Introduction to Probability II

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; only combination of flips leads to event

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, 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

$$E(R) = \sum_{x} x \cdot \Pr(R = x)$$

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

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

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

(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

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

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

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

Very useful !

Linearity of Expectation

- *Very important* tool in randomized algorithms
- Expectation of random variables obey a wonderful rule
- 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 $\alpha_1, \alpha_2, \ldots, \alpha_n$

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

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,

$$E[X] = \sum_x x \cdot \Pr[X = x]$$
$$= 0 \cdot \Pr[X = 0] + 1 \cdot \Pr[X = 1]$$
$$= p$$

- **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'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 $\mathrm{E}[R] = \mathrm{E}\left[\sum_i R_i\right]$

How can we simplify this by LoE?

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 $\mathrm{E}[R] = \mathrm{E}\left[\sum_i R_i\right] = \sum_i \mathrm{E}[R_i]$

How can we simplify this?

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

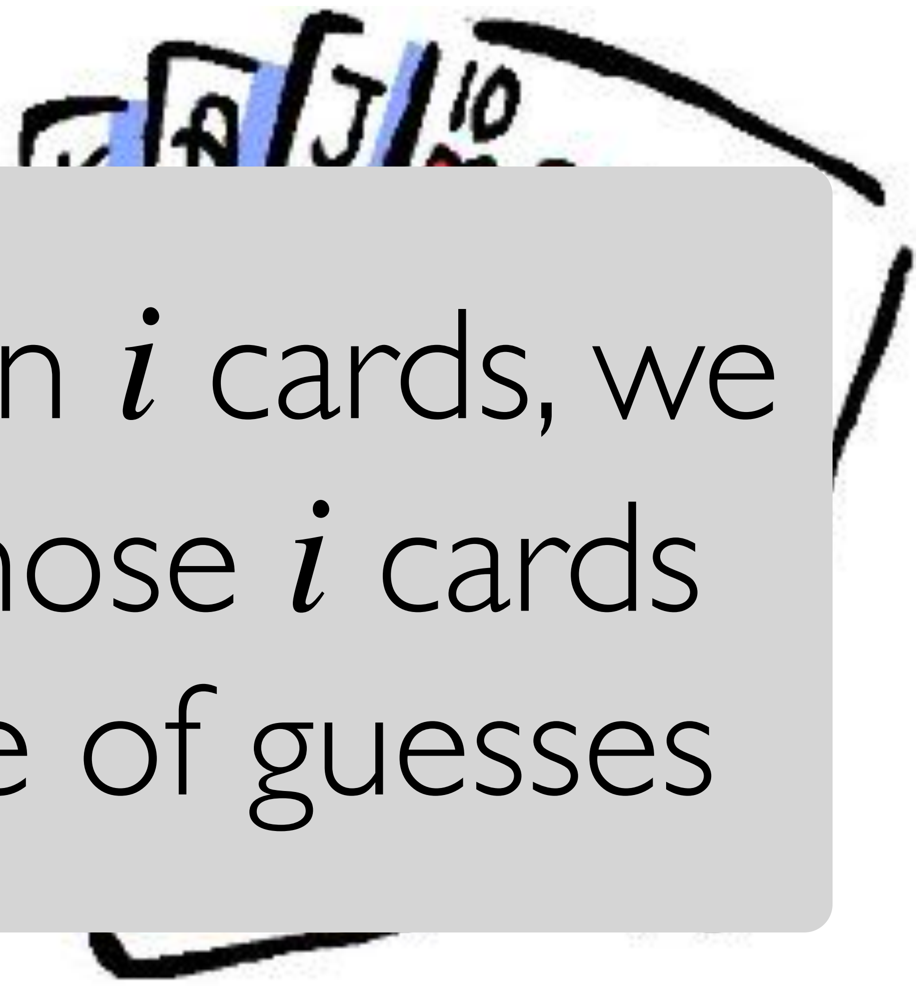
- With a **uniform distribution**, *every outcome is equally likely*
- Examples:
  - fair coin toss (heads and tails are equally likely)
  - fair die roll (all numbers are equally likely)
- Let X be the random variable of the experiment and S be the sample space
  - $\text{Pr}[X = x] = \frac{1}{|S|}$
  - $E[X] = \sum_{x \in S} x \cdot \text{Pr}(X = x) = \frac{1}{|S|} \cdot \sum_{x \in S} x$

Card Guessing: Memory**less**

- 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 or memory to remember cards
- 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 $\mathrm{E}[X] = \mathrm{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathrm{E}[X_i]$
- $\mathrm{E}[X_i] = 0 \cdot \Pr(X_i = 0) + 1 \cdot \Pr(X_i = 1) = \Pr(X_i = 1) = 1/n$
- Thus, $\mathrm{E}[X] = 1$

Card Guessing: Memory**ful**

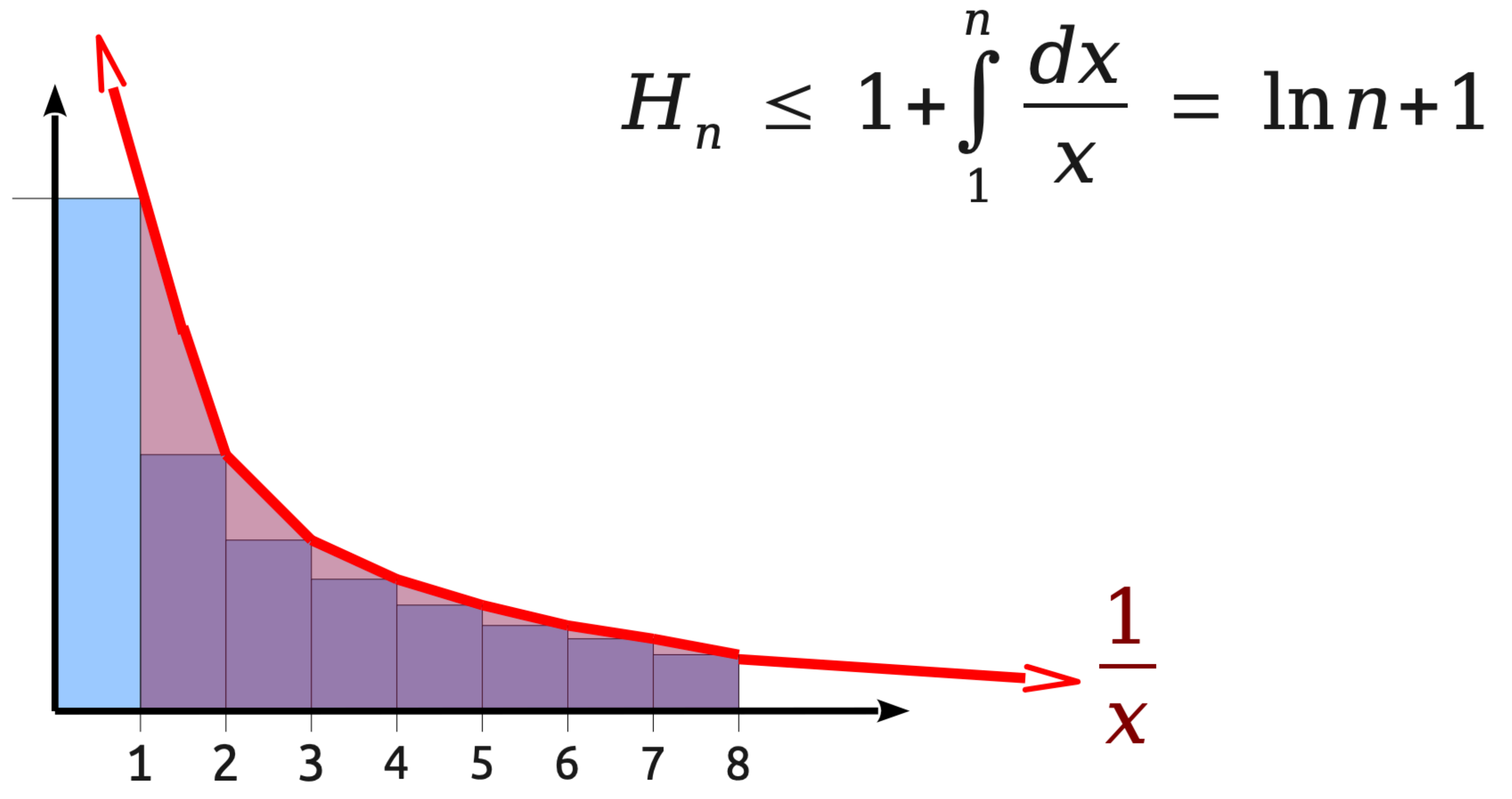
- Suppose we play the same game but now assume you have the ability to remember cards that have already been turned
- Your strategy: guess uniformly at random *among cards that have not been turned over*
- Let X denote the random variable equal to the # of correct guesses and X_i denote the indicator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

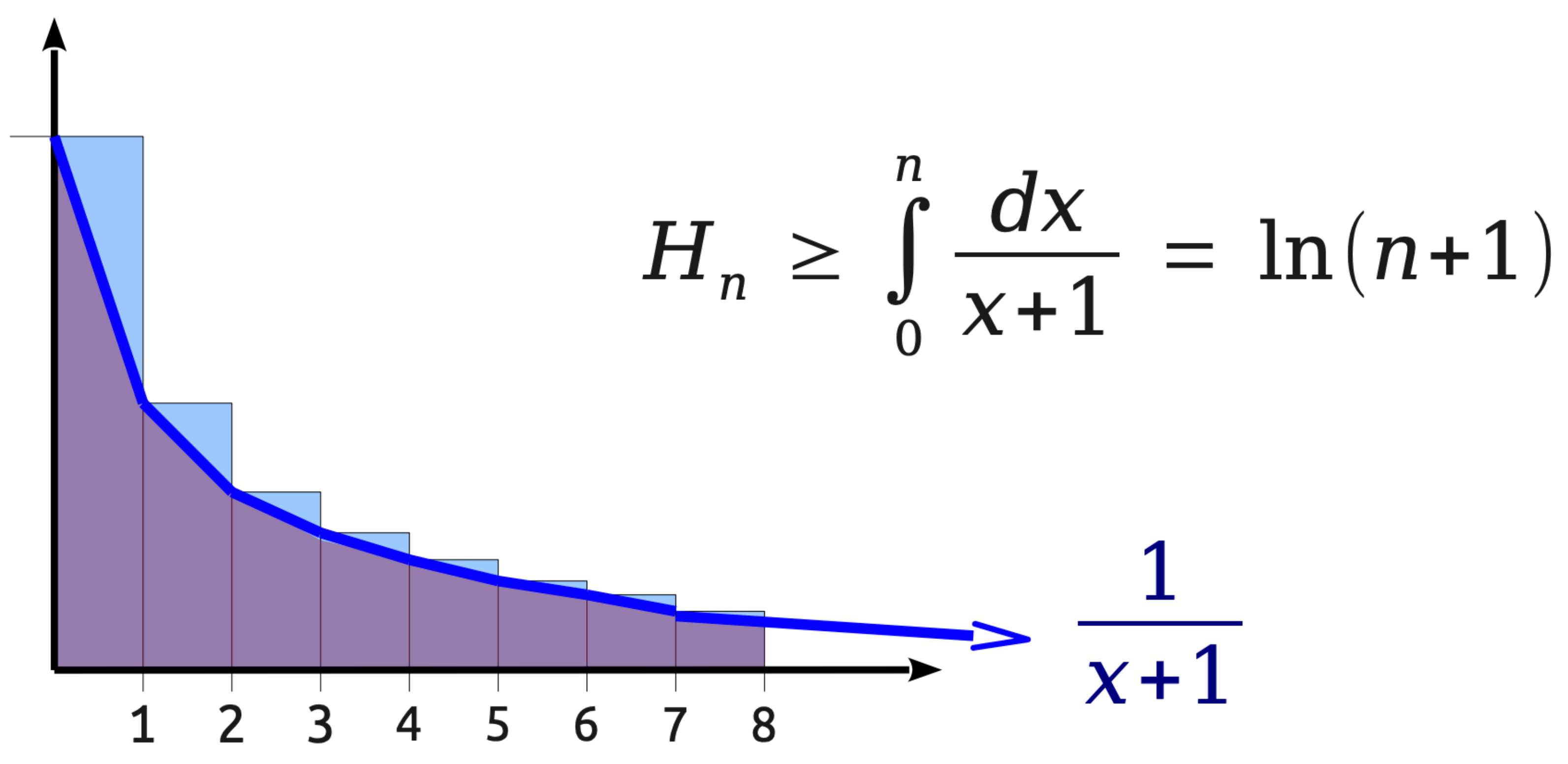


After we've seen i cards, we can rule out those i cards from our range of guesses

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. Upper and lower bound area under the curve





Card Guessing: Memory**ful**

- 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, $E[X] = \sum_{i=1}^{n} \dfrac{1}{n - i + 1} = \sum_{i=1}^{n} \dfrac{1}{i} = \Theta(\log n)$

Geometric Distribution

- Let's say we do a sequence of Bernoulli trials $X_1, X_2, \ldots$ where X_i where each trial is successful ($X_i = 1$) with probability p, and fails ($X_i = 0$) with probability $1 - p$
- **Question**: what is the expected number of trials until the first success?
    - In expectation, what is the value of the first i such that $X_i = 1$?
    - E.g. number of coin flips until heads $(p = 1/2)$
    - E.g. number if times I roll a die until I get a 1 $(p = 1/6)$
- One way to solve it is to just do the sum:
    - $\sum_{i=1}^{\infty} i(1-p)^{i-1}p$

Geometric Expectation (using the sum)

$$\sum_{i=1}^{\infty} i(1-p)^{i-1}p = \sum_{i=1}^{\infty} \sum_{k=1}^{i} (1-p)^{i-1}p =$$

$$\sum_{k=1}^{\infty} \sum_{i=k}^{\infty} (1-p)^{i-1}p = \sum_{k=1}^{\infty} p(1-p)^{k-1} \sum_{i=0}^{\infty} (1-p)^{i} =$$

$$\sum_{k=1}^{\infty} p(1-p)^{k-1} \frac{1}{1-(1-p)} = \sum_{k=1}^{\infty} (1-p)^{k-1} = \sum_{k=0}^{\infty} (1-p)^{k} = \frac{1}{p}$$

Geometric Expectation (using the sum)

- Want to know, how many tries in expectation until first success
- Let's think about this recursively

$$X \leftarrow \begin{cases} 1 \text{ with prob. } p \\ 0 \text{ with prob. } (1-p) \end{cases}$$

FindNumTries:

If $X = 1$

  Return 1

If $X = 0$

  Return $1+$ FindNumTries

If we fail in the first try, we start over from scratch!

- Let F be the number returned by FindNumTries, what want $\mathrm{E}(F)$

Geometric Expectation (using the sum)

- Let F be the number of times FindNumtries is called, what is $\mathrm{E}(F)$?
- $\mathrm{E}(F) = \mathrm{E}(F \mid X_1 = 1) \cdot \Pr(X_1 = 1) + \mathrm{E}(F \mid X_1 = 0) \cdot \Pr(X_1 = 0)$

  $= (1 + 0) \cdot p \quad + \quad (1 + E(F)) \cdot (1 - p)$
- $E(F) = 1/p$

FindNumTries:

If $X = 1$

    Return 1

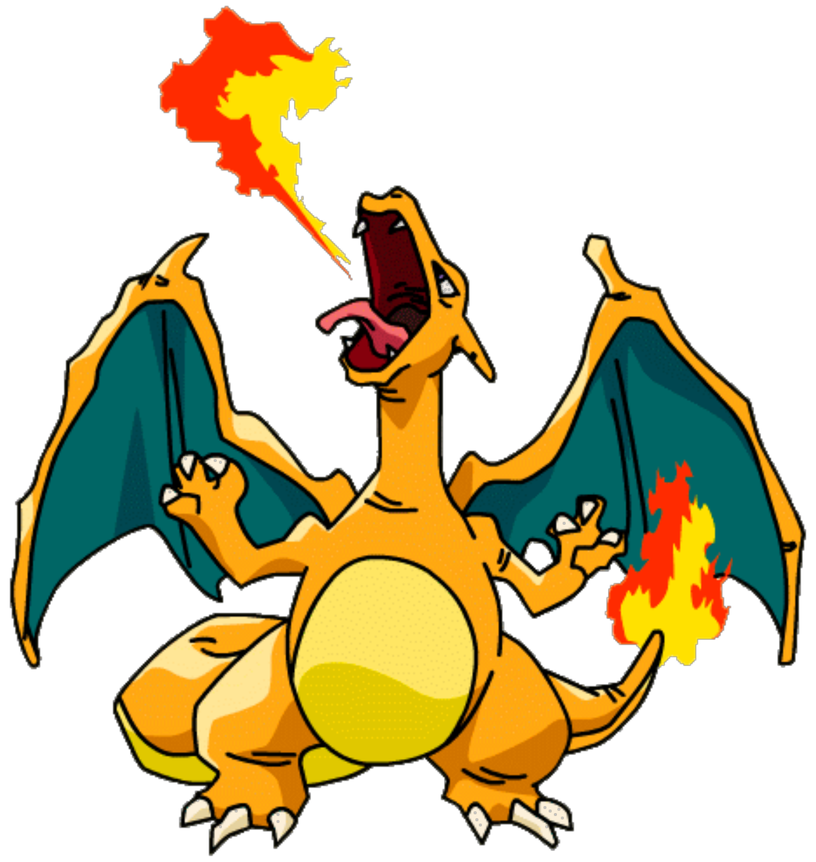
If $X = 0$

    Return $1+$ FindNumTries

Coupon/Pokemon Collector Problem

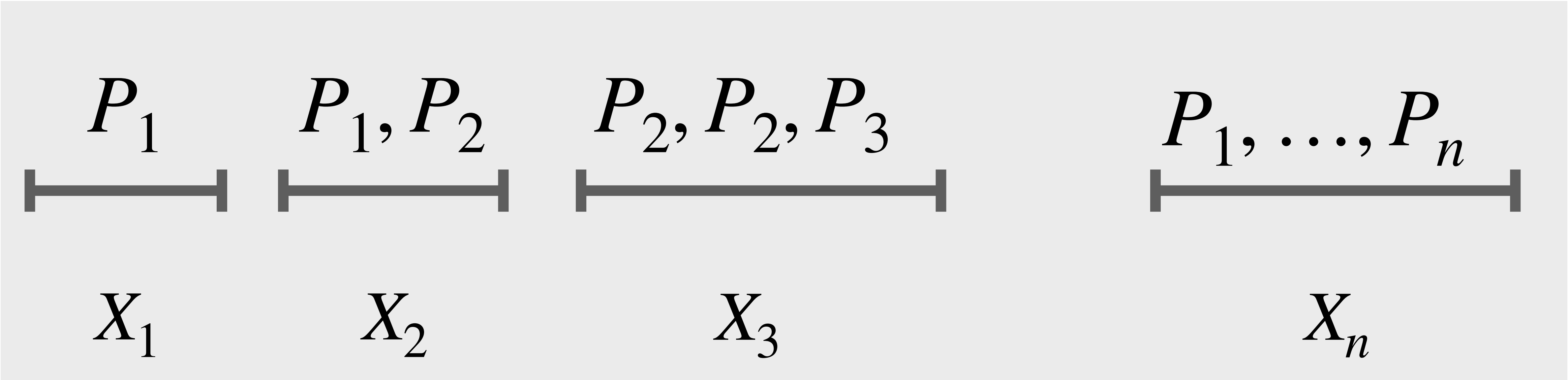
Gotta' Catch 'Em All

- Suppose there are n different types of Pokemon cards
- In each trial we purchase a pack that contains a Pokemon card, where
  - each of the n Pokemon are equally likely to be in a pack
- We repeat until we have at least one of each type of card, how many packs does it take in expectation to collect all?
- Let X be the r.v. equal to the number of packs bought until you first have a card of each type. **Goal**: compute $E[X]$
- We break X into smaller random variables
- **Idea**: we make progress every time we get a card we don't already have



Pokemon Collector Problem

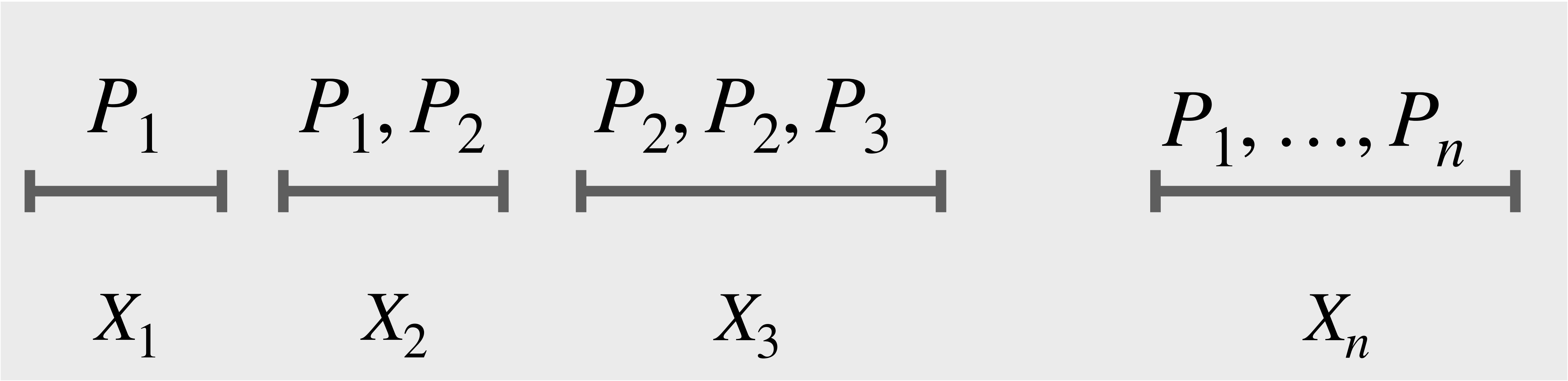
- Let X_i denote the "length of the ith phase", that is, the number of packs bought during the ith phase (ith phase ends as soon as we see the ith distinct card)
- Thus, $X = \sum_{1=1}^{n} X_i$

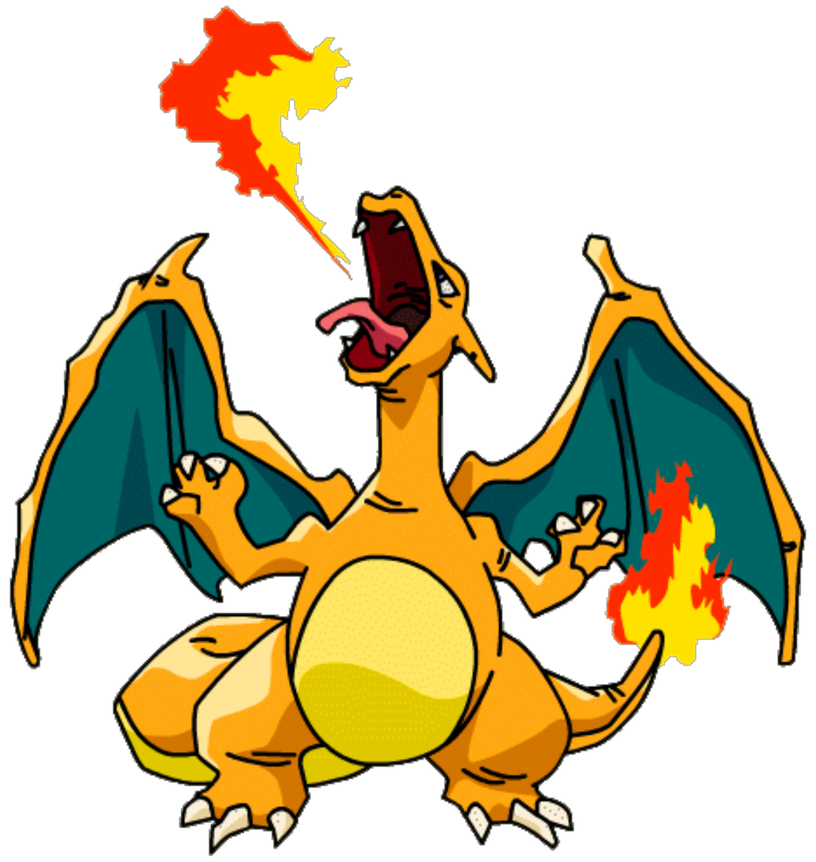


- Each phase can be though of as flipping a biased coin until we see a head, where seeing a head = getting a new card

Pokemon Collector Problem

- $E[X_i]$ is the expected number of coin flips until success (expectation of a geometric r.v.)

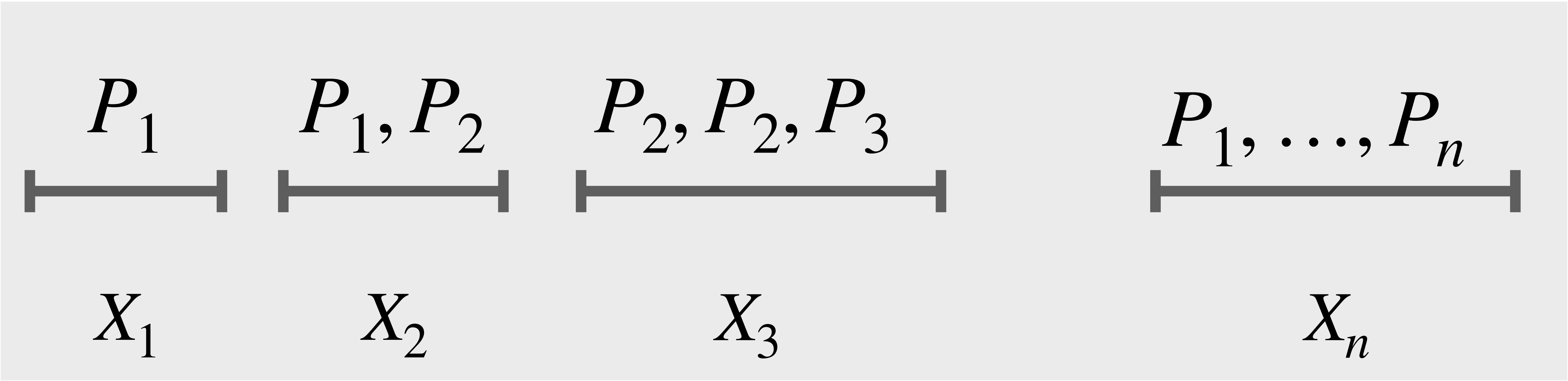


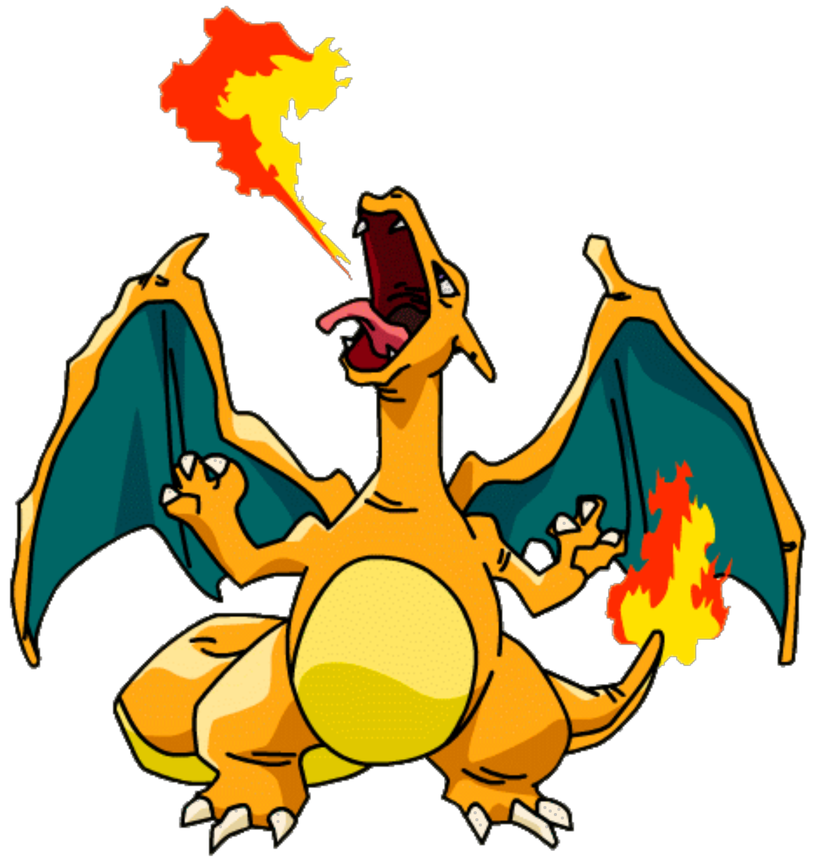


- We know, $E[X_i] = 1/p_i$ where p_i is the probability of success/ probability of seeing a heads during a coin flip in the ith phase
- Before the ith phase starts, we don't have $n - i + 1$ Pokemon
- Remember: each of the n Pokemon are equally likely to be in a pack

Pokemon Collector Problem

- $E[X_i]$ is the expected number of coin flips until success (expectation of a geometric r.v.)

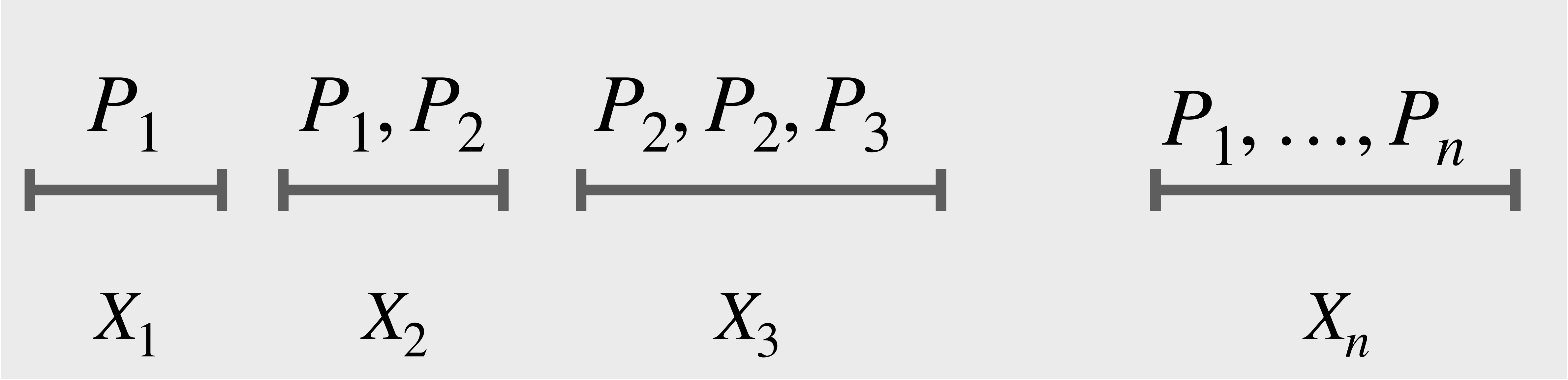




- We know, $E[X_i] = 1/p_i$ where p_i is the probability of success/ probability of seeing a heads during a coin flip in the ith phase
- $p_i = \dfrac{n - i + 1}{n}$

Pokemon Collector Problem

- We know, $E[X_i] = 1/p_i$ where p_i is the probability of success/ probability of seeing a heads during a coin flip in the ith phase



- $E[X_i] =$ Expected[number of flips until first heads] $= 1/p_i = \dfrac{n - i + 1}{n}$

- $$E[X] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{n}{n-i+1} = \sum_{i=1}^{n} \frac{n}{i} = nH_n = \Theta(n \log n)$$

Taking Stock…

- We've run through ***a lot*** of probability vocabulary and rules…
- We've applied some of those rules to answer "interesting" questions
- What's next?
    - Analyzing previously-explored algorithms that have randomness
    - Analyzing previously-explored data structures that have randomness
    - Using randomness to simplify the implementation of common APIs

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