Introduction to Probability

Random Variable

An event either does or does not happen. But what if we want to capture the ***magnitude*** of a probabilistic event?

- Suppose I flip n fair coins: the # of heads is a **random variable**
- Number that comes up when we roll a fair die is a **random variable**
- If an algorithm's behavior is determined by "flipping some coins" then the running time of the algorithm is a **random variable**

- **Definition.** A random variable X is a function from a sample space S (with a probability measure) to some value set (e.g. real numbers, integers, etc.)

Random Variable: Example

- Suppose, for example, I flip a coin 10 times. Let X be the number of heads
  - $\Pr[X = 0] = 1/2^{10}$
  - $\Pr[X = 10] = 1/2^{10}$
  - $\Pr[X = 4]$?
  - $\Pr[X = 4] = \binom{10}{4}\frac{1}{2^4}\frac{1}{2^6} = \frac{105}{512}$

All 10 flips are the same

Many different combinations of H & T

- A random variable that is 0 or 1 (indicating if something happens or not) is called an *indicator random variable* *or* *Bernoulli random variable*

Expectation

Every time you do the experiment, the associated random variable can take a different value

- How can we characterize the average behavior of a random variable?
- **Alternate Definition**. Expected value of a random variable R defined on a sample space S is

$$E(R) = \sum_x x \cdot \Pr(R = x)$$

Sum of values × probabilities that r.v. takes on those values

To get the E to look good in latex, use `\mathrm{E}`

- Let R be the number that comes up when we roll a fair, six-sided die, then the expected value of R is

$$\mathrm{E}(R) = \sum_{i=1}^{6} i \cdot \frac{1}{6} = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2}$$

(We won't use $\mathbb{E}$ in the slides, but if you really want to, it's `\mathbb`)

Conditional Expectation

Definition. If A is an arbitrary event with Pr$[A] > 0$, the conditional expectation of X given A is

$$E[X \mid A] := \sum_x x \cdot \Pr[X = x \mid A]$$

- **(Law of total expectation)** If $\{A_1, A_2, \ldots\}$ is a finite partition of the sample space:

$$E(X) = \sum_i E(X \mid A_i) \cdot \Pr(A_i)$$

Very useful !

Linearity of Expectation

The linearity of expectation (LoE) is an ***important*** tool in randomized algorithms

- The expected value operator for random variables is linear in the sense that:
  $E[X+Y] = E[X] + E[Y]$ and, for any constant α, $E[\alpha X] = \alpha E[X]$
  - Informally, **the expectation of a sum is the sum of the expectations**.
  - Formally, for any random variables $X_1, X_2, \ldots, X_n$ and any coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n$

$$E[\sum_{i=1}^{n} (\alpha_i \cdot X_i)] = \sum_{i=1}^{n} (\alpha_i \cdot E[X_i])$$

Very useful !

- **Note.** Always true! Linearity of expectation **does not require independence** of random variables.

Bernoulli Distribution

- Suppose you run an experiment with probability of success p and failure $1 - p$
    - Example, coin toss where head is success and $\Pr(H) = p$
- Let X be a Bernoulli or indicator random variable that is 1 if we succeed, and 0 otherwise. Then,

$$\begin{aligned} E[X] &= \sum_x x \cdot \Pr[X = x] \\ &= 0 \cdot \Pr[X = 0] + 1 \cdot \Pr[X = 1] \\ &= p \end{aligned}$$

- **Remember this:** expectation of an indicator random variable is exactly the probability of success!

Expected Success: n Bernoulli Trials

- Consider n independent Bernoulli trials (with success probability p). Let R denote the number of successes
  - R is said to follow a *Binomial distribution*
- We want to know expected number of successes $\mathrm{E}(R)$
- Can write R as a sum of indicator random variables
  - $R = \sum_i R_i$ where $R_i = 0$ or $R_i = 1$
- Then $\mathrm{E}[R] = \mathrm{E}\left[\sum_i R_i\right]$ How can we simplify this by Linearity of Expectation?

Expected Success: n Bernoulli Trials

- Consider n independent Bernoulli trials (with success probability p). Let R denote the number of successes
    - R is said to follow a *Binomial distribution* (we'll revisit this)
- We want to know expected number of successes $\mathrm{E}(R)$
- Can write R as a sum of indicator random variables
    - $R = \sum_i R_i$ where $R_i = 0$ or $R_i = 1$
- Then $\mathrm{E}[R] = \mathrm{E}\left[\sum_i R_i\right] = \sum_i \mathrm{E}[R_i] = \sum_{i=1}^{n} p = np$

Uniform Distribution

With a uniform distribution, **every outcome is equally likely**

- Let X be the random variable of the experiment and S be the sample space
- $\text{Pr}[X = x] = \dfrac{1}{|S|}$
- $E[X] = \sum_{x \in S} x \cdot \text{Pr}(X = x) = \dfrac{1}{|S|} \cdot \sum_{x \in S} x$
- Example
    - fair coin toss: heads and tails are equally likely
    - fair die roll: all numbers are equally likely

Card Guessing: Memoryless

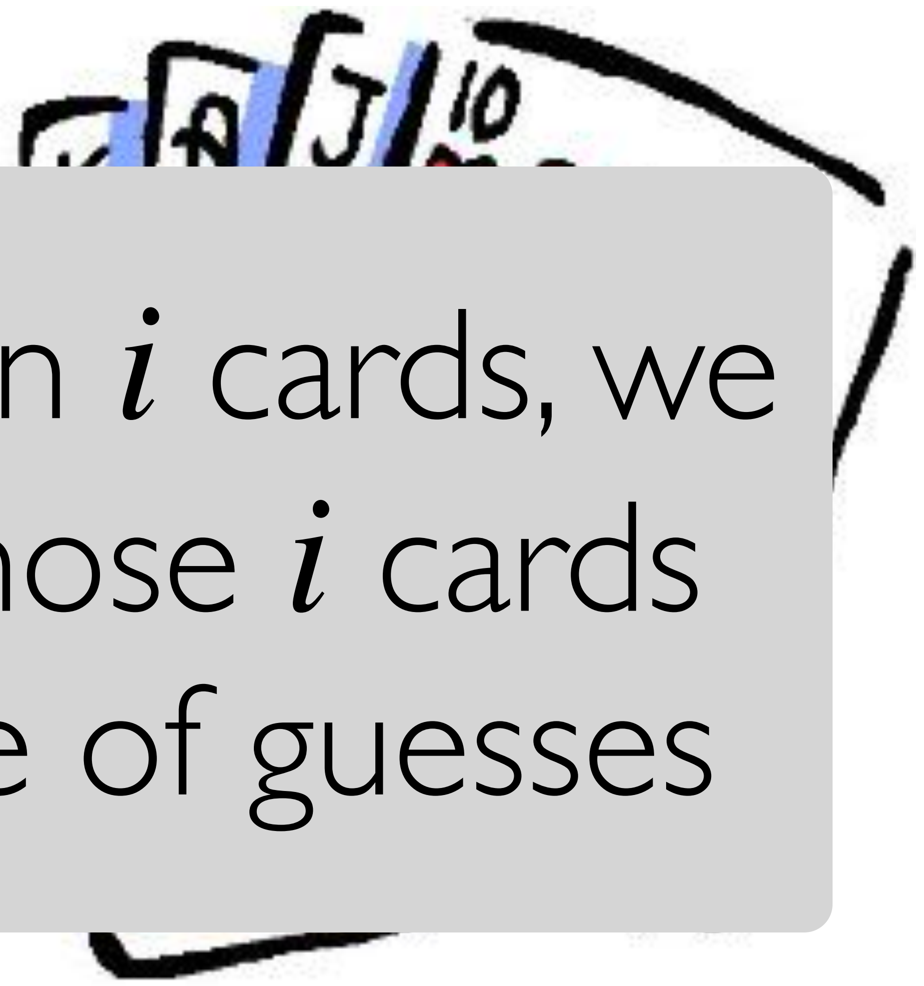
(K&T 13.3)

- To entertain your family you have them shuffle deck of n cards and then turn over one card at a time. Before each card is turned, you predict its identity. Assume you have no psychic abilities and your memory is terrible—you can't remember cards that have been seen
- Your strategy: guess uniformly at random
- How many predictions do you expect to be correct?
- Let X denote the random variable equal to the # of correct guesses and X_i denote the indicator variable that the i^{th} guess is correct
- Thus, $X = \sum_{i=1}^{n} X_i$ and $E[X] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$
- $E[X_i] = 0 \cdot \Pr(X_i = 0) + 1 \cdot \Pr(X_i = 1) = \Pr(X_i = 1) = 1/n$
- Thus, $E[X] = 1$

Card Guessing: Memoryfull

(K&T 13.3)

- Suppose we play the same game but now assume you have the ability to remember cards that have already been turned
- Your strategy: guess uniformly at random *from among the cards that have not yet been turned over*
- Let X denote the random variable equal to the # of correct guesses and X_i denote the indicator variable that the i^{th} guess is correct
- Thus, $X = \sum_{i=1}^{n} X_i$ and $\mathrm{E}[X] = \mathrm{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathrm{E}[X_i]$
- $\mathrm{E}[X_i] = \Pr(X_i = 1) = \dfrac{1}{n-i+1}$
- Thus, $\mathrm{E}[X] = \sum_{i=1}^{n} \dfrac{1}{n-i+1} = \sum_{i=1}^{n} \dfrac{1}{i}$

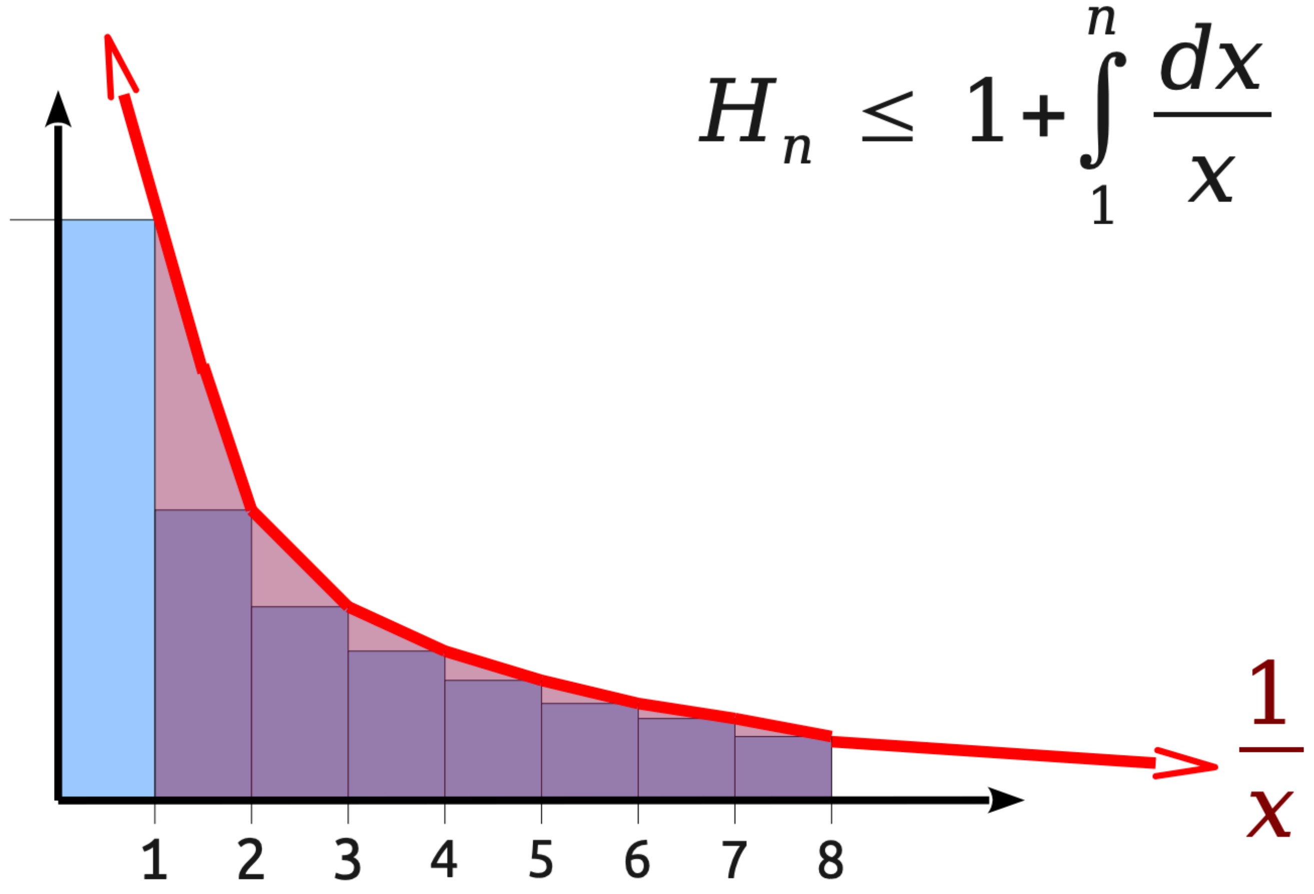


After we've seen i cards, we can rule out those i cards from our range of guesses

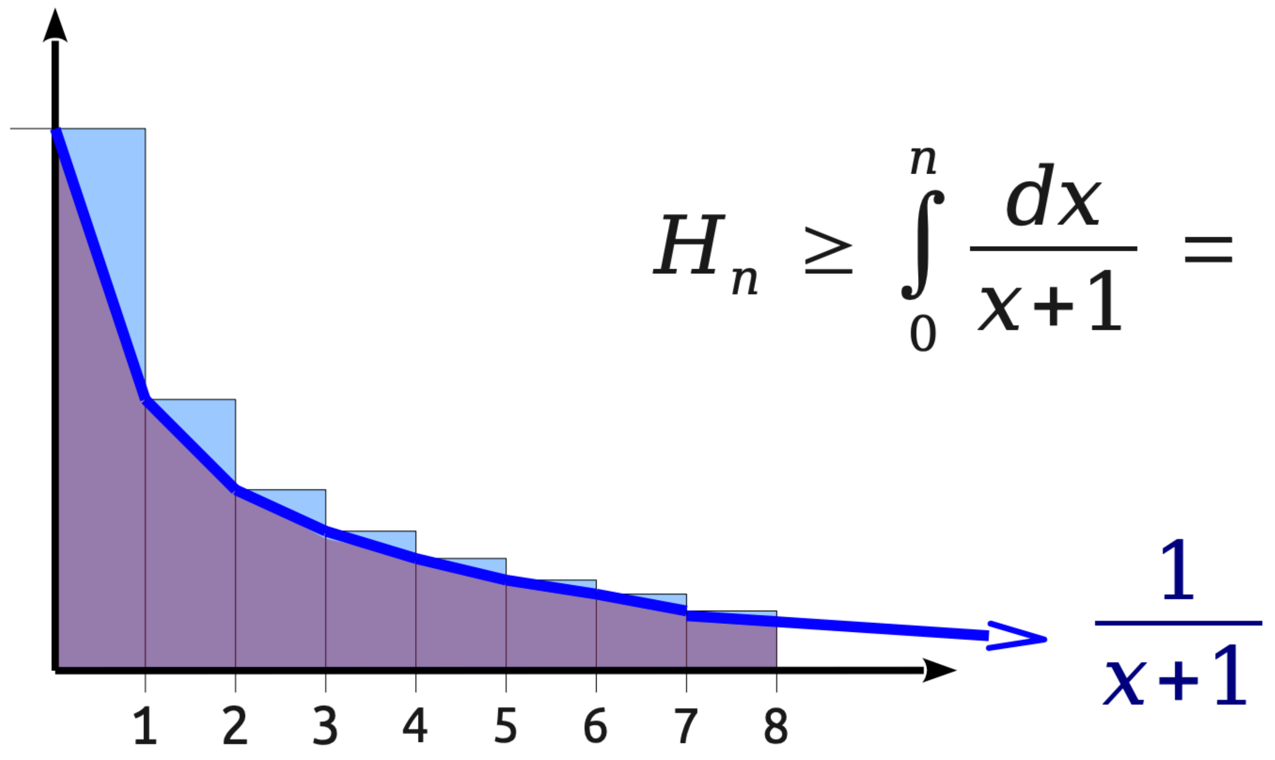
Harmonic Numbers

- The n^{th} harmonic number, denoted H_n is defined as
$$H_n = \sum_{i=1}^{n} \frac{1}{i}$$
- **Theorem**. $H_n = \Theta(\log n)$
- Proof Idea (we won't show in full). Upper and lower bound area under the curve

$$H_n \leq 1 + \int_1^n \frac{dx}{x} = \ln n + 1$$



$$H_n \geq \int_0^n \frac{dx}{x+1} = \ln(n+1)$$



Card Guessing: Memoryfull

(K&T 13.3)

- Suppose we play the same game but now assume you have the ability to remember cards that have already been turned
- Your strategy: guess uniformly at random among cards that have not been turned over
- Let X denote the r.v. equal to the number of correct predictions and X_i denote the indicator variable that the ith guess is correct
- Thus, $X = \sum_{i=1}^{n} X_i$ and $E[X] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$
- $E[X_i] = \Pr(X_i = 1) = \dfrac{1}{n - i + 1}$
- Thus, $E[X] = \sum_{i=1}^{n} \dfrac{1}{n-i+1} = \sum_{i=1}^{n} \dfrac{1}{i} = \Theta(\log n)$

Acknowledgments

- Some of the material in these slides are taken from
  - Shikha Singh
  - Kleinberg Tardos Slides by Kevin Wayne (https://www.cs.princeton.edu/~wayne/kleinberg-tardos/pdf/04GreedyAlgorithmsI.pdf)
  - Jeff Erickson's Algorithms Book (http://jeffe.cs.illinois.edu/teaching/algorithms/book/Algorithms-JeffE.pdf)
  - Hamiltonian cycle reduction images from Michael Sipser's Theory of Computation Book