Choices and Consequences in Computing	**Homework 1**
Jon Kleinberg and Karen Levy	**Due 12 noon, Feb. 8**

The homework is due at 12 noon, Thursday, February 8, 2024. Homework solutions should be submitted through Canvas, and from there use the embedded version of Gradescope inside Canvas. This way, Canvas can keep track of the fact that the assignment was submitted.

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.

Format: 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.

In the Gradescope interface on Canvas, it's important to complete the step in which you highlight where your answer to each question is in the PDF document.

Also, if the PDF you submit for the homework is a scan of a handwritten answer, it is important to make sure the handwritten answer is legible, and for questions involving mathematical notation, that you highlight where any of the main expressions or formulas are for each answer. We need to be able to recognize the answer in order to assign credit to it.

Late Policy: We can accept late homework provided that it is accompanied by a University-approved reason that is conveyed to us prior to the deadline via email. (These include illness and family emergencies.)

We will also accept homework that comes in late without one of these reasons subject to a grade penalty. Homework that comes in after noon on Thursday but before noon on Friday will be accepted with a grade deduction of 10% of the maximum score (e.g. if the homework is out of 70 points, then 7 points will be deducted). There will be an additional deduction of 10% more of the maximum score for each 24 hours after (i.e. 20%, 30%, and 40% before noon on Saturday, Sunday, and Monday respectively), until Monday at noon, after which the homework will not receive credit. Because Gradescope will be open during this time, the homework may be uploaded there directly; you do not need prior approval to do this.

Background: The questions below are primarily based on the material in the lectures through Feb. 2, and the associated readings, including the "Notes on Rankings" in the "Lecture Notes" module on Canvas.

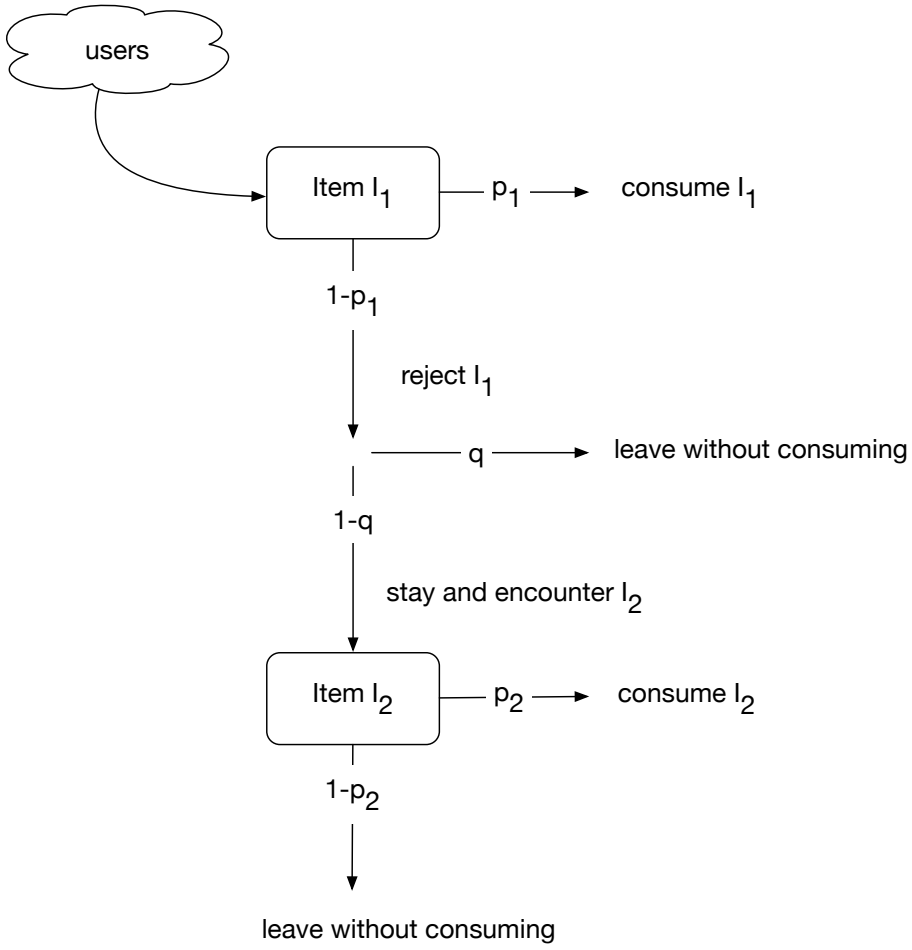

Figure 1: A representation of the progress of a user through the ranking model.

1 Inferring popularity from ranked items

Let's recall the ranking model we discussed in the lectures:

- *The user encounters a list of ranked items in order* $I_1, I_2, \ldots, I_n$.
- *When they encounter item* I_j *in the list:*
  - *They find* I_j *interesting with probability* p_j*, independently of their decisions about previous items.*
  - *If they find* I_j *interesting, they consume it and leave the platform.*
  - *If they don't find* I_j *interesting, they leave the platform anyway with probability* q *(due to their impatience).*
  - *Otherwise they move on to consider item* I_{j+1}.

For ease of reference, in Figure 1 we reproduce a diagram from the class lecture slides, showing how a user moves through a set of two items in this ranking model.

Imagine that you're helping a team of Web designers create a Web page with two videos; we'll call the videos I_1 and I_2, as in the notation from class, and suppose that I_1 is placed above I_2.

Let's use the ranking model to estimate how many users view each video in expectation. Suppose that $p_1 = 1/2$ and $p_2 = 2/5$, and $q = 1/2$, and let's say that 1000 users arrive at the site.

(1.1) *(5 points) Based on the values given above, what is the expected number of users who consume item I_1? What is the expected number of users who consume item I_2? What is the expected number of users who leave the platform without consuming anything? Give brief explanations for your answers.*

Now, suppose the platform doesn't actually have an estimate for the values of p_1, p_2, and q — that is, they're not sure how interesting each of the two videos is, and also not sure how impatient each user is. But they do observe the number of users who watch each video (and let's continuing assuming that each user watches at most one of the two videos).

Suppose that the number of users who watched each of the videos are the numbers you derived in (1.1). One set of values for p_1, p_2, and q that would produce these numbers of consumers of I_1 and I_2 in expectation are the numbers you used in (1.1), and suppose that the people who work at the platform have begun taking those numbers $p_1 = 1/2$, $p_2 = 2/5$, and $q = 1/2$ as their working hypothesis for the values of p_1, p_2, and q. In particular, they seem to be taking as a given that p_1 must be larger than p_2 based on the numbers of viewers they observed, and hence that video I_1 is intrinsically more interesting to viewers than video I_2.

You'd like to convince them that this isn't necessarily true — that it's possible that I_2 is more interesting to viewers than I_1, and it's only getting fewer viewers because of its position in the ranking.

You plan to do this as follows:

(1.2) *(5 points) Give a set of values for p_1, p_2, and q with the properties that (i) $p_2 > p_1$, and (ii) using this set of values, the expected number of users who consume each of I_1 and I_2 respectively is the same as the expected numbers you calculated in (1.1). Show the calculations that establish that the expected numbers of users consuming I_1 and I_2 are the same as they are in (1.1).*

If you're able to do this, you've established that from the numbers of viewers they've observed for the two videos, they don't have enough information to tell which of p_1 or p_2 is larger: it's consistent with the numbers of viewers they've observed that $p_1 = 1/2$ and $p_2 = 2/5$, in which case p_1 is larger; but it's also consistent with the numbers of viewers they've observed that p_1 and p_2 take the values you've provided in this answer to (1.2), in which case p_2 is larger.

2 Investing to Change Position

Let's consider a different case in which a Web page ranks videos in order, and again users move through the ranking according to the model from class (as described at the start of this homework). This time, however, there are three videos I_1, I_2, I_3, produced by three different content creators, and ranked in this order. (Whereas the picture in Figure 1 shows an example with only two videos.)

Suppose that the values from the model in this case are $p_1 = p_2 = p_3 = 1/2$ and also $q = 1/2$. Suppose also that 1600 users per hour arrive at the site.

(2.1) *(6 points) If the creator of each video earns \$1 in expectation from each user who consumes the video, what is the expected revenue of each of I_1, I_2, and I_3 per hour? Give a brief explanation for your answers.*

Now the Web site offers a promotional deal to the creator of video I_2: for \$250 per hour, it will move I_2 forward by one position, into first place.

(2.2) *(2 points) Should the creator of video I_2 accept this deal, if they want to maximize their expected revenue? Give an explanation for your answer.*

Suppose that the Web site offers a similar deal to the creator of video I_3: for \$250 per hour, it will move I_3 forward by one position, into second place. (You can assume that the Web site offers this deal after the deal in (2.2) was either accepted or not, depending on what you concluded in your answer to that question.)

(2.3) *(2 points) Should the creator of video I_3 accept this deal, if they want to maximize their expected revenue? Give an explanation for your answer.*

If you worked out the answers to (2.2) and (2.3) correctly, you should have gotten different answers. The team that runs the Web site hosting the videos is puzzled by this outcome, since, as they explain to you, "Both creators stood to gain the same thing from the deal: moving forward by one position. Why did one of them want the deal and the other one didn't want the deal?"

(2.4) *(3 points) Give an explanation that answers this question posed by the team running the Web site; in particular, why is it not correct to treat the creators of I_2 and I_3 the same for purposes of reasoning about the deal that was offered to each of them, on the grounds that they "stood to gain the same thing from the deal"?*

3 Two-Dimensional Rankings

Our model of rankings is inherently one-dimensional, in that we imagine that the items are placed in a linear order, and then a user reads through the order sequentially.

But there are other kinds of layouts we could consider for online content, and it is interesting to ask what effect these other layouts have on user attention. As an example,

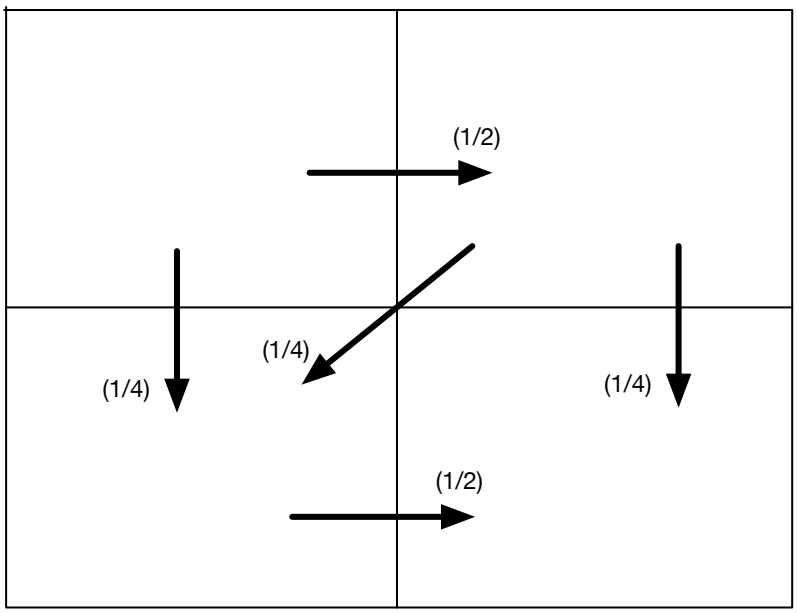

Figure 2: A representation of the progress of a user through a two-dimensional layout of content.

let's consider a simple model in which items are arranged in a two-dimensional grid rather than a one-dimensional list.

Specifically, suppose the designers of a news Web site have four articles I_1, I_2, I_3, I_4 that they would like to arrange in a two-by-two grid. To decide how many people consume each item, we need to establish some basic properties of how users move through this grid, encountering items along the way.

First, let's suppose that the designers put I_1 in the upper-left of the grid, I_2 in the upper-right, I_3 in the lower-left, and I_4 in the lower-right.

Next let's suppose that:

- a user who encounters I_1 decides to consume it with probability $p_1 = 1/3$,
- a user who encounters I_2 decides to consume it with probability $p_2 = 1/3$,
- a user who encounters I_3 decides to consume it with probability $p_3 = 1/5$, and
- a user who encounters I_4 decides to consume it with probability $p_4 = 1/10$.

Moreover, we'll assume that after the user consumes any one article they leave the site without looking at any others.

Finally, let's propose a simplified model for how people move through the grid. We'll suppose that every user starts in the upper-left. They then move through the grid as follows:

- If they are in the upper-left and decide not to consume the item that's there, they move to the upper-right with probability 1/2; they move to the lower-left with probability 1/4; and with the remaining 1/4 probability they leave without consuming anything.
- If they are in the upper-right and decide not to consume the item that's there, they move to the lower-left with probability 1/4; they move to the lower-right with probability 1/4; and with the remaining 1/2 probability they leave without consuming anything.
- If they are in the lower-left and decide not to consume the item that's there, they move to the lower-right with probability 1/2; and with the remaining 1/2 probability they leave without consuming anything.
- Finally, if they are in the lower-right and decide not to consume what's there, they simply leave without consuming anything.

These points about user movement through the grid are depicted schematically in Figure 2. To summarize, a user starts in the upper-left, decides whether to consume I_1, and if not, they move either to the upper-right (with probability 1/2) where they encounter I_2, or they move to the lower-left (with probability 1/4) where they encounter I_3, or they simply leave (with the remaining 1/4 probability). From whichever cell of the grid they end up in, they continue by following the part of the model appropriate to this cell.

Suppose that 9000 users start in the upper-left of the grid.

(3.1) *(5 points) For each of the items I_1, I_2, I_3, I_4, what is the expected number of users who consume this item? What is the expected number of users who leave the platform without consuming anything? Give brief explanations for your answers.*

Now suppose that the designers of the Web site decide to experiment by swapping the positions of items I_2 and I_3 — after the swap, I_2 is in the lower-left and I_3 is in the upper-right. The items I_1 and I_4 stay where they were.

(3.2) *(7 points) After the swap, for each of the items I_1, I_2, I_3, I_4, what is the expected number of users who consume this item? What is the expected number of users who leave the platform without consuming anything? Which items get higher expected consumption after the swap, which get lower expected consumption, and which stay the same? Give brief explanations for your answers.*

4 Videos about a Mayoral Election

Let's dig deeper into the three hypothetical videos described in Question 2. Imagine that the videos are the following:

- I_1 is a news video about two candidates running in a local mayoral election, Candidates A and B. The story is thorough, even-handed, and presents a balanced view of both

candidates' policy positions and qualifications for office, as well as information for readers about how they can register to vote in the election.

- I_2 is also a video about the two candidates running in the local mayoral election. However, this video doesn't contain any substantive information about either candidate's views or qualifications. Instead, it contains several insulting nicknames for both candidates, and features unflattering images of both candidates and their families.

- I_3 is also a video about the two candidates running for mayor, but it treats the two candidates very differently. It discusses Candidate A's accomplishments and qualifications positively, and gives thorough factual information about Candidate A's policy positions. But it includes only negative claims about Candidate B's qualifications, including several seemingly false statements about Candidate B's policy positions and fitness for public office.

Imagine that you work at the website hosting these videos and are providing advice to the company's CEO.

(4.1) *(8 points, please answer in 4-8 sentences) Recall from Question 2 that the website's ranking and display of videos can depend, in part, on content creators' willingness to pay an hourly fee in order to be ranked more highly on the site. Reflect on this policy through the lens of truth theory, drawing on relevant material from class. Explain one way in which this policy can be understood as compatible with the goals of truth theory. Explain one way in which this policy can be understood as in tension with the goals of truth theory.*

The website's CEO would like to remove videos I_2 and I_3 from the website, and gives you the following rationale: "These videos are not giving people high-quality information about the mayoral candidates. I think it's important for people to make well-informed decisions about whom to vote for. Only I_1 is providing high-quality unbiased information. So we're going to take the other videos down."

(4.2) *(8 points, please answer in 4-8 sentences) Reflect on this decision through the lens of democracy theory, drawing on relevant material from class. Explain one way in which the CEO's decision is compatible with the goals of democracy theory. Explain one way in which the CEO's decision is in tension with the goals of democracy theory.*

After their video is removed from the site, the creator of video I_2 (whom we'll call Creator 2) asserts that the website has violated the creator's First Amendment rights. Creator 2 asserts that the website has infringed upon their free speech rights by removing I_2 from the platform.

(4.3) *(8 points, please answer in 4-8 sentences) Drawing on relevant information from class, do you predict that Creator 2's First Amendment claim against the website will be successful? Explain why or why not.*

Imagine that the CEO thinks about the decision some more, and decides that it while they still want to remove video I_2 from the website, they can see some justification for keeping video I_3 up on the website instead of taking it down. After all, video I_3 does contain some valuable substantive information about Candidate A's qualifications and policy positions. However, in light of the seemingly false claims that video I_3 also contains about Candidate B, the CEO decides to place a warning label underneath the video that reads "This video contains false claims about Candidate B's qualifications for political office."

The creator of video I_3 (whom we'll call Creator 3) is unhappy with this decision and asserts that the website has violated Creator 3's First Amendment rights. Creator 3 disputes that video I_3's claims about Candidate B are false, and asserts that the website's warning label has drastically reduced viewers' engagement with Creator 3's videos, which has also reduced their ad revenue.

(4.4) *(8 points, please answer in 4-8 sentences) Drawing on relevant information from class, do you predict that Creator 3's First Amendment claim against the website will be successful? Explain why or why not.*

Imagine that Candidate B is also unhappy about the claims made by Creator 3 in video I_3. Candidate B asserts that the claims made about them in the video are false, that the claims damage Candidate B's reputation, and that Creator 3 knew that those claims were false and made them maliciously.

(4.5) *(3 points, please answer in 2-4 sentences) What kind of legal claim will Candidate B try to make against Creator 3? If Candidate B is able to prove that the elements of that claim exist here, will Creator 3 be able to successfully argue that they have a First Amendment right to make these statements anyway? Why or why not?*