MATH W81: Problem and Projects—1st Iteration (Problem Set 3)

Presentations begin Mon., Jan. 19

None of the following is, as of yet, assigned to a specific group. The first five problems are exercises (mostly) relating to the material discussed in class on Fri. Jan. 16. You should choose at least one of these to work on. Let Prof. Scofield know your group's choice so he can update the document to reflect that the problem has been "taken". Your group should also choose one of the projects listed as numbers 6–15, again apprising me of your choice. (In some instances I may discourage a particular choice given the makeup of your group.)

Group 2: The Polya urn model for contagion is as follows: We start with an urn which contains one white ball and one black ball. At each second we choose a ball at random from the urn and replace this ball and add one more of the color chosen. Write a program to simulate this model, and see if you can make any predictions about the proportion of white balls in the urn after a large number of draws. Is there a tendency to have a large fraction of balls of the same color in the long run?[1]

Group 3: Four women, A, B, C, and D, check their hats, and the hats are returned in a random manner. Let Ω be the set of all possible permutations of A, B, C, D. Let $X_j = 1$ if the jth woman gets her own hat back and 0 otherwise. What is the distribution of X_j? Are the X_i's mutually independent?[2]

Group 4: In London, half of the days have some rain. The weather forecaster is correct 2/3 of the time—i.e., the probability that it rains given that he has predicted rain, and the probability that it does not rain, given that he has predicted that it will not, are both equal to 2/3. When rain is forecast, Mr. Pickwick takes his umbrella. When rain is not forecast, he takes it with probability 1/3. Find

(a) the probability that Pickwick has no umbrella, given that it rains.

(b) the probability that he brings his umbrella, given that it does not rain.[3]

Group 1: Probability theory was used in the famous court case: *People v. Collins*. In this case a purse was snatched from an elderly person in a Los Angeles suburb. A couple seen running from the scene were described as a black man with a beard and a mustache and a blond girl with hair in a ponytail. Witnesses said they drove off in a partly yellow car. Malcolm and Janet Collins were arrested. He was black and, though clean shaven when arrested, there was evidence he recently had had a beard and a mustache. She was blond and usually wore her hair in a ponytail. They drove a partly yellow Lincoln. The prosecution called a professor of mathematics as a witness who suggested that a conservative set of probabilities for the characteristics noted by the witnesses would be as shone below.

man with mustache	1/4
girl with blond hair	1/3
girl with ponytail	1/10
black man with beard	1/10
interracial couple in a car	1/1000
partly yellow car	1/10

The prosecutor then argued that the probability that all of these characteristics are met by a randomly chosen couple is the product of the probabilities or 1/12,000,000, a very small number. He claimed this

[1] Charles M. Grinstead and J. Laurie Snell, *Introduction to Probability: Second Revised Edition* (Providence, RI: American Mathematical Society, 1997), p. 152.

[2] Charles M. Grinstead and J. Laurie Snell, *Introduction to Probability: Second Revised Edition* (Providence, RI: American Mathematical Society, 1997), p. 155.

[3] Charles M. Grinstead and J. Laurie Snell, *Introduction to Probability: Second Revised Edition* (Providence, RI: American Mathematical Society, 1997), p. 153.

was proof beyond a reasonable doubt that the defendants were guilty. The jury agreed and handed down a verdict of guilty of second-degree robbery.

The verdict was appealed, and in the subsequent trial the counsel for the defense argued as follows: Suppose, for example, there are 5,000,000 couples in the Los Angeles area and the probability that a randomly chosen couple fits the witnesses' description is 1/12,000,000. Then the probability that there are two such couples given that there is at least one is not at all small. Find just how "not small". (The California Supreme Court overturned the initial guilty verdict.)[4]

Group 5: Suppose the Registrar is carrying John and Mary's registration cards and drops them in a puddle. When he picks them up he cannot read the names but on the first card he picked up he can make out Mathematics 156 and Business 211, and on the second card he can make out only Mathematics 156. He asks you if you can help him decide which card belongs to Mary. You know that Mary likes business but does not like mathematics. You know nothing about John and assume that he is just a typical student. From this you estimate

$$
\begin{aligned}
P(\text{Mary takes Business 211}) &= 0.5,\\
P(\text{Mary takes Mathematics 156}) &= 0.1,\\
P(\text{John takes Business 211}) &= 0.3,\\
P(\text{John takes Mathematics 156}) &= 0.2.
\end{aligned}
$$

Assume that their choices for courses are independent events. [5]

6. **UMAP Project 340**. Complete the requirements of UMAP Module 340, "The Poisson Random Process," by Carroll O. Wilde. Probability distributions are introduced to obtain practical information on random arrival patterns, interarrival times or gaps between arrivals, waiting line buildup, and service loss rates. The Poisson distribution, the exponential distribution, and Erlang's formulas are used.

   Two problems are posed at the outset. They are

   > **Problem A**. Suppose that you live in an isolated community where fires break out at random at an average of 3 per day. If fires require an average of 1 hour to fight, how many firefighting units should your fire station have to make the community "safe?"
   >
   > **Problem B**. Suppose that you own a hardware store that carries brooms. Your merchandise is restocked only at the close of your business week, each Saturday afternoon. You have limited storage space and therefore wish to keep inventory levels at a minimum. If customers who buy brooms arrive at random times and at an average rate of 10 per week, how many brooms should you have on hand each Monday morning?

7. **UMAP Project 327**. Complete the requirements of UMAP Module 327, "Nuclear Deterrence. Applications of Elementary Probability to International Relations." The abstract says:

   > This module is designed to apply mathematical models to nuclear deterrent problems, and to aid users in developing enlightened skepticism about the use of linear models in stability analyses and long-term predictions. An attempt is made at avoiding overwhelming complexities through concentration on land-based missile forces. It is noted that after the Second World War, the United States' monopoly on nuclear weapons permitted it to attempt

[4] Charles M. Grinstead and J. Laurie Snell, *Introduction to Probability: Second Revised Edition* (Providence, RI: American Mathematical Society, 1997), p. 153, 202, who cite M. W. Gray, "Statistics and the Law," *Mathematics Magazine*, vol. 56 (1983), pp. 67–81.

[5] Charles M. Grinstead and J. Laurie Snell, *Introduction to Probability: Second Revised Edition* (Providence, RI: American Mathematical Society, 1997), p. 159.

> to deter aggression through a policy of "massive retaliation." The history of change from this policy is outlined, and the chances for reaching an equilibrium in the nuclear arms race are examined through probabilistic models.

UMAP Module 311, "The Geometry of the Arms Race," is listed as being related.

Group 2: **UMAP Project 346**. Complete the requirements of UMAP Module 346, "Error Correcting Codes I. Applications of Elementary Algebra to Information Theory." The abstract says:

> It is noted that with the prominence of computers in today's technological society, digital communication systems have become widely used in a variety of applications. Some of the problems that arise in digital communications systems are described. This unit presents the problem of correcting errors in such systems. Error correcting codes are developed as an application of linear algebra and finite field algebra. The models chosen for this module were selected for relative ease of comprehension. The material includes exercises and a model examination.

UMAP Module 336, "Aspects of Coding" and UMAP Modules 337, "A Double-Error Correcting Code," are listed as related.

9. **UMAP Project 538**. Complete the requirements of UMAP Module 538, "Random Walks and Fluctuations." The module begins with the following voting problem: "Suppose Peter won an election over Paul. If Peter received P votes, and Paul received Q votes, what is the probability that Peter, the winner, was always ahead throughout the counting of the votes?" In the abstract, the author expresses his intent (for the module) this way:

> This expository paper on fluctuations in random walks presents new and simplified proofs of several theorems, including the arc sine law for sojourn times. The theorems are proved by simple geometric transformations of the paths. The graphical presentation has visual appeal and makes the results accessible to probability students at any level. This module will enable students to understand some of the counterintuitive results of probability theory in applications common to everyday life.

10. **UMAP Project 649**. Complete the requirements of UMAP Module 649, "Continuous Time, Discrete State Space Markov Chains." The author begins by stating this problem:

> Consider a movie marquee that consists of a large number N of light bulbs. Each of the bulbs burns independently of the others, but will of course burn out at some finite time. The replacement policy is to replace all the burned out bulbs when *fewer than n* bulbs are operating. Among the questions we might want a mathematical analysis to answer are: How much time elapses between replacements? What is the fraction of the time during which 80 percent, say, of the bulbs are operating? As n decreases, how does the time between replacements increase?

Though we have not yet discussed Markov chains, we will be doing so soon, followed soon after by an introduction to differential equations. Thus, while the topics listed as "prerequisites" are probably already understood (mostly) by someone in your group, they should be accessible to all shortly.

Group 4: **UMAP Project 687**. Complete the requirements of UMAP Module 687, "Fire Control Land Management in the Chaparral." The abstract reads:

> Catastrophic wildfires are an ever-present danger in chaparral wilderness areas. This module develops a Markov-chain model that can be used to evaluate strategies for chaparral land management and fire control. The model is applied to determine an optimal prescribed-burning policy.

We will be discussing Markov chains in class this coming Monday.

Group 1: **UMAP Project 698**. Complete the requirements of UMAP Module 698, "Rating Systems for Human Abilities: The Case of Rating Chess Skill." The abstract says:

> This module illustrates how to construct and use a mathematical model for rating chess ability. The model is probabilistic in character, and it is based on Arpad Elo's chess rating system that is used both nationally and internationally to rate chess players. The key assumption is that the probabilities of a win, loss, or draw between two chess players each depend only on the difference between the two players' ratings. When this assumption is suitably augmented with other assumptions, a system is developed that rates newcomers as well as changes the ratings of established tournament players. Some empirical validation of the system is discussed.

Group 3: **UMAP Project 792**. Complete the requirements of UMAP Module 792, "The Spread of Forest Fires." The author describes it thus:

> This is a project that begins with a very approachable model for the spread of a forest fire. The question of interest is whether the fire will die out on its own, and this leads to a hunt for a threshold value for the transmission probability. With the geometric series and a minimal amount of probabiliity, a partial answer is attainable. Namely, some bounds are determined on the critical probability of spread of the fire. This is a common starting point for such questions, and therefore provides a good example of what a research question looks like and how one might start to close in on the solution.

14. **Computer simulation**. You are to construct a simulation for a bank so that its managers can study the time customers have to wait in line. The bank gives you the following assumptions and simulation requirements.

    - The time between customers' arriving varies from a minimum of $\mathit{Arrive}_{\text{min}}$ to a maximum of $\mathit{Arrive}_{\text{max}}$, and any time in between is equally likely (in other words a random number from a uniform distribution).
    - The amount of time required to serve a customer varies from a minimum of $\mathit{Service}_{\text{min}}$ to a maximum of $\mathit{Service}_{\text{max}}$, again with any time in between equally likely.
    - The bank opens its doors and stays open for T minutes. After T minutes, the doors close to prevent new arrivals, but customers currently in line continue to be served. Thus the simulation may run longer than T minutes.
    - There is one bank teller. (But try to design your simulation, if possible, to leave open the possibility of increasing to n tellers.)
    - A typical day begins with the bank opening. The first customer will arrive between $\mathit{Arrive}_{\text{min}}$ and $\mathit{Arrive}_{\text{max}}$ minutes later. The customer goes straight to the teller. The next customer arrives. If the first customer is finished, then the new customer goes straight to the window. If the first customer is still being served, the new customer starts a line or queue. Whenever a customer arrives, that person will go to the teller if the teller is unoccupied or to the end of the line if the

teller is occupied. Whenever a teller finishes serving the current customer, the customer at the front of the queue leaves the queue and goes to the counter. If there is currently no queue formed, then when the current customer is through with service, the teller is idle until the new customer arrives.

(a) Build a simulation that satisfies these requirements. The simulation parameters should be well-named and easy to change so that bank managers can study different scenarios. The simulation should keep track of and report the number of customers, the average wait per customer (the time a customer waits is the time spent in line and does not count the time being served), the percent of time that the teller was occupied, and the length of the simulation (time until the last customer leaves).

Run your model assuming the bank is open $T = 480$ minutes, using the following sets of parameters:

i. $Arrive_{\text{min}} = 1$, $Arrive_{\text{max}} = 3$, $Service_{\text{min}} = 1$, and $Service_{\text{max}} = 15$.
ii. $Arrive_{\text{min}} = 1$, $Arrive_{\text{max}} = 3$, $Service_{\text{min}} = 1$, and $Service_{\text{max}} = 3$.
iii. $Arrive_{\text{min}} = 1$, $Arrive_{\text{max}} = 15$, $Service_{\text{min}} = 1$, and $Service_{\text{max}} = 3$.

Comment on why the results behave as they do.

(b) Revise your model to have n tellers. When more than one teller is open, a decision needs to be made about which teller the next arrival should go to. Number your tellers and always send a customer to the available teller with the smallest number. This corresponds to a bank where the queue starts at one end of a row of tellers. Think about how you would model a queue that begins at the middle of a row of tellers.

Repeat your runs performed above (same parameter sets), once with $n = 2$, and once again with $n = 3$.

(c) Your inter-arrival times have thus far been generated from a uniform distribution. Carry out simulations employing some other realistic manner for generating inter-arrival times.[6]

15. **Hands-on Project**. Pick some random process which you believe might be Poisson in nature, and that you could observe. One possibility might be the arrival of cars during a single cycle of a traffic light. (Should this be the process you observe, I can imagine doing so in the warmth of the cross-walk above the East Beltline.) Observe this process for an extended amount of time (which translates into taking many samples), and do it during several intentionally-chosen periods of the day/night. Keep track of both the random variable you believe to have a Poisson distribution (probably *counts of arrivals* of some sort, as well as a related random variable you believe to have an exponential distribution.

Once you have amassed data for several different periods of time, analyze it. That is, determine specific parameters for the distributions (determine these sets of parameters once for each time frame in which you made observations; then do so again with the data amalgamated into one whole). Plot the data along with the distributions to get a sense of how good is the fit. Follow this up by conducting a goodness-of-fit test (if you do not know, from some previous course, what this is, an internet search will provide a sufficient list of hits describing it) which gives a more objective measure of whether the fit is a good one. To the extent possible, write code (a computer program) to carry out/automate your goodness-of-fit calculations.

[6] Douglas Mooney and Randall Swift, *A Course in Mathematical Modeling* (Washington D. C.: Mathematical Association of America, 1999), p. 354 ff.