Economics 361
Problem Set #1

Jun Ishii *
Department of Economics
Amherst College

Fall 2022

The answers for many of these problems can be found in various texts/websites. But that's no fun (and I'll "get you" come exam time). Consider these practice and solve them in good faith.

Question 1: Warm-up

Using just Kolmogorov's Axioms of Probability and the properties of set operation, prove the following properties of a probability function $P(\cdot)$

- $P(\emptyset) = 0$
- $P(A^c) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- If $A \subset B$ then $P(A) \leq P(B)$

HINT: Think about **partitions**

*Office: Converse Hall 315 Phone: (413) 542-2901 E-mail: jishii@amherst.edu

Question 2: Choice under Uncertainty

(a) "Urn" problems have been a part of probability classes for many years. Their basic structure, consistent with classical notions of probability, allow for the study of some complex concepts – notably those associated with decision theory. The basic structure is as follows: there are balls in some urn; the probability of drawing any one ball in an urn is the same. So, each ball represents a distinct, equally likely outcome to the random experiment of "drawing balls from an urn."

Consider the following "urn" problem

> Two urns contain red and green balls, identical except for color. Urn A has two red and one green ball. Urn B has 101 red and 100 green balls. An urn is randomly chosen, with A and B equally likely to be chosen. You are asked to identify whether the chosen urn is A or B based on two successive draws of balls from the chosen urn. However, after the first ball is drawn and observed, you may choose whether the drawn ball is put back into the urn prior to the second draw – i.e. should the second draw be a "draw with replacement" or a "draw without replacement." To maximize your chances of identifying the urn correctly, how should you (1) choose whether to replace the drawn ball (2) decide which urn was chosen?"

Solve the above problem.

(b) A pharmaceutical firm has developed a new drug that it would like to market. But in order to do so, it must first receive approval from the government. The government will approve the drug if the drug passes both an efficacy test (Test **A**) and a safety test (Test **B**). Additionally, the drug must pass the two different tests in **successive** attempts, although the order of the two tests does not matter (A before B or B before A).

Due to the high costs of these tests, the firm can only afford to conduct three tests. The probability that the drug will pass **A** is higher than the probability that the drug will pass **B**. The tests are independent of each other, in that outcomes of prior tests do not affect the outcomes of later tests. (In industry parlance, the tests are "random trials")

To maximize the probability of the drug receiving government approval, how should the firm schedule the three tests: (AAA), (AAB), (ABB), (BBB), (BBA), (BAA), (ABA), or (BAB)? Explain.

Question 3: Bayes' Rule and Monty Hall

One of the more useful properties of a probability function is **Bayes' Rule**:

$$P(A|B) \quad = \quad \frac{P(B|A)P(A)}{P(B)}$$

(a) Using the definition of conditional probability, prove the above Bayes' Rule.

Arguably the most (in)famous mathematical puzzle involving Bayes' Rule is the "Monty Hall problem," named after the host of the 1960/70s TV game show, "Let's Make a Deal."

Problem: There are 3 doors (A-B-C), behind one of which is the prize. There is an equal probability that the prize will be behind any one of the doors. The host asks you to pick a door. The host then, among the two doors not chosen, reveals the one that does not have the prize. (In case both doors do not hide the prize, the host picks one at random with equal probability.) You are now given the choice either to stick with your original door or switch to the unchosen, unrevealed door. *Should you switch?*

Without loss of generality, suppose you pick door A and the host reveals door B.

(b) Calculate the following probabilities

- P(Opens B | A has prize)
- P(Opens B | B has prize)
- P(Opens B | C has prize)
- P(Opens B)

(c) Now, using Bayes' Rule, calculate

- P(A has prize | Opens B)
- P(C has prize | Opens B)

(d) Should you stick with A or switch to C?

Economic Application: Probability theory is useful in formalizing the process by which we *update* our *belief* structure when given new *information.* In the Monty Hall example, the contestant was no worse *ex ante* choosing **A** over **C**. It was the new information Monty provided *ex post* that made one more desirable. An economic problem similar to the Monty Hall problem is the standards problem: a firm must choose among 3 different standards (think media standards like MPEG formats). Only one of the standards will survive in the long-run. The firm chooses one and finds out that one of the other two standards has failed. Should it stick with its current choice or swap to the other remaining standard? Note that the solution to this will depend on the firm's initial beliefs about the probability of success for each standard: $P(A), P(B), P(C)$.

(e) Now try answering this variant of the Monty Hall problem (credited to the late Harvard statistician Frederick Mosteller) often dubbed as "Prisoner's Dilemma meets Monty Hall"

> "Three prisoners, A, B, and C, with apparently equally good records have applied for parole. The parole board has decided to release two of the three, and the prisoners

know this but not which two. A warder friend of prisoner A knows who are to be released. Prisoner A realizes that it would be unethical to ask the warder if he, A, is to be released, but thinks of asking for the name of one prisoner *other than himself* who is to be released. He thinks that before he asks, his chances of release are $\frac{2}{3}$. He thinks that if the warder says, 'B will be released,' his own chances have gone down to $\frac{1}{2}$, because either A and B or B and C are to be released. And so A decides not to reduce his chances by asking. However, A is mistaken in his calculations. Explain."

Question 4: Real World Events

This question concerns a debate between two prominent economists, Sendhil Mullainathan and Rajiv Sethi, concerning racial bias and police shootings.

Read the following short articles/blog posts

- "Police Killings of Blacks: Here is What the Data Says," Sendhil Mullainathan, NY Times 10/16/2015 https://www.nytimes.com/2015/10/18/upshot/police-killings-of-blacks-what-the-data-says.html
- "Threats Perceived When There Are None," Rajiv Sethi, blog post, 10/16/2015 http://rajivsethi.blogspot.com/2015/10/threats-perceived-when-there-are-none.html
- "It's all about the denominator: Rajiv Sethi and Sendhil Mullainathan in a statistical debate on racial bias in police killings," Andrew Gelman, blog post 10/21/2015 http://andrewgelman.com/2015/10/21/its-all-about-the-denominator-and-rajiv-sethi-and-sendhil-mullainathan-in-a-statistical-debate-on-racial-bias-in-police-killings/

For the purposes of this question, let the relevant random experiment be defined as an encounter between a police officer and some community resident. Some of the relevant events associated with outcomes of this random experiment are

- `Black`: outcomes where the resident was African American
- `Arrested`: outcomes where the resident was arrested by the police officer
- `Shot`: outcomes where the resident was shot by the police officer

For simplicity, assume that $\texttt{Shot} \subset \texttt{Arrested}$; all residents who were shot were also arrested.

Within this framework, the "31.8 percent of people shot by the police were African-American" quoted by Sendhil Mullainathan may be interpreted as an estimate of $\mathrm{P}(\texttt{Black} \mid \texttt{Shot})$. Similarly, the "28.9 percent of arrestees were African-American" as an estimate of $\mathrm{P}(\texttt{Black} \mid \texttt{Arrested})$

(a) Express $\mathrm{P}(\texttt{Black} \mid \texttt{Arrested})$ in terms of $\mathrm{P}(\texttt{Black} \mid \texttt{Shot})$, $\mathrm{P}(\texttt{Black} \mid \texttt{Arrested and Not Shot})$, and $\mathrm{P}(\texttt{Shot} \mid \texttt{Arrested})$

(b) Use the above framework to explain Mullainathan's hypothesis concerning why African Americans are seemingly disproportionately likely to be police shooting victims.

(c) Use the above framework to explain Sethi's critique of Mullainathan's hypothesis.

Note: **(b)** and **(c)** are fairly open-ended problems, with several acceptable answers.

Food for Thought: COVID Testing on Campus

The following problems will not be graded. But you should think about them as we may discuss them in class; they are fair game for quizzes/exams.

Consider a college campus with a student population of S. Public health officials consider a college campus "safe" (with respect to COVID) if no more than $\underline{\theta} \in (0, 0.5)$ share of the student population has COVID. In other words, a "safe" campus is one where no more than $\underline{\theta} \cdot S$ students have COVID.

(a) Suppose the college had access to a COVID test with absolute accuracy. Moreover, assume that the college administered the test sequentially, waiting for the result of the prior test before proceeding to the next. What is the minimum number of students (in terms of $\{S, \underline{\theta}\}$) the college would have to test before it could determine, with absolute accuracy, whether the campus is "safe" ?

(b) Suppose the college only had a test that was accurate with probability $(1-\alpha)$ with $\alpha \in (0, 1)$. In other words, there is a probability α that the test returns the wrong result (negative when positive, positive when negative) for any given student. How many students (in terms of $\{S, \underline{\theta}, \alpha\}$) would the college have to test before it could determine, with absolute accuracy, whether the campus is "safe" ?

(c) How would your answer to **(b)** change if you were able to test students multiple times, with an accuracy of $(1 - \alpha)$ each time, independent of previous test outcomes?

(d) Suppose the college had access to a COVID test with absolute accuracy but that administering the test was costly (time and money)! As such, suppose the college were willing to settle for a campus safety determination with at least $(1 - \alpha)$ accuracy. Could you solve for the minimum number of students (in terms of $\{S, \underline{\theta}, \alpha\}$) the college would need to test in order to achieve this *relaxed* goal? If so, how? If not, what other information would you necessarily need? Explain.

Students live in various residential halls on campus. COVID, as an airborne disease, often leads to infection clusters. Students living in close proximity to an infected student are much more likely to become infected themselves. Students also self-sort into residential halls. In particular, students tend to sort into groups with similar risk attitudes toward COVID.

(e) Given the above – you may also continue to assume a COVID test with absolute accuracy – should the college "target" certain residential halls for testing more than others in order to improve the accruacy with which it determines campus safety? In other words, for a given amount of testing the college has chosen to administer, should the college spread those tests across the various residential halls or concentrate some of those tests on particular residential halls in order to improve the college's testing efficacy? Explain. Note: this is an open-ended problem, one for which there is not a unique "answer."