October 16

Friday, October 16, 2020 4:01 PM

Comp Sci 212

1. Birthday Paradox

2. Random Variables

Announcements

- Midterm October 19 - (up to October 12)
  - Google form if taking make-up
  - Submission by 5:15 (or 9:05)
  - Typed version

Birthday Paradox

In a room of n people, what's the prob. that two people have the same birthday?

- Raise hand if you want to share birthday
- Unmute you, say birthday (or random)
- Hear your birthday called, mention in chat

- View birthdays as a sequence, where i^{th} element is the birthday of person i.
- D be set of birthdays, $|D| = 365$
- Sample space D^n, uniform distribution
- E = set of sequences w/ distinct birthdays

$\Pr[E] = \dfrac{|E|}{|D|^n}$

$|E| = |D| \cdot (|D|-1) \cdot (|D|-2) \cdots (|D|-n+1)$ by generalized product rule

$\Pr[E] = \dfrac{|D| \cdot (|D|-1) \cdot (|D-2) \cdots (|D|-n+1)}{|D|^n}$

$= \prod_{i=0}^{n-1} \dfrac{|D|-i}{|D|} = \prod_{i=0}^{n-1} 1 - \dfrac{i}{|D|}$

multiply for each i from 0 to 1

$\rightarrow \prod_{i=0}^{n-1} 1 - \dfrac{i}{|D|}$

$\le \prod_{i=0}^{n-1} e^{\frac{-i}{|D|}}$

$= e^{-(0 + \frac{1}{|D|} + \frac{2}{|D|} \cdots \frac{n-1}{|D|})}$

$= e^{-\frac{n(n-1)}{2|D|}}$ $\qquad \dfrac{e}{e} \qquad -\dfrac{n(n-1)}{2|D|} \le -1 \qquad e = \dfrac{1}{e}$

$1 - x \le e^{-x}$

If $\dfrac{n(n-1)}{2} \ge |D|$, probability is at most $\frac{1}{e}$.

probability of a pair w/ same birthday is at least $1 - \frac{1}{e}$.

$\rightarrow n \gtrsim \sqrt{2|D|} + 1$

$\hookrightarrow$ approximately greater than

$|D| = 365$, $\quad n \ge 27$, pair w/ same birthday.

Birthday Principle

If there are d days in a year, and $\sqrt{2d}$ people in a room, the probability that two people share a birthday is about $1 - \frac{1}{e}$.

S = sample space, uniform dist. over S.

How to estimate $|S|$, how many samples do we need?

- Keep sampling until you think you've seen all samples - at least $|S|$ sample
- Keep sampling until you see a repeat.
  - $\sqrt{2|S|}$ samples.

Def. A random variable R is a total function on a probability space.

$R: S \rightarrow \mathbb{R}$

Ex. $S = \{H, T\}^{100}$, $R: S \rightarrow \mathbb{R}$, $R(s) = \{$# of H's in s

$\Pr[R = 50] = \Pr[E_1] \quad E_1 = \{s \mid s \in S,\ R(s) = 50\}$.

$\Pr[R \le 25] = \Pr[E_2] \quad E_2 = \{s \mid s \in S,\ R(s) \le 25\}$.

$\Pr[R \text{ is odd}] = \Pr[E_3] \quad E_3 = \{s \mid s \in S,\ R(s) \text{ is odd}\}$

$\Pr[R = 50]$, uniform dist.

$= \dfrac{|E_1|}{|S|} = \dfrac{\binom{100}{50}}{2^{100}}$

$\hookrightarrow$ set of all sequences w/ 50 H's, 50 T's, bijection to subsets of [100] of size 50.

sticky coin dist.

$\Pr[\text{all heads}] = \frac{1}{2}$

$\Pr[\text{all tails}] = \frac{1}{2}$

$\Pr[\text{everything else}] = 0$

$\Pr[R = 50] = 0$.

Ex. $M: S \rightarrow \mathbb{R} \quad M(s) = \begin{cases} 1 & \text{if all coins the same} \\ 0 & \text{o.w.} \end{cases}$

$\Pr[M=1] = \Pr[E] \quad E = \{\text{all heads, all tails}\}$.

uniform dist. $\Pr[M=1] = \dfrac{2}{2^{100}} = 2^{-99}$

sticky dist. $\Pr[M=1] = 1$.

Def. Two random variables A + B are independent if

$R: S \rightarrow \mathbb{R}$, $R(s)$ = # of heads in s

$S = \{H, T\}^{100}$

Thm. R and M are not independent uniform dist.

Proof. $\Pr[R=2 \text{ and } M=1] = 0$

$\Pr[R=2] \neq 0 \quad \Pr[M=1] \neq 0$

$\Pr[R=2] \cdot \Pr[M=1] \neq 0$. $\square$

Thm. R and M are independent under sticky coin dist.

Def. Two random variables A+B are independent if

$$\Pr[A=x_1]\cdot\Pr[B=x_2]=\Pr[A=x_1 \,\&\, B=x_2]$$

for all x_1 & x_2

$\Pr[R=2]\cdot\Pr[M=1]\neq 0.$ $\square$

Thm. R and M are independent under sticky coin dist.

Proof. $\Pr[R=k \text{ and } M=0]=0$ for all k.

$\Pr[M=0]=0$, so $\Pr[R=k]\cdot\Pr[M=0]=0.$

$$\Pr[R=k \text{ and } M=1]=\Pr[R=k]$$
$$=\Pr[R=k]\cdot\Pr[M=1]$$

$\square$