APPLICATION 1: NORMAL NUMBERS

Def A number $x \in \mathbb{R}$ is called normal if, in its Binary expansion,

$$\frac{\#\{1\text{'s in the first } n \text{ digits}\}}{n} \to \frac{1}{2} \quad \text{as } n \to \infty$$

- May be generalized to base 10 (fraction of $\forall$ digit $\to 1/10$)

THM [Borel 1909] Almost all numbers are normal

← i.e. the non-normal #'s form a set of Lebesgue meas $=0$

Proof WLOG for $[0,1]$.

Let $X \sim \text{Unif}[0,1]$. Binary expansion:

$$X = 0.X_1 X_2 X_3 \dots$$

Then X_i are iid Ber($1/2$) r.v's [KW]

SLLN $\Rightarrow$ $\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{= \#\{1\text{'s in the first } n \text{ digits}\}} \longrightarrow \frac{1}{2}$ a.s.

$\square$

OPEN QUESTIONS: Is π normal?

Is e normal?

Is $\sqrt{2}$ normal?

Not even known if all digits occur i.o. in those numbers.

CONJECTURE: $\forall$ irrational algebraic number is normal

↑ root of a nonzero polynomial.

APPLICATION 2: SHANNON'S SOURCE CODING THM

- Imagine a source ("channel") that produces a stream of random iid symbols from some alphabet $S = \{A, B, C, D, E, \dots, Z\}$: symbol A with prob. $p(A)$, symbol B with prob $p(B)$, etc.

  Compress the source?

- WLOG $S \subset \mathbb{R}$, let X be a discrete r.v. with pmf
$$\boxed{p(x) := \mathbb{P}\{X = x\}}$$

- Let $X_1, X_2, \dots$ be iid copies of X (source)

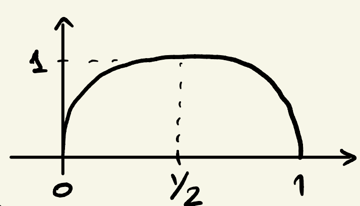
- The likelihood of $\forall$ given word $x_1 x_2 \dots x_n$ (e.g. "LIAM") is
$$p(x_1, \dots, x_n) := \mathbb{P}\{(X_1, \dots, X_n) = (x_1, \dots, x_n)\} = p(x_1) \cdots p(x_n)$$

Take logs, divide by n $\Rightarrow$
$$-\frac{1}{n} \log p(X_1, \dots, X_n) = \frac{1}{n} \sum_{k=1}^{n} -\log p(X_k) \xrightarrow{a.s.} \mathbb{E}[-\log p(X)] \quad \text{(SLLN)}$$
$$= \boxed{-\sum_x p(x) \log p(x) =: H(X)}$$
$\leftarrow$ Def Entropy of X

$\approx$ # bits needed to encode X (next page)

- $\Rightarrow$ $\boxed{p(X_1, \dots, X_n) \approx 2^{-nH(X)}}$ "Asymptotic Equipartition Property" = likelihood of $\forall$ output

Ex ① $X \sim \mathrm{Ber}(p)$ $\Rightarrow$
$$\boxed{H(X) = -p \log p - (1-p) \log(1-p)} \longrightarrow$$



Zero if $p=0$ or $p=1$ (deterministic source)
Maximal(1) if $p = ½$ (most random source) $\Rightarrow$ $p(X_1, \dots, X_n) \approx 2^{-n}$

② $X \sim \mathrm{Unif}\{1, \dots, N\}$ $\Rightarrow$ $\boxed{H(X) = \log N}$ $\leftarrow$ $\log N$ bits needed to specify a # in $\{1, \dots, N\}$

③ X = random letter from English alphabet $\Rightarrow$ $H \approx 2.6$

- Shannon's Thm: "source can be compressed with rate $\approx H(X)$ bits/symbol such that almost no info is lost (i.e. the original symbols can be recovered with prob ≈ 1). Moreover, the rate $H(X)$ is optimal":

Ex: $H \approx 2.6$ bits per English letter (as opposed to $\log_2 26 \approx 4.7$ if all letters were equally likely)

THM (Shannon's Source Coding Thm)

Let X be a r.v. that takes values in a countable set $S \subset \mathbb{R}$.
Let $X_1, X_2, \ldots$ be iid copies of X.
$\forall \varepsilon > 0 \ \exists n_0 > 0 \ \forall n > n_0$:

encoder, decoder

1. $\exists$ pair of maps $S^n \xrightarrow{E} \{0,1\}^{n(H(X)+\varepsilon)} \xrightarrow{D} S^n$ such that

$$\mathbb{P}\{D(E(X_1,\ldots,X_n)) = (X_1,\ldots,X_n)\} > 1-\varepsilon.$$

2. $\forall$ pair of maps $S^n \xrightarrow{E} \{0,1\}^{n(H(X)-\varepsilon)} \xrightarrow{D} S^n$,

$$\mathbb{P}\{D(E(X_1,\ldots,X_n)) = (X_1,\ldots,X_n)\} < \varepsilon.$$

Proof ① (Asymptotic Equipartition Property): By WLLN (p. 97):

$$-\frac{1}{n}\log p(X_1,\ldots,X_n) \xrightarrow{P} H(X)$$

$\Rightarrow \forall \varepsilon > 0 \ \exists n_0 \ \forall n > n_0$:

$$\mathbb{P}\left\{\left|-\frac{1}{n}\log p(X_1,\ldots,X_n) - H(X)\right| \le \varepsilon\right\} > 1-\varepsilon$$

$\Rightarrow$ the "good" subset $A_n \subset S^n$ defined as

$$A_n := \left\{(x_1,\ldots,x_n):\ 2^{-n(H(X)+\varepsilon)} \le p(x_1,\ldots,x_n) \le 2^{n(H(X)+\varepsilon)}\right\} \qquad (*)$$

satisfies

$$\mathbb{P}\{(X_1,\ldots,X_n) \in A_n\} > 1-\varepsilon. \qquad (**)$$

② (Existence of encoding-decoding) We have

$$1 \geq \mathbb{P}\{(X_1,\dots,X_n) \in A\} = \sum_{(x_1,\dots,x_n)\in A} p(x_1,\dots,x_n) \underset{\substack{\uparrow \\ \text{lower bd in } (* \text{ p.98})}}{\geq} |A| \cdot 2^{-n(H(X)+\varepsilon)}$$

$$\Rightarrow |A| \leq 2^{n(H(X)+\varepsilon)}$$

$\Rightarrow \exists$ injective map $E : A_n \to \{0,1\}^{n(H(X)+\varepsilon)}$

let D be the inverse of E on its image.

Extend E and D arbitrarily $\Rightarrow$

$$D\big(E(x_1,\dots,x_n)\big) = (x_1,\dots,x_n) \quad \text{whenever } (x_1,\dots,x_n) \in A_n$$

By (** p. 98),

$$D\big(E(X_1,\dots,X_n)\big) = (X_1,\dots,X_n) \text{ with prob } > 1-\varepsilon.$$

☺

③ (Optimality) WLOG will prove the second part of thm for 2ε instead of ε.

Consider $\forall$ pair of maps $S^n \xrightarrow{E} \{0,1\}^{n(H(X)-2\varepsilon)} \xrightarrow{D} S^n$

and define the subset

$$B := \{(x_1,\dots,x_n) : D(E(x_1,\dots,x_n)) = (x_1,\dots,x_n)\}.$$

By the factorization, $\quad |B| \leq 2^{n(H(X)-2\varepsilon)} \quad (\bigstar)$

Intersecting with the equipartition set A from (* p. 98) $\Rightarrow$

$$\mathbb{P}\{(X_1,\dots,X_n)\in B\} \leq \underbrace{\mathbb{P}\{(X_1,\dots,X_n) \in A \cap B\}}_{= \sum_{(x_1,\dots,x_n)\in A\cap B} p(x_1,\dots,x_n) \overset{(* \text{ p.98})}{\leq} |B| \cdot 2^{n(H(X)+\varepsilon)} \overset{(\bigstar)}{\leq} 2^{-n\varepsilon}} + \underbrace{\mathbb{P}\{(X_1,\dots,X_n) \in A^c\}}_{\overset{(** \text{ p.98})}{\leq} \varepsilon}$$

$$\leq 2^{-n\varepsilon} + \varepsilon \leq 2\varepsilon \quad \text{for sufficiently large } n. \ \square$$