数 学 年 刊 A 辑
2025, 46(4):367–382
DOI: 10.16205/j.cnki.cama.2025.0024

最小熵耦合问题近似解的误差估计*

郝一菲[1]　王颖喆[2]

提要　针对基于 Rényi 熵 H_α 的最小熵耦合问题, 通过引入 profile 方法, 得到了贪心算法近似解与最优解之间熵函数差值的一个上界 $h(\alpha)$. 在两个概率分布耦合的情况下, $h(\alpha)=\frac{1}{\alpha-1}\log_2\left(1+\left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}}-\left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha-1}}\right)$; 在任意有限多个概率分布耦合的情况下, $h(\alpha)=\frac{1}{1-\alpha}\log_2\left(\frac{1}{2}+\frac{1}{\alpha 2^{\alpha}}\right)$. 并将相关结果推广至 Tsallis 熵. 创新之处在于, 给出了误差估计与参数 α 的函数关系, 并且在 $\alpha>0.365$ 时获得了更加精确的结果.

关键词　Rényi 熵, Tsallis 熵, 耦合, 贪心算法
MR (2020) 主题分类　94A17
中图法分类　O211.9, O236, O29
文献标志码　A
文章编号　1000-8314(2025)04-0367-16

§1　引　言

最小熵耦合问题 (Minimum entropy coupling problem, 下文简称 MEC 问题) 是指给定 m 个状态空间有限的离散概率分布, 如何构造熵最小的联合分布. 耦合的熵与边缘分布的交叉信息有密切联系, 寻找最小熵耦合就是寻找最大交叉信息的耦合, 该观点在机器学习领域影响深远 [1]. 全体耦合组成的集合在最优运输问题中被称为运输多面体, Cicalese 等认为寻找最小熵耦合就是寻找运输多面体中的最小熵元素 [2]. Kocaoglu 等在推断因果方向问题中提出一个关键假设: 一对随机变量间正确的因果方向通常具有最小的外部噪声熵, 寻找外部噪声熵即为在一族概率分布中寻找最小熵耦合 [3]. Compton 等进一步指出如果正确的因果方向具有低熵外部噪声, 那么反方向大概率具有高熵的外部噪声 [4]. de Witt 等将 MEC 问题应用于信息隐写, 指出最小熵耦合与最高效的加密程序相对应, 并通过实验展现了基于最小熵耦合的信息隐写的非凡表现 [5]. 以上实例表明, MEC 问题在因果推断、泛函表示、数据降维和通信等领域有着十分广泛的应用, 具有较高的研究价值.

在诸多应用中, 问题的关键在于如何高效准确求解 MEC 问题. Kovačević等指出 MEC 问题是一个 NP-Hard 问题 [6], 直接计算的时间成本高昂, 因此需通过优化算法寻找近似解. Kocaoglu 等设计了一种贪心算法用于求解最小 Shannon 熵的近似解 [7]. Compton 等

本文 2025 年 4 月 16 日收到, 2025 年 10 月 15 日收到修改稿.
[1]北京师范大学数学科学学院, 数学与复杂系统教育部重点实验室, 北京 100875.
E-mail: haoyf@mail.bnu.edu.cn
[2]通信作者. 北京师范大学数学科学学院, 数学与复杂系统教育部重点实验室, 北京 100875.
E-mail: wangyz@bnu.edu.cn
*本文受到国家重点研发计划项目 (No. 2020YFA0712900) 和国家自然科学基金 (No. 11871103) 的资助.

通过引入 profile 方法给出了该近似解与最优解之间的 Shannon 熵差值的一个上界 (1.22 比特)[8].

当前研究成果主要集中在 Shannon 熵, 然而其他熵函数对于信息和系统的刻画更具一般性与灵活性. 朱亚倩将 Shannon 熵的 MEC 问题推广到 Rényi 熵和 Tsallis 熵, 在两个概率分布耦合的情况下, 首先指出联合密度矩阵为上三角矩阵时, 对应的熵最小. 然后得到了贪心算法的近似解与最优解之间熵函数差值的一个上界 (1 比特)[9]. 孙清扬等提出一种新的基于供求匹配的贪心算法, 证明了该算法可以控制在所有系统中求解最小联合 Tsallis 熵的近似误差 [10].

针对现有结论普遍存在的精确程度受限和适用范围狭窄等问题, 本文基于 Rényi 熵和 Tsallis 熵的 MEC 问题, 分别探讨了两个离散型概率分布耦合和任意有限个离散型概率分布耦合的情况, 得到了贪心算法的近似解与最优解熵函数差值的一个上界 $h(\alpha)$. 然后, 利用 Rényi 熵与 Tsallis 熵数学上的等价关系, 将相关结论推广至 Tsallis 熵.

§2 预备知识

§2.1 概率分布的 Rényi 熵和 Tsallis 熵

如果 $\forall i=1,2,\cdots,n$ 有 $p(i)\geqslant 0$ 且 $p(1)+p(2)+\cdots+p(n)=1$ 成立, 则称 $p=(p(1),p(2),\cdots,p(n))$ 为一个概率分布, 其中 $i=1,2,\cdots,n$ 称为状态, $p(i)$ 为状态 i 对应的概率. 不失一般性, 本文始终假设 $p(1)\geqslant p(2)\geqslant\cdots\geqslant p(n)$. 为了便于研究贪心算法运行细节和迭代规律 (详见第 2.3 节), 我们允许概率分布为次随机的, 即总概率 $m(p):=p(1)+p(2)+\cdots+p(n)<1$, 并将状态的数量 n 记作 $|p|$.

概率分布 $p=(p(1),p(2),\cdots,p(n))$ 的 Rényi 熵定义如下:

$$H_\alpha(p):=\frac{1}{1-\alpha}\log_2\Big(\sum_{i=1}^{n}p(i)^\alpha\Big),$$

其中 α 为参数. 如果记

$$F_{\rm cost}^\alpha(p):=\sum_{i=1}^{n}f_{\rm cost}^\alpha(p(i)):=\sum_{i=1}^{n}p(i)^\alpha,\quad \alpha\neq 1,$$

称为 p 的一般熵, 那么

$$H_\alpha(p)=\frac{1}{1-\alpha}\log_2 F_{\rm cost}^\alpha(p).$$

Tsallis 熵定义如下:

$$T_\alpha(p):=\frac{1}{1-\alpha}\Big(\sum_{i=1}^{n}p(i)^\alpha-1\Big),$$

其中 α 为参数.

Rényi 熵和 Tsallis 熵具有如下的转换关系:

$$H_\alpha(p)=\frac{1}{1-\alpha}\log_2[1+(1-\alpha)T_\alpha(p)].$$

注意到 Shannon 熵是当参数 $\alpha \to 1$ 时 Rényi 熵的特殊情况, 即

$$\lim_{\alpha\to 1} H_\alpha(p) = \lim_{\alpha\to 1} \frac{1}{1-\alpha} \log_2\Big(\sum_{i=1}^n p_i^\alpha\Big) = -\sum_{i=1}^n p_i\ \log_2 p_i =: H_1(p),$$

所以 Rényi 熵和 Tsallis 熵均为 Shannon 熵的广义形式, 它们分别带有一个参数, 从而对信息的度量更具一般性和灵活性.

§2.2　概率分布的耦合

设 $p_1, p_2, \cdots, p_s$ 为 s 个概率分布, 它们具有相同的总概率, S 为这 s 个概率分布组成的集合, 记 $m(S) := m(p_1) = \cdots = m(p_s)$. 不妨假定所有概率分布的状态数量相同, 即 $\forall 1 \leqslant k \leqslant s$, 令 $|p_k| = \max\limits_{1\leqslant i\leqslant s} |p_i| = n$, 这是因为对于状态数量小于 n 的概率分布, 可以以概率 0 补齐. 我们把任意以 $p_1, p_2, \cdots, p_s$ 为边缘分布的联合分布 $\mathcal{C}$ 称为 $p_1, p_2, \cdots, p_s$ 的耦合, 即 $\forall 1 \leqslant k \leqslant s, 1 \leqslant i_k \leqslant n$, $\mathcal{C}(\cdot, \cdots, \cdot, i_k, \cdot, \cdots, \cdot) = p_k(i_k)$.

记 $C(p_1, \cdots, p_s)$ 为 $p_1, p_2, \cdots, p_s$ 耦合全体, MEC 问题就是寻找熵最小的耦合, 该耦合称为 MEC 问题的最优耦合或最优解, 记作 OPT_S, 即

$$H(\mathrm{OPT}_S) = \inf_{\mathcal{C}\in C(p_1,\cdots,p_s)} H(\mathcal{C}),$$

其中 H 表示某种熵函数, 如本文研究的 Rényi 熵 H_α 与 Tsallis 熵 T_α.

§2.3　求解最小熵耦合的贪心算法

M 表示以若干概率分布为列向量组成的矩阵, $\mathcal{G}_S$ 表示贪心算法生成的概率分布, 即 MEC 问题的近似耦合或近似解. 贪心算法的伪代码如下表所示. 容易验证 $\mathcal{G}_S$ 同样是不增的且 $|\mathcal{G}_S| \leqslant ns - (s-1)$.

求解最小熵耦合的贪心算法 [7]

Input: s 个状态空间为 $I = \{1, 2, \cdots, n\}$ 的概率分布, 生成矩阵 $M = [p_1^{\mathrm{T}}, p_2^{\mathrm{T}}, \cdots, p_s^{\mathrm{T}}]$.
$\mathcal{G}_S = [\ ]$.
将矩阵 M 的每一列重新排序成为不增序列, 即 $p_i(1) \geqslant p_i(2) \geqslant \cdots \geqslant p_i(n)$.
寻找每一列最大值中的最小值, 记作 r, 即 $r \leftarrow \min\limits_i (p_i(1))$.
while $r > 0$, do
　　将 r 添加至 $\mathcal{G}_S$ 中.
　　更新每一列最大值的状态: $p_i(1) \leftarrow p_i(1) - r,\ \forall i$.
　　将矩阵 M 的每一列重新排序成为不增序列.
　　$r \leftarrow \min\limits_i (p_i(1))$.
end while
return $\mathcal{G}_S$.

Compton 等给出了该贪心算法的近似解精确程度的估计, 在两个概率分布耦合的情况下, 该近似解与最优解之间的 Shannon 熵差值不高于 0.53 比特, 在任意有限多个概率分布耦合的情况下, 该近似解与最优解 Shannon 熵差值不高于 1.22 比特. 朱亚倩指出在两个概率分布耦合的情况下, 该近似解与最优解之间 Rényi 熵与 Tsallis 熵差值均不高于 1 比特.

本文将基于 profile 方法, 给出在两个概率分布耦合和任意有限多个概率分布耦合的情况下, 该近似解与最优解之间的 Rényi 熵和 Tsallis 熵差值的上界估计. 在某些参数范围内, 本文的结果更加精确, 且当参数 $\alpha \to 1$ 时, 本文的结果与 Compton 的结果保持统一.

§2.4 Profile 方法

首先分析成为某概率分布的耦合需满足的必要条件. 以 $p=(0.5,0.4,0.1)$ 为例, 如果 $\mathcal{C}$ 是以 p 为边缘分布的某个耦合, 那么 $\mathcal{C}$ 至少应满足以下 3 个条件:

(1) 至少有概率和为 0.1 的状态, 每个状态的概率均小于等于 0.1, 即

$$\sum_{j=1}^{|\mathcal{C}|}\mathcal{C}(j)\mathbb{1}_{\{j:\mathcal{C}(j)\leqslant 0.1\}}(j)\geqslant 0.1.$$

(2) 至少有概率和为 0.4 的状态, 每个状态的概率均小于等于 0.4, 即

$$\sum_{j=1}^{|\mathcal{C}|}\mathcal{C}(j)\mathbb{1}_{\{j:\mathcal{C}(j)\leqslant 0.4\}}(j)\geqslant 0.4.$$

(3) 至少有概率和为 0.5 的状态, 每个状态的概率均小于等于 0.5, 即

$$\sum_{j=1}^{|\mathcal{C}|}\mathcal{C}(j)\mathbb{1}_{\{j:\mathcal{C}(j)\leqslant 0.5\}}(j)\geqslant 0.5.$$

其中 $\mathbb{1}$ 表示示性函数. 例如, $\mathcal{C}=(0.5,0.2,0.2,0.1)$ 可能成为以 p 为边缘分布的耦合, 但 $\mathcal{C}'=(0.7,0.1,0.1,0.1)$ 不可能是 p 的耦合, 因为 $\sum\limits_{j=1}^{4}\mathcal{C}(j)\mathbb{1}_{\{j:\mathcal{C}(j)\leqslant 0.4\}}(j)=0.1+0.1+0.1<0.4$.

我们将上述分析推广至一般情形.

定义 2.1 对于概率分布 $p=(p(1),p(2),\cdots,p(n))$, 引入 p 的外形 (Sketch), 定义如下:

$$S_p(x):=p(i),\quad 当\ x\in\Big(\sum_{j>i}p(j),\sum_{j\geqslant i}p(j)\Big]时.$$

注意

$$F^{\alpha}_{\rm cost}(p)=\int_0^{m(p)}f^{\alpha}_{\rm unit}(S_p(x)){\rm d}x,\tag{2.1}$$

其中 $f^{\alpha}_{\rm unit}(x):=x^{\alpha-1},\ \alpha\neq 1$.

给定一族概率分布 $S=\{p_1,p_2,\cdots,p_s\}$, 如果 $\mathcal{C}$ 是以 $p_1,p_2,\cdots,p_s$ 为边缘分布的耦合, 自然应满足 $\forall 1\leqslant i\leqslant s,\ S_{\mathcal{C}}(x)\leqslant S_{p_i}(x)$. 由此引入概率分布族 S 的轮廓 (profile), 定义如下.

定义 2.2　$P_S(x) := \min\limits_{p\in S} S_p(x)$.

基于上文分析, 我们有

$$S_{\mathcal{C}}(x) \leqslant P_S(x).$$

(2.1) 通过选取合适的函数 $f_{\rm unit}^{\alpha}(x) = x^{\alpha-1}$, 使得积分值恰好等于 $F_{\rm cost}^{\alpha}(p)$. 尽管 $P_S(x)$ 可能不是某个概率分布的外形, 但仍可仿照 (2.1) 进行如下定义.

定义 2.3　$F_{\rm cost}^{\alpha}(P_S) := \int_0^{m(S)} f_{\rm unit}^{\alpha}(P_S(x))\mathrm{d}x = \int_0^{m(S)} (P_S(x))^{\alpha-1}\mathrm{d}x$.

参考 $F_{\rm cost}^{\alpha}(p)$ 与 Rényi 熵 $H_\alpha(p)$ 的关系, 定义轮廓的 Rényi 熵如下.

定义 2.4　$H_\alpha(P_S) := \frac{1}{1-\alpha}\log_2 F_{\rm cost}^{\alpha}(P_S)$.

§3　主要结论

论文的核心是研究贪心算法的近似解的 Rényi 熵 $H_\alpha(\mathcal{G}_S)$ 与 MEC 问题最优解的 Rényi 熵 $H_\alpha(\mathrm{OPT}_S)$ 之间的关系. 更具体地说, 我们期望得到估计式

$$H_\alpha(\mathcal{G}_S) \leqslant H_\alpha(\mathrm{OPT}_S) + ?$$

中尽量精确的 “?”. 事实上, 直接研究 $H_\alpha(\mathcal{G}_S)$ 与 $H_\alpha(\mathrm{OPT}_S)$ 的关系是困难的, $H_\alpha(P_S)$ 在其中担负着一种中介的作用, 具体来说, 如果搞清楚了 $H_\alpha(\mathcal{G}_S)$ 与 $H_\alpha(P_S)$ 的关系, $H_\alpha(\mathrm{OPT}_S)$ 与 $H_\alpha(P_S)$ 的关系, 那么利用不等式的传递性, $H_\alpha(\mathcal{G}_S)$ 与 $H_\alpha(\mathrm{OPT}_S)$ 的关系也就清楚了. 定理 3.1 首先给出了 $H_\alpha(\mathrm{OPT}_S)$ 与 $H_\alpha(P_S)$ 的关系.

定理 3.1　$H_\alpha(\mathrm{OPT}_S) \geqslant H_\alpha(P_S)$.

结合定理 3.1 的结论, 下文只需研究 $H_\alpha(\mathcal{G}_S)$ 与 $H_\alpha(P_S)$ 的关系. 首先考虑 $s=2$, 即两个概率分布耦合的情况, 结合定理 3.1, 我们有下面的定理.

定理 3.2　当 $s=2$ 时,

$$H_\alpha(\mathcal{G}_S) \leqslant H_\alpha(\mathrm{OPT}_S) + \frac{1}{\alpha-1}\log_2\left(1+\left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} - \left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha-1}}\right).$$

然后考虑任意有限多个概率分布耦合的情况, 结合定理 3.1 得到下面的定理.

定理 3.3　$\forall s\in\mathbb{N}^*$,

$$H_\alpha(\mathcal{G}_S) \leqslant H_\alpha(\mathrm{OPT}_S) + \frac{1}{1-\alpha}\log_2\left(\frac{1}{2}+\frac{1}{\alpha 2^\alpha}\right).$$

定理 3.2 与定理 3.3 的误差估计关于参数 α 的函数关系图象如图 1 所示.

图 1 显示, 误差随参数 α 的增大而减小, 且当 $\alpha > 0.365$ 时, 本文获得了比文 [9] 更强的估计. 值得注意的是

$$\lim_{\alpha\to 1}\frac{1}{\alpha-1}\log_2\left(1+\left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} - \left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha-1}}\right) \approx 0.53,$$

$$\lim_{\alpha\to 1}\frac{1}{1-\alpha}\log_2\Big(\frac{1}{2}+\frac{1}{\alpha 2^{\alpha}}\Big)\approx 1.22,$$

这与文 [8] 的结果相统一.

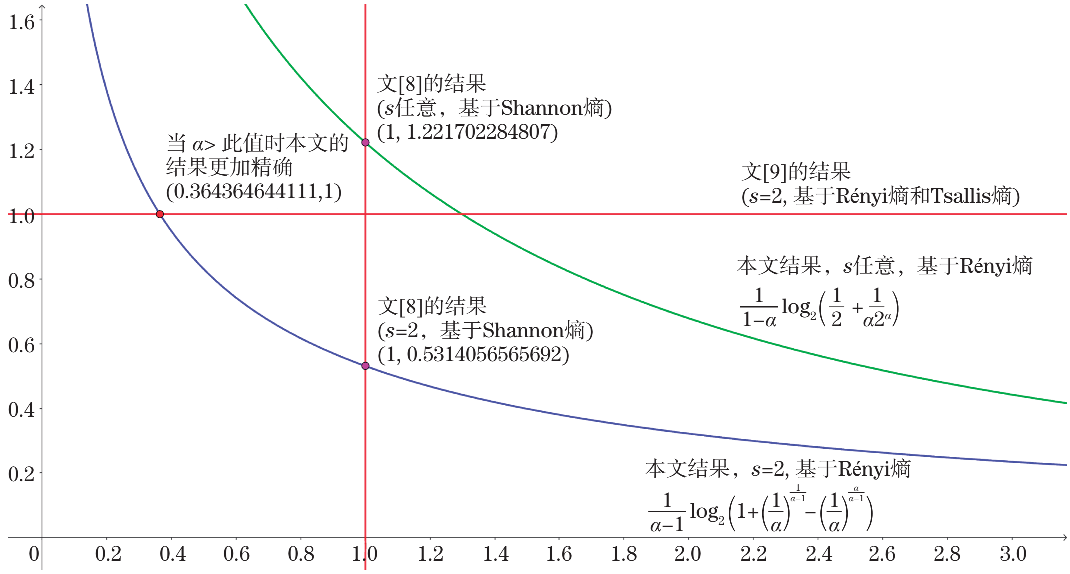

图 1 误差估计 $h(\alpha)$ 关于参数 α 的函数图象, 以及与文 [8–9] 结论对比图

最后, 利用 Rényi 熵与 Tsallis 熵的对应关系, 将上述结论推广至 Tsallis 熵.

定理 3.4 当 $s=2$, 即两个概率分布耦合时,

$$T_\alpha(\mathcal{G}_S)\leqslant\frac{1}{1-\alpha^{\frac{1}{1-\alpha}}(\frac{1}{\alpha}-1)}T_\alpha(\mathrm{OPT}_S)+\frac{\alpha^{\frac{\alpha}{1-\alpha}}}{1-\alpha^{\frac{1}{1-\alpha}}(\frac{1}{\alpha}-1)}.$$

对于一般的 s,

$$T_\alpha(\mathcal{G}_S)\leqslant\Big(\frac{1}{2}+\frac{1}{\alpha 2^{\alpha}}\Big)T_\alpha(\mathrm{OPT}_S)+\frac{1}{(1-\alpha)\alpha 2^{\alpha}}-\frac{1}{2(1-\alpha)}.$$

§4 定理的证明

§4.1 定理 3.1 的证明

为了研究的便利性, 我们引入概率分布 p 的逆外形 (Inverse Sketch) 及概率分布族 S 的逆轮廓 (Inverse Profile).

定义 4.1 对于任意给定的概率分布 p_i, 定义函数

$$S_{p_i}^{-1}(y):=\sum_{j=1}^{|p_i|}p_i(j)\mathbb{1}_{\{p_i(j)\leqslant y\}}$$

为 p_i 的逆外形. 设 $S=\{p_1,p_2,\cdots,p_s\}$, 定义函数

$$P_S^{-1}(y):=\max_{p_i\in S}S_{p_i}^{-1}(y)$$

为 S 的逆轮廓.

我们必须指出的是, 因为外形和轮廓不具有严格单调性质, 所以并不具有严格意义上的反函数. 事实上, $S_p^{-1}(y)$ 和 $P_S^{-1}(y)$ 分别是 $S_p(x)$ 和 $P_S(x)$ 的广义逆函数. 下面我们将熵的概念进一步推广至逆外形与逆轮廓.

定义 4.2 若 H 是某概率分布的逆外形或某概率分布族的逆轮廓, 定义 H 的一般熵

$$F_{\rm cost}^{\alpha}(H):=H(1)+(1-\alpha)\int_0^1H(y)y^{\alpha-2}{\rm d}y.$$

容易验证, 上述定义是合理的. 对于逆外形, 我们有如下结论:

$$\begin{aligned}F_{\rm cost}^{\alpha}(S_p^{-1})&=S_p^{-1}(1)+(1-\alpha)\sum_{j=1}^np_j\Big(\int_{p_j}^1y^{\alpha-2}{\rm d}y\Big)\\&=m(p)-\sum_{j=1}^np_j\Big(\int_{p_j}^1(\alpha-1)y^{\alpha-2}{\rm d}y\Big)\\&=m(p)-\sum_{j=1}^np_j(1-p_j^{\alpha-1})=\sum_{j=1}^np_j^{\alpha}=F_{\rm cost}^{\alpha}(p).\end{aligned}$$

对于逆轮廓, 我们有如下引理.

引理 4.1 $F_{\rm cost}^{\alpha}(P_S)=F_{\rm cost}^{\alpha}(P_S^{-1})$.

证

$$\begin{aligned}P_S(x)&=y,\\x&=P_S^{-1}(y),\\{\rm d}x&=[P_S^{-1}(y)]'{\rm d}y.\end{aligned}$$

从严格意义上讲, $(P_S^{-1})'(y)$ 在不连续点无定义. 但可以令 $[P_S^{-1}(y)]'$ 在这些不连续点取 Dirac δ 函数的若干倍以规避该问题. 将上式代入定义 2.3, 得到

$$\begin{aligned}F_{\rm cost}^{\alpha}(P_S)&=\int_0^{m(S)}(P_S(x))^{\alpha-1}{\rm d}x=\int_0^1y^{\alpha-1}[P_S^{-1}(y)]'{\rm d}y\\&=P_S^{-1}(y)y^{\alpha-1}|_0^1-\int_0^1(\alpha-1)P_S^{-1}(y)y^{\alpha-2}{\rm d}y\\&=P_S^{-1}(1)+(1-\alpha)\int_0^1y^{\alpha-2}P_S^{-1}(y){\rm d}y=F_{\rm cost}^{\alpha}(P_S^{-1}).\end{aligned}$$

引理证毕.

引理 4.2 $\forall y\in[0,1]$, $S_{{\rm OPT}_S}^{-1}(y)\geqslant P_S^{-1}(y)$.

证 根据 P_S^{-1} 的定义, 只要证明 $\forall p\in S$, $S^{-1}_{\mathrm{OPT}_S}(y)\geqslant S_p^{-1}(y)$. 对于概率分布 p 和 p', 如果满足以下条件:

(1) $|p'|=|p|+1$;

(2) $\exists 1\leqslant i\leqslant |p|$, $\exists 1\leqslant j_1<j_2\leqslant |p'|$, 使得 $p(i)=p'(j_1)+p'(j_2)$;

(3) $(p(1),\cdots,p(i-1),p(i+1),\cdots,p(|p|))=(p'(1),\cdots,p'(j_1-1),p'(j_1+1),\cdots,p'(j_2-1),p'(j_2+1),\cdots,p'(|p'|))$,

则称概率分布 p' 由 p 经过一次状态分裂得到. 例如 $(0.5,0.2,0.2,0.1)$ 可以由 $(0.5,0.4,0.1)$ 经过一次状态分裂得到. 容易验证:

(1) $S_{p'}^{-1}(y)=S_p^{-1}(y)$, 当 $y<p'(j_2)$ 时;

(2) $S_{p'}^{-1}(y)>S_p^{-1}(y)$, 当 $p'(j_2)\leqslant y<p'(j_1)+p'(j_2)$ 时;

(3) $S_{p'}^{-1}(y)=S_p^{-1}(y)$, 当 $y\geqslant p'(j_1)+p'(j_2)$ 时.

因此, 如果概率分布 p' 是由概率分布 p 经过一次状态分裂得到, 那么 $\forall y$, $S_{p'}^{-1}(y)\geqslant S_p^{-1}(y)$. 注意到最优耦合 OPT_S 由边缘分布 p 进行若干次状态的分裂得到, 所以 $S^{-1}_{\mathrm{OPT}_S}(y)\geqslant S_p^{-1}(y)$.

引理证毕.

定理 3.1 的证明 当 $0<\alpha<1$ 时, 由定义 4.2 及引理 4.2 知

$$F^{\alpha}_{\mathrm{cost}}(S^{-1}_{\mathrm{OPT}_S})\geqslant F^{\alpha}_{\mathrm{cost}}(S_p^{-1}).$$

又因为

$$H_\alpha(p)=\frac{1}{1-\alpha}\log_2 F^{\alpha}_{\mathrm{cost}}(p),$$

所以

$$H_\alpha(\mathrm{OPT}_S)\geqslant H_\alpha(P_S).$$

当 $\alpha>1$ 时, 注意到此时 $1-\alpha<0$, 所以

$$F^{\alpha}_{\mathrm{cost}}(S^{-1}_{\mathrm{OPT}_S})\leqslant F^{\alpha}_{\mathrm{cost}}(S_p^{-1}).$$

又因为

$$H_\alpha(p)=\frac{1}{1-\alpha}\log_2 F^{\alpha}_{\mathrm{cost}}(p),$$

所以

$$H_\alpha(\mathrm{OPT}_S)\geqslant H_\alpha(P_S).$$

综上, $\forall\alpha\neq 1$, 总有

$$H_\alpha(\mathrm{OPT}_S)\geqslant H_\alpha(P_S)$$

成立.

定理 3.1 证毕.

§4.2　定理 3.2 的证明

设概率分布 p,q 为 MEC 问题的两个边缘分布, 令 p^t,q^t 分别表示在贪心算法中 p,q 经过 t 次循环后得到的概率分布, 并记 $S^t := \{p^t,q^t\}$. 显然 $S^0=S$. 相应地, 令 $\mathcal{G}_S^t := (\mathcal{G}_S(1),\cdots,\mathcal{G}_S(t))$.

情形 1　首先考虑 $0<\alpha<1$, 定义

$$M^t := (1-r)F_{\rm cost}^{\alpha}(\mathcal{G}_S^t)+F_{\rm cost}^{\alpha}(P_{S^t}^{-1}),$$

其中 $r\in(0,1)$ 为参数. 显然 $M^0=F_{\rm cost}^{\alpha}(P_S)$, $M^{|\mathcal{G}_S|}=(1-r)F_{\rm cost}^{\alpha}(\mathcal{G}_S)$. 如果存在适当的 r, 使得 $M^{t+1}\leqslant M^t$ 对于任意的 t 均成立, 根据不等式的传递性, 有

$$(1-r)F_{\rm cost}^{\alpha}(\mathcal{G}_S)=M^{|\mathcal{G}_S|}\leqslant M^0=F_{\rm cost}^{\alpha}(P_S^{-1})$$

成立, 从而得到估计

$$F_{\rm cost}^{\alpha}(\mathcal{G}_S)\leqslant\frac{1}{1-r}F_{\rm cost}^{\alpha}(P_S^{-1})=\frac{1}{1-r}F_{\rm cost}^{\alpha}(P_S). \tag{4.1}$$

注意到随着贪心算法的运行, $F_{\rm cost}^{\alpha}(P_{S^t}^{-1})$ 关于 t 单调递减, $F_{\rm cost}^{\alpha}(\mathcal{G}_S^t)$ 关于 t 单调递增. 故此, 当 r 充分大时, 这样的 r 始终存在. 我们致力于寻找尽量小的 r, 以获得更优估计.

不妨设 $p^t(1)\leqslant q^t(1)$, 那么 $\mathcal{G}_S(t+1)=p^t(1)$. 针对 M^t 的第一部分 $F_{\rm cost}^{\alpha}(\mathcal{G}_S^t)$, 显然有

$$F_{\rm cost}^{\alpha}(\mathcal{G}_S^{t+1})=F_{\rm cost}^{\alpha}(\mathcal{G}_S^t)+p^t(1)^{\alpha} \tag{4.2}$$

成立. 下面研究 $F_{\rm cost}^{\alpha}(P_S^{-1})$ 关于 t 的增减关系. 首先有

(1)

$$S_{p^{t+1}}^{-1}(y)=\begin{cases}S_{p^t}^{-1}(y), & y<p^t(1),\\ S_{p^t}^{-1}(y)-p^t(1), & y\geqslant p^t(1).\end{cases}$$

(2)

$$S_{q^{t+1}}^{-1}(y)=\begin{cases}S_{q^t}^{-1}(y), & y<q^t(1)-p^t(1),\\ S_{q^t}^{-1}(y)+q^t(1)-p^t(1), & q^t(1)-p^t(1)\leqslant y<q^t(1),\\ S_{q^t}^{-1}(y)-p^t(1), & y\geqslant q^t(1).\end{cases}$$

进一步得到

$$P_{S^{t+1}}^{-1}(y)\begin{cases}=P_{S^t}^{-1}(y), & y<q^t(1)-p^t(1),\\ \leqslant P_{S^t}^{-1}(y)+q^t(1)-p^t(1), & q^t(1)-p^t(1)\leqslant y<p^t(1),\\ =P_{S^t}^{-1}(y)-p^t(1), & y\geqslant p^t(1).\end{cases} \tag{4.3}$$

上述第二个等号成立的原因需特别说明. 事实上对于任意的 t, $m(p^t)=m(q^t)=m(S^t)$. 因为 $y\geqslant p^t(1)$, 有

$$S_{p^{t+1}}^{-1}(y)=m(S^{t+1})\geqslant S_{q^{t+1}}^{-1}(y),$$
$$S_{p^t}^{-1}(y)=m(S^t)\geqslant S_{q^t}^{-1}(y).$$

所以

$$P_{S^{t+1}}^{-1}(y)=\max\{S_{p^{t+1}}^{-1}(y),S_{q^{t+1}}^{-1}(y)\}=S_{p^{t+1}}^{-1}(y),$$

$$P_{S^t}^{-1}(y) = \max\{S_{p^t}^{-1}(y), S_{q^t}^{-1}(y)\} = S_{p^t}^{-1}(y).$$

又因为

$$S_{p^{t+1}}^{-1}(y) = S_{p^t}^{-1}(y) - p^t(1),$$

从而当 $y \geqslant p^t(1)$ 时,

$$P_{S^{t+1}}^{-1}(y) = P_{S^t}^{-1}(y) - p^t(1),$$

(4.3) 成立.

结合定义 4.1 及 (4.3), 注意此时 $1-\alpha > 0$, 可得

$$\begin{aligned} F_{\text{cost}}^{\alpha}(P_{S^{t+1}}^{-1}) &= m(S^{t+1}) + \int_0^1 (1-\alpha) P_{S^{t+1}}^{-1}(y) y^{\alpha-2} \mathrm{d}y \\ &\leqslant m(S^t) - p^t(1) + \int_0^1 (1-\alpha) P_{S^t}^{-1}(y) y^{\alpha-2} \mathrm{d}y - \int_{p^t(1)}^1 p^t(1)(1-\alpha) y^{\alpha-2} \mathrm{d}y \\ &\quad + \begin{cases} \displaystyle\int_{q^t(1)-p^t(1)}^{p^t(1)} [q^t(1)-p^t(1)](1-\alpha) y^{\alpha-2} \mathrm{d}y, & q^t(1)-p^t(1) < p^t(1), \\ 0, & \text{否则} \end{cases} \\ &\leqslant F_{\text{cost}}^{\alpha}(P_{S^t}^{-1}) - (p^t(1))^{\alpha} + [q^t(1)-p^t(1)][(q^t(1)-p^t(1))^{\alpha-1} - (p^t(1))^{\alpha-1}]. \end{aligned}$$

令

$$M^{t+1} - M^t \leqslant -r(p^t(1))^{\alpha} + \max_{p^t(1)<q^t(1)<2p^t(1)} [q^t(1)-p^t(1)][(q^t(1)-p^t(1))^{\alpha-1} - (p^t(1))^{\alpha-1}] \leqslant 0,$$

解出 r 需满足的条件

$$r \geqslant \max_{p^t(1)<q^t(1)<2p^t(1)} \Big(\frac{q^t(1)-p^t(1)}{p^t(1)}\Big)^{\alpha} - \frac{q^t(1)-p^t(1)}{p^t(1)}.$$

为了得到尽量小的 r 从而获得更优估计, 令 $x := \frac{q^t(1)-p^t(1)}{p^t(1)} \in (0,1)$, 则

$$r \geqslant \max_{0<x<1} x^{\alpha} - x.$$

考察函数 $g(x) = x^{\alpha} - x$, 其中 $0 < x, \alpha < 1$. 令

$$g'(x) = \alpha x^{\alpha-1} - 1 = 0.$$

因 $g'(x)$ 在 $(0,1)$ 上单调递减, 则当 $x = \big(\frac{1}{\alpha}\big)^{\frac{1}{\alpha-1}}$ 时 $g(x)$ 取得极大值 (同时也是最大值) $\big(\frac{1}{\alpha}\big)^{\frac{\alpha}{\alpha-1}} - \big(\frac{1}{\alpha}\big)^{\frac{1}{\alpha-1}}$. 取

$$r = \Big(\frac{1}{\alpha}\Big)^{\frac{\alpha}{\alpha-1}} - \Big(\frac{1}{\alpha}\Big)^{\frac{1}{\alpha-1}},$$

代入 (4.1) 得到估计

$$F_{\text{cost}}^{\alpha}(\mathcal{G}_S) \leqslant \frac{1}{1 - (\frac{1}{\alpha})^{\frac{\alpha}{\alpha-1}} + (\frac{1}{\alpha})^{\frac{1}{\alpha-1}}} F_{\text{cost}}^{\alpha}(P_S).$$

不等号两边同时取对数, 并乘 $\frac{1}{1-\alpha} > 0$, 得到

$$H_{\alpha}(\mathcal{G}_S) \leqslant H_{\alpha}(P_S) + \frac{1}{\alpha-1} \log_2 \Big(1 + \Big(\frac{1}{\alpha}\Big)^{\frac{1}{\alpha-1}} - \Big(\frac{1}{\alpha}\Big)^{\frac{\alpha}{\alpha-1}}\Big).$$

再结合定理 3.1, 情形 1 证毕.

情形 2 考虑 $\alpha > 1$ 的情形, 注意此时 $f_{\text{cost}}^{\alpha}(x) = x^{\alpha}$ 的凹凸性发生改变, 从而导致相应的不等关系发生变化. 定义

$$M^t := (1+r)F_{\text{cost}}^{\alpha}(\mathcal{G}_S^t) + F_{\text{cost}}^{\alpha}(P_{S^t}^{-1}),$$

其中 $r > 0$ 为参数. 显然 $M^{|\mathcal{G}_S|} = (1+r)F_{\text{cost}}^{\alpha}(\mathcal{G}_S)$. 如果存在适当的 r, 使得 $M^{t+1} \geqslant M^t$ 对于任意的 t 均成立, 根据不等式的传递性, 有

$$(1+r)F_{\text{cost}}^{\alpha}(\mathcal{G}_S) = M^{|\mathcal{G}_S|} \geqslant M^0 = F_{\text{cost}}^{\alpha}(P_S^{-1}),$$

从而得到估计

$$F_{\text{cost}}^{\alpha}(\mathcal{G}_S) \geqslant \frac{1}{1+r}F_{\text{cost}}^{\alpha}(P_S^{-1}) = \frac{1}{1+r}F_{\text{cost}}^{\alpha}(P_S). \tag{4.4}$$

再对不等号两侧取对数, 并乘 $\frac{1}{1-\alpha} < 0$, 得到

$$H_{\alpha}(\mathcal{G}_S) \leqslant H_{\alpha}(P_S) + \frac{1}{\alpha-1}\log_2(1+r).$$

从上式可以看出, 为了获得更优估计, 我们致力于寻找使得 $M^{t+1} \geqslant M^t$ 成立的尽可能小的 r.

假设 $p^t(1) \leqslant q^t(1)$, 重复情形 1 的证明过程知 (4.2)–(4.3) 依然成立. 结合定义 4.1 及 (4.3), 注意此时 $1-\alpha < 0$, 可得

$$\begin{aligned}
F_{\text{cost}}^{\alpha}(P_{S^{t+1}}^{-1}) &= m(S^{t+1}) + \int_0^1 (1-\alpha)P_{S^{t+1}}^{-1}(y)y^{\alpha-2}\mathrm{d}y \\
&\geqslant m(S^t) - p^t(1) + \int_0^1 (1-\alpha)P_{S^t}^{-1}(y)y^{\alpha-2}\mathrm{d}y - \int_{p^t(1)}^1 p^t(1)(1-\alpha)y^{\alpha-2}\mathrm{d}y \\
&\quad + \begin{cases} \displaystyle\int_{q^t(1)-p^t(1)}^{p^t(1)} [q^t(1)-p^t(1)](1-\alpha)y^{\alpha-2}\mathrm{d}y, & q^t(1)-p^t(1) < p^t(1), \\ 0, & \text{否则} \end{cases} \\
&\geqslant F_{\text{cost}}^{\alpha}(P_{S^t}^{-1}) - (p^t(1))^{\alpha} + [q^t(1)-p^t(1)][(q^t(1)-p^t(1))^{\alpha-1} - (p^t(1))^{\alpha-1}].
\end{aligned}$$

令

$$M^{t+1} - M^t \geqslant r(p^t(1))^{\alpha} + \min_{p^t(1)<q^t(1)<2p^t(1)} [q^t(1)-p^t(1)][(q^t(1)-p^t(1))^{\alpha-1} - (p^t(1))^{\alpha-1}] \geqslant 0,$$

解出 r 需满足的条件

$$r \geqslant -\min_{p^t(1)<q^t(1)<2p^t(1)} \Big(\frac{q^t(1)-p^t(1)}{p^t(1)}\Big)^{\alpha} - \frac{q^t(1)-p^t(1)}{p^t(1)}.$$

为了得到尽量小的 r, 从而获得更优估计, 令 $x := \frac{q^t(1)-p^t(1)}{p^t(1)} \in (0,1)$, 则

$$r \geqslant -\min_{0<x<1} x^{\alpha} - x.$$

考察函数 $g(x) = x^{\alpha} - x$, 其中 $\alpha > 1$. 令

$$g'(x) = \alpha x^{\alpha-1} - 1 = 0.$$

因 $g'(x)$ 在 $(0,1)$ 上单调递增, 则当 $x=\left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}}$ 时, $g(x)$ 取得极小值 (同时也是最小值) $\left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha-1}}-\left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}}$. 取

$$r=-\Big(\Big(\frac{1}{\alpha}\Big)^{\frac{\alpha}{\alpha-1}}-\Big(\frac{1}{\alpha}\Big)^{\frac{1}{\alpha-1}}\Big)$$

并代入 (4.4) 得到估计

$$H_\alpha(\mathcal{G}_S)\leqslant H_\alpha(P_S)+\frac{1}{\alpha-1}\log_2\Big(1+\Big(\frac{1}{\alpha}\Big)^{\frac{1}{\alpha-1}}-\Big(\frac{1}{\alpha}\Big)^{\frac{\alpha}{\alpha-1}}\Big).$$

再结合定理 3.1, 情形 2 证毕.

值得注意的是, 情形 1 与情形 2 分别得到的 $\mathcal{G}_S$ 与 OPT_S 之间 Rényi 熵差值的上界 $h(\alpha)$ 具有相同表达式, 故此将二者合并阐述.

定理 3.2 证毕.

§4.3 定理 3.3 的证明

定义 4.3 给定概率分布族 S, 定义函数 $C_S(x):\ [0,1]\to[0,1]$ 满足如下条件:

(1) $C_S(x)$ 单调不降;

(2) 如果 $m(S^t)=x$, 那么 $\mathcal{G}_S(t+1)\geqslant C_S(x)$.

直观来讲, 在贪心算法即将进行第 $t+1$ 次循环时, 如果概率分布族 S 中剩余总概率 $m(S^t)=x$, 那么 $C_S(x)$ 是贪心算法第 $t+1$ 次生成概率值 $\mathcal{G}_S(t+1)$ 的一个下界. 注意 $C_S(x)$ 只与 x 相关, 与 t 无关.

引理 4.3 对于上述定义的 $C_S(x)$, $\forall\alpha\neq1$, 有

$$H_\alpha(\mathcal{G}_S)\leqslant H_\alpha(C_S(x))$$

成立, 其中

$$H_\alpha(C_S(x)):=\frac{1}{1-\alpha}\log_2 F^\alpha_{\mathrm{cost}}(C_S(x)):=\frac{1}{1-\alpha}\log_2\int_0^1(C_S(x))^{\alpha-1}\mathrm{d}x.$$

证 当 $x=1-\sum\limits_{j=1}^{i}\mathcal{G}_S(j)$ 时, 由 $C_S(x)$ 的定义知 $C_S(x)\leqslant S_{\mathcal{G}_S}(x)$. 又因为 $C_S(x)$ 单调不降且 $S_{\mathcal{G}_S}(x)$ 在每个 $\Big(1-\sum\limits_{j=1}^{i+1}\mathcal{G}_S(j),\ 1-\sum\limits_{j=1}^{i}\mathcal{G}_S(j)\Big]$ 上取常值, 所以对于任意的 $x\in(0,1]$,

$$C_S(x)\leqslant S_{\mathcal{G}_S}(x)$$

成立.

当 $0<\alpha<1$ 时, $(S_{\mathcal{G}_S}(x))^{\alpha-1}\leqslant(C_S(x))^{\alpha-1}$, 从而有

$$F^\alpha_{\mathrm{cost}}(\mathcal{G}_S)\leqslant F^\alpha_{\mathrm{cost}}(C_S(x)).$$

不等号两边取对数, 并同时乘 $\frac{1}{1-\alpha}>0$, 得到

$$H_\alpha(\mathcal{G}_S)\leqslant H_\alpha(C_S(x)).$$

当 $\alpha>1$ 时, $(S_{\mathcal{G}_S}(x))^{\alpha-1}\geqslant(C_S(x))^{\alpha-1}$, 从而有

$$F^\alpha_{\mathrm{cost}}(\mathcal{G}_S)\geqslant F^\alpha_{\mathrm{cost}}(C_S(x)).$$

不等号两边取对数, 并同时乘 $\frac{1}{1-\alpha}<0$, 得到

$$H_\alpha(\mathcal{G}_S)\leqslant H_\alpha(C_S(x)).$$

综上, 引理 4.3 证毕.

$C_S(x)$ 的存在性显然, 但不唯一. 下面我们将寻找适当的 $C_S(x)$ 以获得贪心算法的近似解的更优估计. 定义函数

$$R_S(y):=\max_{p\in S}R_p(y):=\max_{p\in S}\sum_{j=1}^{n}\begin{cases}y, & y\leqslant\dfrac{p(j)}{2},\\[2ex] \dfrac{p(j)}{2}, & \dfrac{p(j)}{2}<y<p(j),\\[2ex] p(j), & y\geqslant p(j).\end{cases}$$

引理 4.4 定义 $C_S(x):=\arg\min\limits_{y}[R_S(y)\geqslant x]$, 则 $C_S(x)$ 满足定义 4.3 的条件.

证 在贪心算法即将进行第 $t+1$ 次循环时, 假设概率分布族 S 中剩余总概率 $m(S^t)=x$, 贪心算法下一次创造的概率 $\mathcal{G}_S(t+1)=y$, 则必存在某个概率分布 $p_i^t\in S^t$, 其最大的概率 $p_i^t(1)=y$. 由 $C_S(x)$ 的定义知, 只需证明 $R_{p_i}(y)\geqslant x$.

对于概率分布 p_i 中任一状态 j, 我们给出此时剩余概率的一个上界. 一方面, 因为当前 p_i^t 中最大的概率为 y, 所以剩余概率 $\leqslant\min\{p_i(j),y\}$. 另一方面, 如果 $p_i(j)>y$, 注意到贪心算法生成的概率分布 $\mathcal{G}_S$ 单调不增, 即 $\mathcal{G}_S(t)\leqslant\mathcal{G}_S(t-1)$ 对于任意的 $2\leqslant t\leqslant|\mathcal{G}_S|$ 成立, 所以 $p_i(j)$ 至少有 y 的概率已经在前序循环中被移除并添加至 $\mathcal{G}_S$ 中, 因此剩余概率 $\leqslant p_i(j)-y$.

总之, 结合以上两种情况, 并将 p_i 的所有状态进行累加, 得到如下估计

$$\sum_{j=1}^{n}\begin{cases}p_i(j), & p_i(j)\leqslant y,\\ \min\{y,p_i(j)-y\}, & p_i(j)>y,\end{cases}\geqslant x.$$

特别地, 当 $y\leqslant\frac{p_i(j)}{2}$ 时, $\min\{y,p_i(j)-y\}=y$; 当 $\frac{p_i(j)}{2}<y<p_i(j)$ 时, $\min\{y,p_i(j)-y\}=p_i(j)-y<\frac{p_i(j)}{2}$, 即

$$R_{p_i}(y)=\sum_{j=1}^{n}\begin{cases}y, & y\leqslant\dfrac{p_i(j)}{2},\\[2ex] \dfrac{p_i(j)}{2}, & \dfrac{p_i(j)}{2}<y<p_i(j),\\[2ex] p_i(j), & y\geqslant p_i(j),\end{cases}\geqslant\sum_{j=1}^{n}\begin{cases}p_i(j), & p_i(j)\leqslant y,\\ \min\{y,p_i(j)-y\}, & p_i(j)>y,\end{cases}\geqslant x.$$

引理证毕.

引理 4.5

$$\int_0^1(C_S(x))^{\alpha-1}\mathrm{d}x=\Big(\frac{1}{2}+\frac{1}{\alpha2^\alpha}\Big)F_{\mathrm{cost}}^{\alpha}(P_S^{-1}).$$

证

$$
\begin{aligned}
&\int_0^1 (C_S(x))^{\alpha-1}\mathrm{d}x\\
&=\int_0^1 R_S'(y)y^{\alpha-1}\mathrm{d}y\\
&=R_S(y)y^{\alpha-1}|_0^1+\int_0^1 R_S(y)(1-\alpha)y^{\alpha-2}\mathrm{d}y\\
&=m(S)+\int_0^1\Big[P_S^{-1}(y)+\frac{1}{2}\Big(P_S^{-1}(2y)-P_S^{-1}(y)\Big)+y\int_{2y}^1\frac{[P_S^{-1}(t)]'}{t}\mathrm{d}t\Big](1-\alpha)y^{\alpha-2}\mathrm{d}y\\
&=m(S)+\Big(\int_0^1\frac{P_S^{-1}(y)+P_S^{-1}(2y)}{2}(1-\alpha)y^{\alpha-2}+(1-\alpha)y^{\alpha-1}\Big(\int_{2y}^1\frac{[P_S^{-1}(t)]'}{t}\mathrm{d}t\Big)\mathrm{d}y\Big)\\
&=\frac{1}{2}\Big(m(S)+\int_0^1 P_S^{-1}(y)(1-\alpha)y^{\alpha-2}\mathrm{d}y\Big)+\frac{1}{2}\Big(m(S)+\int_0^1 P_S^{-1}(2y)(1-\alpha)y^{\alpha-2}\mathrm{d}y\Big)\\
&\quad+\int_0^1\frac{[P_S^{-1}(t)]'}{t}\cdot\int_0^{\frac{t}{2}}(1-\alpha)y^{\alpha-1}\mathrm{d}y\mathrm{d}t\\
&:=\mathrm{I}+\mathrm{II}+\mathrm{III}.
\end{aligned}
$$

接下来, 只需对上述三部分分别进行估计. 首先注意到

$$
\mathrm{I}=\frac{1}{2}F_{\mathrm{cost}}^{\alpha}(P_S^{-1}). \tag{4.5}
$$

然后对 II 使用分部积分法并进行换元:

$$
\begin{aligned}
\mathrm{II}&=\frac{1}{2}m(S)-\frac{1}{2}\int_0^1 P_S^{-1}(2y)(\alpha-1)y^{\alpha-2}\mathrm{d}y\\
&=\int_0^1[P_S^{-1}(2y)]'y^{\alpha-1}\mathrm{d}y\\
&=\frac{1}{2}\int_0^2[P_S^{-1}(u)]'\Big(\frac{u}{2}\Big)^{\alpha-1}\mathrm{d}u\\
&=2^{-\alpha}\int_0^1[P_S^{-1}(u)]'u^{\alpha-1}\mathrm{d}u\\
&=2^{-\alpha}F_{\mathrm{cost}}^{\alpha}(P_S^{-1}).
\end{aligned}\tag{4.6}
$$

最后关于 III, 注意到

$$
\int_0^{\frac{t}{2}}(1-\alpha)y^{\alpha-1}\mathrm{d}y=\frac{1-\alpha}{\alpha}y^{\alpha}|_0^{t/2}=\frac{1-\alpha}{\alpha2^{\alpha}}t^{\alpha}, \tag{4.7}
$$

将上述结果代入 III, 得

$$
\frac{1-\alpha}{\alpha2^{\alpha}}\int_0^1[P_S^{-1}(t)]'t^{\alpha-1}\mathrm{d}t=\frac{1-\alpha}{\alpha2^{\alpha}}F_{\mathrm{cost}}^{\alpha}(P_S^{-1}).
$$

综合 (4.5)–(4.7), 有

$$
\int_0^1(C_S(x))^{\alpha-1}\mathrm{d}x=\Big(\frac{1}{2}+\frac{1}{\alpha2^{\alpha}}\Big)F_{\mathrm{cost}}^{\alpha}(P_S^{-1}).
$$

引理证毕.

定理 3.3 的证明 在引理 4.5 的等号两边同时取对数, 并乘 $\frac{1}{1-\alpha}$, 得

$$H_\alpha(C_S(x)) = H_\alpha(P_S) + \frac{1}{1-\alpha}\log_2\left(\frac{1}{2}+\frac{1}{\alpha 2^\alpha}\right).$$

最后, 应用引理 4.3–4.4 及定理 3.1, 有

$$H_\alpha(\mathcal{G}_S) \leqslant H_\alpha(\mathrm{OPT}_S) + \frac{1}{1-\alpha}\log_2\left(\frac{1}{2}+\frac{1}{\alpha 2^\alpha}\right).$$

定理证毕.

§4.4　定理 3.4 的证明

将定理 3.2–3.3 的结论代入

$$H_\alpha(p) = \frac{1}{1-\alpha}\log_2\left[1+(1-\alpha)T_\alpha(p)\right]$$

中, 即可得证.

参 考 文 献

[1] Sun F Y, Hoffmann J, Verma V, et al. InfoGraph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization [J/OL]. 2020, https://arxiv.org/pdf/1908.01000.

[2] Cicalese F, Gargano L, Vaccaro U. Minimum-entropy couplings and their applications [J]. *IEEE Transactions on Information Theory*, 2019, 65(6):3436–3451.

[3] Kocaoglu M, Dimakis A G, Vishwanath S, et al. Entropic causal inference [J/OL]. 2016, https://doi.org/10.48550/arxiv.1611.04035.

[4] Compton S, Kocaoglu M, Greenewald K, et al. Entropic causal inference: Identifiability and finite sample results [J/OL]. 2021, https://doi.org/10.48550/arxiv.2101.03501.

[5] Schroeder de Witt C, Sokota S, Kolter J Z, et al. Perfectly secure steganography using minimum entropy coupling [J/OL]. 2022, https://arxiv.org/pdf/2210.14889.

[6] Kovačević M, Stanojević I, Šenk V. On the entropy of couplings [J]. *Information and Computation*, 2015, 242:369–382.

[7] Kocaoglu M, Dimakis A G, Vishwanath S, et al. Entropic causality and greedy minimum entropy coupling [C]. IEEE International Symposium on Information Theory (ISIT), Aachen, 2017:1465–1469.

[8] Compton S, Katz D, Qi B, et al. Minimum-entropy coupling approximation guarantees beyond the majorization barrier [J/OL]. 2023, https://doi.org/10.48550/arxiv.2302.11838.

[9] 朱亚倩. 最小联合熵的矩阵结构和算法适应性 [D]. 北京: 北京师范大学, 2023.

[10] 孙清扬, 胡海洋, 兰宗谕, 等. 最小联合 Tsallis 信息熵问题算法研究 [J]. 首都师范大学学报 (自然科学版), 2025, 46(6):26–33.

Error Estimation for Approximate Solutions to the Minimum Entropy Coupling Problem

HAO Yifei[1] WANG Yingzhe[2]

[1]School of Mathematical Sciences, Key Laboratory of Mathematics and Complex Systems, Ministry of Education, Beijing Normal University, Beijing 100875, China. E-mail: haoyf@mail.bnu.edu.cn

[2]Corresponding author. School of Mathematical Sciences, Key Laboratory of Mathematics and Complex Systems, Ministry of Education, Beijing Normal University, Beijing 100875, China. E-mail: wangyz@bnu.edu.cn

Abstract For the minimum entropy coupling problem based on Rényi entropy H_α, an upper bound $h(\alpha)$ on the entropy function difference between the greedy algorithm's approximate solution and the optimal solution was derived by introducing the profile method. When the authors couple two probability distributions, $h(\alpha)=\frac{1}{\alpha-1}\log_2\big(1+\big(\frac{1}{\alpha}\big)^{\frac{1}{\alpha-1}}-\big(\frac{1}{\alpha}\big)^{\frac{\alpha}{\alpha-1}}\big)$. When the authors couple finite probability distributions, $h(\alpha)=\frac{1}{1-\alpha}\log_2\big(\frac{1}{2}+\frac{1}{\alpha 2^\alpha}\big)$. These results were further extended to the Tsallis entropy. The innovation of this work lies in establishing the error estimate as an explicit function of the parameter α, and obtaining more precise results for $\alpha>0.365$.

Keywords Rényi entropy, Tsallis entropy, Coupling, Greedy algorithm

2020 MR Subject Classification 94A17

The English translation of this paper will be published in

Chinese Journal of Contemporary Mathematics, Vol. 46 No. 4, 2025

by ALLERTON PRESS, INC., USA