Lecture 2

Note Title 1/28/2016

TODAY

- Definition of Entropy
  (also Prob. notation, conditional entropy)
- Compression
  - Asymptotic (n-shot)
  - Single-shot
  - Prior free Asymptotic

Definition of Entropy

Needs Probability notation

In this course Prob ≡ Prob. over finite sets.

$\Omega = \{1, \ldots k\}$

Prob. distribution P on $\Omega = (P(1), P(2) \ldots P(k))$

- Often speak of random variable $X \sim P$
- Often use P_X to denote distribution of X
- Similarly P_{XY}, P_Y etc.
- Also (if distinguishable with my handwriting) use X (caps) to denote r.v., x (lower case) to denote instantiation

---

Some basic Calculus: (Bayes rule etc.)

- Say $(X, Y) \sim P_{XY}$
- So $P_{XY}(a,b) \triangleq P_r[X = a, Y = b]$
- $P_X(a) = \sum_b P_{XY}(a,b)$; $P_Y(b)$ similar
- $P_{X|b}(a) = \dfrac{P_{XY}(a,b)}{P_Y(b)}$

$P_{X|Y} \triangleq$ Notation! represents $\{P_{X|b}\}_b$

Definition of Entropy

- $X \sim P_X$ has entropy.

$$H(X) \triangleq \sum_{a \in \Omega} P_X(a) \log_2 \frac{1}{P_X(a)}$$

- $$H(X|Y) \triangleq \mathop{\mathbb{E}}_{b} \left[H(X|Y=b) \right]$$

$$= \sum_{b} P_Y(b) \cdot H(X|Y=b)$$

Where does $H(X)$ come from?

- Diff answers from diff sources
- To me ... Stirling's formula.

$$n! \approx \left(\frac{n}{e}\right)^n$$

$t = pn$: int.

$$\Rightarrow \binom{n}{pn} = \frac{n!}{(pn)!\,((1-p)n)!}$$

$$\approx \left(\left(\frac{1}{p}\right)^{p} \cdot \left(\frac{1}{1-p}\right)^{1-p} \right)^{n}$$

$$\approx \left(H(P)\right)^{n}$$

where $P = (P(0), P(1))$

$\Updownarrow$

$(p, 1-p)$.

Why $\binom{n}{pn}$? What is $\binom{n}{pn}$?

– Compression!

Asymptotic Compression Problem

Given P_X on $\Omega = \{1, \dots k\}$

Sender gets $X_1, X_2, \dots X_n$ i.i.d. with

$$X_i \sim P_X$$

- Sender wishes to send $C(X_1 \dots X_n)$ to receiver

$$C : \Omega^n \to \{0,1\}^{\ell_n}$$

- Receiver receives $W = C(X_1 \dots X_n)$

↳ decompresses

$$D : \{0,1\}^{\ell_n} \to \Omega^n$$

∴ Want to minimize ℓ_n s.t.

$$P_{err} = P_r\left[D(C(X_1 \dots X_n)) \neq (X_1 \dots X_n) \right] \leq \epsilon$$

Asymptotic Compression Rate

$$= \lim_{\epsilon \to 0} \lim_{n \to \infty} \left\{ \frac{\ell_n}{n} \right\}$$

where ℓ_n s.t. $\exists$ E, D s.t. $P_{err} \leq \epsilon$

Theorem [Shannon '48]

$\forall$ P_X, Asymptotic Compression Rate $= H(X)$

Proof: (Special Case $\Omega = \{0,1\}$

$P_X = (1-p, p)$

$\uparrow P_X(0), \uparrow P_X(1)$)

Need to show

(1) $\forall \epsilon \exists n_0$ $\forall n \geq n_0$ $\exists E, D$ with $\ell_n \leq (1+o(1)) \cdot H(p) \cdot n$ s.t. $P_{err} \leq \epsilon$

(2) $\forall \epsilon$ $\forall$ suff. large n, $\forall E, D$ w. $\ell_n \leq (1-o(1)) \cdot H(p) \cdot n$. $P_{err} \geq 1 - \epsilon$

- (Elementary) "Structural" Observation

w.h.p. $\geq \left(1 - \frac{\epsilon}{2}\right)$ # 1's in $X_1 \ldots X_n$ $\in \left[(p-\delta)n, (p+\delta)n\right]$

($\forall \epsilon, \delta$, $\forall$ suff. large n)

[Chernoff Bounds ...]

- Assume (with $\frac{\epsilon}{2}$ loss of generality)

(a) # 1's $= t \in \left[(p-\delta)n, (p+\delta)n\right]$

(b) Receiver knows t

→ for upper bound, can transmit t for cost $\log n$ ①

→ for lower bound, assumption strengthens receiver.

- Under Assumption:

Sender enumerates all $\binom{n}{t}$ possibilities of $(X_1 \ldots X_n)$; & sends the right index for achieved sequence.

How many possibilities?

$$\log \binom{n}{t} = \log \binom{n}{pn} \approx H(p) \cdot n \text{ !}$$

(Essentially Proves (1))

6 Under Same Assumption if

$$\frac{2^{ln}}{2^{H(p) \cdot n}} < \frac{\epsilon}{2} \quad \text{Then .....}$$

$$\Pr[\text{any single message}] \leq \frac{1}{\binom{n}{t}}$$

$$\Pr[\text{correct decoding}] \leq 2^{ln} \cdot \frac{1}{\binom{n}{t}} \leq \frac{\epsilon}{2}$$

General Distributions P_X

(AEP): Asymptotic Equipartition Property

$\forall P_X \quad \forall \epsilon \quad \forall$ sufficiently large n

$(X_1 \ldots X_n)$ almost uniformly chosen from universe of size $2^{H(X) \cdot n}$

More formally:

$\exists S \subseteq \Omega^n, \quad |S| \leq 2^{(1+o(1)) \cdot H(X) \cdot n}$

- $\Pr[(X_1 \ldots X_n) \notin S] \leq \epsilon$
- and $\forall (a_1 \ldots a_n) \in S$

$$\Pr[(X_1 \ldots X_n) = (a_1 \ldots a_n)] \leq \frac{2^{o(n)}}{|S|}$$

AEP $\Rightarrow$ [Shannon] Exercise

Why is AEP true? Whither $H(X)$?

Let $P_X = (p_1, \dots p_k)$

- When n large, Expect to see

roughly $\left\{ \begin{array}{ll} p_1 n & 1's \\ p_2 n & 2's \\ \vdots & \\ p_k n & k's \end{array} \right\}$ in $(X_1 \dots X_n)$

- # such $(X_1 \dots X_n)$

$$\approx \binom{n}{p_1 n \;\; p_2 n, \dots \; p_k n}$$

↑ multinomial coefficient

$$\approx 2^{H(X) \cdot n} \quad [\text{Stirling!}]$$

Asymptotic vs. Single-Shot Compression

Single-Shot Problem Defn

prefix-free!

Sender: $X \sim P_X$; $C : \Omega \to \{0,1\}^*$

variable length!

Receiver: $W = C(X)$; $D : \{0,1\}^* \to \Omega$

Want: $\forall x \in \Omega \quad D(C(x)) = x$

Objective: $\min \mathop{\mathbb{E}}_{x \sim P_X} [\,|C(x)|\,]$

Differences: - Error-free
- Variable length
- Expected length minimization
- Prefix-free =? (What? Why?)

- $C : \Omega \to \{0,1\}^*$ is prefix-free if

$\forall\, x \neq y \in \Omega$

$C(x)$ not a prefix of $C(y)$

- Why? Soln to single-shot $\Rightarrow$ soln to n-shot.

Kraft Inequality

Fix $\{\ell_x\}_{x \in \Omega}$.

Say C has pattern $\{\ell_x\}$ if

$$|C(x)| = \ell_x \quad \forall\, x \in \Omega$$

Lemma [Kraft?]: $\exists$ prefix-free C with pattern $\{\ell_x\} \iff \sum_x \frac{1}{2^{\ell_x}} \leq 1$

Proof: Let $\ell = \max_x \ell_x$

Existence Pick $C(1), C(2) \ldots$ greedily.

- if $C(1) = 010$; eliminate all extensions
- Claim: Each choice rules out $2^{-\ell_x}$ fraction of length ℓ strings.
- So as long as $\sum \frac{1}{2^{\ell_x}} < 1$ $\exists$ one string of each length uneliminated.

Converse

- Pick random ℓ bit string w
- Let $\mathcal{E}_x$ = event that $C(x)$ = prefix of w
- $\mathcal{E}_x, \mathcal{E}_y$ = mutually exclusive
- $\Pr[\mathcal{E}_x] = \frac{1}{2^{\ell_x}}$

$$\Rightarrow \sum_x \frac{1}{2^{\ell_x}} = \sum_x \Pr[\mathcal{E}_x] \leq 1$$

⊠

Kraft $\Rightarrow$ Shannon 1-Shot ?

- let $\ell_x = \left\lceil \log \frac{1}{P_X(x)} \right\rceil$

$$\sum 2^{-\ell_x} \leq \sum 2^{-\log \frac{1}{P_x(x)}} \leq \sum P(x) \leq 1$$

So Kraft $\Rightarrow$ prefix-free C exists.

- But $\mathop{\mathbb{E}}_x [\ell_x] \leq \mathop{\mathbb{E}}_x \left[1 + \log \frac{1}{P_x(x)}\right]$

$$\leq 1 + H(P)$$

- Lower Bound for 1-shot compression?

Two approaches:

① Invoke n-shot lower bound + ...

[works, details omitted but you can do it.]

② Consider $q_x = 2^{-\ell_x}$ & show

$$\sum p_x \log \frac{1}{q_x} \geq \sum p_x \log \frac{1}{p_x}$$

↑

[will consider this in later lectures]

In fact

$$\sum p_x \left[\log \frac{1}{q_x} - \log \frac{1}{p_x} \right] \text{ is}$$

important measure in Information Theory.

Next Lecture

- Axioms of Information
- Basic Inequalities

Future lectures

Topic 1: Entropy & Counting

[Need Volunteers / Presenters]

Topic 2: Entropy & Hypothesis Testing

[Need to collect papers]

Source for today's lecture [Kraft ⇐ Tulsiani course]