

Lecture Notes for Probability Theory - Class 7

Yuwei Wu

1 Theorem (MacMillan Theorem)

设集合 $S := \{x_1, x_2, \dots, x_r\}$, 在 S^n 上的有限样本空间 $\Omega_n = \{\omega = (\omega_1, \omega_2, \dots, \omega_n)\}$, 其中 ω_k 都是互相独立且服从相同分布 P_S 从 S 中取值的。令 $P_i = P_S(x_i)$, 记关于分布 P_S 的熵 $H := -\sum_{i=1}^r P_i \log(P_i)$ 。

在概率空间 (Ω_n, P) 中, 对于任意 $\varepsilon > 0$, 在 n 足够大时, 总能找到 $\Omega_n' \subseteq \Omega_n$, 使得:

1. $\exp(n(H - \varepsilon)) \leq |\Omega_n'| \leq \exp(n(H + \varepsilon))$
2. $\lim_{n \rightarrow \infty} P(\Omega_n') = 1$
3. 对于每一个 $\omega \in \Omega_n'$, 有 $\exp(-n(H + \varepsilon)) \leq P(\omega) \leq \exp(-n(H - \varepsilon))$

Proof

首先, 由结论 (2)(3) 可推得 (1):

由于 (2) $\lim_{n \rightarrow \infty} P(\Omega_n') = 1$, 故对任意 $\varepsilon_0 > 0$ 总存在 $N_0(\varepsilon_0)$ 使得 $N > N_0$ 时, $|P(\Omega_N') - 1| < \varepsilon_0$, 即

$$1 - \varepsilon_0 < P(\Omega_N') < 1 + \varepsilon_0$$

又由于 $P(\Omega_N') = \sum_{\omega \in \Omega_N'} P(\omega)$, 故 $|\Omega_N'| \cdot P(\omega)_{\min} \leq P(\Omega_N') \leq |\Omega_N'| \cdot P(\omega)_{\max}$, 即

$$\frac{P(\Omega_N')}{P(\omega)_{\max}} \leq |\Omega_N'| \leq \frac{P(\Omega_N')}{P(\omega)_{\min}}$$

再由 (3) 对每一个 $\omega \in \Omega_N'$, 有 $\exp(-n(H + \varepsilon)) \leq P(\omega) \leq \exp(-n(H - \varepsilon))$, 得

$$P(\omega)_{\min} = \exp(-n(H + \varepsilon)), P(\omega)_{\max} = \exp(-n(H - \varepsilon))$$

故对任意 $\varepsilon_0 > 0$, 总存在 $N_0(\varepsilon_0)$ 使得 $N > N_0(\varepsilon_0)$ 时, 有 $P(\Omega_N') \cdot \exp(N(H - \varepsilon)) \leq |\Omega_N'| \leq P(\Omega_N') \cdot \exp(N(H + \varepsilon))$, 即对任意 $\varepsilon_0 > 0$, 总存在 $N_0(\varepsilon_0)$ 使得 $N > N_0$ 时, 成立

$$(1 - \varepsilon_0) \cdot \exp(N(H - \varepsilon)) \leq |\Omega_N'| \leq (1 + \varepsilon_0) \cdot \exp(N(H + \varepsilon))$$

由于 ε_0 的任意性, 易得 n 充分大时, 成立 (1)

$$\exp(n(H - \varepsilon)) \leq |\Omega_n'| \leq \exp(n(H + \varepsilon))$$

于是只须证明 (2)(3)。对于 (2):

对于任意 $\varepsilon > 0$, 有 $\delta = \frac{\varepsilon}{\sum_{j=1}^r \log P_j}$, 构造 $\Omega_n' = \{\omega \in \Omega_n : |\frac{|V_j(\omega)|}{n} - P_j| \leq \delta, 1 \leq j \leq r\}$, 其中

$$V_j(\omega) = \{i \in [n] : \omega_i = x_j\},$$

由弱大数定律得:

$$\forall 1 \leq j \leq r, \lim_{n \rightarrow \infty} P(|\frac{|V_j(\omega)|}{n} - P_j| > \delta) = 0,$$

于是 $\lim_{n \rightarrow \infty} P(\mathbb{C}_{\Omega_n} \Omega_n^i) = 0$, 即 $\lim_{n \rightarrow \infty} P(\Omega_n^i) = 1$, (2) 得证。

下证上述构造的 Ω_n^i 对 (3) 也成立:

对于每一个 $\omega = (\omega_1, \omega_2, \dots, \omega_n) \subseteq \Omega_n^i$, 有

$$\begin{aligned} P(\omega) &= P(\omega_1) \cdot P(\omega_2) \dots P(\omega_n) \\ &= P_1^{|V_1(\omega)|} \cdot P_2^{|V_2(\omega)|} \dots P_r^{|V_r(\omega)|} \\ &= \exp(\sum_{j=1}^r V_j(\omega) \cdot \log P_j) \\ &= \exp(n \cdot \sum_{j=1}^r \frac{V_j(\omega)}{n} \cdot \log P_j) \end{aligned}$$

由于 $\omega \in \Omega_n^i$, 故有 $P_j - \delta \leq \frac{V_j(\omega)}{n} \leq P_j + \delta$, 代入上式得到:

$$\exp(n \cdot \sum_{j=1}^r (P_j - \delta) \cdot \log P_j) \leq P(\omega) \leq \exp(n \cdot \sum_{j=1}^r (P_j + \delta) \cdot \log P_j),$$

即

$$\exp(n \cdot \sum_{j=1}^r P_j \cdot \log P_j) \cdot \exp(-n \sum_{j=1}^r \delta \cdot \log P_j) \leq P(\omega) \leq \exp(n \cdot \sum_{j=1}^r P_j \cdot \log P_j) \cdot \exp(n \sum_{j=1}^r \delta \cdot \log P_j),$$

故

$$\exp(-n(H + \varepsilon)) \leq P(\omega) \leq \exp(-n(H - \varepsilon)).$$

(3) 得证。

2 Problem (Discrete Memoryless Source(DMS))

对于各类分布 P 的离散无记忆信源, (S, P) 中 S 表示字母表, P_i 表示每个字母出现的概率, S^k 表示长为 k 的字符串。

对于字符串的压缩与解压用过程 $(\text{code})(f, \varphi)$ 表示:

$$S^k \xleftrightarrow[\varphi]{f} \{0,1\}^n$$

定义 $\text{error}(f, \varphi) := P_k(\varphi \circ f(\omega) \neq \omega), \omega \in \Omega_k$ (Ω_k 定义同上文)。

试想: 如果 $|S|^k < 2^n$, 则必存在一一映射使得压缩解压过程不会出错。为了进一步压缩, n 应尽可能小, 即允许出错。

故目标为最小化 $\frac{n}{k}$ 及 $\text{error}(f, \varphi)$ 。

Solution

对于给定 $\varepsilon > 0$, 令 $n(k, \varepsilon)$ 表示满足 $e_{\text{error}}(f, \varphi) \leq \varepsilon$ 最小的 n , 下面证明:

$$\lim_{k \rightarrow \infty} \frac{n(k, \varepsilon)}{k} = - \sum_i P_i \log P_i = H$$

于是, 存在 (f, φ) , 使得 $e_{\text{error}}(f, \varphi) \leq \varepsilon$, 等价于存在 $A \subseteq \Omega_k$ 使得 $P(A) \geq (1 - \varepsilon)$, 且 $|A| \leq 2^n$ (代表 A 内映射不会出错).

设 $S(k, \varepsilon)$ 表示满足上述等价条件的集合 A 的最小基数, 由于 $|A| \leq 2^n$, 有:

$$\left\lceil \log S(k, \varepsilon) \right\rceil = n(k, \varepsilon),$$

$$\text{即 } \lim_{k \rightarrow \infty} \frac{n(k, \varepsilon)}{k} = \lim_{k \rightarrow \infty} \frac{\log S(k, \varepsilon)}{k}.$$

对于任意 $\delta > 0$, 令 $B(k, \delta)$ 表示满足条件 $\omega \in \Omega_k$ 且 $\exp(-k(H + \delta)) \leq P(\omega) \leq \exp(-k(H - \delta))$ 的 ω 的集合。(*)

于是由 MacMillian Theorem 可知:

$$B(k, \delta) \supseteq \Omega_k'$$

即对 $\varepsilon = \delta$, 总能找到 Ω_k' , 使 Ω_k' 里的每个元素都满足 (*), 从而其每个元素也一定在 $B(k, \delta)$ 中。

于是有,

$$\lim_{k \rightarrow \infty} P(B(k, \delta)) \geq \lim_{k \rightarrow \infty} P(\Omega_k') = 1(**)$$

由 (*) 及 (**), 类似 MacMillian Theorem 的 (2)(3) \Rightarrow (1) 的推导, 有:

$$\exp(k(H - \delta)) \leq |B(k, \delta)| \leq \exp(k(H + \delta)).$$

由 $S(k, \varepsilon)$ 最小的性质可知, $S(k, \varepsilon) \leq |B(k, \delta)|$, 于是,

$$\begin{aligned} \overline{\lim}_{k \rightarrow \infty} \frac{1}{k} \cdot \log S(k, \varepsilon) &\leq \overline{\lim}_{k \rightarrow \infty} \frac{1}{k} \cdot \log |B(k, \delta)| \\ &\leq H + \delta \end{aligned}$$

又, 对每一个 $A \subseteq \Omega_k$, 满足 $P(A) \geq 1 - \varepsilon$, 有:

$$\lim_{k \rightarrow \infty} P(A \cap B(k, \delta)) \geq \frac{1 - \varepsilon}{2}$$

于是,

$$\begin{aligned} |A| &\geq |A \cap B(k, \delta)| \\ &\geq P(A \cap B(k, \delta)) \cdot |B(k, \delta)| \\ &\geq \frac{1 - \varepsilon}{2} \cdot \exp(k(H - \delta)) \end{aligned}$$

故,

$$\begin{aligned}\lim_{k \rightarrow \infty} \frac{1}{k} \cdot \log S(k, \varepsilon) &\geq \lim_{k \rightarrow \infty} \frac{1}{k} \cdot \log(\exp(k(H - \delta)) \cdot \frac{1 - \varepsilon}{2}) \\ &\geq H - \delta\end{aligned}$$

由于 δ 的任意性, 有

$$\lim_{k \rightarrow \infty} \frac{1}{k} \cdot \log S(k, \varepsilon) = H$$

3 Problem (DMS 问题推广)

建立映射 $S \rightarrow R^+$, 即给每一个字母赋上权值, 有

$$M(\omega) = M_1(\omega_1) \cdot M_2(\omega_2) \dots M_k(\omega_k), \omega \in \Omega_k$$

定义 $S(k, \varepsilon) := \min M(A) = \min_{\omega \in A} M(\omega), A \subseteq \Omega_k, P_k(A) \geq (1 - \varepsilon)$

类似地有

$$\lim_{k \rightarrow \infty} \left(\frac{\log(S(k, \varepsilon))}{k} - E_k \right) = 0$$

其中 $E_k := \frac{1}{k} \cdot \sum_{i=1}^k \sum_{x \in S} P_i(x) \cdot \log\left(\frac{M_i(x)}{P_i(x)}\right)$

4 Problem (统计问题)

对于概率分布 $P = \{P(x) : x \in X\}$ 及 $Q = \{Q(x) : x \in X\}$, 在样本空间 X 中抽 k 次, 有 $\omega = (\omega_1, \omega_2, \dots, \omega_k)$, 根据已知 k 次事件构成的序列猜测样本空间概率分布是 P 还是 Q 。

要求最小化在实际分布为 Q 时猜错的可能性 (保证在实际分布为 P 时猜错的可能性小于 ε)。即求

$$\beta(k, \varepsilon) = \min Q_k(A), A \subseteq \Omega_k, P_k(A) \geq (1 - \varepsilon)$$

类似上述证明, 有

$$\lim_{k \rightarrow \infty} \frac{1}{k} \cdot \log \beta(k, \varepsilon) = - \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)}$$

5 Problem (Homework)

给定集合 K 到 R 熵函数 $h \in R^{2^{[K]}}$, 对于随机变量 X_1, X_2, \dots, X_k , 使得 $h(A) = H(\{x_i\}_{i \in A})$ 。

证明 h 满足次模函数性质:

$$\begin{cases} h(\emptyset) = 0 \\ h(A) \leq h(B), A \subseteq B \\ h(A \cup B) + h(A \cap B) \leq h(A) + h(B) \end{cases}$$