

大数定律

渐近均分性定理

AEP 的推论：数据
压缩

高概率集与典型集

例题

第 5 讲 渐近均分性

大数定律

渐近均分性定理

AEP 的推论：数据压缩

高概率集与典型集

例题

大数定律

渐近均分性定理

AEP 的推论：数据
压缩

高概率集与典型集

例题

大数定律

渐近均分性定理

AEP 的推论：数据压缩

高概率集与典型集

例题

大数定律

渐近均分性定理

AEP 的推论：数据压缩

高概率集与典型集

例题

定义 1.1: 随机变量的收敛

设 $X_1, X_2, \dots$ 为一随机变量序列. 下面我们考虑序列 $\{X_n\}$ 收敛到 X 的三种不同方式:

1. 如果对任意的 $\epsilon > 0$, $P\{|X_n - X| > \epsilon\} \rightarrow 0$, 则称为依概率收敛.
2. 如果 $E(X_n - X)^2 \rightarrow 0$, 则称为均方收敛.
3. 如果 $P\{\lim_{n \rightarrow \infty} X_n = X\} = 1$, 则称为概率 1 (或称几乎处处) 收敛.

引理 1.2: 马尔可夫不等式

对任意非负随机变量 X 以及任意的 $t > 0$, 我们有

$$P\{X \geq t\} \leq \frac{EX}{t}. \quad (1.1)$$

证明.

如果 X 有分布 $F(x)$,

$$EX = \int_0^\infty x dF = \int_0^t x dF + \int_t^\infty x dF \geq \int_t^\infty x dF \geq \int_t^\infty t dF = tP\{X \geq t\}.$$

从而我们有 (1.1). □

引理 1.3: 切比雪夫不等式

设随机变量 Y 的均值和方差分别为 μ 和 σ^2 . 则有对任意 $\epsilon > 0$,

$$P\{|Y - \mu| > \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}.$$

证明.

在马尔可夫不等式中我们取 $X = (Y - \mu)^2$, 可得

$$\begin{aligned} P\{(Y - \mu)^2 > \epsilon^2\} &\leq P\{(Y - \mu)^2 \geq \epsilon^2\} \\ &= \frac{E(Y - \mu)^2}{\epsilon^2} \\ &= \frac{\sigma^2}{\epsilon^2}, \end{aligned}$$

又注意到 $P\{(Y - \mu)^2 > \epsilon^2\} = P\{|Y - \mu| > \epsilon\}$, 我们有

$$P\{|Y - \mu| > \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}.$$

□

定理 1.4: 弱大数定律

设 $Z_1, Z_2, \dots, Z_n$ 为 i.i.d. 随机变量序列, 其均值和方差分别为 μ 和 σ^2 , 令 $\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i$ 为样本均值. 证明:

$$P\{|\bar{Z}_n - \mu| > \epsilon\} \leq \frac{\sigma^2}{n\epsilon^2}.$$

因此, 当 $n \rightarrow \infty$ 时, $P\{|\bar{Z}_n - \mu| > \epsilon\} \rightarrow 0$. 这就是著名的弱大数定律.

证明.

在切比雪夫不等式中取 $Y = \bar{Z}_n$, 且注意到 $E\bar{Z}_n = \mu$ 以及 $\text{Var}(\bar{Z}_n) = \frac{\sigma^2}{n}$ ($\bar{Z}_n$ 是 n 个独立同分布的随机变量 $\frac{Z_i}{n}$, 每个的方差为 $\frac{\sigma^2}{n^2}$), 我们有

$$P\{|\bar{Z}_n - \mu| > \epsilon\} \leq \frac{\sigma^2}{n\epsilon^2}.$$

从而得证.



大数定律

渐近均分性定理

AEP 的推论：数据压缩

高概率集与典型集

例题

大数定律

渐近均分性定理

AEP 的推论：数据
压缩

高概率集与典型集

例题

定理 2.1: AEP

若 $X_1, X_2, \dots, X_n$ 为 i.i.d. $\sim p(x)$, 则

$$-\frac{1}{n} \log p(X_1, X_2, \dots, X_n) \rightarrow H(X) \text{ 依概率.}$$

证明.

独立随机变量的函数依然是独立随机变量. 因此, 由于 X_i 是 i.i.d., 从而 $\log p(X_i)$ 也是 i.i.d. 的. 从而由大数定律,

$$\begin{aligned} -\frac{1}{n} \log p(X_1, X_2, \dots, X_n) &= -\frac{1}{n} \sum_i \log p(X_i) \\ &\rightarrow -E \log p(X) \text{ 依概率} \\ &= H(X). \end{aligned}$$

这就证明了该定理.



定义 2.2

关于 $p(x)$ 的典型集 $A_\epsilon^{(n)}$ 是序列 $(x_1, x_2, \dots, x_n) \in \mathcal{X}^n$ 的集合, 且满足性质:

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, x_2, \dots, x_n) \leq 2^{-n(H(X)-\epsilon)}.$$

作为渐近均分性的一个推论，可以证明典型集 $A_\epsilon^{(n)}$ 有如下性质：

定理 2.3

1. 如果 $(x_1, x_2, \dots, x_n) \in A_\epsilon^{(n)}$ ，则
$$H(X) - \epsilon \leq -\frac{1}{n} \log p(x_1, x_2, \dots) \leq H(X) + \epsilon.$$
2. 当 n 充分大时， $P\{A_\epsilon^{(n)}\} > 1 - \epsilon.$
3. $|A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$ ，其中 $|A|$ 表示集合 A 中的元素个数.
4. 当 n 充分大时， $|A_\epsilon^{(n)}| \geq (1 - \epsilon)2^{n(H(X)-\epsilon)}.$

证明.

性质 (1) 的证明可直接由 $A_\epsilon^{(n)}$ 的定义, 利用对数函数的单调性得到. 第二个性质由定理 2.1 可以得到. 为证明性质 (3), 我们有

$$\begin{aligned}
 1 &= \sum_{x \in \mathcal{X}^n} p(x) \\
 &\geq \sum_{x \in A_\epsilon^{(n)}} p(x) \\
 &\geq \sum_{x \in A_\epsilon^{(n)}} 2^{-n(H(X)+\epsilon)} \\
 &= 2^{-n(H(X)+\epsilon)} |A_\epsilon^{(n)}|
 \end{aligned}$$

因此 $|A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$.



证明.

为证明 (4), 我们注意到对于充分大的 n , $P\{A_\epsilon^{(n)}\} > 1 - \epsilon$, 从而

$$\begin{aligned} 1 - \epsilon &< P\{A_\epsilon^{(n)}\} \\ &\leq \sum_{\mathbf{x} \in A_\epsilon^{(n)}} 2^{-n(H(X) - \epsilon)} \\ &= 2^{-n(H(X) - \epsilon)} |A_\epsilon^{(n)}|, \end{aligned}$$

其中第二个不等式由典型集的定义可得. 从而, $|A_\epsilon^{(n)}| \geq (1 - \epsilon)2^{n(H(X) - \epsilon)}$, 从而定理得证. $\square$

大数定律

渐近均分性定理

AEP 的推论：数据压缩

高概率集与典型集

例题

大数定律

渐近均分性定理

AEP 的推论：数据
压缩

高概率集与典型集

例题

设 $X_1, X_2, \dots, X_n$ 为 i.i.d $\sim p(x)$ 的随机变量. 我们将 $\mathcal{X}^n$ 中的所有序列划分成两个集合: 典型集 $A_\epsilon^{(n)}$ 及其补集. 我们现在要对 $\mathcal{X}^n$ 中的元素进行 01 编码, 我们希望能够有较短的平均码字长度.

将每个集合中的所有元素按某种顺序 (比如字典序) 排列. 然后给集合中的序列指定下标可以表示 $A_\epsilon^{(n)}$ 中的每个序列. 由于 $A_\epsilon^{(n)}$ 中序列个数 $\leq 2^{n(H+\epsilon)}$, 则表示它们需要的比特数不超过 $n(H+\epsilon) + 1$ (加 1 是因为 $n(H+\epsilon)$ 可能是非整数). 在所有这些序列前加上 0, 于是可知表示 $A_\epsilon^{(n)}$ 需要的比特数不超过 $n(H+\epsilon) + 2$. 类似地, 对不属于 $A_\epsilon^{(n)}$ 中的序列, 我们表示它们需要的比特数不超过 $n \log |\mathcal{X}| + 1$, 再在这些序列前加 1. 这样我们就获得了关于所有序列的一个编码方案.

注意，上述编码方案有如下特征：

- ▶ 编码是 1-1 的，且可以很方便地译码。起始位作为标识位，标明紧随码字的长度。
- ▶ 没有考虑非典型集的元素实际上少于 $\mathcal{X}^n$ 中元素的个数。但这足以产生一个有效的描述。（后面我们会看到，这是由于非典型集占的比例很小）
- ▶ 典型序列具有较短的描述长度。

下面用记号 x^n 表示序列 $x_1, x_2, \dots, x_n$. 设 $l(x^n)$ 表示相应于 x^n 的码字长度. 若 n 充分大, 使得 $P\{A_\epsilon^{(n)}\} \geq 1 - \epsilon$, 于是, 码字长度的数学期望为

$$\begin{aligned}
 E(l(X^n)) &= \sum_{x^n} p(x^n) l(x^n) \\
 &= \sum_{x^n \in A_\epsilon^{(n)}} p(x^n) l(x^n) + \sum_{x^n \in (A_\epsilon^{(n)})^c} p(x^n) l(x^n) \\
 &\leq \sum_{x^n \in A_\epsilon^{(n)}} p(x^n) (n(H + \epsilon) + 2) + \sum_{x^n \in (A_\epsilon^{(n)})^c} p(x^n) (n \log |\mathcal{X}| + 2) \\
 &= P\{A_\epsilon^{(n)}\} (n(H + \epsilon) + 2) + P(\{(A_\epsilon^{(n)})^c\}) (n \log |\mathcal{X}| + 2) \\
 &\leq n(H + \epsilon) + \epsilon n(\log |\mathcal{X}|) + 2\epsilon + 2 \\
 &= n(H + \epsilon').
 \end{aligned}$$

其中 $\epsilon' = \epsilon + \epsilon \log |\mathcal{X}| + \frac{2\epsilon + 2}{n}$, 适当选取 ϵ 和 n 时, ϵ' 可以任意小. 至此, 我们证明了如下的定理.

大数定律

渐近均分性定理

AEP 的推论: 数据压缩

高概率集与典型集

例题

定理 3.1

设 X^n 为服从 $p(x)$ 的 i.i.d. 序列, $\epsilon > 0$, 则存在一个编码将长度为 n 的序列 x^n 映射为比特串, 使得映射是 1-1 的 (因而可逆), 且对于充分大的 n , 有

$$E\left[\frac{1}{n}l(X^n)\right] \leq H(X) + \epsilon.$$

上述定理说明, 在平均意义下, 我们可以用大约 $nH(X)$ 比特来表示序列 X^n .

大数定律

渐近均分性定理

AEP 的推论：数据压缩

高概率集与典型集

例题

大数定律

渐近均分性定理

AEP 的推论：数据
压缩

高概率集与典型集

例题

定义 4.1

对每个 $n = 1, 2, \dots$, 设 $B_\delta^{(n)} \subset \mathcal{X}^n$ 为满足如下条件的元素个数最小的集合:

$$P\{B_\delta^{(n)}\} \geq 1 - \delta.$$

定理 4.2

设 $X_1, X_2, \dots, X_n$ 为服从 $p(x)$ 的 i.i.d. 序列. 对于 $\delta < \frac{1}{2}$ 以及任意的 $\delta' > 0$, 如果 $P\{B_\delta^{(n)}\} > 1 - \delta$, 则对于充分大的 n , 有

$$\frac{1}{n} \log |B_\delta^{(n)}| > H - \delta'.$$

证明.

设 $A_\varepsilon^{(n)}$ 为一个典型集, 则 $P(A_\varepsilon^{(n)}) > 1 - \epsilon$, 从而 $P(A_\varepsilon^{(n)} \cap B_\delta^{(n)}) \geq 1 - \varepsilon - \delta$.
我们有

$$\begin{aligned}
 1 - \varepsilon - \delta &\leq P(A_\varepsilon^{(n)} \cap B_\delta^{(n)}) \\
 &= \sum_{A_\varepsilon^{(n)} \cap B_\delta^{(n)}} p(x^n) \\
 &\leq \sum_{A_\varepsilon^{(n)} \cap B_\delta^{(n)}} 2^{-n(H-\varepsilon)} \\
 &= |A_\varepsilon^{(n)} \cap B_\delta^{(n)}| 2^{-n(H-\varepsilon)} \\
 &\leq |B_\delta^{(n)}| 2^{-n(H-\varepsilon)}.
 \end{aligned}$$

大数定律

渐近均分性定理

AEP 的推论: 数据
压缩

高概率集与典型集

例题

证明.

前式两边取对数, 我们有

$$\frac{1}{n} \log |B_\delta^{(n)}| \geq (H - \varepsilon) + \frac{1}{n} \log(1 - \varepsilon - \delta).$$

取 ε 充分小, 以及 n 充分大, 使得 $\varepsilon - \frac{1}{n} \log(1 - \varepsilon - \delta) < \delta'$, 从而得证. □

下面引进一个新记号以表示在一阶指数意义下的相等概念.

定义 4.3: 记

$a_n \doteq b_n$ 表示

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0.$$

下面引进一个新记号以表示在一阶指数意义下的相等概念.

定义 4.4: 记

$a_n \doteq b_n$ 表示

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0.$$

于是, 我们有: 如果 $\delta_n \rightarrow 0$ 且 $\epsilon_n \rightarrow 0$, 则

$$|B_{\delta_n}^{(n)}| \doteq |A_{\epsilon_n}^{(n)}| \doteq 2^{nH}.$$

例 4.5

为说明 $A_\epsilon^{(n)}$ 与 $B_\delta^{(n)}$ 的区别，考虑一个 Bernoulli 序列 $X_1, X_2, \dots, X_n$ ，其参数 $p = 0.9$ 。此时，典型序列元素中 1 所占的比例近似等于 0.9。然而，这并不包括全部都是 1 的序列，虽然其出现的概率最大。集合 $B_\delta^{(n)}$ 包含所有很可能出现的序列，因而包括全部为 1 的序列。

例 5.1: 课本习题 2.39

设 X, Y 和 Z 为三个服从 $\text{Bernoulli}(\frac{1}{2})$ 的二元随机变量, 且两两独立. 在上述约束条件下, $H(X, Y, Z)$ 的最小值是多少?

证明.
我们有

$$\begin{aligned} H(X, Y, Z) &= H(X, Y) + H(Z|X, Y) \\ &\geq H(X, Y) \\ &= 2 \text{ 比特}. \end{aligned}$$

下面我们验证这个下界可以达到, 从而 $H(X, Y, Z)$ 的最小值为 2 比特.
设 $X, Y \sim \text{Bernoulli}(1/2)$, $Z = X \oplus Y$, 其中 $\oplus$ 表示模 2 加法.

□