[AI2613 随机过程][第一讲] 独立随机变量

张驰豪

最后更新：*2025* 年 *2* 月 *26* 日

目录

随机过程是研究一列随机变量 $X_1, X_2, \ldots$，或者一族随机变量 $\{X(t)\}_{t\geq 0}$ 的学问。我们从独立的随机变量说起。独立的随机变量并不是我们讨论的重点，但我们会引入几个重要的概率模型和数学工具，这个在我们未来的学习中是非常有用的。

1 投球入箱问题（Balls-into-Bins）

投球入箱是一个简单的随机过程：将 m 个球均匀随机地投入 n 个箱子中。我们用 $[m]$ 来编号所有的球，用 $[n]$ 来编号所有的箱子，并用 X_i 来表示第 i 个球落入的箱子的编号。那么，$X_1, X_2, \ldots, X_m$ 就是取值为 $[n]$ 的独立随机变量。假设每一个 X_i 都是均匀的随机变量是最简单的情况。对于这个过程，我们可以提出许多有趣的问题。我们这儿介绍俩经典的。

对于自然数 $n \in \mathbb{N}$，我们用 $[n]$ 表示集合 $\{1, 2, \ldots, n\}$。

1.1 生日悖论（Birthday Paradox）

生日悖论指的是在一个有 30 个同学左右的班级中某些同学很可能共享相同生日这一反直觉现象。将箱子视为日期、球视为学生，两个学生生日相同的事件可以建模为某个箱子包含多于一个球的情况。

注意到每个球的投掷是独立的。在已投掷 $k-1$ 个球且无碰撞的条件下，投掷第 k 个球后仍无碰撞的概率为 $\frac{n-k+1}{n}$。因此，

$$\mathbb{P}\left[\text{无相同生日}\right] = \prod_{k=1}^{m} \frac{n-k+1}{n} = \prod_{k=1}^{m-1} \left(1 - \frac{k}{n}\right).$$

由 $1+x \le e^x$，上面的概率有上界

$$\exp\left(-\frac{\sum_{k=1}^{m-1} k}{n}\right) = \exp\left(-\frac{m(m-1)}{2n}\right).$$

于是，可以看到当 $m = O(\sqrt{n})$ 时，该概率可以任意接近 0。

$1+x \le e^x$ 是一个常用的不等式，它来自于对于 e^x 的线性近似，并且对于任何 $x \in \mathbb{R}$ 均成立。我们有的时候也想用另一方向的不等式，一个实用的并且方便记忆的选择是：

$$\forall k > 0,\ 1 + 1/k \ge e^{1/(k+1)};$$

$$\forall k > 1,\ 1 - 1/k \ge e^{-1/(k-1)}.$$

1.2 奖券收集（Coupon Collector）问题

奖券收集问题是如下问题：如果某品牌麦片每盒包含随机选择的 n 种不同类型奖券之一，需要购买多少盒才能收集全部 n 种奖券？用投球入箱的语言表述，我们用盒子来表示每一种类型的奖券，用每一个球来表示每一包麦片，那么即需要投入多少球才能没有空盒子。

需要购买的盒数是一个随机变量，我们用 Y 来表示。我们首先利用期望的线性性可计算 Y 的期望。设 Y_i 表示在已持有 $i-1$ 种奖券时，收集第 i 种奖券所需的投球次数。那么显然有 $Y = \sum_{i=1}^{n} Y_i$。由期望线性性：

$$\mathbf{E}\left[Y\right] = \sum_{i=1}^{n} \mathbf{E}\left[Y_i\right].$$

显然 $Y_i \sim \mathsf{Geom}(\frac{n-i+1}{n})$，因此 $\mathbf{E}\left[Y_i\right] = \frac{n}{n-i+1}$。从而

$$\mathbf{E}\left[Y\right] = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \cdot H_n.$$

对于几何分布 $X \sim \mathsf{Geom}(p)$，有 $\mathbf{E}\left[X\right] = p^{-1}$。

这儿 $H_n = \sum_{i=1}^{n} \frac{1}{i}$ 为调和级数的前 n 项，满足 $H_n \sim \log n + \gamma$，其中常数 $\gamma = 0.577\ldots$ 被称为欧拉常数。

上述使用期望的线性性进行计算的技巧是非常常用的。此外，奖券收集问题的结论也值得记住，在我们后续课程中会多次的遇到它。

2 集中不等式（Concentration Inequalities）

在奖券收集问题里，从实际的角度来说，我们知道平均 $n \cdot H_n$ 包可以收集全一套奖券是不够的，因为实际开包的过程可能和所谓的“期望值”相去甚远。我们需要知道关于随机变量 Y 更进一步的信息，比如，如果我想保证以至少 99% 的概率收集全所有奖券，我需要开多少包？这就涉及到随机变量的集中不等式。

2.1 马尔可夫不等式（Markov's inequality）

马尔可夫不等式可能是最简单集中不等式，它对于任意非负的随机变量均成立。

定理 1 (马尔可夫不等式). *对任意非负随机变量 $X \geq 0$ 及 $a > 0$，*

$$\mathbb{P}[X \geq a] \leq \frac{\mathbf{E}[X]}{a}.$$

证明. 由于 X 非负，由全期望公式，有 $\mathbf{E}[X] \geq a \cdot \mathbb{P}[X \geq a] + 0 \cdot \mathbb{P}[X < a]$。这便等价于我们想要证明的。 □

全期望公式指的是对于样本空间的一个可数划分 $\{A_i\}$，其中每一个 A_i 是一个事件，有 $\mathbf{E}[X] = \sum_i \mathbf{E}[X \mid A_i] \cdot \mathbb{P}[A_i]$。

示例 2 (奖券收集的集中性). *回忆 Y 为所需球数。应用马尔可夫不等式，对 $c > 0$ 有*

$$\mathbb{P}[X \geq c] \leq \frac{\mathbf{E}[X]}{c} = \frac{n \cdot H_n}{c}.$$

因此，需要开超过 $100 \cdot n \cdot H_n$ 包的概率小于 0.01。

2.2 切比雪夫不等式（Chebyshev's inequality）

马尔可夫不等式的成立仅仅要求随机变量 X 是非负的，但实际上，如果我们考虑 X 和它的期望的偏差程度，即 $\mathbb{P}[|X - \mathbf{E}[X]| \geq a]$，那么马尔可夫不等式总能够给出一个上界，即 $\mathbb{P}[|X - \mathbf{E}[X]| \geq a] \leq \frac{\mathbf{E}[|X - \mathbf{E}[X]|]}{a}$。因此，马尔可夫不等式的成立可以看成是无条件的，但这也说明了它在很多随机变量上并不紧，因为它没有利用足够多的随机变量的信息。

当然，我们还要求 X 是可积的。不过在这门课中大部分时候，我们关心的随机变量总是可积的。

我们用一个简单的技巧即可改进它。对于一个（在我们关心的范围内的）递增的函数 $f: \mathbb{R} \to \mathbb{R}$，我们有 $\mathbb{P}[X \geq a] = \mathbb{P}[f(X) \geq f(a)]$。我们便可以对后者使用马尔可夫不等式。

定理 3 (切比雪夫不等式). *对任意具有有限期望 $\mathbf{E}[X]$ 的随机变量及 $a > 0$，有*

$$\mathbb{P}[|X - \mathbf{E}[X]| \geq a] \leq \frac{\mathbf{Var}[X]}{a^2}.$$

证明. 令 $Y = |X - \mathbf{E}[X]|$，显然 $Y \geq 0$。因此

$$\begin{aligned}\mathbb{P}[|X - \mathbf{E}[X]| \geq a] &= \mathbb{P}[Y \geq a] \\ &= \mathbb{P}\left[Y^2 \geq a^2\right] \\ &\leq \frac{\mathbf{E}\left[Y^2\right]}{a^2} \qquad \text{(马尔可夫不等式)} \\ &= \frac{\mathbf{Var}[X]}{a^2}.\end{aligned}$$

□

示例 4 (再探优惠券收集). 将切比雪夫不等式应用于优惠券收集问题。沿用之前记号，有

$$\mathbb{P}[Y \geq nH_n + t] \leq \mathbb{P}[|Y - \mathbf{E}[Y]| \geq t] \leq \frac{\mathbf{Var}[Y]}{t^2}.$$

我们之前使用 Y_i 表示已有 $i-1$ 种奖券时获得新奖券所需的开包次数。对不同的 i, j，Y_i 与 Y_j 独立。因此

对于两两独立的随机变量，方差和求和可以交换。

$$\mathbf{Var}[Y] = \mathbf{Var}\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} \mathbf{Var}[Y_i].$$

对 $i \in [n]$，$Y_i \sim \mathsf{Geom}\left(\frac{n-i+1}{n}\right)$，故

$$\mathbf{Var}[Y_i] = \frac{1 - \frac{n-i+1}{n}}{\left(\frac{n-i+1}{n}\right)^2} = \frac{(i-1)\cdot n}{(n-i+1)^2} \leq \frac{n^2}{(n-i+1)^2}.$$

因此，我们只需要控制住 $\sum_{i=1}^{n} \frac{1}{(n-i+1)^2} = \sum_{i=1}^{n} \frac{1}{i^2}$。注意到

在这个问题上，通过切比雪夫不等式得到的界比马尔可夫不等式紧得多——为达到相同置信度，马尔可夫不等式说明需选择 $t = \Theta(n \log n)$。

$$\sum_{i=1}^{n} \frac{1}{i^2} \leq 1 + \int_1^{\infty} \frac{\mathrm{d}x}{x^2} = 2$$

成立。因此 $\mathbf{Var}[X] \leq 2n^2$ 且 $\mathbb{P}[X \geq nH_n + t] \leq \frac{2n^2}{t^2}$。切比雪夫不等式告诉我们，需要抽取超过 $\sqrt{200}n + nH_n$ 次的概率小于 0.01。相比马尔可夫不等式，这距离事实要接近很多。

我们发现，在切比雪夫不等式的证明中，我们通过控制 $\mathbf{E}[f(|X - \mathbf{E}[X]|)]$，其中 $f(x) = x^2$，引入了方差。一个很自然的问题是，如果我们选择别的 $f(x)$，能否得到关于“尾分布”$\mathbb{P}[|X - \mathbf{E}[X]| \geq t]$ 的更好的不等式？切比雪夫不等式仅仅用到了随机变量二阶矩的信息。我们首先要来看一下，在什么情况下，它对于尾分布得到的界并不紧。我们考察熟悉的高斯分布 $X \sim \mathcal{N}(0,1)$。使用切比雪夫不等式，我们得到

$$\mathbb{P}[X \geq t] \leq \frac{\mathbf{Var}[X]}{t^2} = \frac{1}{t^2}.$$

而事实上，我们可以使用高斯分布的概率密度函数直接计算出这个尾分布概率：

标准高斯分布的概率密度函数 $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$。

$$\mathbb{P}[X \geq t] = \int_t^\infty \phi(x)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-\frac{x^2}{2}}\,\mathrm{d}x.$$

注意到

$$\int_t^\infty e^{-\frac{x^2}{2}}\,\mathrm{d}x \leq \int_t^\infty \frac{x}{t} \cdot e^{-\frac{x^2}{2}}\,\mathrm{d}x = t^{-1}e^{-t^2/2}.$$

我们有 $\mathbb{P}[X \geq t] \leq \frac{1}{\sqrt{2\pi}t} e^{-\frac{t^2}{2}}$. 这说明，实际上，高斯的尾分布是关于 t^2 指数下降的，对于高斯分布来说，切比雪夫不等式得到的界离真相还差距很远。

如果一个分布它长得有点接近高斯，那我们期待能够得到比切比雪夫不等式更加紧的集中不等式。什么样的分布会接近高斯呢？在概率论课中，我们学习过中心极限定理，我们知道，在一定条件下，很多独立的随机变量之和的分布会收敛到高斯分布。我们接下来两节就讨论这样的随机变量的集中不等式。

2.3 切尔诺夫界（Chernoff bound）

我们接着来考虑一系列独立随机变量的和。中心极限定理说，独立随机变量的和的分布（在一定条件下）会收敛到高斯分布，因此，我们接下来的一系列结论可以看成是“非渐进”版本的中心极限定理，也就是当随机变量的个数 n 是固定的，而不一定趋向于无穷时候的“中心极限定理”。最简单的，就是当每个随机变量是伯努利变量的时候，也对应了棣莫弗-拉普拉斯（De Moivre–Laplace）中心极限定理。我们把这个结论称作切尔诺夫界。

棣莫弗-拉普拉斯中心极限定理说的是对于一列 $X_1, X_2, \ldots$，如果每一个 $X_i \sim \mathsf{Ber}(p)$，并且是相互独立的，那么

$$\frac{S_n - np}{\sqrt{np}} \xrightarrow{D} Y \sim \mathcal{N}(0,1).$$

当然，我们的切尔诺夫界允许每一个 X_i 具有不同的均值 p_i。

定理 5 (切尔诺夫界). *设 $X_1, \ldots, X_n$ 为独立随机变量，其中每个 $X_i \sim \mathsf{Ber}(p_i)$ $(i = 1, 2, \ldots, n)$。令 $X = \sum_{i=1}^n X_i$，并记 $\mu := \mathbf{E}[X] = \sum_{i=1}^n p_i$，则有*

- *对于任意 $\varepsilon > 0$，*
$$\mathbb{P}[X \geq (1+\varepsilon)\mu] \leq \left(\frac{e^\varepsilon}{(1+\varepsilon)^{1+\varepsilon}}\right)^\mu.$$
- *对于 $0 < \varepsilon < 1$，有*
$$\mathbb{P}[X \leq (1-\varepsilon)\mu] \leq \left(\frac{e^{-\varepsilon}}{(1-\varepsilon)^{1-\varepsilon}}\right)^\mu.$$

证明. 我们这儿仅证明第一条（即“上尾界（upper tail bound）”），对于下尾界的证明类似。对任意 $\alpha > 0$，有

$$\mathbb{P}[X \geq (1+\varepsilon)\mu] = \mathbb{P}\left[e^{\alpha X} \geq e^{\alpha(1+\varepsilon)\mu}\right] \leq \frac{\mathbf{E}\left[e^{\alpha X}\right]}{e^{\alpha(1+\varepsilon)\mu}}$$

因此需要估计矩生成函数 $\mathbf{E}\left[e^{\alpha X}\right]$。由于 $X=\sum_{i=1}^{n} X_i$ 是独立伯努利变量之和，可得

如果 X 和 Y 独立，那么 $\mathbf{E}[XY]=\mathbf{E}[X]\,\mathbf{E}[Y]$。

$$\mathbf{E}\left[e^{\alpha X}\right]=\mathbf{E}\left[e^{\alpha\sum_{i=1}^{n} X_i}\right]=\mathbf{E}\left[\prod_{i=1}^{n} e^{\alpha X_i}\right]=\prod_{i=1}^{n}\mathbf{E}\left[e^{\alpha X_i}\right].$$

对 $X_i\sim \mathtt{Ber}(p_i)$，直接计算得：

这里又用了 $1+x\le e^x$。

$$\mathbf{E}\left[e^{\alpha X_i}\right]=p_i e^{\alpha}+(1-p_i)=1+(e^{\alpha}-1)p_i\le \exp\left((e^{\alpha}-1)p_i\right)$$

因此，

$$\mathbf{E}\left[e^{\alpha X}\right]\le\prod_{i=1}^{n}\exp\left((e^{\alpha}-1)p_i\right)=\exp\left((e^{\alpha}-1)\mu\right).$$

从而

$$\mathbb{P}\left[X\ge(1+\varepsilon)\mu\right]\le\left(\frac{\exp\left(e^{\alpha}-1\right)}{\exp\left(\alpha(1+\varepsilon)\right)}\right)^{\mu}$$

此式对任意 $\alpha>0$ 成立。选择 α 使分数最小化，令导数为零：

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\left(\frac{\exp\left(e^{\alpha}-1\right)}{\exp\left(\alpha(1+\varepsilon)\right)}\right)=\exp\left(e^{\alpha}-1-\alpha-\alpha\varepsilon\right)\cdot\left(e^{\alpha}-1-\varepsilon\right)=0.$$

解得 $\alpha=\log(1+\varepsilon)$，代入得：

在本讲义中我们用 log 表示自然对数。

$$\mathbb{P}\left[X\ge(1+\varepsilon)\mu\right]\le\left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{\mu}.$$

□

注意到在上述证明中，我们依旧类似切比雪夫不等式的证明，先使用 $\mathbb{P}[X\ge a]=\mathbb{P}[f(X)\ge f(a)]$，然后使用马尔科夫不等式。我们这儿选取的 $f(x)=e^{\alpha\cdot x}$ 是指数函数，因此 $\mathbf{E}[f(X)]=\mathbf{E}\left[e^{\alpha X}\right]$ 是随机变量 X 的矩生成函数。在一定条件下，矩生成函数唯一确定了随机变量。它可以看成是 X 的所有的 k 阶矩 $\mathbf{E}\left[X^k\right]$ 的加权和，而权重由参数 α 决定。事实上，我们可以选择一个合适的 k，取 $f(x)=x^k$，然后得到类似的结果。但是从证明中可以看到，选用矩生成函数在计算上十分方便。

事实上，定理 5里面给出来的界形式上用起来并不是很方便，也不是很好看出为什么它具有类似于高斯尾分布的下降速度。我们经常使用下面这个推论里给出的界。

推论 6. *对任意 $0<\varepsilon<1$，有更易用（但稍弱）的形式：*

$$\mathbb{P}\left[X\ge(1+\varepsilon)\mu\right]\le\exp\left(-\frac{\varepsilon^2}{3}\mu\right)$$

$$\mathbb{P}\left[X\le(1-\varepsilon)\mu\right]\le\exp\left(-\frac{\varepsilon^2}{2}\mu\right)$$

证明. 仅证上尾。需验证当 $0 < \varepsilon < 1$ 时：

$$\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}} \leq \exp\left(-\frac{\varepsilon^2}{3}\right)$$

取对数等价于：

$$\varepsilon - (1+\varepsilon)\log(1+\varepsilon) \leq -\frac{\varepsilon^2}{3}$$

令 $f(\varepsilon) = \varepsilon - (1+\varepsilon)\log(1+\varepsilon) + \frac{\varepsilon^2}{3}$，其导数满足：

$$f'(\varepsilon) = -\log(1+\varepsilon) + \frac{2}{3}\varepsilon, \quad f''(\varepsilon) = -\frac{1}{1+\varepsilon} + \frac{2}{3}$$

分析凹凸性可知 $f(\varepsilon) \leq 0$，故原式成立。 □

示例 7 (公平硬币检验). 现有一枚硬币，已知其要么为公平硬币（正面概率 *0.5*），要么为有偏硬币（正面概率 $0.5+\varepsilon$）。通过抛掷 T 次后计算正面比例 $\hat{p}$ 进行判断是哪种：若 $\hat{p} \leq 0.5 + \varepsilon/2$ 则判为公平，否则判为有偏。现在请问需要投掷多少次硬币，能够保证以至少 $1-\delta$ 的概率正确判断。

我们不妨假设硬币本身是公平的，即 $X_i \sim \mathsf{Ber}(1/2)$。硬币本身是有偏的场合可以类似说明。那么，总正面数 $X = \sum_{i=1}^{T} X_i$，则误判概率为：

$$\mathbb{P}\left[\hat{p} > 0.5 + \varepsilon/2\right] = \mathbb{P}\left[X > (1+\varepsilon)\mathbf{E}\left[X\right]\right]$$

由切尔诺夫界，当

$$\exp\left(-\frac{\varepsilon^2}{3} \cdot \frac{T}{2}\right) \leq \delta$$

时，解得 $T \geq \frac{6}{\varepsilon^2}\log\frac{1}{\delta}$。事实上在，在忽略常数倍的意义下，这个界是最优的。

我们将在作业里证明 $\Omega\left(\frac{1}{\varepsilon^2}\log\frac{1}{\delta}\right)$ 次是必要的！

2.4 霍夫丁不等式（Hoeffding's inequality）

切尔诺夫界的一个恼人限制是要求每个 X_i 必须是伯努利随机变量。实际上，正如同中心极限定理一样，我们可以大大放松这个要求。正确刻画这类具有类高斯尾分布的概念叫做次高斯分布（sub-Gaussian distribution）。我们这里不详细讨论相关理论，但我们介绍一个用起来比较方便的形式，也是在本门课中最常用的形式，称作霍夫丁（Hoeffding）不等式。

定理 8 (霍夫丁不等式). 设 $X_1, \ldots, X_n$ 为独立随机变量，其中每个 X_i 以概率 1 落在区间 $[a_i, b_i]$ 内（$a_i \leq b_i$）。令 $X = \sum_{i=1}^{n} X_i$ 则对所有 $t \geq 0$ 有

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

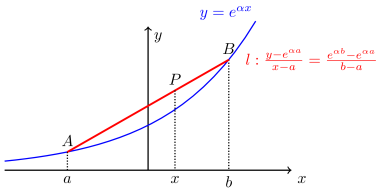
成立。

类似切尔诺夫界的证明，建立此类集中不等式的关键在于对矩生成函数进行有效上界估计。因此，下面的霍夫丁引理将成为证明的主要技术工具。

引理 9 (霍夫丁引理). *设 X 为满足 $\mathbf{E}[X]=0$ 且 $X\in[a,b]$ 的随机变量，则对任意 $\alpha\in\mathbb{R}$ 有*

$$\mathbf{E}\left[e^{\alpha X}\right]\le\exp\left(\frac{\alpha^2(b-a)^2}{8}\right)$$

霍夫丁引理的证明. 我们首先寻找一个线性函数作为 $e^{\alpha x}$ 的上界，以便利用期望的线性性来估计 $\mathbf{E}\left[e^{\alpha X}\right]$。根据指数函数的凸性，对于任何 $x\in[a,b]$，有



$$e^{\alpha x}\le\frac{e^{\alpha b}-e^{\alpha a}}{b-a}(x-a)+e^{\alpha a}.$$

因此，根据期望的线性性，

$$\begin{aligned}
\mathbf{E}\left[e^{\alpha X}\right] &\le \frac{e^{\alpha b}-e^{\alpha a}}{b-a}(\mathbf{E}[X]-a)+e^{\alpha a}\\
(\mathbf{E}[X]=0)\qquad &= \frac{-a}{b-a}e^{\alpha b}+\frac{b}{b-a}e^{\alpha a}\\
&= e^{\alpha a}\left(\frac{b}{b-a}-\frac{a}{b-a}e^{\alpha(b-a)}\right)\\
(\theta=-\tfrac{a}{b-a},\,t=\alpha(b-a))\qquad &= e^{-\theta t}(1-\theta+\theta e^{t})\\
&=: e^{g(t)}.
\end{aligned}$$

其中 $g(t)=-\theta t+\log(1-\theta+\theta e^t)$。根据泰勒定理，对任意实数 t 存在 $\delta\in(0,t)$ 使得

$$g(t)=g(0)+tg'(0)+\frac{1}{2}g''(\delta)t^2.$$

计算各阶导数：

$$\begin{aligned}
&g(0)=0;\\
&g'(0)=-\theta+\left.\frac{\theta e^t}{1-\theta+\theta e^t}\right|_{t=0}=0;\\
&g''(t)=\frac{(1-\theta)\theta e^t}{(1-\theta+\theta e^t)^2}\le\frac{1}{4}.
\end{aligned}$$

因此 $g(t)\le\frac{1}{8}t^2=\frac{1}{8}\alpha^2(b-a)^2$。即得 $\mathbf{E}\left[e^{\alpha X}\right]\le\exp\left(\frac{\alpha^2(b-a)^2}{8}\right)$。 □

借助霍夫丁引理，霍夫丁不等式便很容易证明了。

霍夫丁不等式的证明. 不妨设 $\mathbb{E}[X_i] = 0$（否则用 $X_i - \mathbb{E}[X_i]$ 代替）。由对称性只需证明 $\mathbb{P}[X \geq t] \leq \exp\left(-\frac{2t^2}{\sum(b_i-a_i)^2}\right)$。考虑

$$\mathbb{P}[X \geq t] \leq \frac{\mathbf{E}\left[e^{\alpha X}\right]}{e^{\alpha t}} = \prod_{i=1}^{n} \frac{\mathbf{E}\left[e^{\alpha X_i}\right]}{e^{\alpha t/n}}$$

应用霍夫丁引理并取 $\alpha = \frac{4t}{\sum(b_i-a_i)^2}$ 即得结论。 □

3 多臂老虎机问题（Multi-Armed Bandit）

在本节中，我们讨论在线优化中的一个经典模型：多臂老虎机问题（Multi-Armed Bandit Problem）。我们通过这个问题展示一下集中不等式在算法设计和分析中的作用。假设有一个 n 臂老虎机，每一个臂拉一下会给出一个位于 $[0,1]$ 范围内的随机奖励值。为了方便，我们假设第 i 个臂返回的奖励服从分布 $\texttt{Ber}(\mu_i)$。我们进一步假设 $\mu_1 > \mu_2 \geq \cdots \geq \mu_n$ 成立。我们现在的可以拉动老虎机 T 轮（每一轮选一个臂），目标是使奖励期望最大化。如果我们已知 $\mu_1, \ldots, \mu_k$，最优策略是始终选择臂 1，此时期望奖励为 $T\mu_1$。然而，当我们不知道每个臂的奖励分布及顺序时，就需要设计一种策略来首先探索老虎机。

记 a_t 为第 t 轮拉动的臂，因此第 t 轮的奖励为 $X_t \sim \texttt{Ber}(\mu_{a_t})$。一种衡量策略的好坏的方式是使用**遗憾值**（*regret*）。遗憾值定义为总轮次 T 中始终选择最佳臂 1 所能获得的期望奖励与策略实际期望奖励之间的差值，即未总是选择最佳臂所带来的代价：

$$R(T) := T\mu_1 - \mathbf{E}\left[\sum_{t=1}^{T} X_t\right] \geq 0.$$

在上述表达式中，期望 $\mathbf{E}[\cdot]$ 的随机性通常来源于两个方面：奖励分布 $\texttt{Ber}(\mu_i)$ 的随机性以及策略本身的随机性。对于每个 $i \in [n]$，我们定义 $\Delta_i := \mu_1 - \mu_i \geq 0$，表示第 i 臂的奖励均值与最优臂的奖励均值之间的差距。如果我们的策略是平均的拉每一个臂，那么遗憾值为 $R(T) = \frac{\sum_{i=1}^{n} \Delta_i}{n} \cdot T$。我们可以选择 Δ_i 使得这是一个关于 T 线性增长的，因此，不是一个好的策略。接下来，我们介绍一种名为“探索然后确定”（Explore-then-Commit, ETC）的算法，该算法可以实现次线性遗憾。

3.1 探索然后确定（Explore-then-Commit，ETC）算法

为了使遗憾最小化，策略应尽快识别出最佳臂。最直接的方法是给每个臂尝试一定次数，然后选择奖励的**经验均值**（empirical mean）最大的臂。ETC 算法便是实现了这一思想：每个臂 i 被拉动 L 次（总共进行 nL 次探索），并计算 $\hat{\mu}_i$（在 L 次尝试中获得的平均奖励）。

之后，始终选择 $\hat{\mu}_i$ 最大的臂。其遗憾可以写为：

$$\begin{aligned} R(T) &= L\sum_{i=1}^{n}\Delta_i + \sum_{i=2}^{n}\Delta_i \cdot \sum_{t=nL+1}^{T}\mathbb{P}\left[\hat{\mu}_i > \max_{j\neq i}\hat{\mu}_j\right] \\ &= L\sum_{i=1}^{n}\Delta_i + \sum_{i=2}^{n}\Delta_i \cdot (T-nL)\mathbb{P}\left[\hat{\mu}_i > \max_{j\neq i}\hat{\mu}_j\right]. \end{aligned}$$

当 $i \neq 1$ 时：

$$\mathbb{P}\left[\hat{\mu}_i > \max_{j\neq i}\hat{\mu}_j\right] \leq \mathbb{P}\left[\hat{\mu}_i > \hat{\mu}_1\right].$$

我们用集中不等式来界定上述概率。为此，令 X_j 为探索阶段中第 j 次拉动臂 i 时的奖励，Y_j 为探索阶段中第 j 次拉动臂 1 时的奖励。令 $Z_j = X_j - Y_j \in [-1,1]$，则 $\mathbf{E}\left[Z_j\right] = -\Delta_i \leq 0$。令 $Z = \sum_{j=1}^{L} Z_j$，则 $\mathbf{E}\left[Z\right] = -L\Delta_i$。

根据 Hoeffding 不等式：

$$\mathbb{P}\left[\hat{\mu}_i > \hat{\mu}_1\right] = \mathbb{P}\left[Z > 0\right] = \mathbb{P}\left[Z - \mathbf{E}\left[Z\right] > L\Delta_i\right] \leq \exp\left(-\frac{2(L\Delta_i)^2}{\sum_{j=1}^{L} 2^2}\right) = \exp\left(-\frac{L\Delta_i^2}{2}\right).$$

因此有：

$$\begin{aligned} R(T) &\leq L\sum_{i=1}^{n}\Delta_i + (T-kL)\sum_{i=2}^{n}\Delta_i \exp\left(-\frac{L\Delta_i^2}{2}\right) \\ &\overset{(\Delta_i \leq 1)}{\leq} \sum_{i=1}^{n}\left(L\Delta_i + T\Delta_i \exp\left(-\frac{L\Delta_i^2}{2}\right)\right). \end{aligned}$$

接下来定义：

$$g(L,\Delta_i) := L + T\Delta_i \exp\left(-\frac{L\Delta_i^2}{2}\right).$$

我们希望找到 L 使得 $R(T)$ 的上界最小化，即 $\min_L \max_{\Delta_i} R(T)$。首先计算 $\max_{\Delta_i} g(L,\Delta_i)$：

$$\frac{\partial}{\partial\Delta_i} g(L,\Delta_i) = T(1 - L\Delta_i^2)\exp\left(-\frac{L\Delta_i^2}{2}\right).$$

当 $0 \leq \Delta_i < \frac{1}{\sqrt{L}}$ 时，$\frac{\partial}{\partial\Delta_i} g(L,\Delta_i) > 0$；当 $1 \geq \Delta_i > \frac{1}{\sqrt{L}}$ 时，$\frac{\partial}{\partial\Delta_i} g(L,\Delta_i) < 0$。

因此，对于所有 $L > 1$：

$$g(L,\Delta_i) \leq g\left(L, \frac{1}{\sqrt{L}}\right) = L + \frac{Te^{-1/2}}{\sqrt{L}}. \tag{1}$$

通过令 $L = \Theta(T^{\frac{2}{3}})$，最终可以得到

$$R(T) \leq \sum_{i=1}^{n}\left(L + \frac{Te^{-1/2}}{\sqrt{L}}\right) = \Theta(nT^{\frac{2}{3}}).$$

细心的读者也许会对这里的放缩 $L\Delta_i \leq L$ 产生疑问。如果这里不进行放缩，那么在下面对于(1)进行优化的过程中，第一项会由 L 变成 $L\Delta_i$，对于我们所取之 Δ_i，似乎在阶上会影响整体项的大小。

但实际上这并不会产生常数之外的的影响。因为在忽略常数之后，我们实际上想做的优化是

$$\min_L \max_{\Delta_i} L\Delta_i \vee T\Delta_i \exp\left(-L\Delta_i^2/2\right).$$

但显然这和

$$\min_L \max_{\Delta_i} L \vee T\Delta_i \exp\left(-L\Delta_i^2/2\right).$$

是一样的。（注意这里我们用 $a \vee b$ 表示 $\max\{a,b\}$，类似的，我们会用 $a \wedge b$ 来表示 $\min\{a,b\}$）

ETC 算法实现了 $\Theta(T^{\frac{2}{3}})$ 的次线性遗憾，这是不错的结果，但仍然不是最优的。其主要缺点在于，在探索阶段，该算法对所有臂一视同仁，无论已经获得的奖励如何，每个臂都固定被尝试 L 次。我们可以更加精妙的应用集中不等式，构造一个称之为**上置信界**（*Upper Confidence Bound*，*UCB*）的算法。它可以达到 $\widetilde{O}(T^{\frac{1}{2}})$[1] 的遗憾值，我在这个笔记中有所讨论。

[1] $\widetilde{O}(\cdot)$ 忽略了关于 T 的 log 项