CS 217: Artificial Intelligence and Machine Learning

Lecture 2: Human-Centered AI

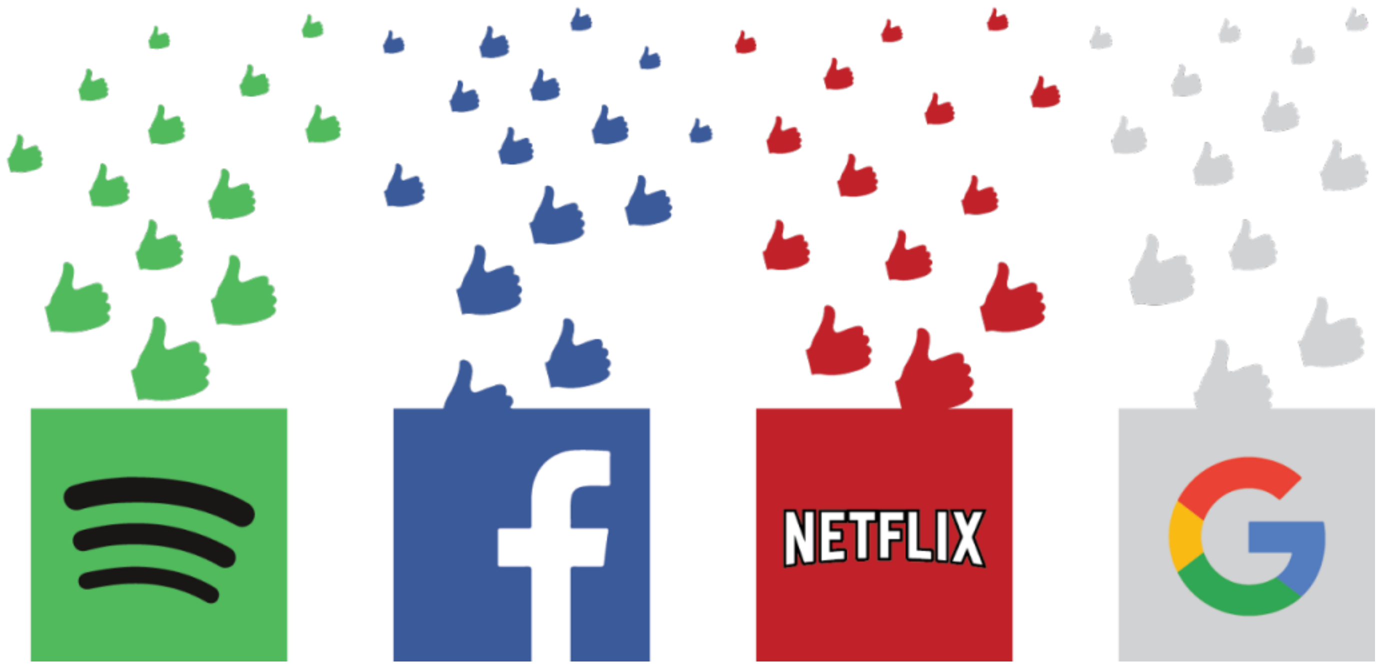
Arpit Agarwal

Computer Science & Engineering,

IIT Bombay

Recommendation Systems

- Understand user-preferences and filter content based on preferences
- **Goal**: Enhance user experience, increase engagement and sales.
- **Applications**: E-commerce (Amazon), Media (Netflix, Spotify), Social Networks (LinkedIn, Facebook).



Netflix Prize

NETFLIX

Netflix Prize

Home | Rules | Leaderboard | Register | Update | Submit | Download

Leaderboard

Display top 40 leaders.

Rank	Team Name	Best Score	% Improvement	Last Submit Time
--	No Grand Prize candidates yet	--	--	--
Grand Prize - RMSE <= 0.8563				
1	PragmaticTheory	0.8584	9.78	2009-06-16 01:04:47
2	BellKor in BigChaos	0.8590	9.71	2009-05-13 08:14:09
3	Grand Prize Team	0.8593	9.68	2009-06-12 08:20:24
4	Dace	0.8604	9.56	2009-04-22 05:57:03
5	BigChaos	0.8613	9.47	2009-06-15 18:03:55
Progress Prize 2008 - RMSE = 0.8616 - Winning Team: BellKor in BigChaos				
6	BellKor	0.8620	9.40	2009-06-17 13:41:48
7	Gravity	0.8634	9.25	2009-04-22 18:31:32
8	Opera Solutions	0.8640	9.19	2009-06-09 22:24:53
9	xlvector	0.8640	9.19	2009-06-17 12:47:27
10	BruceDengDaoCiYiYou	0.8641	9.18	2009-06-02 17:08:31
11	Ces	0.8642	9.17	2009-06-12 23:04:25
12	majia2	0.8642	9.17	2009-06-15 03:35:00
13	xiangliang	0.8642	9.17	2009-06-13 16:35:35
14	Feeds2	0.8647	9.11	2009-06-16 22:21:19
15	Just a guy in a garage	0.8650	9.08	2009-05-24 10:02:54
16	Team ESP	0.8653	9.05	2009-06-16 05:25:11
17	pengpengzhou	0.8654	9.04	2009-05-05 18:18:03
18	NewNetflixTeam	0.8657	9.01	2009-05-31 07:30:22
19	J Dennis Su	0.8658	9.00	2009-03-11 09:41:54
20	Vandelay Industries !	0.8658	9.00	2009-05-11 00:43:14

The general paradigm

Estimate

- Q1 = Pr(click | user, item)
- Q2 = E[rating | user, item]
- Q3 = Pr(share | user, item)
- Q4 = Pr(purchase | user, item)
- ….

Rank items based on the weighted score

W1*Q1 + W2*Q2 + W3*Q3 + W4*Q4 ….

This class

- Focus on various types of biases
- Won’t go into design details of recommendation systems

Feedback Loops

Predict which content will be liked by user → Recommendation Systems → Show more content of that type

Self-fulfilling Prophecies!

Bias in Recommendation: A Simple Example

- Two content types: Red and Blue
- Consider a fixed user (or homogenous set of users)
- Equal preference for Red and Blue

$$\Pr(\text{click} \mid \text{shown red}) = \Pr(\text{click} \mid \text{shown blue}) = \lambda$$

At each time recommender recommends a single item

Bias in Recommendation: A Simple Example

- For t = 1, 2, ……
  - User arrives at the platform
  - Recommend item based on past click

$$P(t) = \frac{R(t)}{N(t)}$$

  - Pr(show red at time t) = $P(t)$
  - Pr(show blue at time t) = 1− $P(t)$

$R(t)$ = Total number of clicks on red before time t

$N(t)$ = Total number of clicks before time t

Bootstrap the new recommender based on legacy system

At t=1, α clicks on red, and β clicks on blue

- $P(1) = \frac{\alpha}{\alpha+\beta}$

Bias in Recommendation: A Simple Example

- What would be $P(t)$ as $t \to \infty$?
  - Will it converge to ½?
  - Will be be biased towards the initial configuration?

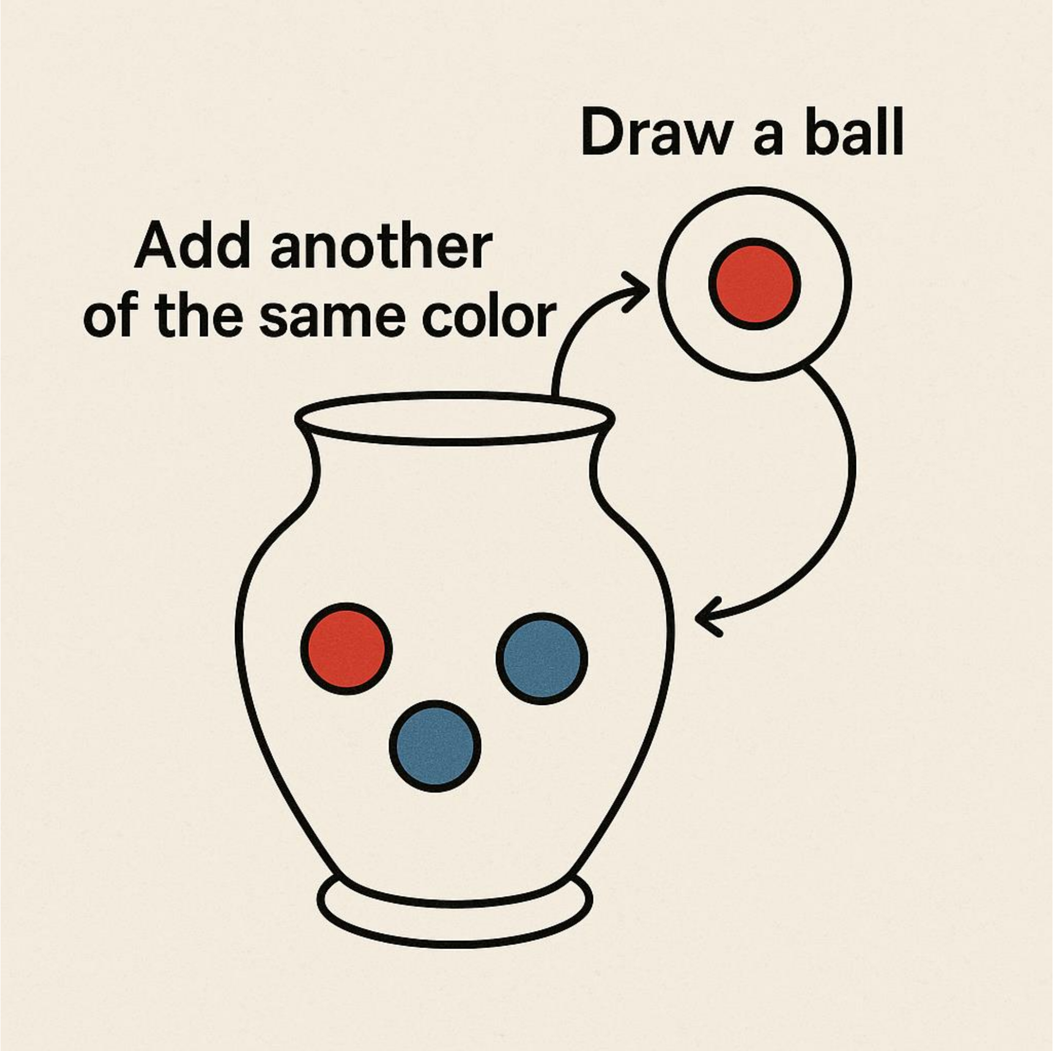
Polya’s Urn

- Initially- Urn with α red balls, and β blue balls

- For t = 1, 2, ……
  - Draw a balls uniformly at random from the Urn
  - **Return it to the urn along with an additional ball of the same color**.
  - $$P(t) = \frac{R(t)}{N(t)}$$

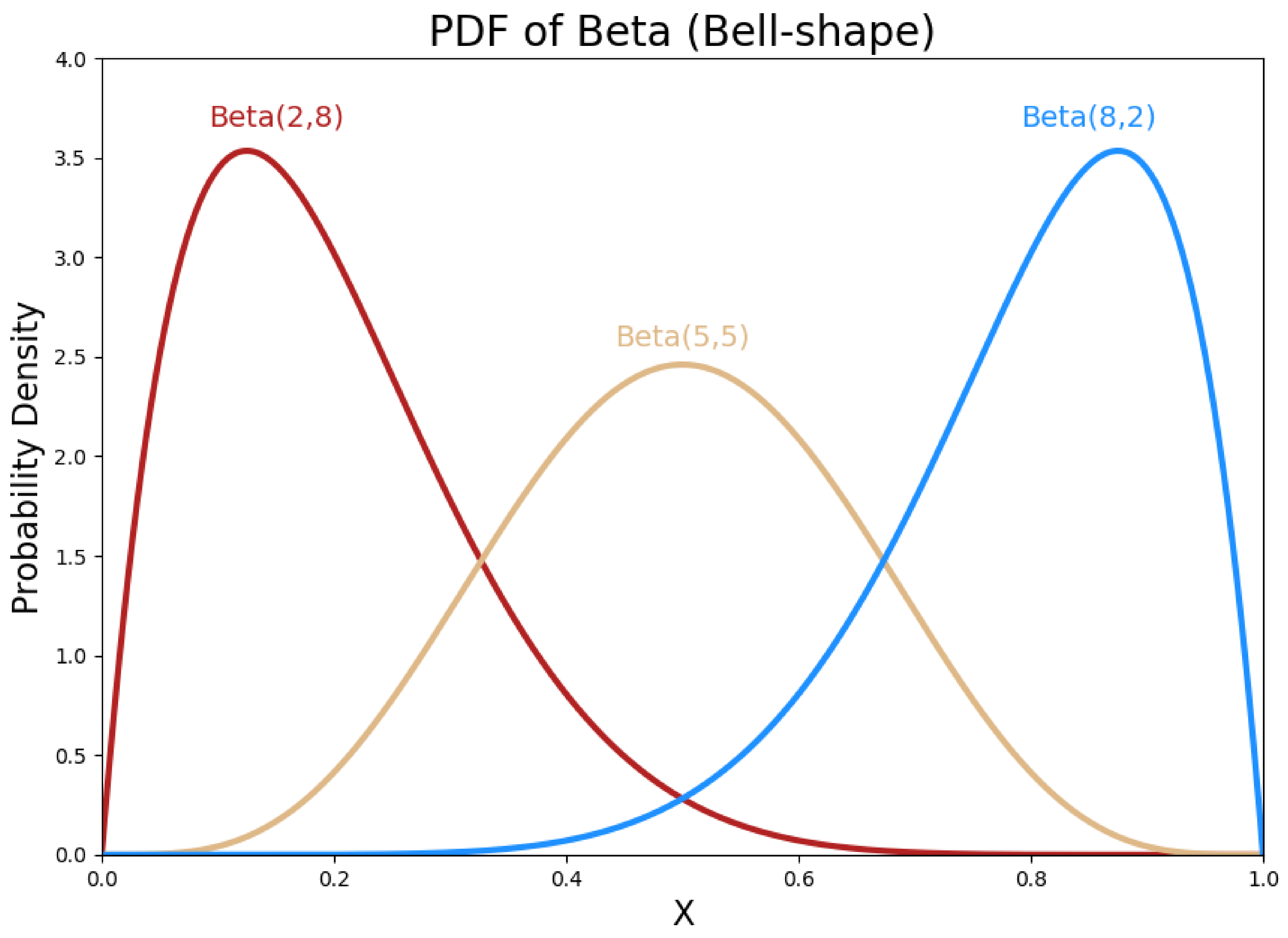
$$N(t) = \alpha + \beta + t - 1$$

$R(t)$ = number of red balls in the urn



Polya’s Urn

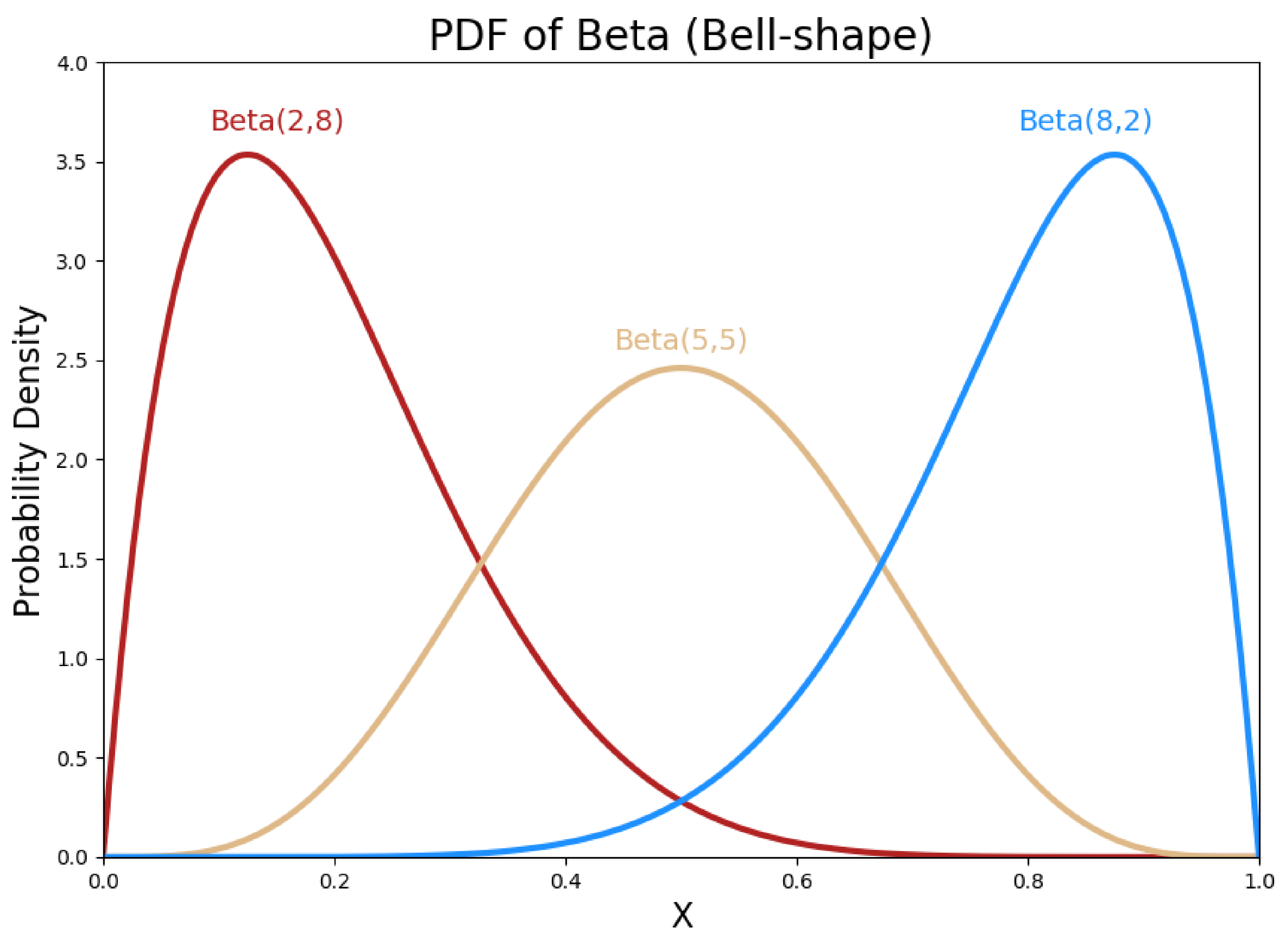
- What would be $P(t)$ as $t \to \infty$?
  - $P(t) \to Beta(\alpha, \beta)$



Bias in Recommendation: A Simple Example

- What would be $P(t)$ as $t \to \infty$?
  - $P(t) \to Beta(\alpha, \beta)$

Small bias in initial configuration will lead to large bias.



What is the problem?

Pr(click | red) is not equal to Pr(click | shown red)!

Does not take into account whether the item was shown or not!

Recommend based on Pr(click | shown red)

What if the opinion/preference shifts?

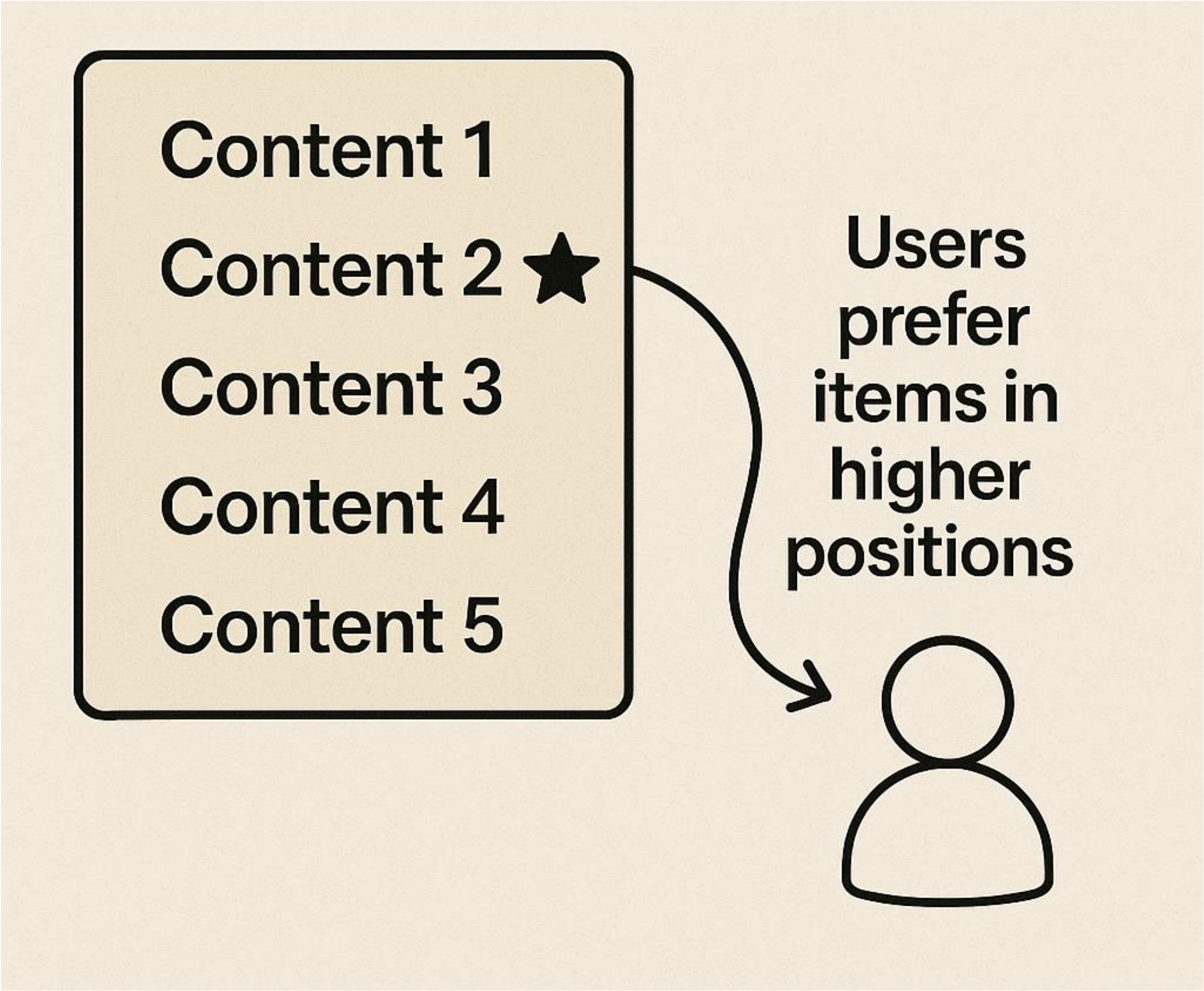
Position Bias in Recommendation

Showing the content is not sufficient!

How do we know whether the content was seen by the user or not?

What if it was placed lower down the order?

Also related to Popularity Bias



Cognitive Bias in Recommendation

The user has two selves “System 1” and “System 2”

- System 1 is impulsive and acts fast
- System 2 acts according to true utilities and exhibits long-term planning

- [Kahneman (2011); Smith and DeCoster (2000); Sloman (1996); Schneider and Shiffrin (1977); Evans (2008)]