## More Probability

## Admin

- Moved Tai office hours (6-8 Wed) to 6-8 Thur (will update on website after class)
- Assignment 9 out Nov 27, Assignment 10 out Dec 4


## Monty Hall Problem

- "Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say number I, and the host, who knows what's behind the doors, opens another door, say number 3, which has a goat. He says to you, "Do you want to pick door number 2?" Is it to your advantage to switch your choice of doors?" --- Craig. F.Whitaker Columbia, MD


## Clarifying the Problem

- The car is equally likely to be hidden behind any of the 3 doors
- The player is equally to pick any of the 3 doors, regardless of the car's location
- After the player picks a door, the host must open a different door with a goat behind it and offer the choice to switch
- If the host has a choice of which door to open, he is equally likely to select each of them




## Find the Sample Space

- Sample space: set of all possible outcomes
- An outcome involves 3 things:
- door concealing the car
- door initially chosen by the player
- door that host opens to reveal a goat
- Every possible combination of this is an outcome
- We can visualize these as a tree diagram
- Sample space $S$ is then:

$$
S=\left\{\begin{array}{llll}
(A, A, B), & (A, A, C), & (A, B, C), & (A, C, B), \\
(B, A, C), & (B, B, A), \\
(B, B, C), & (B, C, A), & (C, A, B), & (C, B, A), \\
(C, C, A), & (C, C, B)
\end{array}\right\}
$$



## Define Events of Interest

- Question. What is the probability that $\qquad$ ?
- Model as an event (subset of the sample space)
- Event that player wins by switching:
- $\{(A, B, C),(A, C, B),(B, A, C),(B, C, A),(C, A, B),(C, B, A)\}$
- Exactly half of the outcomes
- Switching leads to win with probability half?
- No!
$S=\left\{\begin{array}{lllll}(A, A, B), & (A, A, C), & (A, B, C), & (A, C, B), & (B, A, C), \\ (B, B, B, A), \\ (B, B, C), & (B, C, A), & (C, A, B), & (C, B, A), & (C, C, A), \\ (C, C, B)\end{array}\right\}$



## Determine Outcome Probabilities

- Each outcome is not equally likely!
- To determine probability, assign edge probabilities (conditional on previous parts of tree!)
- $\operatorname{Pr}(\mathrm{A}, \mathrm{B}, \mathrm{C})=\frac{1}{18}, \operatorname{Pr}(\mathrm{~A}, \mathrm{~A}, \mathrm{C})=\frac{1}{18}, \operatorname{Pr}(\mathrm{~A}, \mathrm{~B}, \mathrm{C})=\frac{1}{9}$, etc.
- Sum of probabilities of all outcomes is 1
- (Notice) probability is just a function
- Notations. $\operatorname{Pr}[x], \operatorname{Pr}(x)$
- Definition (Probability space). A sample space $S$ together with a probability function $\operatorname{Pr}: S \rightarrow[0,1]$



## Compute Event Probabilities

- We now have a probability of each outcome
- Probability of an event is the sum of the probabilities of the outcomes it contains, i.e., $\operatorname{Pr}(E)=\sum_{x \in E} \operatorname{Pr}(x)$
- $\operatorname{Pr}($ switching wins $)=\frac{1}{9}+\frac{1}{9}+\frac{1}{9}+\frac{1}{9}+\frac{1}{9}+\frac{1}{9}=\frac{2}{3}$
- It is better to switch!
- Takeaway: resist the intuitively appealing answer
$S=\left\{\begin{array}{lllll}(A, A, B), & (A, A, C), & (A, B, C), & (A, C, B), & (B, A, C), \\ (B, B, C), & (B, C, A), & (C, A, B), & (C, B, A), & (C, C, A), \\ (C, C, B)\end{array}\right\}$ Event $($ Switching Wins $)=$

$$
\{(A, B, C),(A, C, B),(B, A, C),(B, C, A),(C, A, B),(C, B, A)\}
$$

## The Birthday Paradox

- Suppose that there are $m$ students in a lecture hall
- Assume for each student, any of the $n=365$ possible days are equally likely as their birthday
- Assume birthday are mutually independent
- Question. What is the likelihood that no two students have the same birthday?
- Let $A_{i}$ be the event that the $i$ th persons birthday is different from the previous $i-1$ people
- $\operatorname{Pr}($ all $m$ different birthdays)
$=\operatorname{Pr}\left(A_{1} \cap A_{2} \cap \ldots \cap A_{m}\right)$
$=\operatorname{Pr}\left(A_{1}\right) \cdot \operatorname{Pr}\left(A_{2} \mid A_{1}\right) \cdot \operatorname{Pr}\left(A_{3} \mid A_{1} \cap A_{2}\right) \ldots \operatorname{Pr}\left(A_{n} \mid A_{1} \cap \ldots \cap A_{n-1}\right)$



## The Birthday Paradox

- $\operatorname{Pr}$ (all $m$ different birthdays)

$$
\begin{aligned}
& =1 \cdot\left(1-\frac{1}{n}\right) \cdot\left(1-\frac{2}{n}\right) \cdot\left(1-\frac{3}{n}\right) \ldots\left(1-\frac{m-1}{n}\right) \\
& =\prod_{j=1}^{m-1}\left(1-\frac{j}{n}\right) \leq \prod_{j=1}^{m-1} e^{-j / n} \approx e^{-m^{2} / 2 n}
\end{aligned}
$$

- $m \approx \sqrt{2 n \ln 2}$ for probability to be $1 / 2$
- For $n=365$, we get $m=22.49$
- Thus, with around 23 people in this class, we have a $50 \%$ chance of two people having the same birthday

Useful Inequality:

$$
(1-x) \leq\left(\frac{1}{e}\right)^{x} \text { for } x \geq 1
$$

## Birthday problem

From Wikipedia, the free encyclopedia

## For yearly variation in mortality rates, see birthday effect. For the mathematical brain teaser that was asked in the Math Olympiad, see Cheryl's Birthday

In probability theory, the birthday problem or birthday paradox concerns the probability that, in a set of $n$ randomly chosen people, some pair of them will have the same birthday. By the pigeonhole principle, the probability reaches $100 \%$ when the number of people reaches 367 (since there are only 366 possible birthdays, including February 29 ). However, $99.9 \%$ probability is reached with just 70 people, and $50 \%$ probability with 23 people. These conclusions are based on the assumption that each day of the year (excluding February 29 ) is equally probable for a birthday

Actual birth records show that different numbers of people are born on different days. In this case, it can be shown that the number of people required to reach the $50 \%$ threshold is 23 or fewer. ${ }^{[1]}$ For example, if half the people were born on one day and the other half on another day, then any two people would have a $50 \%$ chance of sharing a birthday

It may well seem surprising that a group of just 23 individuals is required to reach a probability of $50 \%$ that at least two individuals in the group have the same birthday: this result is perhaps made more plausible by considering that the comparisons of birthday will actually be made between every possible pair of individuals $=23 \times 22 / 2=253$ comparisons, which is well over half the number of days in a year (183 at most), as opposed to fixing on one individual and comparing his or her birthday to everyone else's. The birthday problem is not a "paradox" in the literal logical sense of being self-contradictory, but is merely unintuitive at first glance.

Real-world applications for the birthday problem include a cryptographic attack called the birthday attack, which uses this probabilistic model to reduce the complexity of finding a collision for a hash function, as well as calculating the approximate risk of a hash collision existing within the hashes of a given size of population.

## Random Variables

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


## Random Variable

- Event either does or does not happen, what if we want to capture magnitude of a probabilistic event
- Suppose I flip $n$ independent fair coins, then the number of heads is a random variable
- Number that comes up when we roll a fair die is a random variable
- If an algorithm flips some coins then the running time of the algorithm is a random variable
- A random variable from $S$ to $\{0,1\}$ is called an indicator random variable or Bernoulli random variable


## Expectation

- Every time you do the experiment, associated random variable takes a different value
- How can we characterize the average behavior of a random variable?
- Definition. Expected value of a random variable $R$ defined on a sample space $S$ is

$$
\mathrm{E}(R)=\sum_{w \in S} R(w) \cdot \operatorname{Pr}(w)
$$

- Let $R$ be the number that comes up when we roll a fair, sixsided 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\} or $\backslash E$

(We won't use it like $\mathbb{E}$ in this class, but if you really want to, it's \mathbb)

## Expectation

- We can group together all outcomes for which the random variable takes the same value
- Alternate Definition. Expected value of a random variable $R$ defined on a sample space $S$ is

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

- If $A$ is an arbitrary event with $\operatorname{Pr}[A]>0$, the conditional expectation of $X$ given $A$ is

$$
E[X \mid A]:=\sum_{x} x \cdot \operatorname{Pr}[X=x \mid A]
$$

- (Law of total expectation)
- If $\left\{A_{1}, A_{2}, \ldots\right\}$ is a finite partition of the sample space then

$$
E(X)=\sum_{i} E\left(X \mid A_{i}\right) \cdot \operatorname{Pr}\left(A_{i}\right)
$$

## Linearity of Expectation

- Very important tool in randomized algorithm
- Expectation of random variables obey a wonderful rule
- Informally, it says that 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 $\alpha_{1}, \alpha_{2}, \ldots, \alpha_{n}$

$$
\mathrm{E}\left[\sum_{i=1}^{n}\left(\alpha_{i} \cdot X_{i}\right)\right]=\sum_{i=1}^{n}\left(\alpha_{i} \cdot \mathrm{E}\left[X_{i}\right]\right)
$$

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


## Hat Check Problem

- There is a dinner party where $n$ men check their hats. The hats are mixed up during dinner, so that afterward each man receives a random hat. In particular, each man gets his own hat with probability $1 / n$. What is the expected number of men who get their own hat?
- Let $R$ be the random variable denoting the number of men who get their hat back. Goal: compute $\mathrm{E}(R)$.
- Usual trick. Express random variable $R$ as a sum of indicator random variables $R=\sum_{i=1}^{n} R_{i}$ where $R_{i}$ is 1 if $i$ th person gets their hat back, else it is 0 .
. Use linearity of expectation. $\mathrm{E}(R)=\mathrm{E}\left(\sum_{i=1}^{n} R_{i}\right)=\sum_{i=1}^{n} \mathrm{E}\left(R_{i}\right)$
- What is $\mathrm{E}\left(R_{i}\right)$ ?
. $\mathrm{E}\left(R_{i}\right)=1 / n$, so $\mathrm{E}(R)=\sum_{i=1}^{n} \mathrm{E}\left(R_{i}\right)=1$



## Hat Check Problem

- What is $\mathrm{E}\left(R_{i}\right)$ ?
- $R_{i}=1$ if $i$ gets his hat; $R_{i}=0$ otherwise
- By definition,

$$
\mathrm{E}\left(R_{i}\right)=1 \cdot \operatorname{Pr}(i \text { gets } i \text { 's hat })+0 \cdot \operatorname{Pr}(i \text { gets another hat })
$$

- $\mathrm{E}\left(R_{i}\right)=\operatorname{Pr}(i$ gets $i$ 's hat $)$
- Sample space: all $n$ ! permutations of hats (they are given back in a certain order
- Event we care about: the $i$ th hat given back belong to $i$
- Outcomes in this event: fix $i$ in the $i$ th position. All other hats can be in any order--in fact, the number of outcomes is equal to the number of ways to order all other hats
- So $(n-1)$ ! outcomes in this event. Each occurs with probability $1 / n$ !
- $\mathrm{E}\left(R_{i}\right)=(n-1)!/ n!=1 / n$



## Uniform Distribution

- When every outcome is equally likely
- Let $X$ be the random variable of the experiment
- $\operatorname{Pr}[X=x]=\frac{1}{|S|}$
- $E[X]=\frac{1}{|S|} \cdot \sum_{x \in S} \operatorname{Pr}(X=x)$
- Example, fair coin toss: heads and tails are equally likely



## Card Guessing: Memoryless

- 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. 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 r.v. equal to the number of correct predictions and $X_{i}$ denote the indicator variable that the $i$ th guess is correct
- Thus, $X=\sum_{i=1}^{n} X_{i}$ and $\mathrm{E}[X]=\mathrm{E}\left[\sum_{i=1}^{n} X_{i}\right]=\sum_{i=1}^{n} \mathrm{E}\left[X_{i}\right]$
- $\mathrm{E}\left[X_{i}\right]=0 \cdot \operatorname{Pr}\left(X_{i}=0\right)+1 \cdot \operatorname{Pr}\left(X_{i}=1\right)=\operatorname{Pr}\left(X_{i}=1\right)=1 / n$
- Thus, $\mathrm{E}[X]=1$



## Card Guessing: Memoryfull

- 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 $i$ th guess is correct
. Thus, $X=\sum_{i=1}^{n} X_{i}$ and $\mathrm{E}[X]=\mathrm{E}\left[\sum_{i=1}^{n} X_{i}\right]=\sum_{i=1}^{n} \mathrm{E}\left[X_{i}\right]$
- $\mathrm{E}\left[X_{i}\right]=\operatorname{Pr}\left(X_{i}=1\right)=\frac{1}{n-i+1}$
- Thus, $\mathrm{E}[X]=\sum_{i=1}^{n} \frac{1}{n-i+1}=\sum_{i=1}^{n} \frac{1}{i}$



## 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

$$
\hat{\hat{1}} \quad H_{n} \leq 1+\int_{1}^{n} \frac{d x}{x}=\ln n+1
$$



## Card Guessing: Memoryfull

- 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 $i$ th guess is correct
- Thus, $X=\sum_{i=1}^{n} X_{i}$ and $E[X]=E\left[\sum_{i=1}^{n} X_{i}\right]=\sum_{i=1}^{n} E\left[X_{i}\right]$
- $E\left[X_{i}\right]=\operatorname{Pr}\left(X_{i}=1\right)=\frac{1}{n-i+1}$
. Thus, $E[X]=\sum_{i=1}^{n} \frac{1}{n-i+1}=\sum_{i=1}^{n} \frac{1}{i}=\Theta(\log n)$


## Acknowledgments

- Some of the material in these slides are taken from
- Kleinberg Tardos Slides by Kevin Wayne (https:// www.cs.princeton.edu/~wayne/kleinberg-tardos/pdf/ 04GreedyAlgorithmsl.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

