arXiv:2601.04384v1 [math.PR] 7 Jan 2026

关于随机排列的反集中

AARON BERGER, ROSS BERKOWITZ, PAT DEVLIN, VAN VU

Abstract. 经典反集中结果主要关注随机和 $S := \sum_{i=1}^n \xi_i v_i$，其中 ξ_i是独立随机变量， v_i 是实数。本文证明了关于随机和 $S := \sum_{i=1}^n w_{\pi_i} v_i$的新集中结果，其中 w_i， v_i 是实数， π 是随机排列。

1. 引言

反集中是概率论中一种广为人知且重要的现象。反集中不等式断言，在某些条件下，随机变量 S 在一个小区域或单点上的概率质量不会太大。我们沿用 Levy [6] 的做法，将实随机变量 Y 的集中函数定义为

$$Q(S,t) = \sup_{|I|=t} \mathbb{P}(S \in I), \tag{1}$$

其中 I 遍历所有长度为 t的闭区间。

包括 Littlewood、Offord、Erd˝os、Levy、Kolmogorov、Ro- gozin 和 Esseen 在内的许多著名数学家，为早期对集中函数和反集中现象 [5, 2, 9, 3, 4, 6, 1]的研究作出了重要贡献。自 21 世纪初以来，反集中研究重新焕发生机，其动机和应用来自多个领域，包括随机矩阵理论、组合学、随机函数和数据科学。人们发现了新的、更强的反集中不等式，这些不等式在解决许多长期悬而未决的问题中发挥着关键作用，例如随机矩阵理论中的圆定律猜想；综述见 [10] 。

在上述大多数研究中，所讨论的随机变量是许多相互独立的原子随机变量的线性组合。Erd˝ os-Littlewood-Offord 不等式是这一领域最早且最著名的结果之一，其形式如下。

定理 1.1 (Erd˝os-Littlewood-Offord)。设 $v \in \mathbb{R}^n$ 满足：对所有 $i \in [n]$， $|v_i| \geq 1$；并设 ξ_i 为独立同分布的 *Rademacher* 变量。那么对于任意区间 I

$$\mathbb{P}(\sum_{i=1}^n \xi_i v_i \in I) = O\left(\frac{|I|+1}{\sqrt{n}}\right).$$

注意，随机变量 $\sum_{i=1}^n \xi_i v_i$ 可以写成随机向量 w 与确定性向量 $v =$的内积：$=(\xi_1$、……、 $\xi_n)$ 和 $v = (v_1$、 ……、 $v_n)$。

在这个定理及许多其他定理中（见 [10]）， w 是一个随机向量，其分布由独立原子随机变量生成的乘积测度决定。在应用中，另一个常用的概率空间是由随机排列生成的空间。本文旨在开启对随机排列空间中反集中结果的研究。

设 $w = (w_1$、……、 $w_n)$ 是一个固定向量， π 是 S_n中的一个排列。我们将通过排列 π置换 w 的坐标后得到的向量记为 $w_\pi(w_{\pi(1)}$、……、 $w_{\pi(n)})$。固定 $v \in \mathbb{R}^n$。我们研究的对象是随机变量 $w_\pi \cdot v = \sum_i w_i v_{\pi(i)}$， 其中 π 从 S_n 中均匀随机选取。

在 Soze 最近的一篇论文中，受一项关于随机多项式的应用启发，研究了 $v = (1$， 2，……， $n)$ 的特殊情形，并证明了

定理 1.2。 设 $v=(1，2，\ldots\ldots，n)$ 且 $w\in\mathbb{R}^n$ 为满足 $w\cdot\mathbf{1}=0$ 的单位向量。那么对于任意$L\in\mathbb{R}$

$$\mathbb{P}\left(|w_\pi\cdot v-Ln|\le 1\right)=O\left(\frac{1}{n}e^{-c|L|}\right),$$

对于某个绝对常数 $c>0$。

为了让读者对假设 $w\cdot\mathbf{1}=0$ 有一些直观认识，我们指出，如果 w 的所有坐标都相同，那么 $w_\pi\cdot v$ 为常数，也就不存在任何非平凡的反集中结论。因此，在涉及随机排列的结论中， $w_\pi\cdot v$ 的反集中性质将取决于 w 和 v 与 $\mathbf{1}$ 平行的程度。

作为定理 1.1 的类似结果，有以下推论。

推论 1.3。 设 $v=(1、2、\ldots、n)$，且 $w\in\mathbb{R}^n$ 是满足 $w\cdot\mathbf{1}=0$ 的单位向量。那么对于任意点 $x\in\mathbb{R}$

$$\mathbb{P}(w_\pi\cdot v=x)=O\left(\frac{1}{n}\right).$$

这里的上界 $O(1/n)$ 是紧的，正如 $w=(1、-1、0、\ldots、0)$ 所示。Huang、McKinnon 和 Satriano [11]，最近研究了这个点质量的上界问题，并在 $n=p$ 或 $n=p+1$ 的情形下（其中 p 是素数）得到了正确的渐近阶和最优界。随后，Pawlowski [12] 对所有 n 得到了最优界，证明在以下假设下，如果 v 的坐标两两不同且 $w\cdot\mathbf{1}\neq 0$，那么对所有 $x\in\mathbb{R}$ 我们有

$$\mathbb{P}[w_\pi\cdot v=x]\le\begin{cases}\frac{1}{n} & \text{若 } n \text{ 为奇数,}\\ \frac{1}{n-1} & \text{若 } n \text{ 为偶数。}\end{cases}$$

在这两种情形下，都存在一些 v， w 的取值，使得等号成立。

我们将推导一个适用于任意向量 v 和 w的上界。我们采用的方法将与 Soze 的方法完全不同；Soze 的方法专门针对特殊情形 $v=(1,2、\ldots、n)$。如果 w 满足 $w_1.\le w_2\le\cdots\le w_n$，我们称其为递增。不失一般性，假设 w 和 v 均为递增，并定义

$$\Delta(v)=\min_{i\neq j}|v_i-v_j|.$$

我们的主要结果如下

定理 1.4 (随机排列的反集中)。 设 w、 $v\in\mathbb{R}^n$ 为递增序列，且$\Delta(v)>0$。对于任意区间 I 以及满足 $i_1+i_2\le n$ 和 $w_{n-i_2}-w_{i_1}>0$的指标 i_1、 i_2，有

$$\mathbb{P}(w_\pi\cdot v\in I)=O\left(1+\frac{|I|}{\Delta(v)(w_{n-i_2}-w_{i_1})}\right)\frac{1}{(i_1+i_2)\sqrt{\min(i_1,i_2)}}.$$

如果能通过某种巧妙的选择使 i_1、 i_2和 $w_{n-i_2}-w_{i_1}$ 同时足够大，那么这就给出了一个很强的上界。当 v_i中存在重复项，或者 $\Delta(v)$较小时，仍然可以通过对一个具有较大 Δ的子集进行条件化来得到上界。例如，（在下文中选择 $i_i=i_2=\varepsilon n/3$）可得到推论1.3的如下改进

推论 1.5。 假设 v的坐标各不相同，且没有任何 w_i重复超过$(1-\epsilon)n$次，其中 $\epsilon>0$。那么对任意 $x\in\mathbb{R}$,

$$\mathbb{P}(w_\pi\cdot v=x)=O\left(\frac{1}{\epsilon\ n^{3/2}}\right).$$

Thro稍加一些额外工作（见第 3 节中的证明），我们还可以得到以下结果

推论 1.6. 设 $w, v \in \mathbb{R}^n$ 满足 $\Delta(v) > 0$。设 $w = n^{-1}\sum w_i$ 且 $\sigma^2 = \sum(w_i - w)^2$。那么对任意区间 I,

$$\mathbb{P}(w_\pi \cdot v \in I) = O\left(\frac{|I|}{n\sigma\Delta(v)}\sqrt{\log n} + \frac{1}{n}\right).$$

定理的证明中使用的主要工具1.4 如下，它们本身也具有独立的研究意义

引理 1.7 （有放回随机子集和）。 设 $A: = \{a_1, \ldots, a_n\}$ 为一个由互不相同的实数组成的集合。令 S 为从 A中有放回地均匀随机选取 k 个元素之和。则对于任意 $x \in \mathbb{R}$,

$$\mathbb{P}(S = x) = O\left(\frac{1}{n\sqrt{k}}\right).$$

引理 1.8 （无放回随机子集和）。 设 $\{a_1, \ldots, a_n\}$ 为一个由互不相同的实数组成的集合。令 S 为从 A中无放回地均匀随机选取 k 个元素之和。则对于任意 $x \in \mathbb{R}$,

$$\mathbb{P}(X = x) = O\left(\frac{1}{n\sqrt{\min(k, n-k)}}\right).$$

备注： 我们最初于 2019–2020 年左右证明了本文的主要结果，将其作为研究随机多项式的工具。然而，我们当时设想的应用最终证明比预期更加技术化，因此当时没有发表我们的研究结果。此后，我们了解到近期进展 [11, 12] ，并决定先行发表本文，因为这些结果本身就具有独立的研究意义。

2. 主要引理的证明

我们将用到 Rogozin [9]的以下结果。设 Y 是一个实随机变量；回顾集中函数的概念

$$Q(Y, t) := \max_y \mathbb{P}(y \le Y \le y + t).$$

定理 2.1 (Rogozin)。 存在一个常数 $C_0 > 0$， 使得以下结论成立。设 X_1，…， X_k是独立随机变量，并令 $X = \sum_I^k X_i$。则

$$Q(X, t) \le C_0(\sum_{i=1}^{k}(1 - Q(X_i, t)))^{-1/2}.$$

有了这些准备，我们可以证明相关引理了。

引理 *1.7*的证明。 不失一般性，我们可以假设 $n \ge 10$。我们对 归纳法进行证明，归纳变量为 k。令 $C = 6(C_0 + 1)$。由于 $\mathbb{P}(X = x) \le n^{-1}$ 对任意 x均成立，因此当 $k \le 4$时，结论显然成立。从现在起，我们可以假设 $k > 4$。

现在令 I 为满足 $|I \cap v| = \lceil n/2 \rceil$ 的最短闭区间，并设 $t = |I|$ （若有多个这样的区间，任取一个）。根据 I 和 t的定义， $Q(X_j, t) \le \lceil n/2 \rceil/n < 2/3$，对所有$1 \le j \le k$均成立。注意到

$$\mathbb{P}(X = x) = \sum_{i=1}^{n}\mathbb{P}(X' = x - a_i \wedge X_k = a_i) = n^{-1}\sum_{i=1}^{n}\mathbb{P}(X' = x - a_i)$$

令 $J := x - I := \{x - y, y \in I\}$. 则

$$\sum_{i=1}^{k}\mathbb{P}(X' = x - a_i) = \mathbb{P}(X' \in J) + \sum_{j, a_j \notin I}\mathbb{P}(X' = x - a_j).$$

根据定理 2.1， $\mathbb{P}(X' \in J) \leq C_0\sqrt{3/(k-1)}$。此外，根据归纳假设

$$\mathbb{P}(X' = x - a_j) \leq C(n\sqrt{k-1})^{-1}.$$

因此，

$$n^{-1}\sum_{j,a_j \notin I} \mathbb{P}(X' = x - a_j) \leq \frac{C}{2}(n\sqrt{k-1})^{-1},$$

根据 $k \geq 5$ 这一事实以及 C 的定义，可得

$$\mathbb{P}(X = x) \leq \frac{1}{n\sqrt{k-1}}(C_0\sqrt{3} + C/2) \leq C\frac{1}{n\sqrt{k}},$$

正如所需。 □

引理 *1.8* 的证明。我们分两种情形讨论。

情形 1（$k \leq n^{0.49}$）：设 $k \leq n^{.49}$.。令 $X_1, \ldots, X_k$ 从 A 中均匀随机选取（允许重复）。令 $\mathcal{E}$ 表示 $X_1, \ldots, X_k$ 的取值互不相同这一事件。由于 $k = o(\sqrt{n})$，根据生日悖论，$\mathbb{P}(\mathcal{E}) = 1 - o(1)$。另一方面，若 $\mathcal{E}$ 发生，则由 $X_1, \ldots, X_k$ 构成的集合在所有大小为 k 的子集中均匀随机。因此

$$\mathbb{P}(S = x) = \mathbb{P}(\sum_{i=1}^{k} X_i = x | \mathcal{E}) = \mathbb{P}(\sum_{i=1}^{k} X_i = x \wedge \mathcal{E})\mathbb{P}(\mathcal{E})^{-1} \leq (1 + o(1))\mathbb{P}(\sum_{i=1}^{k} = x),$$

该结论由引理1.7得证。

情形 2（$k > n^{0.49}$）：现在我们转而考虑 $k > n^{.49}$ 的情形。不失一般性，我们可以假设 $k \leq n/2$。我们使用逆 Littlewood-Offord 理论中的以下结果（关于一般性讨论和详细说明，参见 [10]）。

引理 2.2。（稀疏逆 *Erd˝os-Littlewood-Offord*）[10] 设 $c < 1$ 和 C 为正常数。存在常数 R、A，它们依赖于 c、C，使得以下结论成立。

设 x_i 为独立同分布的 $(0, 1)$ 随机变量，且 $n^{-c} \leq \mathbb{P}(x_i = 1) := \alpha \leq 1/2$。设 b_1、……、b_n 为满足以下条件的整数

$$\rho := \sup_a \mathbb{P}(b_1x_1 + \cdots + b_nx_n = a) \geq n^{-C}.$$

那么，存在一个秩为 $r \leq R$ 的真对称广义等差数列 Q，其中至少包含 v_i 的 $0.9n$（按重数计），并且 $|Q| \leq A\rho^{-1}(\alpha n)^{-r/2} + 1$.

为了使用上述结果，我们令 $b_i = a_i + N$ 其中 N 是一个稍后选取的大整数。为导出矛盾，假设存在某个 x

$$\mathbb{P}(S = x) \geq B\frac{1}{n\sqrt{k}}, \tag{2}$$

，其中 $B > 0$ 是一个待选取的大常数。

定义 x_i 如引理 2.2 中所述，其中 $\mathbb{P}(x_i = 1) = \alpha k/n$。恰有 k 个 x_i 非零的概率至少为 $c_1k^{-1/2}$，其中常数 $c_1 > 0$。因此，(2) 蕴含

$$\rho \geq \mathbb{P}(b_1x_1 + \cdots + b_nx_n = x + Nk) \geq B\frac{1}{n\sqrt{k}}c_1k^{-1/2} = Bc_1(nk)^{-1}.$$

现在我们应用引理 2.2。通过选取 N 足够大（例如 $N \geq 2^n$），很容易看出，不存在长度为 $O(n^2)$ 且至少包含一半 b_i的对称算术级数。因此，本引理中的秩 r 至少为 2。根据引理的结论

$$|Q| \leq AB^{-1}c_1^{-1}nk(\alpha n)^{-1} + 1.$$

令 $B \geq 2Ac_1^{-1}$，并回想 $\alpha = k/n$，我们得到

$$|Q| \leq \frac{Ac_1}{B}n + 1 \leq \frac{n}{2} + 1 < 0.9n,$$

矛盾。 □

3. 定理1.4和推论1.6

定理*1.4*。令 $\beta(n,\ k) := \frac{1}{n\sqrt{\min(k,n-k)}}$。~~回忆一下~~，$w = (w_1,\ \ldots,\ w_n)$ 和 $v = (v_1,\ v_2,\ \ldots,\ v_n)$ 是递增的，且 $\Delta = \min_{i\neq j} |v_i - v_j| \neq 0$。

对于任意 t，令 $A(t) = \{i : w_i < t\}$，$B(t) = \{i : w_i \geq t\}$。对于一个排列 σ，令 $S(\sigma) = \sum_{i=1}^n w_{\sigma(i)}v_i$。固定两个数 $t_1 < t_2$，并令 $A := A(t_1)$ 和 $B := B(t_2)$。然后我们通过以下过程定义一个排列 σ：

(i) 对每个 $i \notin A \cup B$ 选取 $\sigma(i)$；(ii) 对每个 $\{i,j\} \subseteq A$ 和每个 $\{i,j\} \subseteq B$，确定 $\sigma(i)$ 与 $\sigma(j)$ 的相对顺序；(iii) 确定集合 $\{\sigma(i):\ \ i \in A\}$。

注意，如果从 S_n 中均匀随机地选取 σ，那么给定 (i) 和 (ii) 的任意结果，(iii) 中的集合在 $\sigma(A) \cup \sigma(B)$ 的所有 $|A|$ 元子集中均匀分布。

假设 (i) 和 (ii) 的结果已知。令 $X = \sigma(A) \cup \sigma(B)$，对于每个满足 $|U| = |A|$ 的 $U \subseteq X$，令 $S(U)$ 表示总和 $\sum_{i=1}^n w_i v_{\sigma(i)}$，其中 σ 是在步骤 (iii) 中选择 $\sigma(A) = U$ 后得到的排列。

定义 P 上 $|A|$ 元子集的如下偏序：$U \preceq V$ 当且仅当对所有 x，都有 $|U \cap (-\infty, x]| \leq |V \cap (-\infty, x]|$。Stanley [8] 证明了这个偏序集具有 Sperner 性质，这意味着其宽度为

$$width(P) = \max_t \left| \left\{ U \subseteq \{1, 2, \ldots, n\} \ : \ |U| = k, \text{ and } \sum_{u\in U} u = t \right\} \right|.$$

此外，引理 1.8 推出 P 的宽度为 $O(\binom{n}{k}\beta(k,n))$。

接下来，注意到 $U \preceq V$ 当且仅当对所有 i，i^{th} 中最大的元素不小于 U 中最大的元素。因此，若 $U \prec V$，且不存在严格介于 U 和 V 之间的集合 W，那么 U 必定具有形式 $U = \{y\} \cup V \setminus \{x\}$，其中某个 $y > x$。此时，我们有

$$S(V) - S(U) = (v_x - v_y)(w_i - w_j),$$

其中某些元素为 $i \in A$ 和 $j \in B$。由此可知，若 $U \prec V$，则有

$$S(V) - S(U) \geq \Delta(t_2 - t_1) > 0. \tag{3}$$

令 E 为满足 $S(U) \in I$ 的集合族 U。根据狄尔沃斯定理，我们可以将E分解为链 C_1、C_2、.....、C_l，其中 l是 E的宽度。根据（3），我们知道E 中的每条链的大小至多为 $1 + \frac{|I|}{\Delta(t_2-t_1)}$。另一方面，如上所述，宽度至多为$O(\binom{|X|}{|A|})\beta(|A|, |A| + |B|)$。

因此，

$$|E| = O(\left(1 + \frac{|I|}{\Delta(t_2 - t_1)}\right)\binom{|X|}{|A|}\beta(|A|, |A| + |B|))$$

这意味着

$$\mathbb{P}(S(U) \in I) = O(\left(1 + \frac{|I|}{\Delta(t_2 - t_1)}\right)\beta(|A|, |A| + |B|)).$$

通过一个常规的条件化论证，即可证明该定理。 □

推论的证明 *1.6.* 不失一般性，设 $w = 0$，于是 $\sigma^2 = \sum w_i^2$。令 j 为 w中负项对应的最大指标，并不失一般性地假设 $j \leq n/2$。首先，我们可能会猜测选择 $i_1 = k < j + 1 = i_2$。若存在某个指标 $k \leq j$ 使得 $w_k^2 > \frac{\sigma^2}{100k \log n}$，那么根据定理 1.4，此选择可得所述上界。否则，我们有

$$\sum_{k \leq j} w_k^2 \leq \sum_{k \leq j} \frac{\sigma^2}{100k \log n} \leq \frac{\sigma^2}{100}. \tag{4}$$

由柯西-施瓦茨不等式可得 $\left|\sum_{k \leq j} w_k\right| \leq \sigma\sqrt{j/10} \leq \sigma\sqrt{n/10}$。另一方面，由于 $w = 0$，我们有

$$\frac{\sigma\sqrt{n}}{10} \geq \left|\sum_{k \leq j} w_k\right| = \left|\sum_{k > j} w_k\right| \geq \frac{n}{2} w_{n/2}.$$

因此我们看到 $w_{n/2} \leq \sigma/(5\sqrt{n})$。由于 $w_{n/2}$ 非常小，我们接下来猜测取 $i_1 = w_{n/2}$，$i_2 = w_{n-k}$，其中 $k < n/2$。如果 $(w_k - w_{n/2})^2 > \frac{\sigma^2}{10k \log n}$，那么根据定理 1.4，此选择可得出推论。否则，$w_{n-k} \leq \frac{\sigma}{5\sqrt{n}} + \frac{\sigma}{\sqrt{10k \log n}}$，并且

$$\sum_{k \leq n/2} w_{n-k}^2 \leq \sum_{k \leq n/2} \frac{\sigma^2}{10n} + \frac{\sigma^2}{5k \log n} \leq \frac{3\sigma^2}{10}. \tag{5}$$

最后，注意到

$$\sum_{k \in [j, n/2]} w_k^2 \leq \frac{1}{2} \sum_{k \geq j} w_k^2 \leq \frac{\sigma^2}{2}. \tag{6}$$

结合（4）、（5）和（6），我们看到

$$\sigma^2 = \sum_k w_k^2 \leq \frac{\sigma^2}{100} + \frac{\sigma^2}{2} + \frac{3\sigma^2}{10} < \sigma^2,$$

这就产生了矛盾，证明完毕。

□

参考文献

[1] P. Erd˝os, *On a lemma of Littlewood and Offord,* 《美国数学会通报》51 (1945), 898-902. [2] C. G. Ess´een, *On the Kolmogorov-Rogozin inequality for the concentration function,* 《概率论及相关领域》5 (1966), 210-216. [3] A. Kolmogorov, *Two uniform limit theorems for sums of independent random variables,* 《概率论及其应用》1 (1956), 384-394. [4] A. Kolmogorov, *Sur les propriétés des fonctions de concentrations de M. P. Lévy,* Ann. Inst. H. Poincar´e 16 (1958), 27-34. [5] J. E. Littlewood and A. C. Offord, *On the number of real roots of a random algebraic equation.* III. 《数学文集》N.S. 12 , (1943). 277–286. [6] P. Levy, 随机变量加法理论, Gauthier-Villars 1937, 1954. [7] K. S ¨oze, *Real zeroes of random polynomials I. Flip-invariance Tur´an's lemma and the Newton-Hadamard polygon* ,,,,《以色列数学杂志》220 (2017), 817-836. [8] R. Stanley, *Weyl groups, the hard Lefschetz theorem, and the Sperner property,* 《工业与应用数学学会代数与离散数学杂志》, 1 168 184, 1980 : – B. A. Rogozin, *An estimate for concentration functions,* 《概率论及其应用》6 (1961), 94-97. [9]

[10] H. Nguyen and V. Vu, 小球概率、逆定理及其应用, Erd˝os 百年诞辰 Bolya 学会数学研究 25 (2013), 409–463.
[11] J.Huang, D. McKinnon and M. Satriano, *What fraction of an S_n-orbit can lie on a hyperplane ?* 线性代数及其应用, 613 (2021) 1-23. [12] B. Pawlowski, *The fraction of an S_n-orbit on a hyperplane,* 线性代数及其应用 702 (2024), 98-111.