

[AI2613 随机过程][第十一讲] 鞅与集中不等式

张驰豪

最后更新：2025 年 5 月 15 日

目录

1 Hoeffding 不等式	2
2 Azuma-Hoeffding 不等式	3
2.1 McDiarmid 不等式	4
3 Azuma-Hoeffding 不等式的证明	6

1 Hoeffding 不等式

我们在第一课就介绍过如下的 Hoeffding 不等式, 用于约束一系列独立随机变量之和与其期望的距离。

定理 1 (Hoeffding 不等式). 设 $X_1, \dots, X_n$ 为独立随机变量, 每个 $X_i \in [a_i, b_i]$, 其中某些 $a_i \leq b_i$ 以概率 1 成立。令 $X = \sum_{i=1}^n X_i$, 以及 $\mu := \mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}[X_i]$, 则对于任意 $t \geq 0$ 有

$$\mathbb{P}[|X - \mu| \geq t] \leq 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

作为一个简单应用, 我们考虑如下袋中取球问题。一个袋子中有 g 个绿色球和 r 个红色球。这些球除了颜色外完全相同。我们希望通过抽取球来估计红球比例 $\frac{r}{r+g}$ 。

带放回抽取 令 $X_i = \mathbf{1}[\text{第 } i \text{ 次抽到红球}]$, 令 $X = \sum_{i=1}^n X_i$ 。显然, 每个 $X_i \sim \text{Ber}\left(\frac{r}{r+g}\right)$, 并且 $\mathbf{E}[X] = n \cdot \frac{r}{r+g}$ 。由于所有 X_i 是独立的, 我们可以直接应用 Hoeffding 不等式并得到

$$\mathbb{P}[|X - \mathbf{E}[X]| \geq t] \leq 2 \exp\left(-\frac{2t^2}{n}\right).$$

我们这里可以使用 Hoeffding 不等式的关键便是 X_i 之间的独立性。如果我们稍微修改一下设定, 考虑不放回抽取的情形, 那么这种独立性就没有了。

不带放回抽取 同样, 我们令 $Y_i = \mathbf{1}[\text{第 } i \text{ 次抽到红球}]$ 。但与带放回抽取的情形不同, $\{Y_i\}$ 中的变量是相关的。令 $Y = \sum_{i=1}^n Y_i$ 。我们首先计算 $\mathbf{E}[Y]$ 。对于任意 $i \geq 1$, $\mathbf{E}[Y_i]$ 是第 i 次抽到红球的概率。注意到不带放回抽取等价于先对 $r+g$ 个球进行一个均匀排列, 然后按照该顺序逐个抽取。因此, $Y_i = 1$ 的概率为 $\frac{r \cdot (r+g-1)!}{(r+g)!} = \frac{r}{r+g}$ 。因此, 有 $\mathbf{E}[Y] = n \cdot \frac{r}{r+g}$ 。

然而, 由于 $\{Y_i\}$ 是相关的, 我们无法直接应用 Hoeffding 不等式来获得其概率集中的界, 但是直观上, 概率集中现象在这个模型中应该也是存在的。这便促使我们通过去除独立性要求来推广 Hoeffding 不等式。

2 Azuma-Hoeffding 不等式

推广 Hoeffding 不等式的一个方法是把独立随机变量的和推广成条件独立的随机变量的和。等价的，这些条件独立的随机变量的部分和便构成了一个鞅。

定理 2 (Azuma-Hoeffding 不等式). 设 $\{Z_n\}_{n \geq 0}$ 是关于滤链 $\{\mathcal{F}_n\}$ 的一个鞅。如果对于每个 $i \geq 1$ ，有 $|Z_i - Z_{i-1}| \leq c_i$ 以概率 1 成立，则

$$\mathbb{P}[|Z_n - Z_0| \geq t] \leq 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right).$$

该定理的证明见 Section 3。定理 2 中对 $\{Z_n\}$ 是鞅的要求虽然推广了独立随机变量之和的要求，但看起来也不是什么特别容易满足的要求。在我们上面提到的无放回抽球的问题里，这个鞅结构藏在哪儿并不是很显然。这儿，我们要使用之前定义过的 Doob 鞅。

定义 3 (Doob 鞅). 令 $X_1, \dots, X_n$ 为一列 (不要求独立的) 随机变量，且 $f(\bar{X}_{1,n}) = f(X_1, \dots, X_n) \in \mathbb{R}$ 为一个函数。对于 $i \geq 0$ ，令 $Z_i \triangleq \mathbb{E}[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i}]$ 。则我们称 $\{Z_n\}_{n \geq 0}$ 为一个 Doob 鞅。

回顾我们上面讨论的袋中取球问题。定义 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 为 $f(y_1, y_2, \dots, y_n) = \sum_{i=1}^n y_i$ 。那么在不带放回抽取的情形中， $Y = \sum_{i=1}^n Y_i = f(Y_1, Y_2, \dots, Y_n)$ 。现在我们为 f 构造 Doob 鞅。

令 $Z_i = \mathbb{E}[f(\bar{Y}_{1,n}) \mid \bar{Y}_{1,i}]$ 。我们知道 $Z_0 = \mathbb{E}[f(\bar{Y}_{1,n})] = \mathbb{E}[Y] = n \cdot \frac{r}{r+g}$ 且 $Z_n = f(\bar{Y}_{1,n})$ 。为了应用 Azuma-Hoeffding 不等式，我们需要对鞅的“宽度” $|Z_i - Z_{i-1}|$ 进行界定。

根据定义，

$$Z_i - Z_{i-1} = \mathbb{E}[f(\bar{Y}_{1,n}) \mid \bar{Y}_{1,i}] - \mathbb{E}[f(\bar{Y}_{1,n}) \mid \bar{Y}_{1,i-1}].$$

如果我们用 S_i 表示前 i 个球中红球的数量，即 $S_i = \sum_{j=1}^i Y_j$ ，那么

$$\mathbb{E}[f(\bar{Y}_{1,n}) \mid \bar{Y}_{1,i}] = \mathbb{E}[f(\bar{Y}_{1,n}) \mid S_i] = S_i + (n-i) \cdot \frac{r - S_i}{g + r - i}.$$

因此， $S_i = S_{i-1} + Y_i$ ，并且

$$\begin{aligned} Z_i - Z_{i-1} &= \left(S_i + (n-i) \cdot \frac{r - S_i}{g + r - i} \right) - \left(S_{i-1} + (n-i+1) \cdot \frac{r - S_{i-1}}{g + r - i + 1} \right) \\ &= \frac{g + r - n}{g + r - i} \left(Y_i + \frac{S_{i-1} - r}{g + r - i + 1} \right). \end{aligned}$$

注意到 $r \geq S_{i-1}$ 且 $g \geq (i-1) - S_{i-1}$ ，我们有

$$\begin{aligned} Z_i - Z_{i-1} &\leq \frac{g + r - n}{g + r - i} \left(1 + \frac{S_{i-1} - r}{g + r - i + 1} \right) \leq \frac{g + r - n}{g + r - i} \leq 1, \\ Z_i - Z_{i-1} &\geq \frac{g + r - n}{g + r - i} \left(\frac{S_{i-1} - r}{g + r - i + 1} \right) \geq -\frac{g + r - n}{g + r - i} \geq -1. \end{aligned}$$

和 Hoeffding 不等式的条件相比，Azuma-Hoeffding 里对于每一个 $k \in [n]$ ， Z_k 对应于 $\sum_{i=1}^k (X_i - \mathbb{E}[X_i])$ ，即平移成期望为 0 的随机变量之后的部分和。 c_i 就类比为 $b_i - a_i$ 。虽然这里看起来差了两倍 ($|X_i| \leq c_i \iff -c_i \leq X_i \leq c_i$)，但实际上对于期望为 0 的随机变量，Hoeffding 不等式也可以证明一个 c_i 的版本。

实际上，满足一定好的收敛性质的鞅，它一定能够被写成 Doob 鞅的形式。

因此 $-1 \leq X_i \leq 1$ ，我们可以对 $Z_n - Z_0$ 应用 Azuma-Hoeffding 不等式，得到

$$\mathbb{P}[|Y - \mathbb{E}[Y]| \geq t] \leq 2 \exp\left(-\frac{t^2}{2n}\right).$$

2.1 McDiarmid 不等式

敏锐的读者一定注意到了，我们在袋中取球问题中使用的 Doob 鞅 + Azuma-Hoeffding 的方法是一个非常强大且通用的工具，可以对很一般的高维分布上的函数证明集中界。对于由 n 个随机变量 $X_1, \dots, X_n$ 定义的模型以及我们希望估计的量 $f(X_1, \dots, X_n)$ ，我们可以将 Azuma-Hoeffding 不等式应用到 f 的 Doob 鞅上。如前例所示，这个集中界的质量依赖于鞅的宽度，即 $|Z_i - Z_{i-1}|$ 的大小。如果函数 f 和变量 $\{X_i\}_{1 \leq i \leq n}$ 具有某些良好的性质，那么确定每个 $|Z_i - Z_{i-1}|$ 的宽度会比较简单。

定义 4 (c -Lipschitz 函数). 一个函数 $f(x_1, \dots, x_n)$ 满足 c -Lipschitz 条件如果

$$\forall i \in [n], \forall x_1, \dots, x_n, \forall y_i: |f(x_1, \dots, x_i, \dots, x_n) - f(x_1, \dots, y_i, \dots, x_n)| \leq c.$$

McDiarmid 不等式是将 Azuma-Hoeffding 不等式应用于 Lipschitz 函数 f 和独立 $\{X_i\}$ 的结果。

定理 5 (McDiarmid 不等式). 设 f 是一个满足 c -Lipschitz 条件的函数， $X_1, \dots, X_n$ 是 n 个独立随机变量。那么我们有

$$\mathbb{P}[|f(X_1, \dots, X_n) - \mathbb{E}[f(X_1, \dots, X_n)]| \geq t] \leq 2 \exp\left(-\frac{2t^2}{nc^2}\right).$$

证明. 我们用 f 和 $\{X_i\}_{i \geq 1}$ 定义一个 Doob 鞅 $\{Z_i\}_{i \geq 1}$:

$$\forall i: Z_i = \mathbb{E}[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i}].$$

于是

$$Z_i - Z_{i-1} = \mathbb{E}[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i}] - \mathbb{E}[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i-1}].$$

接下来我们尝试确定 $Z_i - Z_{i-1}$ 的宽度。显然

$$Z_i - Z_{i-1} \geq \inf_x \left\{ \mathbb{E}[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i-1}, X_i = x] - \mathbb{E}[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i-1}] \right\},$$

以及

$$Z_i - Z_{i-1} \leq \sup_y \left\{ \mathbb{E}[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i-1}, X_i = y] - \mathbb{E}[f(\bar{X}_{1,n}) \mid \bar{X}_{1,i-1}] \right\}.$$

上下界之间的间隔为

$$\sup_{x,y} \left\{ \mathbf{E} [f(\bar{X}_{1,n}) \mid \bar{X}_{1,i-1}, X_i = y] - \mathbf{E} [f(\bar{X}_{1,n}) \mid \bar{X}_{1,i-1}, X_i = x] \right\}.$$

对于任意 x, y 以及 $\sigma_1, \dots, \sigma_{i-1}$,

$$\begin{aligned} & \mathbf{E} \left[f(\bar{X}_{1,n}) \mid \bigwedge_{1 \leq j \leq i-1} X_j = \sigma_j, X_i = y \right] - \mathbf{E} \left[f(\bar{X}_{1,n}) \mid \bigwedge_{1 \leq j \leq i-1} X_j = \sigma_j, X_i = x \right] \\ &= \sum_{\sigma_{i+1}, \dots, \sigma_n} \left(\mathbb{P} \left[\bigwedge_{i+1 \leq j \leq n} X_j = \sigma_j \mid \bigwedge_{1 \leq j \leq i-1} X_j = \sigma_j, X_i = y \right] \cdot f(\sigma_1, \dots, \sigma_{i-1}, y, \sigma_{i+1}, \dots, \sigma_n) \right. \\ & \quad \left. - \mathbb{P} \left[\bigwedge_{i+1 \leq j \leq n} X_j = \sigma_j \mid \bigwedge_{1 \leq j \leq i-1} X_j = \sigma_j, X_i = x \right] \cdot f(\sigma_1, \dots, \sigma_{i-1}, x, \sigma_{i+1}, \dots, \sigma_n) \right) \\ &\stackrel{(\heartsuit)}{=} \sum_{\sigma_{i+1}, \dots, \sigma_n} \mathbb{P} \left[\bigwedge_{i+1 \leq j \leq n} X_j = \sigma_j \right] \cdot (f(\sigma_1, \dots, \sigma_{i-1}, y, \sigma_{i+1}, \dots, \sigma_n) - f(\sigma_1, \dots, \sigma_{i-1}, x, \sigma_{i+1}, \dots, \sigma_n)) \\ &\stackrel{(\spadesuit)}{\leq} c, \end{aligned}$$

其中 $(\heartsuit)$ 利用了 $\{X_i\}$ 的独立性, $(\spadesuit)$ 利用了 f 的 c -Lipschitz 性质。

应用 Azuma-Hoeffding 不等式, 我们有

$$\mathbb{P} [|Z_n - Z_0| \geq t] = \mathbb{P} [|f(X_1, \dots, X_n) - \mathbf{E} [f(X_1, \dots, X_n)]| \geq t] \leq 2e^{-\frac{t^2}{nc^2}}.$$

□

我们接着介绍 McDiarmid 不等式的两个应用。

示例 6 (模式匹配). 令 $P \in \{0, 1\}^k$ 为一个固定字符串。对于一个随机字符串 $X \in \{0, 1\}^n$, P 在 X 中出现的期望次数是多少?

我们可以利用期望的线性性质轻松计算出出现次数的期望。我们定义 n 个独立随机变量 $X_1, \dots, X_n$, 其中 X_i 表示 X 的第 i 个字符。令 $Y = f(X_1, \dots, X_n)$ 为 P 在 X 中的出现次数。注意到 P 在 X 中最多出现 $n - k + 1$ 次, 我们可以枚举每次出现的第一个位置。通过期望的线性性质, 我们有

$$\mathbf{E} [f] = \frac{n - k + 1}{2^k}.$$

然后我们可以利用 McDiarmid 不等式来证明 f 是良好集中的。为此, 我们注意到 $\{X_i\}$ 中的变量是独立的, 并且函数 f 是 k -Lipschitz 的: 如果我们改变 X 的一个比特, 出现次数最多改变 k 。

因此

$$\mathbb{P} [|Z_n - Z_0| \geq t] = \mathbb{P} [|f - \mathbf{E} [f]| \geq t] \leq 2e^{-\frac{2t^2}{nk^2}}.$$

McDiarmid 不等式的另一个应用是建立 Erdős-Rényi 随机图 $\mathcal{G}(n, p)$ 的染色数的集中性。

示例 7 ($\mathcal{G}(n, p)$ 的染色数). 我们之前使用过 $\mathcal{G}(n, p)$ 表示在所有 n 个顶点的无向简单图上的一个分布。在该模型中，每对可能的顶点间的边以概率 p 独立存在。

对于一个图 $G \sim \mathcal{G}(n, p)$ ，我们用 $\chi(G)$ 表示其染色数，即使用 q 种颜色对 G 进行正确染色的最少颜色数。有不同的方法来用随机变量表示 G 。

最自然的方法是为每对顶点 $e = \{u, v\} \subseteq V$ 引入一个变量 X_e ，其中 $X_e = 1$ [边 e 存在于 G]。那么 $\{X_e\}$ 是独立的，并且染色数可以写成一个函数 $\chi(X_{e_1}, X_{e_2}, \dots, X_{e_{\binom{n}{2}}})$ 。很容易看出 χ 是 1-Lipschitz 的，因为添加或移除一条边最多只会改变染色数 1。因此，根据 McDiarmid 不等式，我们有

$$\mathbb{P}[|\chi - \mathbb{E}[\chi]| \geq t] \leq 2e^{-2t^2 \binom{n}{2}^{-1}}.$$

然而，这个界并不理想，因为我们需要设置 $t = \Theta(n)$ 才能将右侧上界为一个常数。

我们可以用一种更高效的方法对图 G 进行编码，同时保留 Lipschitz 和独立性属性。假设 G 的顶点集合为 $\{v_1, \dots, v_n\}$ 。我们定义 n 个随机变量 $Y_1, \dots, Y_n$ ，其中 Y_i 编码了 v_i 和 $\{v_1, \dots, v_{i-1}\}$ 之间的边。一旦 $Y_1, \dots, Y_n$ 给定，图就完全确定了，因此染色数可以写成一个函数 $\chi(Y_1, \dots, Y_n)$ 。由于 Y_i 仅涉及 v_i 与 $v_1, \dots, v_{i-1}$ 之间的连接，这 n 个变量是独立的。

同样容易看出，如果 Y_i 发生变化，染色数最多变化 1。因此 χ 也是 1-Lipschitz 的。应用 McDiarmid 不等式，我们有

$$\mathbb{P}[|\chi - \mathbb{E}[\chi]| \geq t] \leq 2e^{-\frac{2t^2}{n}}.$$

通过这种方式，我们只需要 $t = \Theta(\sqrt{n})$ 来界定右侧。

3 Azuma-Hoeffding 不等式的证明

我们简要证明一下 Azuma-Hoeffding 不等式，其过程与我们第一节课讲过的 Hoeffding 不等式的证明非常相似。

Azuma-Hoeffding 不等式的证明 (定理 2). 回顾在尝试证明 Hoeffding 不等式时，最困难的部分是估计项

$$\mathbb{E}[e^{\alpha Z_n}] = e^{\alpha Z_0} \cdot \mathbb{E}\left[\prod_{i=1}^n e^{\alpha(Z_i - Z_{i-1})}\right].$$

在 Azuma-Hoeffding 的不等式中，我们可以利用鞅的性质代替独立性来获得这一项的

界。使用条件期望的 chain rule，我们有

$$\begin{aligned} \mathbf{E} \left[\prod_{i=1}^n e^{\alpha(Z_i - Z_{i-1})} \right] &= \mathbf{E} \left[\mathbf{E} \left[\prod_{i=1}^n e^{\alpha(Z_i - Z_{i-1})} \mid \mathcal{F}_{n-1} \right] \right] \\ &= \mathbf{E} \left[\prod_{i=1}^{n-1} e^{\alpha(Z_i - Z_{i-1})} \mathbf{E} \left[e^{\alpha(Z_n - Z_{n-1})} \mid \mathcal{F}_{n-1} \right] \right]. \end{aligned}$$

回忆我们第一次课对于 Hoeffding 引理的证明，我们可以完全一样的得到一个条件期望的版本：

$$\mathbf{E} \left[e^{\alpha(Z_n - Z_{n-1})} \mid \mathcal{F}_{n-1} \right] \leq e^{-\frac{\alpha c_I^2}{8}}.$$

使用这个引理做归纳便能完成证明。

□