Nov5

Thursday, November 5, 2020 11:47 PM

Recall

$$\Omega \xrightarrow{X} \mathbb{R}.$$

takes values in $x_1, \ldots, x_r$.

The expected value of X is

$$E(X) = \sum_{i=1}^{r} x_i \cdot \underbrace{f_X(x_i)}$$

$$\hookrightarrow Pr(X = x_i) = Pr(\omega : X(\omega) = i).$$

Example

Coin prob of heads $= p$

Toss over & over again forever.

$\Omega =$ (potential lists of results)

$$X : \Omega \longrightarrow \mathbb{R}$$

$X(\omega) = n$ if n is first heads flipped

$$X(TTTH\text{---}) = 4$$

Last time compute $f_X(n)$

$$\begin{aligned} f_X(n) &= Pr(X(\omega) = n) \\ &= Pr(\text{first heads } n) \\ &= (1-p)^{n-1} \cdot p. \end{aligned}$$

$$E(X) = \sum_{n=1}^{\infty} n \cdot f_X(n) \leftarrow$$

∞ sum....
(likely to converge...)

$$= \sum_{n=1}^{\infty} \underline{n(1-p)^{n-1}} p \qquad \underline{-\frac{d}{dp}(1-p)^n}$$

$$= -p \sum_{n=1}^{\infty} \frac{d}{dp}(1-p)^n$$

$$= -p \frac{d}{dp} \underline{\sum_{n=1}^{\infty} (1-p)^n}$$

Recall $0 < k < 1$

$$\frac{1}{1-k} = \sum_{n=0}^{\infty} k^n$$

$$= -p \cdot \frac{d}{dp}\left(\frac{1}{p} - 1\right)$$

$$= -p\left(\frac{-1}{p^2}\right) = \frac{1}{p}$$

e.g. $p = 1/2$

$E(\text{first heads}) = \frac{1}{1/2} = 2$

Rmk

Similar Ω = choices random #s. (in some range)

$$X : \Omega \longrightarrow \mathbb{R}$$

$$\omega \longmapsto \left(\begin{array}{l}\text{amount of guesses} \\ \text{until prime}\end{array}\right)$$

$E(X)$ = when should I expect a prime.

computing this gives us expectation for when findPrime(Low, High) terminates.

Probabilistic Attacks: Collision

Main Point

Much easier to find matching objects than a particular object.

Example (Birthday Paradox)

Random group of 40 ppl.

$\boxed{1}$ Does anyone share my birthday?

$\boxed{2}$ Do 2 of them share a birthday?

Pr(one them has my bday)

$\boxed{1}$ ⚠ Naive (incorrect guess).

$\rightsquigarrow \frac{40}{365}$ ← overcounts if multiple people have my bday.

(rmk do for 366 ppl, not $\frac{366}{365} \cancel{>} 1$)

E = at least one person shares my bday.

E^c = No one shares my birthday.

$\Rightarrow F_i$ = Person i doesn't share my bday $(i = 1, \dots, 40)$. F_i are independent events.

$$\Pr(F_i) = \frac{364}{365}$$

$$\begin{aligned}\Pr(E) &= 1 - \Pr(E^c) \\ &= 1 - \prod_{i=1}^{40} \Pr(F_i) \\ &= 1 - \left(\frac{364}{365}\right)^{40} \approx 10.4\%.\end{aligned}$$

$\boxed{2}$ E = Prob 2 ppl share bday

E^c = Prob no 2 ppl do.

F_i = i'th person doesn't share a bday w/ persons $1, 2, \dots, i-1$.

$$\rightsquigarrow E^c = F_1 \cap F_2 \cap F_3 \cap \dots \cap F_{40} \leftarrow \checkmark$$

$$Pr(E) = 1 - Pr(E^c)$$

$$= 1 - Pr(F_1) \cdot Pr(F_2 | F_1) \cdot Pr(F_3 | F_2 \cap F_1) \cdot \dots$$

$$\dots Pr(F_{40} | F_{39} \cap \dots \cap F_1). *$$

what is $Pr(\underset{\uparrow}{F_{i+1}} | \underbrace{F_i \cap F_{i-1} \cap \dots \cap F_1}) = \dfrac{365 - i}{365}$

$$* = 1 - \underline{\frac{365}{365} \cdot \underset{\uparrow}{\frac{364}{365}} \cdot \frac{363}{365} \dots \underset{\uparrow}{\frac{326}{365}}} \approx 89.1\%$$

Exercise

23 ppl $\Longrightarrow$ ≈ 50% chance 2 share a bday.

Collision Algorithms / Meet-in-the-middle Attacks

Setup: Bob has a box of N numbers

Choose n of them & puts them

in a list

$x_1, x_2, \ldots, x_n$ ← replaces

Then randomly (w/ replaces) chooses m more for a second list

$y_1, \ldots, y_m$

Fact $n, m \approx \sqrt{N}$ or bigger then there likely exist $y_i = x_j$ (an overlap).

Theorem

An urn w/ N balls.

n are red.

$N - n$ are blue. ←

Bob randomly selects a ball. Then replaces it. And repeats a total of m times.

(i) $\Pr(\text{at least } 1 \text{ red}) = 1 - \left(1 - \frac{n}{N}\right)^m$ ←

(ii) A lower bound for this probability ↙ is $\Pr(\text{at least } 1 \text{ red}) \geq 1 - e^{-mn/N}$ *

(iii) If $N >> 0$, $n, m \approx \sqrt{N}$ (not too much larger)

Then (ii) is almost an equality.

Proof

$E = $ (at least one red)

$E^c = $ (all blue).

$F_i = $ (i'th ball is blue). ← indep. $E^c = \bigcap F_i$

$$\Pr(E) = 1 - \Pr(E^c)$$
$$= 1 - \prod_{i=1}^{m} \Pr(F_i)$$
$$= 1 - \left(1 - \frac{n}{N}\right)^m$$

$$\Pr(F_i) = \frac{N-n}{N} = \left(1 - \frac{n}{N}\right)$$

✓ (i)

(ii) $e^{-x} \geq 1-x$ for all x

(iii) $e^{-x} \approx 1-x$ for small x.

Numerical Computation HW

(ii) & (iii) HW

Why are these the same question?

N numbers	①	Urn w/ N blue numbered balls.
list 1 of n numbers	②	Pick n blue numbered balls & paint them red & replace them
List 2 of m numbers	③	Picking m numbered balls (w/ replacement).
Overlap	4	Picking a red ball in ~~st~~ep 3.

Examples

- Shuffle a deck & deal 8 cards face up
- shuffle a second deck. Draw 8 cards, replacing each time.
- Prob that one of cards I drew in second deck matched a face up one?

Urn ⟷ Deck

Red ⟷ 8 face up.

Blue ⟷ 44 remaining.

Picking Balls ⟷ Second deal.

$$\Pr(\text{Match}) = 1-\left(1-\frac{8}{52}\right)^{8} \approx 73.7\%.$$

(approx $1-e^{-64/52} \approx 70.8\%$).

Instead deal 10 cards first ← $n=10$

draw 5 in second. ← $m=5$

$$\Pr(\text{match}) = 1 - \left(1 - \frac{\overset{\downarrow}{10}}{52}\right)^{\overset{\swarrow}{5}} \approx 65.6\%.$$

Big Number Example

- 10 billion labelled objects.
- Random selection of 100,000. (list 1)
- Random selection (w/ replacement) of 100,000 (list 2).

$$\Pr(\text{match}) = 1 - \left(1 - \frac{100000}{10^{10}}\right)^{100000} \approx .632122$$

55

$$\geq 1 - e^{-\frac{100{,}000^2}{10^{10}}} = 1 - e^{-1} = .632121$$

Notice ~~double~~ increase lengths of lists,

$$1 - e^{\underbrace{-mn/N}} \rightarrow$$ exponentially approaches 0.

i.e. lists have 200,000 elts.

$$\Pr \geq 1 - e^{-4} \approx 98.2\%$$

" " 300,000 "

$$\Pr \geq 1 - e^{-9} \approx 99.988\% \ldots$$

Question:

How long should I make my lists If I want a good chance of a match?

Answer w/ $n = m$

N objects

2 lists length n.

How big is n to have 50% at match?

99.99% ?

Answer $1-\left(1-\frac{n}{N}\right)^n = p$ for p harder

$\Pr \approx 1-e^{-n^2/N} = p \leftarrow$

$$e^{-n^2/N} = 1-p$$

$$-n^2/N = \ln(1-p)$$

$$n = \sqrt{N\ln(1-p)} = \sqrt{N\ln((1-p)^{-1})}$$

$p=\frac{1}{2}$ $\quad n = \sqrt{N\ln(2)} = .83\sqrt{N}$

$p=.9999$ $\quad n = \sqrt{N\ln(10^4)} = 3.035\sqrt{N}$.

Collision algs have $\approx O(\sqrt{N})$ length lists in order to find an intersection.

⇝ Probably also takes $\underline{O(\sqrt{N})}$ to make the list.

Discretete Log Collision Alg

Prop G a group. $g \in G$ elt of order N.
Suppose $g^x = h$ has a solution.
Then the solution can be found
in $O(\sqrt{N})$ steps (exponentiation in G).

Proof

Look for y, z s.t.

$$g^y = h \cdot g^z$$

Then $g^{y-z} = h \implies x = \log_g h = y - z$.

- List $y_1, y_2, \dots, y_n$ in $\{1, \dots, n-1\}$
- $\{g^{y_1}, g^{y_2}, \dots, g^{y_n}\}$ ← List 1 (red balls)
- Choose random $z_1, \dots, z_n$
- $\{g^{z_1}\cdot h, g^{z_2}\cdot h, \dots, g^{z_n}\cdot h\}$ ← List 2

(urn is $\{1, g, g^2, \dots, g^{N-1}\}$

$$\Pr(\text{a match}) = 1 - \left(1 - \frac{n}{N}\right)^n$$

If we have a match we win

$$g^{y_i} = h g^{z_j}$$

$$h = g^{y_i - z_j} \longrightarrow \log_g h = y_i - z_j$$ ☺

$$n \approx 3\sqrt{N} \implies \Pr \geq 99.98\%$$

$n \approx 5\sqrt{N} \Longrightarrow \Pr \geq 1 - 10^{-10}$

lists are $O(\sqrt{N})$ for good prob of winning! $n = O(\sqrt{N})$

Time Complexity

* $\underline{2n}$ steps to compute lists

↳ Each expotiation g^{x_i} $O(\log(N))$ steps

Total making lists takes

$$O(\sqrt{N} \log N)$$

* Find match. $\Leftarrow \underset{\substack{\uparrow \\ \text{Hash}}}{O(\sqrt{N})}$ or $\underset{\substack{\uparrow \\ \text{sort.}}}{O(\sqrt{N} \log N)}$

In total $O(\sqrt{N} \log N)$.

Memory constraint of $O(\sqrt{N})$.

N big $\Longrightarrow$ BAD. (For baby Giant).

Next time we will introduce Pollard that avoids memory issues.