Lecture 16

- Wrap up PHP
- Intro to Probability & Probabilistic Method.

Return of the Error Correcting Codes

Error Correcting Codes (ECC) are widely used in practice. When we were discussing $\mathbb{Z}_p$ and polynomials, we saw how to encode $m \in \mathbb{Z}_p^n$ in order to correct 't' errors.

But in many real world applications, we transmit bits, not numbers in $\mathbb{Z}_p$.

We can express $\mathbb{Z}_p$ elements as bit strings, but we will see later why this is suboptimal.

$$m \in \{0,1\}^n \xrightarrow{\text{Encode}} z \in \{0,1\}^{\ell}$$

$z \in \{0,1\}^{\ell}$ $\downarrow$ t-noisy channel can corrupt t bits

$z' \in \{0,1\}^{\ell}$

Goal: Recover m from z'

How large SHOULD ℓ be?

$\ell \geq n(2t+1)$ suffices.

But this is a lot of redundancy!
Suppose $t = n/4$. To communicate n bits of information, we are sending $\Theta(n^2)$ bits over the noisy channel.

QN: For $t = n/4$, what is the smallest ℓ that might possibly work?

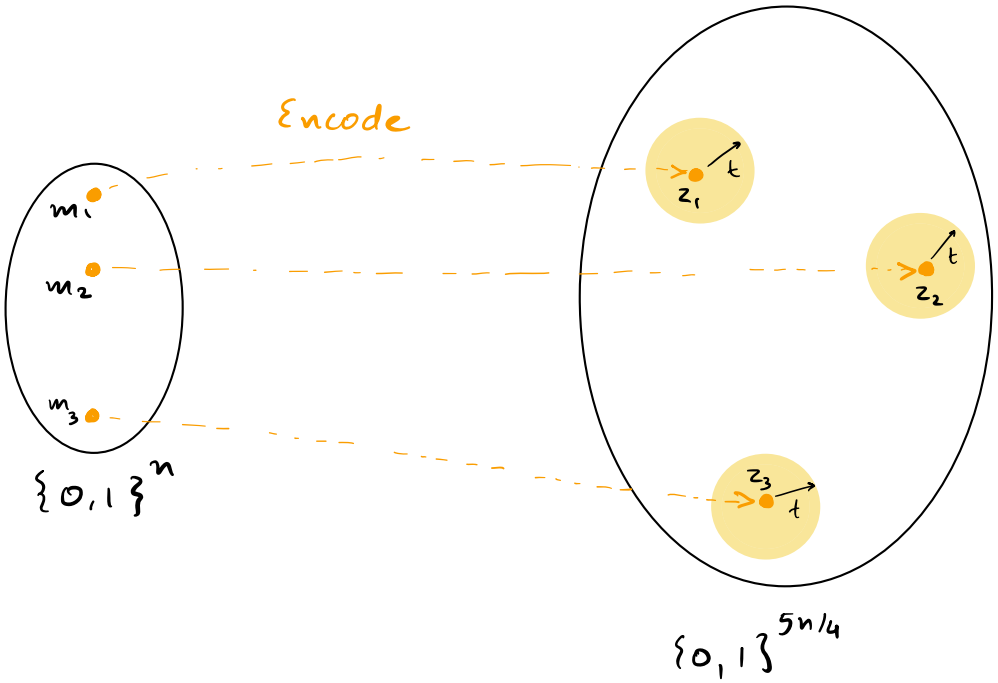
Easier Qn: For $t = n/4$, will $\ell = n + t = 5n/4$ work? Is it possible to define
Encode : $\{0,1\}^n \longrightarrow \{0,1\}^{5n/4}$ and
Decode : $\{0,1\}^{5n/4} \longrightarrow \{0,1\}^n$ s.t.
we can reliably recover the message after t corruptions?

Answer : No.

Proof by contradiction.

Suppose $\exists$ Encode : $\{0,1\}^n \longrightarrow \{0,1\}^{5n/4}$
and Decode : $\{0,1\}^{5n/4} \longrightarrow \{0,1\}^n$
s.t. for all $m \in \{0,1\}^n$,

if $z = \text{Encode}(m)$, $z' = t\text{-noisy}(z)$,
then $\text{Decode}(z') = m$.



Suppose $z_i = \text{Encode}(m_i)$ for each $m_i \in \{0,1\}^n$.

Then, $\forall\, i$, any z' that is at distance at most t from z_i, $\text{Decode}(z') = m_i$.

Let $B_i = \{ z' : \text{Distance}(z_i, z') \leq t \}$
($z_i = \text{Encode}(m_i)$)

Observation: $B_i \cap B_j = \phi$ for all $i \neq j$.

Suppose $z' \in B_i \cap B_j$. By correctness of Decode,
$\text{Decode}(z') = m_i$ (since $z' \in B_i$)
$\text{Decode}(z') = m_j$ (since $z' \in B_j$)
Decode is a fixed fn. Hence contradiction.

Claim: $|B_i| > 4^{n/4}$ for all i.

Proof: $|B_i| = 1 + \binom{l}{1} + \binom{l}{2} + \cdots + \binom{l}{t}$

No. of pts. at dist. i from z_i is $\binom{l}{i}$.

$$\binom{\ell}{i} \leq \binom{\ell}{t} \quad \text{since} \quad t = n/4 < \ell/2$$

$$\binom{\ell}{t} = \frac{\ell!}{t!\,(\ell-t)!} = \frac{(n+1)\cdots(n+n/4)}{1\cdots(n/4)}$$

$$> \frac{n^{n/4}}{(n/4)^{n/4}} = 4^{n/4}$$

$$\therefore \quad |B_i| > 4^{n/4}$$

■

Each $B_i \subseteq \{0,1\}^\ell$, so $\bigcup B_i \subseteq \{0,1\}^\ell$

$B_i \cap B_j = \phi$, so $\left|\bigcup B_i\right| = \sum_i |B_i|$

$$> 2^n \cdot 4^{n/4}$$

$$|\{0,1\}^\ell| = 2^\ell = 2^n \cdot 2^{n/4}.$$

Since $\bigcup_i B_i \subseteq \{0,1\}^\ell$, $\left|\bigcup B_i\right| \leq |\{0,1\}^\ell|$.

But this is a contradiction since $|\{0,1\}^\ell|$

$$= 2^n \cdot 2^{n/4}$$
$$< 2^n \cdot 4^{n/4}$$
$$< \left|\bigcup_i B_i\right|$$

■

QN: Is it possible to reliably send n bits of information over a $(n/4)$ noisy channel, using only $\Theta(n)$ bits of communication?

in other words, is it possible for $\ell = \Theta(n)$ when the channel is $(n/4)$-noisy?

We will answer this question using the PROBABILISTIC METHOD.

Another QN: Let A be any set of integers, $|A| = n$. We want a subset of A that is 3-sum free. That is, we want $S \subseteq A$ s.t. $\nexists\, a, b, c \in S$, $a + b \neq c$. How large can S be?

If $A = \{1, 2, \ldots, n\}$, then we can take $S = \{1, 3, 5, \ldots\}$ and this ensures that it is 3-sum free. Can we show such a large subset for any $A \subseteq \mathbb{N}$, $|A| = n$?

large: $\Theta(n)$

Both these questions have no 'probability' involved. But we will resolve them using probability.

Defining Discrete Probability Formally:

Ω : finite / countably infinite set

sample space

$p : \Omega \to \underbrace{\mathbb{R}^{\geq 0}}_{\text{non negative real numbers}}$ s.t. $\sum_{x \in \Omega} p(x) = 1.$

(Ω, p) define a discrete probability dist.

An event $\mathcal{E}$ is any subset of Ω.

$$Pr[\mathcal{E}] = \sum_{x \in \mathcal{E}} p(x).$$

Using properties of finite sets, we can arrive at simple but useful probability bounds.

For any sets A, B, $|A| + |B| = |A \cup B| + |A \cap B|$

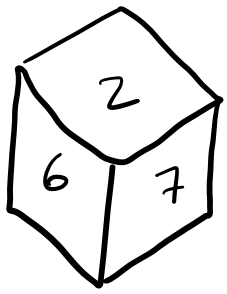
Thm [Union bound]:

For any events $\mathcal{E}_1, \mathcal{E}_2 \subseteq \Omega$,

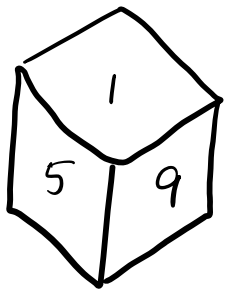
$$Pr[\mathcal{E}_1 \cup \mathcal{E}_2] \leq Pr[\mathcal{E}_1] + Pr[\mathcal{E}_2]$$

Examples of discrete probability dist.ⁿ :

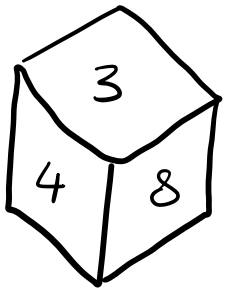
Ex 1 : See Section 17.3 from [LLM18]



A



B



C

Given : Three dice A, B, C s.t. each dice has same numbers on opposite faces.

Two player game. Player 1 picks a dice, player 2 picks another dice.

Event : Both players roll their dice once. Player 1's dice has higher number than Player 2's dice.

Player 1 wins if $Pr[\text{Event}] \geq 1/2$.

Claim: For every choice of dice by Player 1, $\exists$ choice of dice by Player 2 s.t. $\Pr[\text{Event}] = \frac{4}{9}$.

In other words, Player 1 always loses.

This is somewhat surprising, because
dice A is 'better' than dice B
" B " " " " C
" C " " " " A.

Transitivity does not hold.

Here, I will only list the sample space for the case where Player 1 picks A, Player 2 picks 'C'.

$$\Omega = \{(2,3), (2,4), (2,8), (6,3), (6,4), (6,8), (7,3), (7,4), (7,8)\}$$

$p(x) = 1/9$ for all $x \in \Omega$.

E: "Player 1 wins" corresponds to $\{(6,3), (6,4), (7,3), (7,4)\}$

$\therefore \Pr[E] = 4/9.$

Suppose we consider a different experiment:
Player 1 picks a dice and rolls it twice
Player 2 picks a dice and rolls it twice.

Event: sum of the two rolls of Player 1
> " " " " " " " 2.

Suppose Player 1 picks A
" 2 " C.

Here, $\Omega = \left\{ (a,b,c,d) : \begin{array}{l} a,b \in \{2,6,7\} \\ c,d \in \{3,4,8\} \end{array} \right\}$

$$\text{Event} = \left\{ (a,b,c,d) \in \Omega \text{ s.t. } a+b > c+d \right\}$$

Compute $\Pr[\text{Event}]$.

It may be easier to write a short program that computes this.

■

Qn: Generalize the above for $(2k)$-sided dice.

Ex 2 : Birthday Paradox (Sec. 17.4 from [LLM18])

Experiment : t numbers are sampled uniformly at random from $\{1, 2, \ldots, n\}$. What is the probability that at least two of the sampled numbers are equal ?

$$\Omega = \left\{ (a_1 \ldots a_t) : \; a_i \in \{1, 2, \ldots, n\} \right\}$$

$$\mathcal{E} = \left\{ (a_1 \ldots a_t) \in \Omega \text{ s.t. } \exists\, i, j \in \{1, \ldots, t\},\ i \neq j,\ a_i = a_j \right\}$$

$$? \leq \Pr[\mathcal{E}] \leq ?$$

Upper bd. on $\Pr[\mathcal{E}]$:

For any i, j, $i < j$, let $\mathcal{E}_{ij} = \left\{ (a_1 \ldots a_n) \text{ s.t. } a_i = a_j \right\}$

$$\mathcal{E} = \bigcup_{i<j} \mathcal{E}_{ij} \qquad \Pr[\mathcal{E}_{ij}] = 1/n .$$

$\therefore$ Using union bd, $\Pr[\mathcal{E}] \leq \sum_{i<j} \Pr[\mathcal{E}_{ij}]$

$$= \frac{t \cdot (t-1)}{2n} .$$

■

Next, we will compute a lower bound on $\Pr[\mathcal{E}]$. Instead of computing $\Pr[\mathcal{E}]$, we will compute $\Pr[\bar{\mathcal{E}}] = \Pr[\Omega \setminus \mathcal{E}]$. Observe $\Pr[\mathcal{E}] = 1 - \Pr[\bar{\mathcal{E}}]$, and therefore it suffices to compute an upper bound on $\Pr[\bar{\mathcal{E}}]$.

$$\bar{\mathcal{E}} = \left\{ (a_1 \cdots a_t) : \forall\, i < j,\ a_i \neq a_j \right\}$$

$$|\bar{\mathcal{E}}| = n \cdot (n-1) \cdot \cdots \cdot (n-t+1)$$

$$\Pr[\bar{\mathcal{E}}] = \frac{n-1}{n} \cdot \cdots \cdot \frac{(n-t+1)}{n}$$

$$= \left(1 - \tfrac{1}{n}\right)\left(1 - \tfrac{2}{n}\right) \cdots \left(1 - \tfrac{t-1}{n}\right)$$

Fact : $1 - x \leq e^{-x}$ for all x

$$\Pr[\bar{\mathcal{E}}] \leq e^{-1/n} \cdot e^{-2/n} \cdot \ldots \cdot e^{-(t-1)/n}$$

$$= e^{-[(t-1)t/2n]}$$

$$\Pr[\mathcal{E}] \geq 1 - e^{-[(t-1)t/2n]}$$

If $(t-1) > 2\sqrt{n}$, then $\frac{(t-1)t}{2n} > 2$

$$-\frac{(t-1)t}{2n} < 2$$

$$e^{-(t-1)t/2n} < e^{-2}$$

$$1 - e^{-(t-1)t/2n} > 1 - e^{-2}$$

$$\Pr[\mathcal{E}] > 1 - e^{-2}$$ ∎

This is called birthday "paradox" because people, somehow, expect the probability of collision to be very low if $t \sim 2\sqrt{n}$. This flawed intuition was the source of several crypto vulnerabilities.

Qn: Given a sequence of t numbers $(a_1, \ldots, a_t)$ where each $a_i \in \{1, \ldots, n\}$, we say that $(a_1, \ldots, a_t)$ has a k-collision if $\exists\ i_1 < \ldots < i_k$

s.t. $a_{i_1} = a_{i_2} = \cdots = a_{i_k}$.

Experiment : sample $a_1, a_2, \ldots, a_t$ unif. at rand. from $\{1, 2, \ldots, n\}$.

Event : k-collision exists.

How large must t be (as a fn. of n, k) so that $\Pr[\text{Event}] > 1/2$?

SUMMARY, QUESTIONS :

Discrete Prob. dist^n defined using a finite / countably infinite set Ω, together with $p : \Omega \to \mathbb{R}^{\geq 0}$, $\sum_{x \in \Omega} p(x) = 1$

Event $\mathcal{E}$: subset of Ω.

$\Pr[\mathcal{E}] = \sum_{x \in \mathcal{E}} p(x)$

$$\forall \ \mathcal{E}_1, \ \mathcal{E}_2, \quad Pr[\mathcal{E}_1 \cup \mathcal{E}_2] \leq Pr[\mathcal{E}_1] + Pr[\mathcal{E}_2].$$

Birthday Paradox / Bound :

t numbers sampled unif. at rand. from $\{1, \ldots, n\}$.
Event $\mathcal{E}$: at least 2 of the sampled numbers are equal.

$$1 - e^{-[t(t-1)/2n]} \leq Pr[\mathcal{E}] \leq \frac{t(t-1)}{2n}$$

Questions :

1. Let $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ be a unif. rand. permutation. Let i, k be fixed numbers in $\{1, 2, \ldots, n\}$. We are interested in the probability that $\sigma(i) = k$.

Define the sample space Ω, and event $\mathcal{E}$. What is $Pr[\mathcal{E}]$?

ans : Ω = { set of all permutations $\sigma : [n] \to [n]$ }
$|\Omega| = n!$
$\mathcal{E} = \{ \sigma \in \Omega \text{ s.t. } \sigma(i) = k \}$
$Pr[\mathcal{E}] = 1/n$.

2. Let $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ be a unif. rand. permutation. We write down the numbers s.t. number k appears at position $\sigma(k)$.

Example : $n = 4$.

$\sigma(1) = 4 \quad \sigma(2) = 2 \quad \sigma(3) = 1 \quad \sigma(4) = 3$

$$(3 \quad 2 \quad 4 \quad 1)$$

Let i be a fixed number in $\{1, 2, \ldots, n\}$.

Event $\mathcal{E}_i$: After writing down the n numbers, the i^{th} number is the largest among the first i numbers.

Define the sample space Ω, and event $\mathcal{E}_i$. What is $\Pr[\mathcal{E}_i]$?

$\Omega = \{$ set of all permutations $\sigma : [n] \to [n]\}$

$\mathcal{E}_i = \{\sigma \in \Omega$ s.t. $\sigma^{-1}(i) > \sigma^{-1}(j)$ for all $j < i\}$

$\Pr[\mathcal{E}_i] = ?$

(*) Qn: Given a sequence of t numbers $(a_1, \ldots, a_t)$ where each $a_i \in \{1, \ldots, n\}$, we say that $(a_1, \ldots, a_t)$ has a k-collision if $\exists\ i_1 < \ldots < i_k$ s.t. $a_{i_1} = a_{i_2} = \ldots = a_{i_k}$.

Experiment: sample $a_1, a_2, \ldots, a_t$ unif. at rand. from $\{1, 2, \ldots, n\}$.

Event: k-collision exists.

How large must t be (as a fn. of n, k) so that $\Pr[\text{Event}] > 1/2$?

(*) Qn: Let $\Omega = \{1, 2, \ldots, n\}$, $p : \Omega \to \mathbb{R}^{\geq 0}$, $\sum_{x \in \Omega} p(x) = 1$

p defines a probability dist.^{n} over Ω.

Suppose we sample $a_1 \ldots a_t$, each a_i sampled independently from (Ω, p).

Let $t = \Theta(\sqrt{n})$. Can we conclude that $\exists$ constant c s.t. $\Pr[\text{collision}] \geq c$?

(**) Qn: Generalize the strange dice phenomenon for $(2k)$-sided dice.