SUPPLEMENTARY MATERIALS: MATHEMATICS + CANCER: AN UNDERGRADUATE “BRIDGE” COURSE IN APPLIED MATHEMATICS∗

TRACY L. STEPIEN†, ERIC J. KOSTELICH‡, AND YANG KUANG‡

In these supplementary materials, we provide a sample syllabus, homework problem sets, and computer lab descriptions for use in developing a Mathematics + Cancer course. We also provide information and data on student evaluations of the course.

SM1. Sample Syllabus.

Course Description. This course addresses some of the mathematical questions regarding the dynamics of cancer growth and treatment. We will discuss some classical models of cancer growth and then consider several case studies, which may include prostate cancer, melanoma, chronic myeloid leukemia, and glioblastoma. We will begin with a review of some basic concepts from ordinary differential equations (ODEs), then discuss cancer growth and treatment models that can be formulated as ODEs. Limitations of the models, including an introduction to uncertainty quantification, will also be highlighted. In the latter half of the course, we will study a very important partial differential equation, the diffusion equation, and its connection to models of cancer growth. Our investigations will include analytical, qualitative, and numerical methods, including the use of ODE solvers in MATLAB. During most weeks, one or more research papers will be assigned for reading. Short quizzes on their content will be given at the start of class during the week following the reading assignment. The course will be a lecture-lab format: We will meet in the regular classroom on one day a week for lecture and in the computer lab on the other day. Previous MATLAB programming experience is helpful but not essential. At the end of the course, you will present a 15-minute talk and write an original paper of 8–12 pages in length about a research paper of your choice. The instructor will suggest some suitable papers, but they are happy to consult with you if you wish to suggest another paper that interests you.

Course Outcomes. You will learn how to formulate, solve, parameterize, and validate ordinary differential equation models of certain kinds of cancer and treatment; some of the basic biology of cancer and its treatment; the use and limitations of ode45 and similar solvers in MATLAB; and how to give a talk on mathematical research.

Prerequisites. Three semesters of calculus and one semester of differential equations. Linear algebra is recommended as a pre- or co-requisite.

Assessment. Assessment will be determined from regular written homework assignments, short quizzes based on the content of assigned reading, computer labs, a midterm exam, and a final presentation. During the second half of the semester, you will have the opportunity to work as part of a team (usually 2 people) on a project involving some topic discussed in the course. You will discuss one or two research papers and/or do some numerical simulations of a relevant mathematical model. The

∗This work was supported by the National Science Foundation’s Mentoring through Critical Transition Points program under grant number DMS-1148771.
†School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85287-1804. Current address: Department of Mathematics, University of Arizona, Tucson, AZ 85721 (stepien@math.arizona.edu).
‡School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85287-1804 (kostelich@asu.edu, kuang@asu.edu).
instructor will suggest some papers, but you are welcome to propose other research topics that interest you. In any case, the final topic and scope of work will be negotiated with the instructors well in advance. In lieu of a final exam, you will give an oral presentation and write a 8–12 page paper (due at the final exam time).

**Tentative Schedule.**

- **Week 1.** Exponential and logistic growth models; direction fields; difference equations.
- **Week 2.** Autonomous differential equations; phase diagrams; the logistic and von Bertalanffy’s growth models. (Reading: Hanahan and Weinberg [SM12])
- **Week 3.** Autonomous differential equations; phase diagrams (Reading: Couzin [SM6])
- **Week 4.** Mathematical models of chemotherapy and surgery. (Reading: Kohandel et al. [SM16])
- **Week 5.** Other formulations of Gompertzian growth; conditional probability. (Reading: Jannini et al. [SM14], Hirata et al. [SM13], Portz et al. [SM20], Swanson et al. [SM24])
- **Weeks 7–8.** Statistical sampling, hypothesis testing, and uncertainty quantification. (Reading: Cohen [SM5], Goodman [SM10])
- **Week 9.** Midterm exam. How to give a scientific presentation. Overview of final talks and papers, including suggested references.
- **Week 10.** The diffusion equation; separation of variables. (Reading: Byrne [SM4])
- **Week 11–12.** Glioblastoma biology and treatment; mathematical models of glioblastoma. (Reading: Swanson et al. [SM23])
- **Week 13.** Open office and lab hours. Discussion of final talks, papers, and associated computations.
- **Week 14.** Mathematical model of TBA cancer growth.
- **Week 15.** Oral presentations.

**SM2. Homework Problems.** The following sets of problems may be used in conjunction with the content modules for assigned written homework.

**SM2.1. Dynamical Models of General Tumor Growth I.**

1. One breeding pair of cockroaches can produce 1,000 fertile offspring (500 new breeding pairs). If every new breeding pair survives and produces 1,000 offspring, after how many generations would the entire earth be covered 1 meter deep in cockroaches? (*The surface area of the earth is about $5 \times 10^8$ km$^2$. Assume each cockroach occupies a volume of 1 cm$^3$. You may also assume that the parents die off when the juveniles reach maturity.*)

2. Let $L(t)$ denote the length of a worm at time $t$. von Bertalanffy proposed a growth model for a typical worm given by $L' = r(L_\infty - L(t))$, where $L_\infty$ is the length of a fully grown individual and $r$ is the growth rate. Find a general solution for $L(t)$, where $L_0$ is the initial length of the baby worm.

3. Consider two applications of the separation of variables for ordinary differential equations related to the von Bertalanffy growth model and the Gompertzian growth model:
(a) Show how to derive the solution \( S(t) = \frac{a}{\beta} + \left( S_0 - \frac{a}{\beta} \right) e^{-\frac{\alpha}{\beta} t} \) from the ordinary differential equation \( \frac{dS}{dt} = \frac{1}{3}(\alpha - \beta S) \).

(b) Show how to derive the solution \( L(x) = L_0 \exp\left( \frac{\alpha}{\ln(q)} (1 - q^x) \right) \) from the ordinary differential equation \( \frac{dL}{dx} = -\alpha q^x L \).

4. Show how the substitution of the function \( x(t) = \ln\left( \frac{K}{N(t)} \right) \) into the differential equation \( \frac{dN}{dt} = \alpha N(t) \ln\left( \frac{K}{N(t)} \right) \) produces the exponential decay equation \( \frac{dx}{dt} = -\alpha x \).

5. Draw a phase diagram for the differential equation \( x' = x(x-1)(2-x) \). Identify each equilibrium point and state whether it is stable or unstable.

SM2.2. Dynamical Models of General Tumor Growth II.

1. Consider an exponential growth process, \( x' = rx \) (\( r > 0 \)). The doubling time is the time \( \Delta t \) required such that \( x(t + \Delta t) = 2x(t) \). Find \( \Delta t \) as a function of \( r \) and show that it does not depend on the value of \( t \).

2. Consider the general von Bertalanffy model \( \frac{dW}{dt} = \alpha W^\lambda - \beta W^\mu \). Show that the origin is a source and \( \frac{W}{\alpha \beta} \) is globally attracting if and only if \( \lambda < \mu \). Analyze completely the case in which \( \lambda \geq \mu \). Hints:
   - The origin is a source if initial conditions near the origin eventually move away from the origin. To prove that the origin is a source, you need to show that \( W' > 0 \) if \( W(0) \) is suitably small and positive and likewise that \( W' < 0 \) if \( W(0) \) is suitably small and negative.
   - Show by substitution that if \( W_{eq} = \left( \frac{\alpha}{\beta} \right)^{1/(\mu - \lambda)} \), then \( W' = 0 \). In other words, an initial population that is exactly equal to \( W_{eq} \) remains at a constant size, thus \( W_{eq} \) is an equilibrium point.
   - We say that \( W_{eq} \) is globally attracting if for every positive choice of \( W(0) \), the corresponding solution asymptotes to \( W_{eq} \). Demonstrate that \( W' < 0 \) whenever \( W(t) > W_{eq} \) and that \( W' > 0 \) whenever \( W(t) < W_{eq} \).
   - It is helpful to sketch a phase diagram.

3. Let \( y = W^{1-\lambda} \). Assuming \( W \) satisfies the equation

\[
\frac{dW}{dt} = \alpha W^\lambda - \beta W,
\]

show that \( \frac{dy}{dt} = (1-\lambda)(\alpha - \beta y) \). Then show that equation (SM1) with initial condition \( W(0) = W_0 \) has the solution

\[
W(t) = \left( \frac{\alpha}{\beta} - \left[ \frac{\alpha}{\beta} - W_0^{1-\lambda} \right] e^{-(1-\lambda)\beta t} \right)^{1/(1-\lambda)}, \quad \lambda \neq 1.
\]

4. Let

\[
A = \begin{pmatrix} \frac{1}{2} & \frac{3}{2} \\ \frac{3}{2} & \frac{1}{2} \end{pmatrix}
\]

(a) Calculate the eigenvalues and eigenvectors of \( A \).

(b) Consider the system \( x = Ax \). Is the origin a sink, a source, or a saddle?
SM2.3. Mathematical Models of Chemotherapy and Surgery. The goal of these exercises is to work through the simplest version of the Kohandel et al. [SM16] model of ovarian cancer treatment. In these exercises, we suppose that an ovarian tumor of size \(x(t)\) grows exponentially starting at \(t = 0\) from an initial seed of size \(X_0\); that is, \(x' = ax\) for some positive constant \(a\). Also suppose that, when chemotherapy is given, it kills cells at a rate proportional to the current tumor size (the log-kill hypothesis). In other words, the chemotherapy kills cells at the rate of \(-cx(t)\), where \(c\) is another positive constant.

1. What differential equation describes \(dx/dt\) while chemotherapy is given? What condition on \(a\) and \(c\) must hold for chemotherapy to produce a response (i.e., shrink the tumor)?

2. Consider a situation in which the tumor grows (untreated) from time \(t = 0\) to time \(t = t_0\). What is \(x(t_0)\)?

3. Assume that the tumor is diagnosed at \(t_0\) and treatment begins. Suppose that chemotherapy is given first and continues from \(t = t_0\) to \(t = t_f\). Find \(x(t_f)\) in terms of \(a, c, t_0, t_f\), and \(X_0\).

4. Surgery is performed at \(t = t_f\) and removes \(e^{-k}\) of the tumor cells remaining after chemotherapy. Find \(\lim_{t\to t_f^+} x(t)\), that is, the tumor population immediately after surgery.

5. As in Problem 2, suppose the tumor grows (untreated) from time \(t = 0\) to time \(t = t_0\). Now suppose that surgery is performed when the tumor is first diagnosed and removes \(e^{-k}\) of the tumor cells. Find \(\lim_{t\to t_0^+} x(t)\), the cell population immediately after surgery, in terms of \(a, c, t_0, t_f\), and \(X_0\).

6. Next assume that chemotherapy follows immediately after surgery and continues to \(t = t_f\). Find \(x(t_f)\) in terms of \(a, c, t_0, t_f\), and \(X_0\).

7. According to this model, does the sequencing of surgery and chemotherapy affect the final tumor cell population? Justify your answer.

Now assume that the tumor grows according to the logistic growth model \(x' = ax(1 - x/K)\).

8. Show that \(x_{CS} < x_{SC}\), where \(x_{CS}\) and \(x_{SC}\) are as given in the main article in equations (19) and (16), respectively.

SM2.4. Conditional Probability.

1. These exercises consider a way to quantify the effectiveness of drug tests.
   (a) Suppose a test for illicit drug use correctly identifies 90% of current users but has a false positive rate of 2.5%. What fraction of people who test positive actually are current users, using data from the 2015 Annual Survey on Drug Use and Health [SM21] that found that about 10.1% of Americans 12 years and older had used an illicit drug in the month prior to the survey (i.e., \(P(D) = 0.101\))?
   (b) Supposing a 90% detection rate as before, what must the false positive rate be to be 95% certain that a person who tests positive uses illicit drugs?
   (c) Repeat the previous problem assuming that the test correctly identifies 99% of current users (i.e., \(P(+|D) = 0.99\)).

2. These exercises explore the likelihood that test results for a rare genetic mutation actually imply that you have the mutation.
   (a) Suppose that 1% of the population carries a particular genetic mutation. Also suppose that the test correctly identifies 95% of the people who
carry the mutation and correctly rules out 99% of the population that does not carry the mutation. What is \( P(M|+) \), the probability that you carry the mutation given that you have tested positive for it?

(b) What is the probability that you do not carry the mutation given that you test negative?

(c) As in (a), but suppose that 10% of the population carries the mutation.

(d) As in (b), but suppose that 10% of the population carries the mutation.

**SM2.5. Uncertainty Quantification.** Recall the following two definitions of error: Suppose the true (nonzero) value of some quantity is \( x \) and we measure \( \hat{x} \). The **absolute error** in the measurement is \( |\hat{x} - x| \) and the **relative error** is

\[
\frac{|\hat{x} - x|}{x}.
\]

These exercises ask you to estimate errors in solutions of differential equations that arise from various sources.

1. Consider an exponential growth process, \( x' = rx \), where \( r > 0 \). Suppose that we have a correct model (i.e., the value of \( r \) is known) but that the initial condition, \( x(0) = x_0 \), is subject to measurement error. If we measure \( \hat{x}_0 \) instead, then our predicted value is \( \hat{x}(t) = \hat{x}_0 e^{rt} \), but the truth is \( x(t) = x_0 e^{rt} \).

Show that the relative error in the prediction at time \( t \) is constant.

2. Next suppose that we estimate the growth rate as \( \hat{r} \) instead of the true value, \( r \). Explain what happens to the relative error as a function of time (there are two cases).

3. Finally, consider a logistic growth model, \( x' = rx(1-x/K) \), where there are two parameters, the growth rate \( r \) and the carrying capacity, \( K \). You may use an analytical or graphical argument to justify each of your answers.

(a) If \( r \) and \( K \) are known but the initial condition \( \hat{x}_0 \) is subject to measurement error, what do you expect will happen to the relative error in any prediction as \( t \to \infty \)?

(b) As in (a), but suppose that the growth rate \( \hat{r} \) is subject to error.

(c) As in (a), but suppose that the carrying capacity \( \hat{K} \) is subject to error.

**SM2.6. Solving the Heat Equation.**

1. Consider the ODE

\[
\frac{d^2 \phi}{dx^2} + \lambda \phi = 0.
\]

For each set of boundary conditions in the list below, determine the conditions that \( \lambda \) must satisfy so that there is a unique and real-valued solution. You may assume that \( \lambda > 0 \) in your analysis.

(a) \( \phi(0) = 0 \) and \( \phi(\pi) = 0 \)

(b) \( \phi(0) = 0 \) and \( \phi(1) = 0 \)

(c) \( \phi(0) = 0 \) and \( \frac{d\phi}{dx}(L) = 0 \)

(d) \( \frac{d\phi}{dx}(0) = 0 \) and \( \phi(L) = 0 \)

(e) \( \frac{d\phi}{dx}(0) = 0 \) and \( \frac{d\phi}{dx}(L) = 0 \)

2. Consider the heat equation on a rod of length \( L \),

\[
\frac{\partial u}{\partial t} = k \frac{\partial^2 u}{\partial x^2},
\]
with the boundary conditions \( u(0, t) = u(L, t) = 0 \). For each initial condition in the list below, find the solution \( u(x, t) \).

(a) \( u(x, 0) = 6 \sin \left( \frac{2\pi x}{L} \right) \)

(b) \( u(x, 0) = 3 \sin \left( \frac{\pi x}{L} \right) - \sin \left( \frac{3\pi x}{L} \right) \)

3. Show how to apply the separation of variables procedure to solve the heat equation

\[
\frac{\partial u}{\partial t} = k \frac{\partial^2 u}{\partial x^2}
\]

subject to the boundary conditions

\[
\frac{\partial u}{\partial x}(0, t) = \frac{\partial u}{\partial x}(L, t) = 0.
\]

These no-flux boundary conditions imply that there is no flow of heat across the ends of the rod (i.e., the ends are perfectly insulated). Find a general solution of the heat equation \( u_n \).

SM3. Computer Labs. The following sets of computer labs may be used in conjunction with the content modules. All of the labs are designed for MATLAB [SM25], but the labs can be implemented in other mathematical computing programs such as GNU Octave [SM9], Maple [SM17], Mathematica [SM26], and CPython [SM7].

SM3.1. MATLAB Preliminaries. In the case where students need to be (re)oriented with MATLAB, it is suggested that the first computer lab be an introduction. Since students will write a lot of function files throughout the course, it is important to develop an understanding of entering commands, writing conditionals and loops, plotting, and creating script files. Many materials are available online that can serve as a basis for an introductory lab, for example, through the MathWorks web site [SM25].

SM3.2. Solving First-Order Ordinary Differential Equations. To introduce students to function files and solving first-order ordinary differential equations with \texttt{ode45} in MATLAB, we focus a lab on the exponential equation, logistic equation, and the Gompertz equation which are all of the form

\[
\frac{dy}{dt} = f(t, y).
\]

First, create function files for the right hand sides of the equations \( f(t, y) \), where in particular we have

(a) \( f(t, y) = \alpha y \)

(b) \( f(t, y) = \alpha y \left( 1 - \frac{y}{\beta} \right) \)

(c) \( f(t, y) = \alpha y \ln \left( \frac{\beta}{y} \right) \)

with \( \alpha \) and \( \beta \) as parameters. For some of the functions, choose exact values for \( \alpha \) and \( \beta \), but for at least one of the functions, let the parameters be unspecified so that they are inputs into the function file.
Once the function files are written, the ordinary differential equation (SM2) can be numerically solved with ode45 given an initial condition and interval of integration. Some variations that can be explored are:

- plot both the numerical and true solutions;
- solve with various initial conditions and plot the results; and
- given sample data and the exponential equation, find an initial condition and parameter value \( \alpha \) for which the solution best fits the data using guess and check (limit possible values to integers).

**SM3.3. Analyzing and Solving Systems of First-Order Ordinary Differential Equations.** After examining single equations, another lab can focus on systems of first-order ordinary differential equations. A good system of equations that can be studied is the Gyllenberg–Webb model [SM11]. Gyllenberg and Webb observed that actively proliferating cells in tumors can enter a quiescent state when they stop dividing and quiescence tends to be more common in large tumors compared to small tumors. They developed a mathematical model of the transition of cells from a proliferative state to a quiescent state and back. The model equations are

\[
\begin{align*}
\frac{dP}{dt} &= (\beta - \mu_p - r_{0}(N))P + r_{i}(N)Q, \\
\frac{dQ}{dt} &= r_{0}(N)P - (r_{i}(N) + \mu_q)Q,
\end{align*}
\]

where \( P(t) \) is the number of proliferative cells, \( Q(t) \) is the number of quiescent cells, and \( N(t) = P(t) + Q(t) \) is the total number of cells. The parameters are the proliferation rate \( \beta > 0 \), death rate for proliferative cells \( \mu_p \geq 0 \), and death rate for quiescent cells \( \mu_q \geq 0 \). The proliferative cells transition to and from being quiescent by rates \( r_{0}(N) \) and \( r_{i}(N) \), respectively. We let the transition rates be

\[
\begin{align*}
r_{0}(N) &= \frac{KN}{aN + 1}, \quad r_{i}(N) = \frac{r}{N + m},
\end{align*}
\]

and note that \( k, a, r, \) and \( m \) are parameters.

It is useful to determine what effects the parameters and initial conditions \( P(0) \) and \( Q(0) \) have on the solutions, and this can be analyzed by investigating phase portraits (for example with pplane [SM19]) and then numerically solving the system of differential equations (SM4). Parameter values can be doubled and halved one at a time and the solutions can be plotted to compare and contrast the results.

**SM3.4. Mathematical Models of Chemotherapy and Surgery.** After introducing the model of chemotherapy and surgery treatment sequencing in ovarian cancer of Kohandel et al. [SM16] in lecture, students can numerically simulate and analyze the different results when comparing adjuvant versus neoadjuvant chemotherapy and the three cell-kill hypotheses. One could also look at the differences between the different tumor growth model functions \( f(x) \), but it may be more beneficial to choose just one. Since the students will have already written function files for the exponential, logistic, and Gompertz right hand side functions in the lab described in Subsection SM3.2, they should be able to implement them into their code for this lab.

The next step is for students to write function files for the right hand side of equations (15) and (18) for each of the log-kill hypotheses given in equations (11)–(13). Then for each of the log-kill hypotheses, a script file can be written to examine adjuvant chemotherapy and neoadjuvant chemotherapy.
For adjuvant therapy, equation (14) is solved first on the interval of integration $[0, t_s]$. Then the last entry of the $x$ solution vector (multiplied by $e^{-k}$) is used in the initial condition to solve equation (15) on the interval of integration $[t_s, t_f]$. The final tumor size (16) is just the last entry of the $x_c$ solution vector. The overall solution vectors from the initial value problems (14) and (15) can be combined into one longer vector for plotting.

For neoadjuvant therapy, the only change that needs to be made is within the initial conditions for the second half of the simulation. After the first part (17) is solved, the initial condition for the second equation (18) is just the last entry of the $x_c$ solution vector. However, the final tumor size (19) uses the last entry of the $x_c$ solution vector, but it needs to be multiplied by $e^{-k}$.

To compare the different simulations, students can:

- plot the adjuvant and neoadjuvant solutions on the same graph for each of the cell-kill hypotheses;
- determine if the order of sequencing of surgery and chemotherapy affects the final number of tumor cells; and
- determine which of the three cell-kill hypotheses results in the smallest tumors by time $t_f$.

For each of the situations where neither the adjuvant or neoadjuvant chemotherapy treatment methods resulted in a small tumor size compared to the initial size, students can find parameters (out of $c$, $k$, and $\delta$) that result in a larger increase in tumor size.

**SM3.5. Conditional Probability.** An example of conditional probability occurs in testing a drug’s efficacy in treating a disease. In this lab, we simulate a clinical trial for a hypothetical cancer drug designed to reduce tumor size. This lab was based off of the “Biological Example of Conditional Probability: Drug Testing” section and DrugTesting.m MATLAB code in Bodine et al. [SM3].

The hypothetical clinical trial has 200 patients, and half receive the cancer drug and half receive a placebo. The number of patients who received the cancer drug and experienced a decrease in tumor size versus increase and the number of patients who received the placebo and experienced a decrease in tumor size versus increase are given to the students. Ultimately we would like to know: what is the probability that if you take the cancer drug ($A$) then the size of your tumor will decrease ($B$), i.e., what is $P(B|A)$? The true value of $P(B|A) = P(A \cap B)/P(A)$ can be calculated to compare to results from the lab.

To simulate the clinical trial, $N$ experiments are run $N$ times, where $N$ is the number of patients in the clinical trial, in which we randomly pick one patient and record if they were given the cancer drug or placebo and if the size of their tumor increased or decreased. The probability of “took cancer drug” ($A$) and the probability of “took cancer drug” and “tumor decrease” ($A \cap B$) for each set of experiments is added to a running sum. These summed values approximate $P(A)$ and $P(A \cap B)$, which can be divided to approximate $P(B|A)$.

Using the estimate of $P(B|A)$, the lab can be extended to determine $P(A|B)$ with Bayes’ Theorem (2). While Bayes’ Theorem does not present the quickest way to code the probabilities, it allows the students to gain more experience with conditional statements (if, then, else) by estimating $P(B)$, $P(A^c)$, and $P(A^c \cap B)$.

**SM3.6. Statistical Sampling.** Suppose we want to know the average height of a group of individuals, but we are unable to measure the height of every single person in the group. How many people do we need to measure to get a reasonable estimate?

To explore sample means, we use a data set that lists the height in a sample of 1794
pregnant women (Bland [SM2]; data available for download at http://www-users.york.ac.uk/~mb55/datasets/datasets.htm as “Height”). We will explore how the mean, median, variance, and standard deviation changes based on the sample size of the data, and plot histograms and box plots to help visualize what is going on. First, besides having the full data set, the data should be also grouped into smaller samples of various sizes (e.g., 10, 20, 40, 80, 120, 320, 640, and 1280). This can be done ahead of time so that there is uniformity between the students’ results, or students can create the data subsets by randomly selecting the data points.

For each sample set, have students calculate the mean, median, variance, and standard deviation as well as plot a histogram with the normal distribution probability density function overlaid on top. In one figure, plot the box plot for all of the data sets. With this information, students can infer what the minimum sample size is that gives a reasonable estimate of the average height.

**SM3.7. Hypothesis Testing.** A result is called *statistically significant* if it has been predicted as unlikely to have occurred by chance alone, according to a predetermined threshold probability, the significance level. We study survival data for advanced lung cancer patients (Kalbfleisch and Prentice [SM15]; data available for download at ftp://ftp.wiley.com/public/sci tech_med/failure_time/ as “Data Set I” or at http://lib.stat.cmu.edu/DASL/). The main purpose of the study was to compare the effects of two chemotherapy treatments in prolonging survival time, termed “standard” and “test.” The patients can have different types of tumors and they have been classified into four categories: squamous, small, adeno, and large. We compare the chemotherapy treatments for the four tumor size categories separately.

The **null hypothesis** to test with this data set is that there is no difference in average survival time between receiving “standard” treatment or “test” treatment, i.e. $H_0 : \mu_{test} = \mu_{standard}$, while the **alternative hypothesis** is that there is a difference in average survival time between receiving “standard” treatment or “test” treatment, i.e. $H_a : \mu_{test} \neq \mu_{standard}$. We choose to test at a significance level $\alpha = 0.05$.

The statistic of interest is the $t$-statistic

\[
t = \frac{\bar{x}_{test} - \bar{x}_{standard}}{\sqrt{\frac{s^2_{test}}{n_{test}} + \frac{s^2_{standard}}{n_{standard}}}},
\]

where $\bar{x}_T$ is sample mean, $s_T$ is the sample standard deviation, and $n_T$ is the sample size (number of individuals) for groups $T$ = “test” and “standard.”

After students count the number of patients with squamous, small, adeno, and large tumors with survival $\leq$ 100 days, 400 days $<$ survival $\leq$ 300 days, and survival $> 300$ days (12 categories total), then sample means, sample standard deviations, and sample sizes can be calculated. Then the $t$-statistic (SM6) can be calculated to determine whether to reject or fail to reject the null hypothesis. Students can comment on for which size tumors there was a statistically significant difference in average survival time between the “standard” and “test” treatments and explain what it means in terms of what the results of the clinical trial imply about the effectiveness of the treatments for the different size tumors. Then students can vary $\alpha$ to find cases where more of the tests results in rejecting the null hypothesis.

**SM3.8. Simulating a Prostate Cancer Growth Model.** The goal of this lab is to reproduce the figures in Portz et al. [SM20], which models the effects of androgen-deprivation therapy in prostate cancer and compares numerical simulations with actual patient data. This lab can be split into 2 or 3 sessions.
SM3.8.1. Model Equations. The model of Portz et al. [SM20] is motivated by the Droop [SM8] model as discussed in Subsection 3.6. Population dynamics of androgen-dependent (AD) cells and androgen-independent (AI) cells are described by the following equations,

\[
\frac{dX_1}{dt} = \mu_m \left(1 - \frac{q_{\min,1}}{Q_1}\right) X_1 - \frac{d_1 X_1}{\text{cell death}} - \lambda_1(Q_1) X_1 + \lambda_2(Q_2) X_2,
\]

\[
\frac{dX_2}{dt} = \mu_m \left(1 - \frac{q_{\min,2}}{Q_2}\right) X_2 - \frac{d_2 X_2}{\text{cell death}} - \lambda_2(Q_2) X_2 + \lambda_1(Q_1) X_1,
\]

where \(X_1\) is the number of AD cells, \(X_2\) is the number of AI cells, androgen is the cell quota \(Q\) in cells \(i\) for \(i = 1, 2\), \(d_1\) and \(d_2\) are the constant rates of cell death, \(\lambda_1(Q_1)\) is the transition rate from AD to AI due to mutations, and \(\lambda_2(Q_2)\) is the transition rate from AI to AD. AD cells grow faster as \(Q(t)\) increases and while AI cells need androgen, they can grow at much lower levels (hence \(q_{\min,1} > q_{\min,2}\)). It is assumed that the transition rates due to mutations are of the form of Hill functions,

\[
\lambda_1(Q) = c_1 \left(\frac{K_1^3}{K_1^3 + Q^3}\right),
\]

\[
\lambda_2(Q) = c_2 \left(\frac{Q^3}{K_2^3 + Q^3}\right),
\]

where \(c_1, c_2, K_1,\) and \(K_2\) are parameters. With these assumptions, the AD→AI transition rate yields a low mutation rate for normal androgen levels and a high rate for low androgen levels, while the AI→AD transition rate is high for normal androgen levels and low for low levels.

The Portz et al. [SM20] model also includes prostate-specific antigen (PSA), as this is a quantity that can be clinically measured, unlike the amount of AD and AI cells. It is assumed that PSA is produced at a rate proportional to the cell populations \(X_1\) and \(X_2\) and cell quotas \(Q_1\) and \(Q_2\) and is cleared from the blood at a constant rate, thus

\[
\frac{dP}{dt} = \sigma_0(X_1 + X_2) + X_1 \left(\frac{\sigma_1 Q_1^3}{Q_1^3 + \rho_1^3}\right) + X_2 \left(\frac{\sigma_2 Q_2^3}{Q_2^3 + \rho_2^3}\right) - \delta P,
\]

where \(P\) is PSA, \(\delta\) is the removal rate, and \(\sigma_0, \sigma_1, \sigma_2, \rho_1,\) and \(\rho_2\) are parameters. Furthermore, it is assumed that the cell quota \(Q_i(t)\) changes over time for each cell population and it depends on the androgen suppression therapy given to each patient. The derivative \(Q_i(t)\) depends on the cell quota, drug dosage, and androgen consumption with the cells, and the equations for \(Q_i(t)\) are

\[
\frac{dQ_i}{dt} = v_m \left(\frac{q_m - Q_i}{q_m - q_{\min,i}}\right) \left(\frac{A}{A + v_h}\right) - \mu_m(Q_i - q_{\min,i}) - b Q_i,
\]

where \(v_m, q_m, A, v_h,\) and \(b\) are parameters that are assumed to be the same for each cell population.
SM3.8.2. Simulations. The first step is to obtain the prostate-specific antigen (PSA) data for Case #1–7 (Figures 4–10, solid dots, “serum PSA”) and the androgen data (open dots, “serum testosterone”) from Akakura et al. [SM1]. This can be done for the students or it could be part of the lab. An application such as Plot Digitizer [SM18] can be used. Note that the data in Akakura et al. [SM1] is given in months while the simulations in Portz et al. [SM20] are done in days, so the extracted data should be converted to days.

The androgen data $A(t)$ appears in the cell quota $Q_1$ and $Q_2$ equations, and the data needs to be interpolated for use in numerical simulations. One method is to use an exponential fit as given in Equation (13) in Portz et al. [SM20], however, it is suggested to make use of the `pchip` command in MATLAB for purposes of this lab.

The equations to be simulated are (SM7)–(SM10). In particular, the variables are $X_1$, $X_2$, $Q_1$, $Q_2$, and $P$, and the parameters are $\mu_m$, $q_{min,1}$, $q_{min,2}$, $d_1$, $d_2$, $c_1$, $c_2$, $K_1$, $K_2$, $v_m$, $v_h$, $b$, $\sigma_0$, $\sigma_1$, $\sigma_2$, $\rho_1$, $\rho_2$, and $\delta$. For all 7 cases, the parameter values can be used, as given in Table I in Portz et al. [SM20], and the remaining parameter values that can be used are given in Table SM1.

Each case should be simulated for the time in days that there is data. In terms of initial conditions, for the cell quotas the following can be used for all 7 cases,

$$Q_1(0) = Q_2(0) = 0.4.$$  

For the initial PSA concentration $P(0)$, the MATLAB command `pchip` can be used to extrapolate the PSA data at $t = 0$. Lastly, it is assumed that the PSA concentration is a linear combination of cell concentrations $X_1$ and $X_2$, thus we can write this as

$$X_1(0) = \alpha \beta P(0), \quad X_2(0) = (1 - \alpha) \beta P(0),$$

for some $\alpha$ and $\beta$, which for all 7 cases let $\alpha = \frac{14}{15},$ while $\beta = 1$ for Cases #1-4, $\beta = 0.6$ for Case #5, $\beta = 0.7$ for Case #6, and $\beta = 0.8$ for Case #7.

At this point, simulations can be run to reproduce Figures 6–7 in Portz et al. [SM20]. It is suggested that students attempt to reproduce the figures such that all colors, line styles, axes, and legends are the same. Since the parameter values given in Table SM1 and (SM11) are truncated from the exact values used in Portz et
al. [SM20], note that the reproduced curves will be not be exactly the same as the originals.

The last topic that can be covered in this lab is error estimates, in particular the error in the PSA concentration $P$ found in the numerical simulation compared to the data from Akakura et al. [SM1]. After interpolating the numerical solution for $P$, students can calculate absolute error

$$(\text{SM14}) \| \hat{x} - x \|,$$

relative error

$$(\text{SM15}) \left\| \frac{\hat{x} - x}{x} \right\|,$$

and mean square error

$$(\text{SM16}) \frac{1}{N} \| \hat{x} - x \|^2,$$

where $x$ is a vector of values from data, $\hat{x}$ is a vector of values from simulations, and $N$ is the length of the vectors. The mean square error can be compared to the values obtained in Table II in Portz et al. [SM20].

Further questions to discuss regarding the error include: does looking at figures or the absolute, relative, or mean square error give the best representation of the error for the simulations in this lab? Does the case with the smallest calculated error match with which figure appears to match the data the best?

**SM3.9. Solving the Heat Equation.** While there are many examples of how to use MATLAB to solve partial differential equations like the heat equation, we explore one aspect of the numerical details using a finite differences scheme. In this lab, the heat equation on the interval $[0, L]$ will be converted into a system of ordinary differential equations, which will be solved using `ode45`. Before coding anything in MATLAB, students should be instructed to write down the details for the derivation of the numerical ordinary differential equations. The outline of the scheme is:

1. Subdivide interval $[0, L]$ into $n$ equal subintervals; noting that the resulting grid has $n+1$ points. Let $\Delta x$ be the length of each subinterval, and let $x_1 = 0$ be the leftmost grid point and $x_{n+1} = L$ be the rightmost grid point.
2. For a given fixed time $t$, let $u_i = u(x_i, t)$. Approximate $\partial u / \partial x$ at $x_i$ by the forward finite difference

$$(\text{SM17}) \Delta u_i = \frac{u_{i+1} - u_i}{\Delta x}, \quad i = 1, 2, \ldots, n,$$

and approximate $\partial^2 u / \partial x^2$ by the finite difference

$$(\text{SM18}) \Delta^2 u_i = \frac{\Delta u_i - \Delta u_{i-1}}{\Delta x}, \quad i = 2, 3, \ldots, n.$$

Expand the formula on the right-hand side of equation (SM18) using equation (SM17) to find an expression for $\Delta^2 u_i$ in terms of $u_{i+1}$, $u_i$, and $u_{i-1}$.

3. Assuming the Dirichlet boundary conditions $u(0, t) = u(L, t) = 0$, the heat equation is approximated by the system of ordinary differential equations,

$$(\text{SM19}) u_i' = k \Delta^2 u_i, \quad i = 2, 3, \ldots, n,$$

where the prime denotes the time derivative, and $u_1 = u_{n+1} = 0$. 

This manuscript is for review purposes only.
To analyze the results, students can:

- compare the numerical solution to the true solution;
- change the number of subintervals \( n \); and
- plot the solution versus \( x \) and determine what happens over time.

**SM3.10. Simulating a Model of Brain Tumor Growth.** The brain tumor growth model of Swanson et al. [SM22, SM23] is described by a reaction–diffusion equation. In one dimension on the spatial domain \([0, L]\), this equation is

\[
\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2} + \rho u (1 - u), \quad 0 \leq x \leq L, \quad t > 0,
\]

where \( u(x, t) \) is the concentration of tumor cells (as a percentage of the carrying capacity). \( D \) is the diffusion constant that gives a measure of undirected (random) motion of the cells and \( \rho \) is the proliferation rate.

Considering brain tumor growth in one dimension, a biologically reasonable domain is a nerve fiber. In this case, we will assume the two ends of the nerve fiber are insulated, which corresponds to no-flux boundary conditions, i.e., the Neumann conditions

\[
\frac{\partial u(0, t)}{\partial x} = 0, \quad \frac{\partial u(L, t)}{\partial x} = 0, \quad t \geq 0.
\]

We also assume that a certain concentration of tumor cells exists at the “initial” time, which would correspond to the time at diagnosis, for example

\[
u(x, 0) = \begin{cases} 
1 \frac{L}{3} \leq x \leq 2L \frac{3}{3} \\
0, \text{ otherwise}
\end{cases}
\]

The two main differences between the lab described in Subsection SM3.9 are the proliferation term \( \rho u (1 - u) \) and the boundary conditions. The proliferation term is approximated by adding \( \rho u_i (1 - u_i) \) to the discretization (SM19).

For the boundary conditions, we use ghost points \( u_0 \) and \( u_{n+2} \). Discretize the boundary conditions (SM21) with the centered differences

\[
\Delta u_i = \frac{u_{i+1} - u_{i-1}}{2\Delta x}, \quad i = 1,
\]

\[
\Delta u_i = \frac{u_{i-1} - u_{i+1}}{2\Delta x}, \quad i = n + 1.
\]

Then setting (SM23a) and (SM23b) equal to zero, substitute \( u_0 \) and \( u_{n+1} \) into the discretizations for \( u'_i \) and \( u'_{n+1} \). Students can visualize the results of the simulation by plotting the solution at 5 equally spaced times. Parameters can be varied to see how smaller or larger diffusion or proliferation rates can affect the solution.

**SM4. Student Evaluations of Course.** The Mathematics and Cancer course has been taught three times: in Fall 2013, 2015, and 2017. Table SM2 summarizes some of the student responses on a standardized course evaluation form. Answers are reported as the mean value on a 5-point Likert scale (5=strongly agree, 1=strongly disagree). Approximately 40 percent of the students have been women, and, on average, about 10 percent of the initial cohort of students have dropped the course by mid-semester. Overall, 40 to 50 percent of each student cohort have been mathematical sciences majors; 12 to 20 percent, engineering majors; and the remainder from chemistry, life sciences, economics, and other majors.
Table SM2

<table>
<thead>
<tr>
<th></th>
<th>Fall ’13</th>
<th>Fall ’15</th>
<th>Fall ’17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final enrollment</td>
<td>17</td>
<td>16</td>
<td>25</td>
</tr>
<tr>
<td>Number of survey respondents</td>
<td>6</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Assignments/exams promote learning</td>
<td>4.2</td>
<td>4.7</td>
<td>4.4</td>
</tr>
<tr>
<td>Class sessions are well planned</td>
<td>3.5</td>
<td>5.0</td>
<td>4.1</td>
</tr>
<tr>
<td>Pace/level are appropriate</td>
<td>2.8</td>
<td>4.8</td>
<td>4.5</td>
</tr>
<tr>
<td>Course engages my interest</td>
<td>4.0</td>
<td>4.7</td>
<td>4.4</td>
</tr>
<tr>
<td>I learned something valuable</td>
<td>4.2</td>
<td>4.7</td>
<td>4.5</td>
</tr>
</tbody>
</table>


