Choices and Consequences in Computing	Homework 3
Jon Kleinberg and Karen Levy	Due 12 noon, March 10

The homework is due at 12 noon, Thursday, March 10, 2022. As noted on the course home page, homework solutions should be submitted by upload to Canvas. The file you upload must be in PDF format. You can create a separate file with the solutions (you don't need to repeat the questions); it is fine to create the homework using any format as long as it's handed in as a single PDF file.

The point-value for each part is given with the questions. We will give partial credit for progress toward an answer, even if it isn't complete.

We can accept late homework without a grade penalty provided that it is accompanied by a University-approved reason that is conveyed to us prior to the deadline. (These include illness and family emergencies.) Late homework with grade penalties are covered by the late-homework policy posted on Canvas.

Background: The questions below are primarily based on the material in the lectures through Feb. 25, and the associated readings including the "Notes on Bayesian Models" in the "Lecture Notes" module on Canvas.

---

1 Choosing a Media Source

We've seen in class how media sources can differentiate themselves in the information they provide, including when the information being provided is an interpretation of an underlying set of shared observations. Bayes' Rule can be a way to build models for the choice among these sources. Here we'll look at how this plays out in an example involving hypothetical medical information.

Suppose that you're following the progress of a new vaccine that's being developed, and it's currently in the middle of a trial whose results are not yet public. The vaccine will either turn out to be chemically active or not in combatting the disease it's designed for. We'll use "active" as shorthand for "chemically active in combatting the disease," which is the good outcome that people are hoping for. Whether it's active is what the vaccine trial is attempting to determine (though of course it can never provide an answer with absolute certainty).

There's a federal agency F that's monitoring the vaccine trial, and based on how things have gone, they've formed an *internal evaluation* of the trial whose outcome is not yet public. The possible outcomes for this internal evaluation are *Low*, *Medium*, or *High*, based on the

strength of evidence offered by the trial. Again, the outcome is known inside the agency F, but you don't know it.

One thing you do know about how the outcomes work is the following:

- If the vaccine is active, then the probability of a Low outcome is .1, the probability of a Medium outcome is .4, and the probability of a High outcome is .5.
- If the vaccine is not active, the the probability of a Low outcome is .85, the probability of a Medium outcome is .1, and the probability of a High outcome is .05.

You start with a prior probability of .25, in the absence of any of the upcoming evidence from the trial, that the vaccine is active.

Now suppose you learn that the internal evaluation from the federal agency F was High. You'd like to use Bayes' Rule to work out the conditional probability that the vaccine is active, given this High evaluation. To do this, you could declare A to be the claim that the vaccine is active, and E to be the evidence that the trial received a High evaluation; in this case, the value you're looking for is $\Pr[A \mid E]$.

(1.1) *(4 points) Use Bayes' Rule to determine* $\Pr[A \mid E]$*, the probability that the vaccine is active given that the vaccine trial received a High evaluation. Give an explanation in which you show how you set up the calculation.*

(1.2) *(4 points) Now let's ask the same question in the case of a Medium evaluation: Use Bayes' Rule to determine the probability that the vaccine is active given that the vaccine trial received a Medium evaluation. Give an explanation in which you show how you set up the calculation.*

(1.3) *(4 points) And finally the same question in the case of a Low evaluation: Use Bayes' Rule to determine the probability that the vaccine is active given that the vaccine trial received a Low evaluation. Give an explanation in which you show how you set up the calculation.*

All of this isn't directly useful to you at the moment, because the federal agency F's internal evaluation isn't yet public. So instead you turn to social media. In particular, there are two Twitter accounts X and Y, each run by people who have contacts at the federal agency F that is monitoring the vaccine trial. These contacts informally let them know the outcome of the internal evaluation (Low, Medium, or High).

The two accounts handle this information differently.

- Account X tends to see the good news in things, and so if they hear that the evaluation was either Medium or High, they tweet something like, "Things are looking promising for the vaccine trial." If they hear that the evaluation was Low, they tweet something

like, "Things don't look good for the vaccine trial." In neither case does the account explicitly share the actual outcome of the internal evaluation; it just uses this outcome to determine what to tweet.

- Account Y is more cautious, so they tweet that things look promising for the vaccine trial only if the evaluation was High. If they hear the evaluation was either Medium of Low, they tweet that things don't look good for the vaccine trial. This account also doesn't explicitly share the actual outcome of the evaluation; it too just uses it to determine what to tweet.

You'd like to follow one of these two Twitter accounts in order to gain information about the vaccine trial.

(1.4) *(4 points) You want to follow one of these Twitter accounts provided that it has the following property: for each value of the evaluation (Low, Medium, High) that causes you to have a conditional probability at least .5 that the vaccine is active, you'd like the account to say that things look promising; and for each value of the evaluation (Low, Medium, High) that causes you to have a conditional probability less than .5 that the vaccine is active, you'd like the account to say that things don't look good. (You can use the conditional probabilities that you worked out in (1.1), (1.2), and (1.3) as part of this.)*

If this is what you're looking for, which of the two Twitter accounts should you follow? Give an explanation for why you would choose this account, and also for why you would not choose the other account.

2 Inferring Genres

We saw in class that Bayes' Rule can be used to make inferences about why users in a latent factor model chose particular items. In this question, let's try out some more examples of this principle.

Suppose you have a summer internship with a large news site that hosts several million videos; these videos are viewed by the several hundred million users of the site. The news site traditionally hasn't done much analysis of the viewing data for their videos, but they've started to try this, and as part of that they have built a latent factor model. Unlike the examples in class, this model is very large, since it includes values for the several hundred million users and the several million videos on the site, and it contains 300 genres.

As usual, in the latent factor model, a user first selects a genre using a given probability distribution over the set of genres. Once they have selected a genre, they then select a video *via* this genre, by selecting according to probabilities specified by the genre. As in our examples from class, multiple genres might lead a user to select the same video, if this video has positive probability within multiple genres.

To take a concrete example from the news site's video collection, one of their videos is a clip from a Cornell hockey game, which can be reached both via Genre 72, which roughly corresponds to "winter sports," as well as via Genre 96, which roughly corresponds to "videos about college life." Given the large number of videos on the site, any one video has a very low probability of being chosen, and that's true of the Cornell hockey clip. It has the following probabilities of being chosen via different genres:

- It has probability .00012 (which we can abbreviate as 1.2×10^{-4}) of being chosen by someone selecting from Genre 72.
- If has probability .00036 (which we can abbreviate as 3.6×10^{-4}) of being chosen by someone selecting from Genre 96.
- It has a probability of 0 of being chosen via any other genre (which is why it's not crucial for this question to discuss what the other genres in their model correspond to).

Now, when a user arrives at a video like this Cornell hockey clip, the site would like to guess which genre the user is currently selecting from (so that the site can decide what else to recommend to them at the moment). In other words, for a specific video that the user is watching, and a specific genre, the site would like to answer the question, *"What is the probability that the user is selecting from the specific genre, given that they are watching the specific video?"*

Let's try this for the Cornell hockey clip. We use the information given above, and a user who has the following probabilities over genres:

- *User genre probabilities:* The user we'll consider has a probability .9 of selecting according to Genre 72 (winter sports) and a probability .1 of selecting according to Genre 96 (college life). (For simplicity, we're considering a user who only has these two genres as interests, though in general of course a user could have positive probability for many genres.)

Suppose that this user watches the Cornell hockey clip.

(2.1) *(2 points) We'd like to compute the probability that the user is selecting from Genre 72, given that they're watching the Cornell hockey clip, by writing it as a conditional probability* $\Pr[A \mid E]$*, for some claim* A *and some evidence* E*. What would you choose for* A *and* E *so that* $\Pr[A \mid E]$ *is the value we're seeking?*

(2.2) *(4 points) Using your choice of* A *and* E *from (2.1), use Bayes' Rule to compute* $\Pr[A \mid E]$*, the probability that the user is selecting from Genre 72 given that they're watching the Cornell hockey clip. Give an explanation in which you show how you set up the calculation.*

To your surprise, your calculation for this example causes some concern at the Web site. It turns out that the data analysts working for the site have made up a simpler formula that they believe (incorrectly, it now turns out) is giving them the right answer to the probability that the user is selecting from a certain genre given that they're watching a particular video.

Their formula, for videos that can be chosen within two different genres, works as follows. If the video has a probability $a > 0$ of being chosen via Genre i, and a probability $b > 0$ of being chosen via Genre j, then when a user clicks on the video, they assume there is an $a/(a+b)$ probability that the user is selecting from Genre i.

When they sanity-checked the formula on some examples, they were reassured that it seemed to give intuitively sensible answers; in particular, if a is much larger than b, then the formula says that there is a high probability that the user was selecting from Genre i (reflecting its relatively greater popularity in Genre i), and if a is much smaller than b, then the formula says that there is a low probability (reflecting its relatively lower popularity in Genre i).

Despite this intuition, the formula isn't actually doing the right thing.

(2.3) *(2 points) With the information given by the latent factor model, what does the Web site's (incorrect) formula give as the probability that the user is selecting from Genre 72? (To confirm, this question is asking for a value that isn't in fact the correct probability, but it's useful to figure out what this is so that we can compare it with the correct value.)*

You explain to the people who run the site that you're getting a different probability than their formula gives in this example. They ask, "But what was wrong with our data analysts' formula?"

(2.4) *(2 points) Since the people who run the site don't have experience with Bayes' Rule calculations, give an explanation that describes at a high level (and without explicit use of the Bayes' Rule formula) what their formula is failing to take into account, leading it to produce incorrect answers in general. As part of this, describe (again at a high level) how your explanation applies to the user from parts (2.1) through (2.3).*

Assuming you worked out the probabilities correctly in (2.2), then you'll have found that although the "winter-sports" genre assigns a lower selection probability than the "college-life" genre to the Cornell hockey clip, the user in this question (with probabilities of .9 and .1 for these two genres respectively) is nevertheless more likely to be selecting from the winter-sports genre than from the college-life genre when they view the Cornell hockey clip.

(2.5) *(2 points) This type of phenomenon is widespread. Give an example from your own consumption of media where something similar happens: some item of content that you view, read, or listen to where (i) the item would naturally be viewed as belonging to two different*

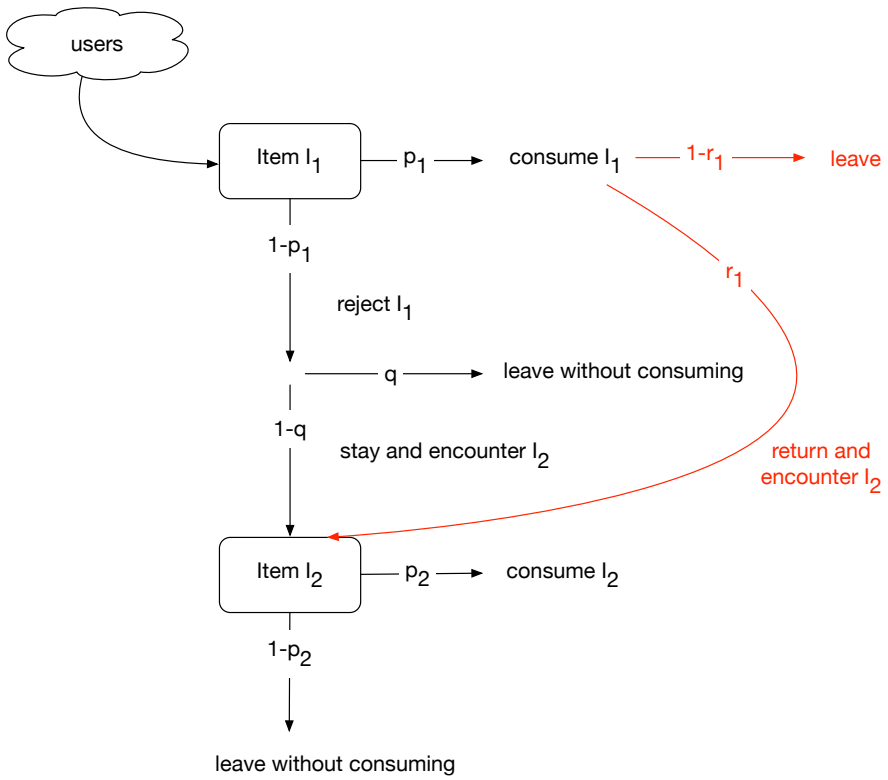

Figure 1: A representation of the progress of a user through the version of the ranking model with return from Homework 1.

genres i and j in a latent factor model; (ii) Genre i would arguably assign a higher selection probability than Genre j to this item; but (iii) you usually tend to consume the item via Genre j, which is the genre that assigns it a lower general selection probability.

As part of your answer, say what the piece of content is, what reasonable choices for Genre i and Genre j would be, and why you think conditions (i), (ii), and (iii) would be satisfied for this example. Clearly we can't know what an actual latent factor model would do in your example, so we're looking primarily for a qualitative explanation and argument to accompany your example.

3 Inferring a User's Past Browsing History

Bayes' Rule can also be used to analyze aspects of a user's behavior in the ranking model from earlier in the semester. Let's go back to Question 3 from Homework 1, in which we extended the ranking model to incorporate the idea that with some probability r_1, the user might return to the ranked list after consuming item I_1 and in this way encounter item I_2 as well.

To recap what we saw in this question from Homework 1, we had a ranking with two

items I_1 and I_2 in order, and users behaved as shown in Figure 1. Specifically:

- A user who encounters item I_1 has a probability $p_1 = 1/3$ of consuming I_1, and a probability $1 - p_1 = 2/3$ of not consuming it.
- If the user consumes I_1, then they have a probability $r_1 = 1/5$ of returning to the ranked list and directly encountering item I_2.
- If the user does not consume I_1, then they have a probability $q = 1/2$ of leaving the site due to impatience, and a probability $1 - q = 1/2$ of staying on the site to encounter item I_2.
- Finally, a user who encounters item I_2 has a probability $p_2 = 1/4$ of consuming I_2, and a probability $1 - p_2 = 3/4$ of not consuming it.

So in this model, for a user currently encountering item I_2, there are two possibilities for how they got to I_2: either they are here because (i) they encountered I_1, rejected it, and continued on to encounter I_2; or (ii) they encountered I_1, consumed it, and then returned to the list to encounter I_2 as well. It would therefore be natural to ask the following question, which is different from the kinds of questions that we asked on Homework 1:

> *(*) What is the probability that a user has consumed item I_1, given that they are currently encountering item I_2?*

This is a useful thing to estimate, given that you might want to know how likely it is that the user encountering I_2 has already experienced the contents of I_1. Our goal is to build up to an answer to this question in a couple of steps.

(3.1) *(2 points) You'd like to write the answer to Question (*) as a conditional probability* $\Pr[A \mid E]$*, for some claim A and some evidence E. What would you choose for A and E so that* $\Pr[A \mid E]$ *is the value we're seeking?*

(3.2) *(5 points) Using the numbers from Question 3 on Homework 1 (also summarized above), what is the value of the probability specified in Question (*)? That is, apply Bayes' Rule to your choice of A and E from (3.1), using these numbers from Homework 1, to compute* $\Pr[A \mid E]$*, the probability that a user encountering item I_2 has previously consumed item I_1.*

4 A Tale of Three Platforms

Suppose there are two dominant social media platforms, Platforms A and B. People use Platforms A and B primarily to share and comment on videos. They both have tens of

millions of users and their primary source of revenue is from advertisers whose ads appear just before user-generated video content.

Imagine that the CEO of Platform A begins to notice that its users seem to have lots of mistaken beliefs about the First Amendment and its role in society. The CEO of Platform A believes it's important that people understand the truth about the First Amendment. As a result, Platform A implements a policy in which the platform adds warning labels to all content that it deems to be "about" the First Amendment. It does this by flagging all videos with descriptions that contain the phrases "free speech," "First Amendment," or "1st Amendment" with a label that says "Hey there! This video seems to be about the First Amendment. A lot of people don't understand the First Amendment, so here are some resources to help you understand it better [with a link to a list of authoritative sources]." The CEO of Platform A has come to you to consult about the likely consequences of this approach.

(4.1) *(8 points) In 4-6 sentences, explain three possible issues that Platform A might encounter. These issues can be about the success of Platform A's policy/implementation, or the effects of the policy on the information ecosystem.*

Meanwhile, Platform B has decided to take a different approach. Platform B has a "heart" feature that people can use to endorse certain posts, and it has noticed that posts containing misinformation about the First Amendment tend to receive an unusually large number of hearts very quickly after they are posted. They think the source of this activity may be a coordinated misinformation campaign that is attempting to sow doubt about the First Amendment and is trying to boost these misinformation messages. As a result, Platform B decides to implement a new policy in which it temporarily throttles the display of content that is receiving a very large number of hearts soon after it is posted—that is, it slows down the rate at which these posts are shown to users on the platform. The CEO of Platform B has come to you to consult about the likely consequences of this approach.

(4.2) *(8 points) In 4-6 sentences, explain three possible issues that Platform B might encounter. These issues can be about the success of Platform B's policy/implementation, or the effects of the policy on the information ecosystem.*

Despite these efforts, misinformation about the First Amendment on Platforms A and B continues to receive attention as a serious public problem that is causing significant harm and misunderstanding. The CEOs of the platforms are called before Congress to testify about what they have been doing to combat these misunderstandings.

Shortly thereafter, Platforms A and B decide to try something new. Platforms A and B decide to develop a joint database they can both use for moderating misinformation about the First Amendment. The database will consist of a variety of assertions about the First Amendment, each labeled as "true" or "false" by an independent legal scholar. They

are hoping that this approach allows them to more specifically flag information about the First Amendment that is false (rather than just information about the First Amendment generally).

(4.3) *(8 points) In 4-6 sentences, describe at least three possible consequences of the approach taken by Platforms A and B. These consequences can be positive or negative.*

Suppose that shortly after this, a new small startup, Platform C, emerges. Platform C is very similar in functionality to Platforms A and B: it also facilitates sharing and commenting on videos, and is similarly ad-supported. But Platform C also has made some improvements based on the shortcomings of Platforms A and B; for example, it can load videos faster, has some additional commenting features, and has a less buggy mobile app than Platforms A and B do.

However, unlike Platforms A and B, Platform C plans to refrain from moderating misinformation about the First Amendment or any other topic. "Who are we to decide? We don't want to be in the business of being arbiters of truth, and we think our approach—not to mention our superior features!—will draw tens of millions of users who feels the same way," says Platform C's CEO.

(4.4) *(3 points) What would the Johnson and Post model of the "market for rules," described in Lecture 11, suggest about Platform C's likelihood of success? Explain in 2-3 sentences.*

(4.5) *(8 points) Explain at least three reasons we might not expect Platform C to be successful in its goal of reaching tens of millions of users, in 4-6 sentences total.*