强弱典型集

May 27, 2023

Yimin Zhao
ym-zhao.com

记号

- X, Y 等表示离散随机变量。
- x, y 等表示离散随机变量取得的确定值。
- $\mathcal{X}, \mathcal{Y}$ 表示有限集，其包含X, Y的所有可能取值，也被称作字母表（alphabet）。
- 给定离散随机变量X, $g(X)$是以X为定义域的一个函数，$\mathbb{E}(g(X))$ 表示$g(X)$的数学期望。
- $p(\cdot)$表示事件$\cdot$的概率, $p(X = x)$可以简写为$p(x)$, $p(X)$指代X的概率密度函数。
- "i.i.d" 表示 independent, identically distributed (独立同分布)。

随机变量的收敛

定义：给定一个随机变量数列$X_1, X_2, ...$, 我们称这个数列 **收敛** 到随机变量X当且仅当满足下列条件任意一个 (这三种定义互相等价，第一种最常见)：

1. 概率角度：$\forall \varepsilon > 0, \varepsilon' > 0, \exists N > 0$, 当$n > N$ 时，$p(|X_n - X| > \varepsilon) < \varepsilon'$
2. 均方差角度：$\forall \varepsilon > 0, \exists N > 0$, 当$n > N$ 时 $E(X_n - X)^2 < \varepsilon$
3. 概率极限角度：$p(\lim_{n \to \infty} X_n = X) = 1$

大数定律

对于i.i.d随机变量 $X_1, X_2, ..., X_n \sim p(x), \overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i$:

- 强大数定律定义：$p\left(\lim_{n \to \infty} \overline{X_n} = E(X_1)\right) = 1$, 即
$$\forall \varepsilon > 0, \exists N > 0, \forall n > N, p\left(\overline{X_n} = E(X_1)\right) = 1$$
- 弱大数定律定义：$\lim_{n \to \infty} p\left(\overline{X_n} = E(X_1)\right) = 1$, 即
$$\forall \varepsilon > 0, \exists N > 0, \forall n > N, p\left(\overline{X_n} = E(X_1)\right) > 1 - \varepsilon$$

弱大数定律的意义在于有些时候随机变量未必存在期望值，此时强大数定律定义不存在。大数定律的直观解释为"样本数量越多，则其算术平均值就有越高的概率接近期望值"。

对于通信中的编码而言，当码长序列足够长时，其中一部分序列就会显示出某种固定性质，（可类比抛硬币当次数越大正反各接近一半）：各个符号出现的频率接近于概率，而这些序列的的概率则趋近于相等，且它们的和非常接近于1。这些序列就是 **典型序列** 。其余不具备这种性质的序列，我们称之为非典型序列，这些非典型序列的出现概率之和接近于零。序列的长度越长，典型序列的总概率就越接近于1，它的各个序列的出现概率越趋近于相等。而我们就把这种现象称之为 **渐近均分性** 。

弱渐进均分性（Weak EAP）

定义：对于i.i.d随机变量 $X_1, X_2, ... \sim p(x)$, 我们用X去代表这些随机变量，那么有

$$-\frac{1}{n} \log p(X_1 X_2 ... X_n) \to H(X)$$

其中箭头表示"概率渐进"。

证明：

$$
\begin{aligned}
-\frac{1}{n}\log p(X_1X_2...X_n) &= -\frac{1}{n}\sum_{i=1}^{n}\log p(X_i)\\
&\to \underbrace{-E\log p(X)}_{\text{弱大数定律}}\\
&= H(X)
\end{aligned}
$$

严格意义上: $\forall\varepsilon>0,\exists N>0,$ 当$n>N$ 时，在概率意义上，

$$
H(X)-\varepsilon\le-\frac{1}{n}\log p(X_1X_2...X_n)\le H(X)+\varepsilon
$$

在信息论中底数是2, 因此上式等价于:

$$
2^{-n(H(X)+\varepsilon)}\le p(X_1X_2...X_n)\le 2^{-n(H(X)-\varepsilon)}
$$

弱典型集（Weak Typical Set）

对于 $p(x)$, 弱典型集$W^n_{[X]\varepsilon}$ 为满足下式条件的所有序列（字符串） $\mathtt{x}=(x_1,x_2,...,x_n)\in\mathcal{X}^n$ 所构成:

$$
2^{-n(H(X)+\varepsilon)}\le p(\mathtt{x})\le 2^{-n(H(X)-\varepsilon)}
$$

性质:

1. $W^n_{[X]\varepsilon}$ 是 $\mathcal{X}^n$ 的子集，一般认为码元是2，因此$|\mathcal{X}^n|=2^n$
2. $\forall\varepsilon>0,\exists N>0,$ 当$n>N,\left|W^n_{[X]\varepsilon}\right|\le 2^{n(H(X)+\varepsilon)}$, 证明:

$$
\begin{aligned}
1 &= \sum_{x\in\mathcal{X}^n}p(x)\\
&\ge \sum_{x\in W^n_{[X]\varepsilon}}p(x)\\
&\ge \sum_{x\in W^n_{[X]\varepsilon}}2^{-n(H(X)+\varepsilon)}\\
&= 2^{-n(H(X)+\varepsilon)}\left|W^n_{[X]\varepsilon}\right|
\end{aligned}
$$

3. $\forall\varepsilon>0,\exists N>0,$ 当$n>N,\left|W^n_{[X]\varepsilon}\right|\ge(1-\varepsilon)2^{n(H(X)-\varepsilon)}$, 证明:

$$
\begin{aligned}
1-\varepsilon &< p\left(W^n_{[X]\varepsilon}\right)\\
&\le \sum_{x\in W^n_{[X]\varepsilon}}2^{-n(H(X)-\varepsilon)}\\
&= 2^{-n(H(X)-\varepsilon)}\left|W^n_{[X]\varepsilon}\right|
\end{aligned}
$$

4. $\frac{\left|W^n_{[X]\varepsilon}\right|}{|\mathcal{X}^n|}\le 2^{n(H(X)-\log|X|)}$, 因此当$n$足够大时，弱典型集内元素数量占比可以忽略不计，证明:

$$
\begin{aligned}
\frac{\left|W^n_{[X]\varepsilon}\right|}{|\mathcal{X}^n|} &= \frac{\left|W^n_{[X]\varepsilon}\right|}{2^n}\\
&\le 2^{n(H(X)-1+\varepsilon)}\\
&= \underbrace{2^{n(H(X)-\log|X|)}}_{\log|X|=\log_2^2=1}
\end{aligned}
$$

5. $\forall\varepsilon>0,\exists N>0,$ 当$n>N,1-\varepsilon<p\left(\mathtt{x}\in W^n_{[X]\varepsilon}\right)\le 1,$ 因此当n足够大时，任意字符串属于弱典型集的概率都接近1

强典型集（Strong Typical Set）

对于 $p(x)$, 弱典型集$T_{[X]\delta}^{n}$ 为满足下式条件的所有序列（字符串）$\mathtt{x}=(x_1,x_2,...,x_n)\in\mathcal{X}^n$ 所构成：

$$\sum_x\left|\frac{1}{n}N(x;\mathtt{x})-p(x)\right|\leq\delta$$

其中$N(x;\mathtt{x})$表示字符x在字符串$\mathtt{x}$中的出现次数。因此这个定义的含意是“如果一个字符串中每种字符出现频率都近似概率，那么这个字符串就属于强典型集”。

性质：

1. 对于 $\mathtt{x}\in T_{[X]\delta}^{n}$

$$2^{-n(H(X)+\eta)}\leq p(\mathtt{x})\leq 2^{-n(H(X)-\eta)}$$

证明：

$$\begin{aligned}
\log p(\mathtt{x}) &= \log\sum_x p(x)^{N(x,\mathtt{x})}\\
&=\sum_x N(x,\mathtt{x})\log p(x)\\
&=n\sum_x\left[\left(\frac{N(x,\mathtt{x})}{n}-p(x)+p(x)\right)\log p(x)\right]\\
&=n\sum_x\left[\left(\frac{N(x,\mathtt{x})}{n}-p(x)\right)\log p(x)+p(x)\log p(x)\right]\\
&=-n\sum_x\left[\left(\frac{N(x,\mathtt{x})}{n}-p(x)\right)(-\log p(x))\right]+nH(X)
\end{aligned}$$

由于

$$\begin{aligned}
\left|\sum_x\left[\left(\frac{N(x,\mathtt{x})}{n}-p(x)\right)(-\log p(x))\right]\right| &\leq\sum_x\left|\frac{N(x,\mathtt{x})}{n}-p(x)\right|(-\log p(x))\\
&\leq\delta\times\max_x(-\log p(x))\\
&=\eta
\end{aligned}$$

其中$\delta\to0$时$\eta\to0$, 因此

$$-n(H(X)+\eta)\leq\log p(x)\leq-n(H(X)-\eta)$$

2. $\forall\delta,\eta>0,\exists N>0,$ 当$n>N$

$$(1-\delta)2^{n(H(X)-\eta)}\leq\left|T_{[X]\delta}^{n}\right|\leq 2^{n(H(X)+\eta)}$$

证明同弱典型集

3. $\forall\delta>0,\exists N>0,$ 当$n>N,1-\delta<p\left(\mathtt{x}\in T_{[X]\delta}^{n}\right)\leq1,$ 因此当n足够大时，任意字符串属于强典型集的概率都接近1, 证明：

$$
\begin{aligned}
p\left(\mathtt{x} \in T^n_{[X]\delta}\right) &= p\left(\sum_x \left|\frac{1}{n}N(x;\mathtt{x}) - p(x)\right| \le \delta\right) \\
&= 1 - p\left(\sum_x \left|\frac{1}{n}N(x;\mathtt{x}) - p(x)\right| > \delta\right) \\
&\ge 1 - p\left(\left|\frac{1}{n}N(x;\mathtt{x}) - p(x)\right| > \frac{\delta}{|\mathcal{X}|} \text{ for some } x\right) \\
&> 1 - \delta
\end{aligned}
$$

补证

$$
p\left(\left|\frac{1}{n}N(x;\mathtt{x}) - p(x)\right| > \frac{\delta}{|\mathcal{X}|} \text{ for some } x\right) < \delta
$$

定义指标随机变量 $B_k = 1 \text{ if } \mathtt{x}_k = x \text{ else } 0$, 可得 $N(x,\mathtt{x}) = \sum_{k=1}^n B_k(x)$, 同时 $B_k(x)$ 对于不同 k i.i.d, 因此 $E(B_k(x)) = p(x)$, 由弱大数定律可得

$$
p\left(\left|\frac{1}{n}\sum_{k=1}^n B_k(x) - p(x)\right| > \frac{\delta}{|\mathcal{X}|}\right) < \frac{\delta}{|\mathcal{X}|}
$$

从而

$$
\begin{aligned}
p\left(\left|\frac{1}{n}N(x;\mathtt{x}) - p(x)\right| > \frac{\delta}{|\mathcal{X}|} \text{ for some } x\right) &= p\left(\left|\frac{1}{n}\sum_{k=1}^n B_k(x) - p(x)\right| > \frac{\delta}{|\mathcal{X}|} \text{ for some } x\right) \\
&= p\left(\cup_x \left|\frac{1}{n}\sum_{k=1}^n B_k(x) - p(x)\right| > \frac{\delta}{|\mathcal{X}|}\right) \\
&\le \sum_x p\left(\left|\frac{1}{n}\sum_{k=1}^n B_k(x) - p(x)\right| > \frac{\delta}{|\mathcal{X}|}\right) \\
&< \delta
\end{aligned}
$$

Reference

1. https://zhuanlan.zhihu.com/p/149188816
2. https://www.cnblogs.com/hitgxz/p/12127539.html