprehension	45 mins.	H	15 mins.
A	45 mins.	I	25 mins.
B	15 mins.	J	20 mins.
C	20 mins.	K	60 mins.
D	45 mins.	L	5 mins.
E	5 mins.	M	15 mins.
F	20 mins.	Report preparation	120 mins.
G	5 mins.		