Lecture Notes for Probability Theory - Class 7

Yuwei Wu

1 Theorem (MacMillan Theorem)

设集合 $S := \{x_1, x_2, ..., x_r\}$, 在 S^n 上的有限样本空间 $\Omega_n = \{\omega = (\omega_1, \omega_2, ..., \omega_n)\}$, 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. $\lim_{n\to\infty} P(\Omega_n^{'}) = 1$
3. 对于每一个 $\omega \subseteq \Omega_n^{'}$, 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) $\lim_{n\to\infty} P(\Omega_n^{'}) = 1$, 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 $\omega \subseteq \Omega_N^{'}$, 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 $\delta = \frac{\varepsilon}{\sum_{j=1}^{r} logP_j}$, 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:

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

于是 $\lim\limits_{n\to\infty} P(\complement_{\Omega_n}\Omega_n^{'}) = 0$, 即 $\lim\limits_{n\to\infty} P(\Omega_n^{'}) = 1$,(2) 得证。
下证上述构造的对于 $\Omega_n^{'}$ 对 (3) 也成立:
对于每一个 $\omega = (\omega_1, \omega_2, ..., \omega_n) \subseteq \Omega_n^{'}$, 有

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

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

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

即

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

故

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

(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 (code)(f, φ) 表示:

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

定义 $e_{rror}(f, \varphi) := P_k(\varphi \circ f(\omega) \neq \omega), \omega \in \Omega_k(\Omega_k$ 定义同上文).
试想: 如果 $|\mathrm{S}|^k < 2^n$, 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 $e_{rror}(f, \varphi)$.

Solution

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

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

于是, 存在 (f,φ), 使得 $e_{rror}(f,\varphi) \leq \varepsilon$, 等价于存在 $A \subseteq \Omega_k$ 使得 $P(A) \geqslant (1-\varepsilon)$, 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:

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

即 $\lim\limits_{k\to\infty} \frac{n(k,\varepsilon)}{k} = \lim\limits_{k\to\infty} \frac{\log S(k,\varepsilon)}{k}$.

对于任意 $\delta > 0$, 令 $B(k,\varepsilon)$ 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:

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

于是,

$$\begin{cases} \lim\limits_{k\to\infty} P(B(k,\delta)) \geqslant \lim\limits_{k\to\infty} P(\Omega_k{}') = 1 \\ |B(k,\delta)| \leq exp(k(H+\delta)) \end{cases}$$

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

$$\begin{aligned} \overline{\lim_{k\to\infty}} \frac{1}{k} \cdot \mathrm{logS}(k,\varepsilon) &\leqslant \overline{\lim_{k\to\infty}} \frac{1}{k} \cdot \log|B(k,\delta)| \\ &\leqslant H+\delta \end{aligned}$$

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

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

于是,

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

故,

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

由于 δ 的任意性, 有

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

3 Problem (DMS 问题推广)

建立映射 $S\to R^{+}$, 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 $S(k,\varepsilon):=\min M(A)=min\sum_{\omega\in A}M(\omega),A\subseteq\Omega_k,P_k(A)\geqslant(1-\varepsilon)$

类似地有

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

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

4 Problem (统计问题)

对于概率分布 $P=\{P(x):x\in X\}$ 及 $Q=\{Q(x):x\in X\}$, 样本空间中抽 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

$$\beta(k,\varepsilon)=minQ_k(A),A\subseteq\Omega_k,P_k(A)\geqslant(1-\varepsilon)$$

类似上述证明, 有

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

5 Problem (Homework)

给定集合 K 到 R 熵函数 $h\in R^{2^{[K]}}$, 对于随机变量 $X_1,X_2,...,X_k$, 使得 $h(A)=H(\{x_i\}_{i\in A})$。证明 h 满足次模函数性质:

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