arXiv:2301.03096v1 [math.PR] 8 Jan 2023

无放回抽样与Hoeffding统计量的集中界

BARTŁOMIEJ POLACZYK

摘要。我们证明了基于无放回抽样的经验过程上确界的贝内特型集中界，并给出了任意Hoeffding统计量情形下的相应界。与此前同类结果相比，我们得到了更精确的集中性轮廓。

关键词：测度集中、抽样、经验过程、Hoeffding统计量。**美国数学学会分类号**：60E15, 60C05

1. 预备知识

在这篇简短的札记中，我们研究均匀随机置换的特定泛函的集中性。具体而言，我们关注无放回抽样时经验过程的上确界。这类过程可视为具有特殊形式（即含有重复行）的矩阵的 Hoeffding统计量。我们还为一般底层矩阵的单个 Hoeffding统计量得到了相应的界。文献中已对此类界进行了广泛研究，例如 [2, 5, 15], ; 它们在各种应用中都发挥着重要作用，例如直推式学习 [16], 或统计检验 [1]。

1.1. 本文结构。

本节余下部分将介绍一些核心记号。在第 2 节中，我们将给出关于无放回抽样时经验过程上确界的集中性结果。在第 3 节中，我们将给出单个 Hoeffding统计量的类似结果。第 4节将给出其余集中性估计的证明。辅助事实的证明及一些补充讨论则放在附录中。

1.2. 基本记号。

对于 $n \in \mathbb{N}$，考虑集合 $[n]$ 上的置换对称群 $S_n := \{1, \dots, n\}$，并赋予其均匀概率测度 π_n。它是交换过程 的平稳分布，该过程由生成元 L 定义，其公式为

$$Lf(\sigma) = \frac{1}{n(n-1)} \sum_{i,j=1}^{n} \big(f(\sigma \circ \tau_{ij}) - f(\sigma)\big) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \big(f(\sigma \circ \tau_{ij}) - f(\sigma)\big),$$

其中 τ_{ij} 表示元素 i 和 j的换位。用 $\mathbb{E}$表示关于 π_n的期望。此外，对于函数 $f\colon S_n \to \mathbb{R}$，简记 $f_{ij}(\cdot)$ 为 $= f(\cdot \circ \tau_{ij})$。相应的狄利克雷形式可表示为

$$\begin{aligned}\mathcal{E}(f,g) &= \frac{1}{2n(n-1)} \mathbb{E} \sum_{i,j=1}^{n} (f_{ij} - f)(g_{ij} - g) \\ &= \frac{1}{n(n-1)} \mathbb{E} \sum_{1 \le i < j \le n} (g_{ij} - g)(f_{ij} - f).\end{aligned}$$

若 f 和 g 具有相同的单调性，则由 L 的可逆性可知，我们也有

$$\mathcal{E}(f,g) = \frac{1}{n(n-1)} \mathbb{E} \sum_{i,j=1}^{n} (g_{ij} - g)_+ (f_{ij} - f)_+.$$

若修正的对数-索博列夫 不等式以常数 $\rho_0 > 0$ 成立，则称

$$\rho_0 \operatorname{Ent}_\mu(f) \le \mathcal{E}(f, \log f) \tag{1.1}$$

对于所有正函数 f，其中 $\operatorname{Ent}_\mu(f) = \int f \log \frac{f}{\int f \, d\mu} \, d\mu - \int f \, d\mu \log(\int f \, d\mu)$ 为熵泛函。对于此过程，$\rho_0 \ge \frac{1}{n-1}$ 由 Gao–Quastel [8] 和 Bobkov–Tetali [3] 独立得出（注意，不同参考文献中生成元 L 的归一化方式有所不同——此处给出的缩放常数与我们的设定相匹配）。

本研究部分得到波兰国家科学中心通过编号为 2020/37/N/ST1/02667 的 Preludium项目资助。

2. 无放回抽样——上确界的集中性

考虑一组向量 $\mathcal{X} \subset \mathbb{R}^n$。令 $I_1, \dots, I_n$ 为不放回均匀抽取的样本，$J_1, \dots, J_n$ 为从集合 $[n]$中有放回抽取的样本。对于 $m \le n$，定义

$$Z = \sup_{x\in\mathcal{X}} \sum_{k=1}^{m} x_{I_k}, \qquad Z' = \sup_{x\in\mathcal{X}} \sum_{k=1}^{m} x_{J_k} \tag{2.1}$$

因此， Z' 可视为关于独立随机变量的经验过程的上确界 J_k。关于 Z' 尾部的研究由来已久，始于 Talagrand 的工作[15]。

为了分析 Z 的尾部，通常将其表示为矩阵族上的 Hoeffding统计量的上确界会更方便。具体而言，对于 $x \in \mathcal{X}$，记 $a^x \in \mathbb{R}^{n\times n}$，使得 a 的前 m 行均由向量 x 的副本构成，其余各行仅含零元素，即 $a_{ij} = x_j$ （$i \le m$、$j \in [n]$）且 $a_{ij} = 0$ （$i > m$、$j \in [n]$）。那么

$$Z = \sup_{x\in\mathcal{X}} \sum_{k=1}^{n} a^x_{k\sigma_k},$$

其中 $\sigma = (I_1, I_2, \dots, I_n) \sim \pi_n$。此外，对任意 i、$j \in [n]$，记 $\sigma_{ij} = \sigma \circ \tau_{ij}$，并且

$$Z_{ij} = \sup_{x\in\mathcal{X}} \sum_{k=1}^{n} a^x_{k\sigma_{ij}(k)},$$

因此，将修正的对数 Sobolev 不等式（1.1）应用于 Z 的拉普拉斯变换后可得

$$\operatorname{Ent}(e^{\lambda Z}) \le \frac{\lambda}{n}\,\mathbb{E}\, e^{\lambda Z} \sum_{ij} (1 - e^{-\lambda(Z - Z_{ij})})_+ (Z - Z_{ij})_+.$$

接下来，我们用以下量来表述关于 Z 的集中性结果

$$\Sigma^2 = \sup_{x\in\mathcal{X}} \sum_{k=1}^{m} x^2_{I_k}, \qquad \widetilde{\Sigma}^2 = \sup_{x\in\mathcal{X}} \sum_{k=1}^{m} x^2_{J_k}.$$

正如 [9]，中所指出的，根据 Hoeffding 的一个论证 [11] （另见 [14]），若 E 是一个赋范空间且 $g\colon [n] \to E$，则对于任意凸函数 $\Psi\colon E \to \mathbb{R}$，

$$\mathbb{E}\,\Psi\Big(\sum_{k=1}^{m} g(I_k)\Big) \le \mathbb{E}\,\Psi\Big(\sum_{k=1}^{m} g(J_k)\Big). \tag{2.2}$$

式（2.2）中关于 Z、Z' 及相关量的含义，将在下面的引理中说明；该引理尤其蕴含 $\mathbb{E}\, Z \le \mathbb{E}\, Z'$ 和 $\mathbb{E}\,\Sigma^2 \le \mathbb{E}\,\widetilde{\Sigma}^2$。为完整起见，我们在附录 A 中给出证明。

L引理 2.1。设 $\phi\colon \mathbb{R} \to \mathbb{R}$ 为凸且单调递增函数，并令 Z、Z' 由 （2.1）给出。则

$$\mathbb{E}\,\phi(Z) \le \mathbb{E}\,\phi(Z').$$

关于 Z 集中性的主要结果如下文定理所示，该定理给出了一个 Bennett 型界。

定理 2.2。设 Z 由 (2.1) 给出，并假设 $\mathcal{X} \subset [-1,1]^n$。那么，存在某些绝对常数$C_1$， $C_2 > 0$,

$$\forall\, t \ge 0 \qquad \mathbb{P}(Z \ge \mathbb{E}\, Z + t) \le 2\exp\Big(-\frac{t}{C_1}\log\Big(1 + \frac{t}{C_2\,\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\Big),$$

其中 $\widetilde{\Sigma}^2 = $上确界$_{x\in\mathcal{X}} \sum_{k=1}^{m} x^2_{J_k}$。可取 $C_1 = 36$, $C_2 = 46$。

注 2.3。假设 $\mathcal{X} \subset \{\, x \in [-1,1]^n\colon \sum_i x_i = 0 \,\}$，并记 $v = m \sup_{x\in\mathcal{X}} \operatorname{Var}(x_{J_1}) + 2\,\mathbb{E}\, Z'$。那么，Tolstikhin–Blanchard–Kloft [16, 定理 2] 证明了

$$\forall\, t \ge 0 \qquad \mathbb{P}(Z \ge \mathbb{E}\, Z' + t) \le \exp\Big(-t\log\big(1 + \frac{t}{v}\big) + t - v\log\big(1 + \frac{t}{v}\big)\Big). \tag{2.3}$$

回顾一下，根据 Hoeffding论证（2.2），参见引理 2.1， $\mathbb{E}\, Z \le \mathbb{E}\, Z'$ 而在许多情况下，后一个量可能显著更大。利用对称化和针对 Rademacher平均的 Talagrand压缩原理，参见例如 [13], 我们可以估计

$$\mathbb{E}\,\widetilde{\Sigma}^2 \le m \sup_{x\in\mathcal{X}} \operatorname{Var}(x_{J_1}) + 8\,\mathbb{E} \sup_{x\in\mathcal{X}} \sum_{k=1}^{m} \varepsilon_k x_{J_k},$$

其中 $\varepsilon_1, \dots, \varepsilon_m$ 是与 $J_1, \dots, J_m$ 相互独立的 i.i.d. Rademacher变量。因此，当集合 $\mathcal{X}$ 关于原点对称时，我们得到

$$\mathbb{E}\,\widetilde{\Sigma}^2 \le m \sup_{x\in\mathcal{X}} \operatorname{Var}(x_{J_1}) + 16\,\mathbb{E}\, Z' \le 8v$$

因此，我们在定理 2.2 中得到的估计，与（2.3）相比，给出了相对于“恰当”均值的偏差界，同时指数项的尺度变化不更差（除数值常数外）。

然而，在一般情况下，并不一定满足 $\widetilde{\Sigma}^2 \le \Sigma^2$，因此可知，定理 2.2 给出的界无法直接比较。还值得注意的是，[16]的作者给出了一个界 $\mathbb{E}\,Z' \le \mathbb{E}\,Z + 2\frac{m^3}{n}$，表明在概率估计中，可以用 $\mathbb{E}\,Z'$ 替换 $\mathbb{E}\,Z$，而当 m的值较小时，损失并不大。附录 D中，我们给出了一个例子，说明在非对称集合 $\mathcal{X}$的一般情况下，我们的估计仍优于（2.3）。

为证明定理 2.2中的贝内特型不等式，我们需要以下由 Ledoux 给出的估计 [12]。为完整起见，我们在附录 B中给出证明。

Lemma 2.4 （ 中定理 2.4 的证明）。设 Z' 由 （2.1）给出，并假设 $\mathcal{X} \subset [0,1]^n$。该 n

$$\forall\, \lambda \ge 1/4 \qquad \log \mathbb{E}\, e^{\lambda Z'} \le \frac{1}{16} e^{8\lambda}\, \mathbb{E}\, Z'.$$

我们还需要以下命题，以给出 Z 的伯恩斯坦不等式。我们将其证明推迟到第 4 节。

命题 2.5。 设 Z 由 (2.1) 给出，并假设 $\mathcal{X} \subset [-1,1]^n$。则

$$\forall\, t \ge 0 \qquad \mathbb{P}(Z \ge \mathbb{E}\, Z + t) \le \exp\Big(-\min\Big(\frac{t}{32}, \frac{t^2}{128\,\mathbb{E}\,\Sigma^2}\Big)\Big),$$

其中 $\Sigma^2 = $上确界$_{x\in\mathcal{X}} \sum_k x_{I_k}^2$。

定理 *2.2* 的证明。 若 $32t < C_1C_2\,\mathbb{E}\,\widetilde{\Sigma}^2$，则应用命题 2.5 并估计 对数$(1+x) \le x$，可得只要 $128 \le C_1C_2$，

$$\begin{aligned}
\mathbb{P}(Z \ge \mathbb{E}\, Z + t) &\le \exp\Big(-\min\Big(\frac{t}{32}, \frac{t^2}{128\,\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\Big), \\
&\le \exp\Big(-\min\Big(\frac{t}{32}, \frac{t^2}{C_1C_2\,\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\Big) \\
&= \exp\Big(-\frac{t^2}{C_1C_2\,\mathbb{E}\,\widetilde{\Sigma}^2}\Big) \le \exp\Big(-\frac{t}{C_1}\log\Big(1+\frac{t}{C_2\,\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\Big)
\end{aligned}$$

在这种情况下结论即得。

若 $32t \ge C_1C_2\,\mathbb{E}\widetilde{\Sigma}^2$，然后令

$$\rho^{-1} = \alpha \log\big(1 + \beta \frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\big)$$

取某个 $\alpha, \beta > 0$ （稍后确定），并记

$$Z^{\downarrow} = \sup_{x\in\mathcal{X}} \sum_{k=1}^{m} x_{I_k} \mathbf{1}_{\{|x_{I_k}| \le \rho\}}$$

and

$$Z^{\uparrow} = \sup_{x\in\mathcal{X}} \sum_{k=1}^{m} |x_{I_k}| \mathbf{1}_{\{|x_{I_k}| > \rho\}}$$

使得 $Z \le Z^{\downarrow} + Z^{\uparrow}$。我们估计 $Z^{\downarrow}$ 和 $Z^{\uparrow}$的尾概率。B

y the est imate $\log(1+x) \le x$, by the definition of ρ and as long as $\alpha\beta \le 1/4$,

$$\frac{t}{32\rho} \le \alpha\beta \cdot \frac{t^2}{32\,\mathbb{E}\,\widetilde{\Sigma}^2} \le \frac{t^2}{128\,\mathbb{E}\,\widetilde{\Sigma}^2} \le \frac{t^2}{128\,\mathbb{E}\,\Sigma^2},$$

由命题 2.5 应用于 $Z^{\downarrow}/\rho$,

$$\begin{aligned}
\mathbb{P}(Z^{\downarrow} \ge \mathbb{E}\, Z^{\downarrow} + t) &\le \exp\Big(-\min\Big(\frac{t}{32\rho}, \frac{t^2}{128\,\mathbb{E}\,\Sigma^2}\Big)\Big) \\
&= \exp\Big(-\frac{t}{32\rho}\Big) = \exp\Big(-\frac{\alpha t}{32}\log\big(1+\beta\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\big)\Big).
\end{aligned} \tag{2.4}$$

下面转而讨论 $Z^{\uparrow}$ 的尾部。记

$$Z'_{\rho} = \sup_{x\in\mathcal{X}} \sum_{k=1}^{m} |x_{J_k}| \mathbf{1}_{\{|x_{J_k}| > \rho\}}.$$

引理 2.1 应用于 $\{(|x_i|\mathbf{1}_{\{|x_i|>\rho\}})_{i=1}^n : \ x \in \mathcal{X}\}$ 以 $\mathcal{X}$ 并结合应用于 2.4 的引理 Z'_{ρ} 得到

$$\log \mathbb{E}\, e^{\lambda Z^{\uparrow}} \le \log \mathbb{E}\, e^{\lambda Z'_{\rho}} \le \frac{1}{16} e^{8\lambda}\, \mathbb{E}\, Z'_{\rho} \tag{2.5}$$

对所有 $\lambda \ge 1/4$ 均成立。取

$$\lambda^* = \frac{1}{8}\log\Big(1+\beta\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\Big).$$

由于根据假设 $32t \ge C_1C_2\,\mathbb{E}\,\widetilde{\Sigma}^2$，因此 $\lambda^* \ge \frac{1}{8}$对数$(1+\frac{\beta C_1C_2}{32}) \ge \frac{1}{4}$，只要 $\beta C_1C_2 \ge 32(e^2-1)$。此外，注意到

$$\mathbb{E}\,Z'_\rho \le \rho^{-1}\,\mathbb{E}\,\widetilde{\Sigma}^2 \le \frac{32t\rho^{-1}}{C_1C_1}.$$

因此，结合（2.5），由 Chernoff 界可得：

$$\begin{aligned}\mathbb{P}(Z^\uparrow \ge t) &\le \exp\Big(-t\lambda^* + \frac{e^{8\lambda^*}}{16}\,\mathbb{E}\,Z'_\rho\Big)\\ &= \exp\Big(-\frac{t}{8}\log\Big(1+\beta\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\Big) + \frac{1}{16}\Big(\mathbb{E}\,Z'_\rho + \frac{\mathbb{E}\,Z'_\rho}{\mathbb{E}\,\widetilde{\Sigma}^2}t\beta\Big)\Big)\\ &\le \exp\Big(-\frac{t}{8}\log\Big(1+\beta\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\Big) + \frac{1}{16}\Big(\frac{32t\rho^{-1}}{C_1C_1} + t\beta\rho^{-1}\Big)\Big)\\ &= \exp\Big(-t\log\Big(1+\beta\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\cdot\Big(\frac{1}{8} - \frac{2\alpha}{C_1C_2} - \frac{\alpha\beta}{16}\Big)\Big).\end{aligned} \tag{2.6}$$

利用估计 $\log(1+x) \le x$，我们得到

$$|\mathbb{E}\,Z^\downarrow - \mathbb{E}\,Z| \le \mathbb{E}\,Z^\uparrow \le \frac{\mathbb{E}\,\widetilde{\Sigma}^2}{\rho} \le \alpha\beta t. \tag{2.7}$$

因此，结合（2.4）、（2.6）和（2.7），并且只要 $\alpha\beta \le 1/4$ 且 $\beta C_1C_2 \ge 32$（e^2-1），我们便得到

$$\begin{aligned}\mathbb{P}(Z \ge \mathbb{E}\,Z + 2t + \alpha\beta t) &\le \mathbb{P}(Z^\uparrow + Z^\downarrow \ge \mathbb{E}\,Z + 2t + \alpha\beta t)\\ &\le \mathbb{P}(Z^\uparrow + Z^\downarrow \ge \mathbb{E}\,Z^\downarrow - |\,\mathbb{E}\,Z - \mathbb{E}\,Z^\downarrow| + 2t + \alpha\beta t)\\ &\le \mathbb{P}(Z^\uparrow + Z^\downarrow \ge \mathbb{E}\,Z^\downarrow + 2t)\\ &\le \mathbb{P}(Z^\uparrow \ge t) + \mathbb{P}(Z^\downarrow \ge \mathbb{E}\,Z^\downarrow + t)\\ &\le 2\exp\Big(-\min\Big(\frac{\alpha}{32}, \frac{1}{8} - \frac{\alpha\beta}{16} - \frac{2\alpha}{C_1C_2}\Big)\cdot t\log\Big(1+\beta\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\Big).\end{aligned}$$

代入 $t \leftarrow (2+\alpha\beta)^{-1}t$ 并估计 $\frac{1}{2+\alpha\beta} \ge \frac{4}{9}$，得到

$$\begin{aligned}\mathbb{P}(Z \ge \mathbb{E}\,Z + t) &\le 2\exp\Big(-\frac{1}{2+\alpha\beta}\min\Big(\frac{\alpha}{32}, \frac{1}{8} - \frac{\alpha\beta}{16} - \frac{2\alpha}{C_1C_2}\Big)\cdot t\log\Big(1+\frac{\beta}{2+\alpha\beta}\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\Big)\\ &\le 2\exp\Big(-\frac{4}{9}\min\Big(\frac{\alpha}{32}, \frac{1}{8} - \frac{\alpha\beta}{16} - \frac{2\alpha}{C_1C_2}\Big)\cdot t\log\Big(1+\frac{4\beta}{9}\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\Big)\\ &\le 2\exp\Big(-\frac{4}{9}\min\Big(\frac{\alpha}{32}, \frac{1}{8} - \frac{\alpha\beta}{16} - \frac{2\alpha}{C_1C_2}\Big)\cdot t\log\Big(1+\frac{t}{C_2\,\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\Big).\end{aligned}$$

只要 $\alpha\beta \le 1/4$、$\beta C_1C_2 \ge 32(e^2-1)$ 和 $4\beta C_2 \ge 9$ 成立即可。令 $\alpha = 2$ 和 $\beta = \frac{1}{8}$，便得到结论，其中 $C_1 = 36$ 和 $C_2 = 46$。□

3. 单个 Hoeffding 统计量的集中性

本节给出单个 Hoeffding 统计量的集中性界，推广了 Chatterjee [6], Bercu–Deylon–Rio [2] 和 Albert [1] 的结果。下文中，f 表示某个 Hoeffding 统计量，即

$$f(\sigma) = \sum_{k=1}^n a_{k\sigma(k)}, \tag{3.1}$$

其中 $(a_{ij})_{i,\,j=1}^n \in \mathbb{R}^{n\times n}$ 是某个实矩阵。本节的主要结果如下定理。据我们所知，这是首个同时刻画 Hoeffding 统计量的次高斯行为和泊松行为的结果。

定理3.1。设 f 由下式给出：(3.1)。若 $a_{ij} \in [-1,1]$ 对所有 i，j 且 $\sum_{ij} a_{ij} = 0$，则存在某些绝对常数 C_1，$C_2 > 0$，

$$\forall\, t \ge 0 \qquad \mathbb{P}(f \ge t) \le 2\exp\Big(-\frac{t}{C_1}\log\Big(1+\frac{t}{C_2\,\mathbb{E}\,\Sigma^2}\Big)\Big),$$

其中 $\Sigma^2 = \sum_k a^2_{k\sigma_k}$，使得 $\mathbb{E}\,\Sigma^2 = \frac{1}{n}\sum_{ij} a_{ij}^2$。可以取 $C_1 = C_2 = 36$

注记 3.2. 与 Bercu–Deylon–Rio 中一样 [2], 注意，令

$$d_{ij} = a_{ij} - \frac{1}{n}\sum_{k=1}^{n}(a_{ik}+a_{kj}) + \frac{1}{n^2}\sum_{k,l=1}^{n} a_{kl}$$

可得 $\operatorname{Var}(f) = \frac{1}{n-1}\sum_{ij} d_{ij}^2$ 且 $f - \mathbb{E}f = \sum_{k=1}^{n} d_{k\sigma(k)}$. 因此，将定理 3.1 应用于$(f-\mathbb{E}f)/2$而非 f(注意 $\sum_{ij} d_{ij} = 0$，而 $a_{ij} \in [-1,1]$ 是任意的)，可得

$$\forall\, t \geq 0 \qquad \mathbb{P}(f \geq \mathbb{E}f + t) \leq 2\exp\Big(-\frac{t}{2C_1}\log\Big(1 + \frac{t}{2C_2\operatorname{Var}(f)}\Big)\Big). \tag{3.2}$$

正如Hoeffding在[10] 中所示（另请参见Bolthausen [4]，其采用了基于Stein方法的方法），只要

$$\lim_{n\to\infty} \frac{\max_{i,j\in[n]} d_{ij}}{\operatorname{Var}(S_n)} = 0,$$

那么 f 满足中心极限定理，即

$$\frac{f - \mathbb{E}f}{\sqrt{\operatorname{Var}(f)}} \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0,1)$$

依分布收敛。显然，当 t 取较小值时，(3.2)中的界呈次高斯形式，因此与上述中心极限定理的行为相符（相差数值常数）。类似地，若取 $a_{ij} = \mathbf{1}_{\{i=j\}}$，则 f成为随机排列 σ的不动点个数。在这种情况下，f 的精确尾分布众所周知，参见 [7, 第四节。4], 当 $\exp(-Ct$对数 $t)$ 且 t较大、某个 $C > 0$时，其阶为此，与界(3.2)一致。这表明定理 3.1 在数值常数范围内是最优的。

为了证明定理 3.1中的 Bennett不等式，我们先在下述定理中对非负统计量推导该不等式。

定理 3.3。设 f由 (3.1)给出。如果 $a_{ij} \in [0,1]$ 对所有 i，j均成立，则

$$\forall\, t \geq 0 \qquad \mathbb{P}(f > \mathbb{E}f + t) \leq \exp\Big(-\frac{t}{4}\log\Big(1 + \frac{t}{4\,\mathbb{E}f}\Big)\Big).$$

注 3.4。定理 3.3 已经改进了（指数中的数值常数除外）Bernstein 型界

$$\forall\, t \geq 0 \qquad \mathbb{P}(f > \mathbb{E}f + t) \leq \exp\Big(-\frac{t^2}{4\,\mathbb{E}f + 2t}\Big)$$

由 Chatterjee [6, 命题 1.1] 得到。

定理 *3.3*的证明。由于 $a_{ij} \in [0,1]$，因此对任意 i，j，

$$\sum_{ij}(f_{ij} - f)_+ = \sum_{ij}(a_{i\sigma_j} + a_{j\sigma_i} - a_{i\sigma_i} - a_{j\sigma_j})_+ \leq \sum_{ij}(a_{i\sigma_j} + a_{j\sigma_i}) = 2\sum_{ij} a_{ij} = 2n\,\mathbb{E}f. \tag{3.3}$$

利用修正的对数 Sobolev 不等式，结合式（3.3）以及 $x \mapsto e^{2x}$的凸性，可得

$$\begin{aligned}
\operatorname{Ent}(e^{\lambda f}) &\leq \frac{\lambda}{n}\,\mathbb{E}\,e^{\lambda f}\sum_{ij}(e^{\lambda(f_{ij}-f)_+} - 1)(f_{ij} - f)_+ \\
&\leq \frac{\lambda}{n}(e^{2\lambda} - 1)\,\mathbb{E}\,e^{\lambda f}\sum_{ij}(f_{ij} - f)_+ \\
&\leq 2\lambda(e^{2\lambda} - 1)\,\mathbb{E}f\,\mathbb{E}\,e^{\lambda f} \\
&\leq 4\lambda^2 e^{2\lambda}\,\mathbb{E}f\,\mathbb{E}\,e^{\lambda f}
\end{aligned}$$

对所有 $\lambda \geq 0$成立。因此，利用命题 C.1 及 $a = 4\,\mathbb{E}f$，$b = 2$即可得出结论。 □

最后，为了证明定理 3.1，我们需要以下命题。其证明推迟到第 4节。

命题 3.5. 设 f由 (3.1)给出。若 $a_{ij} \in [-1,1]$ 对所有 i，j，则

$$\forall\, t \geq 0 \qquad \mathbb{P}(f \geq \mathbb{E}f + t) \leq \exp\Big(-\min\Big(\frac{t}{32}, \frac{t^2}{128\,\mathbb{E}\Sigma^2}\Big)\Big),$$

其中 $\Sigma^2 = \sum_k a_{k\sigma_k}^2$ 使得 $\mathbb{E}\Sigma^2 = \frac{1}{n}\sum_{ij} a_{ij}^2$。

定理 *3.1*的证明。对固定的 $t > 0$，令

$$\rho^{-1} = 2\log\big(1 + \frac{t}{16\,\mathbb{E}\Sigma^2}\big)$$

并记

$$f^{\downarrow}(\sigma) = \sum_i a_{i\sigma_i}\mathbf{1}_{\{|a_{i\sigma_i}|\leq\rho\}}$$

and

$$f^{\uparrow}(\sigma)=\sum_i |a_{i\sigma_i}|\mathbf{1}_{\{|a_{i\sigma_i}|>\rho\}}$$

使得 $f\le f^{\downarrow}+f^{\uparrow}$。我们估计 $f^{\downarrow}$和 $f^{\uparrow}$的尾概率。

由估计式 $\log(1+x)\le x$ and 以及 ρ 的定义,

$$\frac{t}{32\rho}\le\frac{t^2}{256\,\mathbb{E}\,\Sigma^2}\le\frac{t^2}{128\,\mathbb{E}\,\Sigma^2},$$

因此，由命题 3.5 应用于 $f^{\downarrow}/\rho$,

$$\begin{aligned}\mathbb{P}(f^{\downarrow}\ge\mathbb{E}\,f^{\downarrow}+t)&\le\exp\Big(-\min\Big(\frac{t}{32\rho},\frac{t^2}{128\,\mathbb{E}\,\Sigma^2}\Big)\Big)\\&=\exp\Big(-\frac{t}{32\rho}\Big)=\exp\Big(-\frac{t}{16}\log\big(1+\frac{t}{16\,\mathbb{E}\,\Sigma^2}\big)\Big).\end{aligned}\tag{3.4}$$

由 $f^{\uparrow}$、ρ 的定义以及估计式 $\log(1+x)\le 2\log(1+\sqrt{x})\le 2\sqrt{x}$,

$$\mathbb{E}\,f^{\uparrow}\le\frac{\mathbb{E}\,\Sigma^2}{\rho}=2(\mathbb{E}\,\Sigma^2)\log\Big(1+\frac{t}{16\,\mathbb{E}\,\Sigma^2}\Big)\le\sqrt{t\,\mathbb{E}\,\Sigma^2},$$

因此，由定理 3.3 应用于 $f^{\uparrow}$,

$$\begin{aligned}\mathbb{P}(f^{\uparrow}\ge\mathbb{E}\,f^{\uparrow}+t)&\le\exp\Big(-\frac{t}{4}\log\Big(1+\frac{t}{4\,\mathbb{E}\,f^{\uparrow}}\Big)\Big)\\&\le\exp\Big(-\frac{t}{4}\log\Big(1+\frac{1}{4}\sqrt{\frac{t}{\mathbb{E}\,\Sigma^2}}\Big)\Big)\\&\le\exp\Big(-\frac{t}{8}\log\Big(1+\frac{t}{16\,\mathbb{E}\,\Sigma^2}\Big)\Big),\end{aligned}\tag{3.5}$$

其中在最后一步中，我们再次使用了估计式 $2\log(1+\sqrt{x})\ge\log(1+x)$。利用假设$\mathbb{E}\,f=0$、三角不等式，并估计 $\log(1+x)\le x$，可得

$$|\mathbb{E}\,f^{\downarrow}|=|\mathbb{E}\,f^{\downarrow}-\mathbb{E}\,f|\le\mathbb{E}\,f^{\uparrow}\le\frac{\mathbb{E}\,\Sigma^2}{\rho}\le\frac{1}{8}t.\tag{3.6}$$

结合（3.4）、（3.5）和（3.6），可得

$$\begin{aligned}\mathbb{P}(f\ge 9t/4)&\le\mathbb{P}(f^{\downarrow}\ge 9t/8)+\mathbb{P}(f^{\uparrow}\ge 9t/8)\\&\le\mathbb{P}(f^{\downarrow}\ge\mathbb{E}\,f^{\downarrow}+t)+\mathbb{P}(f^{\uparrow}\ge\mathbb{E}\,f^{\uparrow}+t)\\&\le 2\exp\Big(-\frac{t}{16}\log\Big(1+\frac{t}{16\,\mathbb{E}\,\Sigma^2}\Big)\Big).\end{aligned}$$

代入 $t\leftarrow 4t/9$ 即可得到结果。 □

4. 命题 2.5 和 3.5 的证明

这两个命题都是下文给出的关于 Hoeffding统计量上确界的一个更一般结果的特例。设 $R\subset\mathbb{R}^{n\times n}$ 为一组实矩阵。记

$$S=\sup_{r\in R}\sum_{k=1}^{n} r_{k\sigma_k}.\tag{4.1}$$

本节的主要结果是以下估计。

命题 4.1。 设 S 由 （4.1） 给出，并假设 $R\subset[-1,1]^{n\times n}$。则

$$\forall\, t\ge 0\qquad \mathbb{P}(S\ge\mathbb{E}\,S+t)\le\exp\Big(-\min\Big(\frac{t}{32},\frac{t^2}{128\,\mathbb{E}\,\Sigma_R^2}\Big)\Big),$$

其中 $\Sigma_R^2=$上确界$_{r\in R}\sum_k r_{k\sigma_k}^2$。

命题 2.5 和 3.5 是命题 4.1 的特例，如下所示。

命题 2.5的证明。 应用命题 4.1，令 $R=\{a^x\colon\ x\in\mathcal{X}\}$（回顾第 a^x 节开头引入的矩阵 2 的定义）。 □

命题的证明 3.5。 应用命题 4.1，其中 $R=\{a\}$。 □

为证明命题4.1，我们先陈述关于 1.1 的修正的对数 Sobolev 不等式，其针对 S 的拉普拉斯变换。对于任意 $i,\ j\in[n]$，记

$$S_{ij}=\sup_{r\in R}\sum_{k=1}^{n} r_{k\sigma_{ij}(k)}.$$

于是，修正的对数 Sobolev 不等式（1.1）意味着

$$\operatorname{Ent}(e^{\lambda S})\le\frac{\lambda}{n}\,\mathbb{E}\Big[e^{\lambda S}\sum_{ij}(1-e^{-\lambda(S-S_{ij})})_+(S-S_{ij})_+\Big],$$

对 $1-e^{-x}\le x$ 进行估计后，可进一步将其特化为

$$\operatorname{Ent}(e^{\lambda S})\le\frac{\lambda}{n}\,\mathbb{E}\Big[e^{\lambda S}\sum_{ij}(S-S_{ij})_+^2\Big]. \tag{4.2}$$

我们还需要以下辅助事实。

引理 4.2. 设 S 由 (4.1) 给出，并假设 $R\subset[0,1]^{n\times n}$。则

$$\forall\,\lambda\in[0,1/4]\qquad \log\mathbb{E}\,e^{\lambda S}\le 2\lambda\,\mathbb{E}\,S.$$

证明。 不失一般性，假设 R 是有限的。设 $\hat{}r$ 是一个取值于 R 的随机矩阵，使得 $S=\sum_{k=1}^{n}\hat{}r_{k\sigma_k}$。我们有

$$\begin{aligned}\sum_{ij}(S-S_{ij})_+^2&\le\sum_{ij}(\hat r_{i\sigma_i}+\hat r_{j\sigma_j}-\hat r_{i\sigma_j}-\hat r_{j\sigma_i})_+^2\\&\le\sum_{ij}(\hat r_{i\sigma_i}+\hat r_{j\sigma_j})^2\le 2n\sum_{i}(\hat r_{i\sigma_i})^2\le 2nS,\end{aligned} \tag{4.3}$$

其中，最后一个不等式使用了 $R\in[0,1]^{n\times n}$。

根据修正的对数 Sobolev 不等式（4.2），结合（4.3），我们得到

$$\operatorname{Ent}(e^{\lambda S})\le\frac{\lambda^2}{n}\,\mathbb{E}\Big[e^{\lambda S}\sum_{ij}(S-S_{ij})_+^2\Big]\le 2\lambda^2\,\mathbb{E}[e^{\lambda S}S]$$

对所有 $\lambda\ge 0$。应用命题 C.2 于 $a=2,\ b=0$ 可得

$$(1-2\lambda)\log\mathbb{E}\,e^{\lambda S}\le\lambda\,\mathbb{E}\,S,$$

对所有 $\lambda\ge 0$ 均成立，由此得出结论。 □

现在可以证明命题 4.1。

命题 *4.1* 的证明。 令 $\hat{}r$ 为一个取值于 R 的随机矩阵，满足 $S=\sum_{k=1}^{n}\hat{}r_{k\sigma_k}$。根据 ℓ^2中的三角不等式，

$$\begin{aligned}\sum_{ij}(S-S_{ij})_+^2&\le\sum_{ij}(\hat r_{i\sigma_i}+\hat r_{j\sigma_j}-\hat r_{i\sigma_j}-\hat r_{j\sigma_i})_+^2\\&\le 8\sum_{ij}\hat r_{i\sigma_i}^2+8\sum_{ij}\hat r_{i\sigma_j}^2\le 8n\Sigma_R^2+8\sum_{ij}\hat r_{i\sigma_j}^2.\end{aligned} \tag{4.4}$$

注意到

$$\sum_{ij}\hat r_{i\sigma_j}^2=\sum_{ij}\hat r_{ij}^2=n\,\mathbb{E}\sum_{i}\hat r_{i\sigma_i}^2\le n\,\mathbb{E}\sup_{r\in R}\sum_{i}r_{i\sigma_i}^2=n\,\mathbb{E}\,\Sigma_R^2,$$

由此（4.4）可进一步特化为

$$\sum_{ij}(S-S_{ij})_+^2\le 8n(\Sigma_R^2+\mathbb{E}\,\Sigma_R^2). \tag{4.5}$$

结合（4.2）中的修正的对数 Sobolev 不等式与（4.5），可得

$$\operatorname{Ent}(e^{\lambda S})\le\frac{\lambda^2}{n}\,\mathbb{E}\Big[e^{\lambda S}\sum_{ij}(S-S_{ij})_+^2\Big]\le 8\lambda^2\big((\mathbb{E}\,e^{\lambda S})(\mathbb{E}\,\Sigma_R^2)+\mathbb{E}[e^{\lambda S}\Sigma_R^2]\big). \tag{4.6}$$

回忆熵的变分公式 $\operatorname{Ent}(h)=\sup\{\mathbb{E}\,hg:\ \mathbb{E}\,e^{g}\le 1\}$，由此可知，对任意 h,g

$$\mathbb{E}\,hg\le\operatorname{Ent}(h)+(\mathbb{E}\,h)\log(\mathbb{E}\,e^{g}). \tag{4.7}$$

先应用（4.7），结合 $h=e^{\lambda S},\ g=\Sigma_R^2/4$，再应用引理 4.2 可得

$$\mathbb{E}\big[e^{\lambda S}\Sigma_R^2\big]\le 4\operatorname{Ent}(e^{\lambda S})+4\big(\mathbb{E}\,e^{\lambda S}\big)\big(\log\mathbb{E}\,e^{\Sigma_R^2/4}\big)\le 4\operatorname{Ent}(e^{\lambda S})+2\big(\mathbb{E}\,e^{\lambda S}\big)\big(\mathbb{E}\,\Sigma_R^2\big),$$

将其与（4.6）结合，得到

$$(1-32\lambda^2)\operatorname{Ent}(e^{\lambda S}) \le 24\lambda^2\big(\mathbb{E}\,{\Sigma_R}^2\big)\big(\mathbb{E}\,e^{\lambda S}\big)$$

对所有 $\lambda \ge 0$，使得

$$\operatorname{Ent}(e^{\lambda S}) \le \frac{192}{7}\lambda^2\big(\mathbb{E}\,{\Sigma_R}^2\big)\big(\mathbb{E}\,e^{\lambda S}\big) \le 32\lambda^2\big(\mathbb{E}\,{\Sigma_R}^2\big)\big(\mathbb{E}\,e^{\lambda S}\big)$$

对所有 $\lambda \in [0, 1/16]$。我们通过应用命题 C.3 以及 $\varepsilon = \frac{1}{16}$ 和 $b = 32\,\mathbb{E}\,\Sigma$ ^2_R. □

5. Acknowledgements

我谨向 Radosław Adamczak 致谢，感谢他认真阅读了本手稿的初稿，并提出了许多显著提升其质量的建议。

参考文献

1. Mélisande Albert, 《随机置换和的集中不等式》, 《高维概率 VIII》（尚）（Nathael Gozlan、Rafał Latała、Karim Lounici 和 Mokshay Madiman 编）, 施普林格国际出版公司, 2019 年, 第 341–383 页。1, 4 2. Bernard Bercu、Bernard Delyon 和 Emmanuel Rio, 《和与鞅的集中不等式》, 《施普林格数学简明读本》, 施普林格, 尚, 2015 年。MR 3363542 1, 4, 5 3. Sergey G. Bobkov 和 Prasad Tetali, 《离散情形中的修正对数 Sobolev 不等式》, 《理论概率论杂志》19 2283379 1 4 2006 2, 289 336 ( ), no. – . MR . E. Bolthausen, 《组合中心极限定理中余项的估计》, 《概率论及其应用》66 (1984 751577 5 5 3, 379 386), no. – . MR . Stéphane Boucheron、Gábor Lugosi 和 Pascal Massart, 《集中不等式：独立性的非渐近理论》, 牛津大学出版社, 牛津, 2013 年。MR 3185193 1, 9 6. Sourav Chatterjee, 《用于集中不等式的 Stein 方法》, 《概率论及相关领域》138 2, 305 2007 1 321 ( ), no. - – . MR 2288072 4, 5 7. Willliam Feller, 《概率论及其应用导论》, 第 2 卷, 5 8 2008 约翰·威立父子出版公司, . . Fuqing Gao 和 Jeremy Quastel, 《随机换位模型和伯努利–拉普拉斯模型中熵的指数衰减》, 《应用概率年刊》13 2023890 1 9 2003 4, 1591 1600 ( ), no. – . MR . David Gross 和 Vincent Nesme, 《关于从有限矩阵集合中无放回抽样的注记》, 2010 年。2 10 44058 5 11 22 1951 558 566 58 1963 13。Wassily Hoeffding, 《组合中心极限定理》, 《数学统计年刊》（）, –。MR ., 《有界随机变量之和的概率不等式》, 《美国统计学会杂志》（）, –30。MR 144363 2 12. Michel Ledoux, 《关于乘积测度的 Talagrand 偏差不等式》, 《ESAIM 概率与统计》1 (1995/97), 63–87。MR 1399224 3 13. Michel Ledoux 和 Michel Talagrand, 《Banach 空间中的概率》, 《数学及其边界领域成果》(3) [《数学及相关领域成果》(3)], 第 23 卷, 施普林格出版社, 柏林, 1991 年, 《等周理论与随机过程》。MR 1102015 2 14. Kyle Luh 和 Nicholas Pippenger, 《无放回抽样的大偏差界》, 《美国数学月刊》121 3193733 2 15 2014 5, 449 454 ( ), no. – . MR . Michel Talagrand, 《乘积空间中的新集中不等式》, 《数学发明》126 (1996), 第 3 期, 505–563。MR 1419006 1, 2, 12 16. Ilya Tolstikhin、Gilles Blanchard 和 Marius Kloft, 《直推式学习的局部复杂度》, 《学习理论会议》, PMLR, 2014 年, 第 857–884 页。1, 2, 3, 11, 12

附录A. 引理2.1的证明

设 $E=\mathbb{R}^n$ 和 $g(i)=e_i$，其中 $e_i\in\mathbb{R}^n$ 是一个向量，其第 i 个坐标为 1，其余坐标为 0。此外，对任意 $v\in\mathbb{R}^n$，令

$$\Psi(v)=\phi\Big(\sup_{x\in\mathcal{X}}\langle x,v\rangle\Big),$$

其中，$\langle\cdot,\cdot\rangle$ 是标准点积。于是，

$$\phi(Z)=\phi\Big(\sup_{x\in\mathcal{X}}\langle x,\sum_{k=1}^m e_{I_k}\rangle\Big)=\Psi\Big(\sum_{k=1}^m g(I_k)\Big)$$

且恒等地有 $\phi(Z')=\Psi(\sum_{k=1}^m g(J_k))$。最后，对于任意 v、$w\in\mathbb{R}^n$ 和 $t\in[0,1]$

$$\begin{aligned}\Psi(tw+(1-t)v)&=\phi\Big(\sup_{x\in\mathcal{X}}\langle x,tw+(1-t)v\rangle\Big)\\&\le\phi\Big(t\sup_{x\in\mathcal{X}}\langle x,w\rangle+(1-t)\sup_{x\in\mathcal{X}}\langle x,v\rangle\Big)\le t\Psi(w)+(1-t)\Psi(v),\end{aligned}$$

其中，在第一个不等式中我们用到了 ϕ 是递增的；在第二个不等式中我们用到了 ϕ 是凸函数。最后，对2.2中那一对(g,Ψ)应用Hoeffding论证。

Appendix B. Proof of Lemma 2.4

我们回顾一些关于熵的事实。对于任意关于 $\sigma(J_1,..,J_m)$ 可测的随机变量 Y，以及任意 $k\in[m]$，令 $\mathbb{E}^{(k)}$ 表示仅关于 J_k 的期望，即:

$$\mathbb{E}^{(k)}[Y]=\mathbb{E}\big[Y\,|\,J_1,\dots,J_{k-1},J_{k+1},\dots,J_m\big].$$

对于这样的正值 Y，回顾熵的张量化公式（参见例如 [5, 定理 4.10]）

$$\mathrm{Ent}(Y)\le\mathbb{E}\sum_{k=1}^m\mathrm{Ent}^{(k)}(Y),\tag{B.1}$$

其中

$$\mathrm{Ent}^{(k)}(Y)=\mathbb{E}^{(k)}\big[Y\log Y\big]-\mathbb{E}^{(k)}\big[Y\big]\log\mathbb{E}^{(k)}\big[Y\big]$$

是与 $\mathbb{E}^{(k)}$ 对应的熵泛函。此外，回顾熵的以下变分公式

$$\mathrm{Ent}(Y)=\inf_{c>0}\mathbb{E}\Big[Y(\log Y-\log c)-(Y-c)\Big].\tag{B.2}$$

引理 2.4 的证明。令 $k\in[m]$，设

$$Z'_k=\sup_{x\in\mathcal{X}}\sum_{l=1,l\neq k}^m x_{J_l}$$

(若 $m=1$，则令 $u_1=0$）。由熵的张量化公式（B.1）以及（B.2），

$$\begin{aligned}\mathrm{Ent}(e^{\lambda Z'})&\le\mathbb{E}\sum_{k=1}^m\mathrm{Ent}^{(k)}(e^{\lambda Z'})\\&=\mathbb{E}\sum_{k=1}^m\inf_{c_k>0}\mathbb{E}^{(k)}\Big[e^{\lambda Z'}(\lambda Z'-\log c_k)-(e^{\lambda Z'}-c_k)\Big]\\&\le\mathbb{E}\sum_{k=1}^m\mathbb{E}^{(k)}\Big[e^{\lambda Z'}(\lambda Z'-\lambda Z'_k)-(e^{\lambda Z'}-e^{\lambda Z'_k})\Big]\\&\le\mathbb{E}\Big[e^{\lambda Z'}\sum_{k=1}^m\phi(-\lambda(Z'-Z'_k))\Big],\end{aligned}\tag{B.3}$$

其中 $\phi(z)=e^z-z-1$。

注意，

$$\sum_{k=1}^m(Z'-Z'_k)\le Z'$$

并且，对于任意的 $z\in[0,1]$ 和 $\lambda\ge1/4$，根据函数 $z\mapsto e^{-z/4}-1$ 的凸性

$$\phi(-\lambda z)=e^{-\lambda z}-1+\lambda z\le e^{-z/4}-1+\lambda z\le-\frac{z}{4}e^{-1/4}+\lambda z\le\Big(\lambda-\frac{1}{8}\Big)z.$$

由于 $\mathcal{X} \subset [0,1]^n$ 根据假设，因此 $0 \le Z' - Z'_k \le 1$ 由此我们可以进一步估计（B.3），对于任意 $\lambda \ge 1/4$ 如下：

$$\operatorname{Ent}(e^{\lambda Z'}) \le \Big(\lambda - \frac{1}{8}\Big)\,\mathbb{E}\Big[e^{\lambda Z'} \sum_{k=1}^{m} (Z' - Z'_k)\Big] \le \Big(\lambda - \frac{1}{8}\Big)\,\mathbb{E}\big[e^{\lambda Z'} Z'\big],$$

经过重新排列后可得

$$\mathbb{E}\big[e^{\lambda Z'} Z'\big] \le 8\,\mathbb{E}\, e^{\lambda Z'} \log \mathbb{E}\, e^{\lambda Z'},$$

这又等价于

$$\frac{d}{d\lambda}\big(\log \mathbb{E}\, e^{\lambda Z'}\big) \le 8 \log \mathbb{E}\, e^{\lambda Z'}$$

对任意 $\lambda \ge 1/4$。对 λ y 积分可得

$$\log \mathbb{E}\, e^{\lambda Z'} \le e^{8\lambda - 2} \log \mathbb{E}\, e^{Z'/4}. \tag{B.4}$$

接下来估计项 log $\mathbb{E}\, e^{Z'/4}$。再次利用 $0 \le Z' - Z'_k \le 1$，可得

$$\sum_{k=1}^{m} (Z' - Z'_k)^2 \le Z'.$$

此外，通过比较导数可得，对任意 $z \ge 0$,

$$\phi(-z) \le \frac{z^2}{2}$$

因此我们还可以进一步估计（B.3）如下

$$\operatorname{Ent}(e^{\lambda Z'}) \le \frac{\lambda^2}{2}\,\mathbb{E}\Big[e^{\lambda Z'} \sum_{k=1}^{m} (Z' - Z'_k)^2\Big] \le \frac{\lambda^2}{2}\,\mathbb{E}\big[e^{\lambda Z'} Z'\big].$$

应用命题 C.2 以及 $a = \frac{1}{2}$ 和 $b = 0$ 可得

$$\forall\, \lambda \ge 0 \qquad \Big(1 - \frac{\lambda}{2}\Big) \text{对数}\ \mathbb{E}\, e^{\lambda Z'} \le \lambda\, \mathbb{E}\, Z'$$

从而

$$\forall\, \lambda \in [0, 1/4] \qquad \log \mathbb{E}\, e^{\lambda Z'} \le \frac{8}{7} \lambda\, \mathbb{E}\, Z',$$

将其与（B.4）结合可得

$$\log \mathbb{E}\, e^{\lambda Z'} \le \frac{2}{7e^2} e^{8\lambda}\, \mathbb{E}\, Z' \le \frac{1}{16} e^{8\lambda}\, \mathbb{E}\, Z'$$

正如所需。 □

Appendix C. Variants of the Herbst argument

在整 个章节中， X 是一个随机变量，其拉普拉斯变换 F 有定义 n $[0,\infty)$。在 这种情况下，回顾一下

$$\operatorname{Ent}(e^{\lambda X}) = \lambda F'(\lambda) - F(\lambda) \log F(\lambda)$$

对所有 $\lambda \ge 0$均成立。下面汇总 Herbst 论证的一些变体。

命题 C.1。若对于任意 $\lambda \ge 0$,

$$\lambda F'(\lambda) - F(\lambda) \log F(\lambda) \le a\lambda^2 e^{b\lambda} F(\lambda) \tag{C.1}$$

对于某个 a， $b > 0$，则

$$\forall\, \lambda \ge 0 \qquad \log \mathbb{E}\, e^{\lambda(X - \mathbb{E}\, X)} \le \frac{a}{b} \lambda (e^{b\lambda} - 1) \tag{C.2}$$

特别是

$$\forall\, t \ge 0 \qquad \mathbb{P}\big(X \ge \mathbb{E}\, X + t\big) \le \exp\Big(-\frac{t}{2b} \log\Big(1 + \frac{b}{2a} t\Big)\Big). \tag{C.3}$$

证明。令 $H(\lambda) = \text{~~对数 } F(\lambda)~~ \lambda$对 $\lambda > 0$。那么， （C.1）推出 $H'(\lambda) \le a e^{b\lambda}$。由于 $H(0^+) = \mathbb{E}\, X$，那么对于任意 $\lambda > 0$,

$$H(\lambda) \le \mathbb{E}\, X + \frac{a}{b}(e^{b\lambda} - 1),$$

这等价于（C.2），因此根据 Chernoff 界

$$\mathbb{P}\big(X \ge \mathbb{E}\, X + t\big) \le \inf_{\lambda > 0} \exp\Big(-\lambda t + \frac{a}{b} \lambda (e^{b\lambda} - 1)\Big)$$

对所有 $t \ge 0$。选择 $\lambda = \frac{1}{b}$对数$(1 + \frac{b}{2a} t)$ 得到 (C.3)。 □

命题 C.2。 假设对所有 $\lambda \ge 0$,

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \le \lambda^2\big(aF'(\lambda) + bF(\lambda)\big) \tag{C.4}$$

对于某个 a, $b \in \mathbb{R}$。于是

$$\forall\, \lambda \ge 0 \qquad (1 - a\lambda)\log \mathbb{E}\, e^{\lambda X} \le \lambda\, \mathbb{E}\, X + b\lambda^2. \tag{C.5}$$

此外，若 $a > 0$ 和 X 不是常数，则 $a\,\mathbb{E}\,X + b > 0$ 并且

$$\forall\, t \ge 0 \qquad \mathbb{P}\big(X \ge \mathbb{E}\,X + t\big) \le \exp\Big(-\min\Big(\frac{t}{4a}, \frac{t^2}{8(a\,\mathbb{E}\,X + b)}\Big)\Big). \tag{C.6}$$

证明。 令 $H(\lambda) = \log F(\lambda) - \lambda$ 对于 $\lambda > 0$。于是，（C.4）蕴含

$$H'(\lambda) \le a\frac{F'(\lambda)}{F(\lambda)} + b = \frac{d}{d\lambda}\big(a\log F(\lambda) + b\lambda\big).$$

因此，对于任意 $\lambda > 0$,

$$H(\lambda) \le H(0^+) + a\log F(\lambda) + b\lambda,$$

这等价于（C.5），因为 $H(0^+) = \mathbb{E}\,X$。从两边减去 $(1 - a\lambda)\lambda\,\mathbb{E}\,X$ 得到

$$(1 - a\lambda)\log \mathbb{E}\, e^{\lambda(X - \mathbb{E}\,X)} \le \lambda^2(a\,\mathbb{E}\,X + b). \tag{C.7}$$

由詹森不等式以及 X 不是常数这一事实可得，对数 $\mathbb{E}\, e^{\lambda(X - \mathbb{E}\,X)} > 0$。若 $\lambda \le 1/2a$，则$1/2 \le 1 - a\lambda$, 因此（C.7）蕴含

$$\forall\, \lambda \in [0, 1/2a] \qquad 0 < \log \mathbb{E}\, e^{\lambda(X - \mathbb{E}\,X)} \le 2\lambda^2(a\,\mathbb{E}\,X + b).$$

因此，根据 Chernoff 界

$$\mathbb{P}\big(X \ge \mathbb{E}\,X + t\big) \le \inf_{0 \le \lambda \le 1/2a} \exp\Big(-\lambda t + 2\lambda^2(a\,\mathbb{E}\,X + b)\Big)$$

对所有 $t \ge 0$。取 $\lambda = t/(4(a\,\mathbb{E}\,X + b))$ 若 $t \le 2(a\,\mathbb{E}\,X + b)/a$ 且 $\lambda = \frac{1}{2a}$ 否则，即得（C.6）。 □

命题 C.3。 假设存在某个 ε, $b > 0$ 使得对所有 $\lambda \in [0, \varepsilon]$,

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \le b\lambda^2 F(\lambda). \tag{C.8}$$

Then

$$\forall\, t \ge 0 \qquad \mathbb{P}\big(X \ge \mathbb{E}\,X + t\big) \le \exp\Big(-\min\Big(\frac{\varepsilon t}{2}, \frac{t^2}{4b}\Big)\Big). \tag{C.9}$$

证明。 将（C.8）除以 $\lambda^2 F(\lambda)$ 并对 λ 积分，得到

$$\frac{\log \mathbb{E}\, e^{\lambda X}}{\lambda} \le \mathbb{E}\,X + \lambda b$$

对所有 $\lambda \in [0, \varepsilon]$。因此，根据 Chernoff 界

$$\mathbb{P}\big(X \ge \mathbb{E}\,X + t\big) \le \inf_{0 \le \lambda \le \varepsilon} \exp\big(-\lambda t + b\lambda^2\big)$$

对所有 $t \ge 0$。取 $\lambda = \frac{t}{2b}$ 若 $t \le 2b\varepsilon$ 且 $\lambda = \varepsilon$ 否则可得（C.9）。 □

Appendix D. Example

本节给出一个示例，说明在非对称集合 [16] 的情形下，我们的定理 $\mathcal{X}$ 的结果如何改进 Tolstikhin–Blanchard–Kloft 2.3 中的界，参见注释 2.3。

对于某些 $k, l \in \mathbb{N}$（稍后确定），使得 $0 < l \le k \le n/2$，令 $A, B \subset [n]$ 为两个基数分别为 k 和 $k/2$的不相交集合，并设

$$\mathcal{X} = \{\, \mathbf{1}_S - \mathbf{1}_B : S \subset A,\ |S| \le l \,\}.$$

对于任意集合 $S \subset [n]$，记

$$R_S = |\{\, j \in [m] : I_j \in S \,\}|, \quad \widetilde{R}_S = |\{\, j \in [m] : J_j \in S \,\}|$$

于是 $Z = \min(R_A, l) - R_B$ 且 $\Sigma^2 = \min(R_A, l) + R_B$。注意，对于任意集合$S \subset [n]$，都有 $\mathbb{E}\, R_S = \mathbb{E}\, \widetilde{R}_S = \frac{m|S|}{n}$，因此

$$\frac{mk}{n} = \mathbb{E}\, R_B \le \mathbb{E}\, \Sigma^2 \le \mathbb{E}\, \widetilde{\Sigma}^2 \le \mathbb{E}\, R_A + \mathbb{E}\, R_B = \frac{3mk}{2n}.$$

此外，令

$$W = |\{\, i \in A : \exists\, j \in [m]\ \ J_j = i \,\}|$$

表示在有放回抽样方案中从集合 A 抽取的元素数量。于是，在集合 $\{W \le l\}$、$Z' = \widetilde{R}_A - \widetilde{R}_B$ 和 $\widetilde{\Sigma}^2 = \widetilde{R}_A + \widetilde{R}_B$ 上成立。任取 $m \simeq \frac{n}{2}$，其中若 $x_n = y_n(1+o(1))$，我们使用记号 $x_n \simeq y_n$。首先证明 $\{W \le l\}$ 以高概率发生。我们有

$$\mathbb{E}\, W = k - k\big(1-\frac{1}{n}\big)^m \simeq k\big(1-e^{-1/2}\big) < \frac{k}{2} \simeq \mathbb{E}\, R_A = \mathbb{E}\, \widetilde{R}_A.$$

因此，根据 Azuma不等式

$$\mathbb{P}\Big(W \le k\big(1-e^{-1/2}\big) + t\Big) \ge 1 - e^{-ct^2/m} \simeq 1 - e^{-2ct^2/n}$$

对于任意 $t \ge 0$，以及某个普适常数 $c > 0$。任取 $l \simeq \frac{k}{2}\big(1-e^{-1/2}+\frac{1}{2}\big)$，使得 $\mathbb{E}\, W \le l \le \mathbb{E}\, R_A$。那么，上述 Azuma不等式意味着，事件 $W \le l$ 发生的概率至少为 $1 - \exp(-2c'\frac{k^2}{n})$，其中 $c' > 0$ 是某个普适常数。再取 $k = \Theta\left(n^{1/2+\varepsilon}\right)$，其中某个 $\varepsilon \in$ 满足 $(0, 0.5]$ （回想一下，我们还假设 $k \le n/2$），使得

$$\mathbb{E}\,|Z'|\mathbf{1}_{\{W>l\}} \le 2m\,\mathbb{P}(W > l) \lesssim 2me^{-2c'n^{2\varepsilon}} = o(1),$$

其中，如果 $x_n \le Cy_n$，则 $x_n \lesssim y_n$，其中 $C > 0$ 是某个普适常数，因此

$$\mathbb{E}\, Z' \simeq \mathbb{E}\, \widetilde{R}_A - \mathbb{E}\, \widetilde{R}_B = \frac{k}{4}.$$

另一方面，

$$\mathbb{E}\, Z \le l - \mathbb{E}\, R_B \simeq \frac{k}{2}\big(1-e^{-1/2}+\frac{1}{2}\big) - \frac{k}{4} = \frac{k}{2}\big(1-e^{-1/2}\big)$$

因此

$$\mathbb{E}\, Z' - \mathbb{E}\, Z \ge \mathbb{E}\, Z' - l + \mathbb{E}\, R_B \simeq \frac{k}{2}(e^{-1/2} - \frac{1}{2}) \ge 0.05k.$$

因此，Tolstikhin–Blanchard–Kloft 得到的界 [16, 定理 2]，

$$\forall\, t \ge 0 \qquad \mathbb{P}(Z \ge \mathbb{E}\, Z' + t) \le \exp\Big(-t\log\big(1+\frac{t}{v}\big) + t - v\log\big(1+\frac{t}{v}\big)\Big),$$

无法提供超过 $\mathbb{E}\, Z + t$ 的偏差估计，对任意参数 $t \in [0, 0.05k]$均如此。另一方面，我们的定理 2.2 给出的界为

$$\forall\, t \ge 0 \qquad \mathbb{P}(Z \ge \mathbb{E}\, Z + t) \le 2\exp\Big(-\frac{t}{C_1}\log\Big(1+\frac{t}{C_2\,\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\Big),$$

当 $t = \alpha k$ 时，回忆 $\mathbb{E}\,\widetilde{\Sigma}^2 = \Theta(k)$，可写为

$$\mathbb{P}(Z \ge \mathbb{E}\, Z + \alpha k) \le 2\exp\big(-c''\alpha k\big),$$

其中 $c'' > 0$ 是某个绝对正常数。最后，我们注意到，后一个不等式也可以由对称群 [15] 上的 Talagrand凸距离不等式得到。

华沙大学数学研究所，波兰电子邮箱：B.Polaczyk@mimuw.edu.pl