CS258: Information Theory

Fan Cheng
Shanghai Jiao Tong University

http://www.cs.sjtu.edu.cn/~chengfan/
chengfan@sjtu.edu.cn
Spring, 2020

Outline

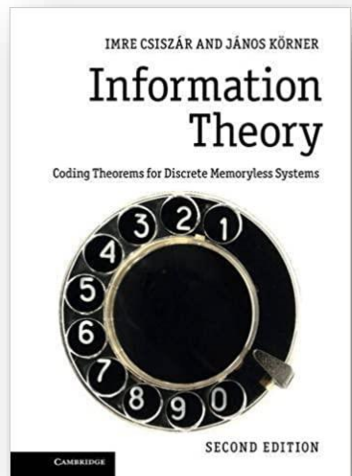
- Types
- Applications of Types

Types: brief history

Weak typicality
C. E. Shannon

Strong typicality
Toby Berger (02)


Theory of types
Imre Csiszar (96) and Janos Korner (14)



Types: idea

The central problem of science is to understand

$$\Pr(\boldsymbol{X_1}, \dots, \boldsymbol{X_n})$$

where $X_1, X_2, \dots, X_n$ are i.i.d $\sim p(x)$ over $\mathcal{X} = \{x_1, x_2, \dots, x_m\}$

- Main idea: Count the number of x_i appears in $X_1, \dots, X_n$.

Let $\mathcal{X} = \{1,2,3\}, p = (\frac{1}{3}, \frac{1}{6}, \frac{1}{2})$

$$\Pr(1,2,3,3,2,1,1) = \left(\frac{1}{3}\right)^3 \left(\frac{1}{6}\right)^2 \left(\frac{1}{2}\right)^2 \Rightarrow (3,2,2)$$

According to their events, X_i can be rewritten as

$$\mathbf{Pr}(\boldsymbol{X_1}, \dots, \boldsymbol{X_n}) = \boldsymbol{p(x_1)^{n_1} p(x_2)^{n_2} \dots p(x_m)^{n_m}}$$

where $n_i's$ are nonnegative integers and

$$\boldsymbol{\sum_{i=1}^{m} n_i = n.}$$

- $\Pr(X_1, X_2, \dots, X_n)$ can be grouped by the corresponding $(n_1, \dots, n_m)$

$$q_i = \frac{n_i}{n}$$

When n is given, $q = (q_1, \dots, q_m)$ is a probability distribution.

***q* is the type of $\Pr(X_1, \dots, X_n) \leftrightarrow p, q$**

$$\boldsymbol{\sum_{i=1}^{m} n_i = n, q = \left(\frac{n_1}{n}, \dots, \frac{n_m}{n}\right)}$$

Types: basic

Let $\mathcal{P}_n$ denotes the set of types with denominator n.

$$\boldsymbol{\mathcal{P}_n = \left\{\left(\frac{n_1}{n}, \dots, \frac{n_m}{n}\right) : \sum_{i=1}^{m} n_i = n, n_i \in \mathbb{N}\right\}}$$

For a sequence $\mathrm{x} \in \mathcal{X}^n$, denote its type by P_{x} ($m = |\mathcal{X}|$)

Let $P \in \mathcal{P}_n$, the set of sequences of length n and type P is called the **type class of P**, denoted $T(P)$:

$$T(P) = \{\mathrm{x} \in \mathcal{X}^n : P_{\mathrm{x}} = P\}.$$

- $$|\mathcal{P}_n| \leq (n+1)^{|\mathcal{X}|} \text{ (a polynomial in } n\text{)}$$

If $X_1, X_2, \dots, X_n$ are drawn i.i.d. according to $Q(x)$, the probability of x depends on its types and is given by

$$\boldsymbol{Q^n(\mathrm{x}) = 2^{-n(H(P_{\mathrm{x}}) + D(P_{\mathrm{x}}||Q))}}$$

$$a^b = 2^{b \log a}$$

$$Q^n(x) = \prod_{a \in \mathcal{X}} Q(a)^{nP_{\mathrm{x}}(a)} = \prod_{a \in \mathcal{X}} 2^{nP_{\mathrm{x}}(a) \log Q(a)} = 2^{\sum_{a \in \mathcal{X}} nP_{\mathrm{x}}(a) \log Q(a)} = 2^{n(-D(P_{\mathrm{x}}||Q) - H(P_{\mathrm{x}}))}$$

- If x is in the type class of Q, then

$$Q^n(\mathrm{x}) = 2^{-nH(Q)} \Leftarrow P_{\mathrm{x}} = Q$$

Types: cardinality and probability

- **(Cardinality)** For any type $P \in \mathcal{P}_n$,

$$\frac{1}{(n+1)^{|\mathcal{X}|}} 2^{nH(P)} \leq |T(P)| \leq 2^{nH(P)}$$

- **(Probability)** For any $P \in \mathcal{P}_n$ and any distribution Q, the probability of the type class $T(P)$ under Q^n is $2^{-nD(P||Q)}$ to the first order in the exponent. More precisely,

$$\frac{1}{(n+1)^{|\mathcal{X}|}} 2^{-nD(P||Q)} \leq Q^n(T(P)) \leq 2^{-nD(P||Q)}$$

- There are only a polynomial number of types, and an exponential number of sequences of each type
- $(n+1)^{|\mathcal{X}|}$ is tiny compared to $2^{nH(P)}$ or $2^{nD(P||Q)}$
- If $P \neq Q$, $Q^n\big(T(P)\big) \to 0$

Reference: Ch. 11.1of T. Cover

Types ⇒ LLN

- (Typical sets) Given an $\epsilon > 0$, we can define a typical set T_Q^ϵ of sequences for the distribution Q^n as

$$T_Q^\epsilon = \{x^n : D(P_{x^n}||Q) \le \epsilon\}.$$

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. $\sim P(x)$. Then

$$\Pr\{D(P_{x^n}||P) > \epsilon\} \le 2^{-n\left(\epsilon - \frac{|\mathcal{X}|\log(n+1)}{n}\right)},$$

and consequently, $D(P_{x^n}||P) \to 0$ with probability 1.

- $D(P'||P) = 0$ iff $P' = P$
- $P' \to P \Rightarrow \text{WLLN}$

Reference: Ch. 11.2 of T. Cover