根据 Young 不等式可证明著名的 Hölder 不等式.

引理6.3 (Hölder 不等式)**.** 对任意随机变量 X 和 Y 以及实数 $p>0$ 和 $q>0$ 满足$1/p+1/q=1$, 有

$$E(|XY|) \leq (E(|X|^p))^{\frac{1}{p}}(E(|Y|^q))^{\frac{1}{q}}.$$

特别地, 当 $p=q=2$ 时 *Hölder* 不等式变成为 *Cauchy-Schwartz* 不等式.

Proof. 设 $c=(E(|X|^p))^{\frac{1}{p}}$ 和 $d=(E(|Y|^q))^{\frac{1}{q}}$, 根据 Young 不等式有

$$\frac{|XY|}{cd} = \frac{|X|}{c}\frac{|Y|}{d} \leq \frac{1}{p}\frac{|X|^p}{c^p} + \frac{1}{q}\frac{|Y|^q}{d^q}.$$

对上式两边同时取期望有

$$\frac{E(|XY|)}{cd} \leq \frac{1}{p}\frac{E(|X|^p)}{c^p} + \frac{1}{q}\frac{E(|Y|^q)}{d^q} = \frac{1}{p} + \frac{1}{q} = 1,$$

从而完成证明. □

6.3 Chernoff 不等式

首先给出随机变量的矩生成函数 (Moment Generating Function) 的定义.

定义6.1. 定义随机变量 X 的矩生成函数为

$$M_X(t) = E[e^{tX}].$$

下面给出关于矩生成函数的一些性质:

定理6.3. 设随机变量 X 的矩生成函数为 $M_X(t)$, 对任意 $n \geq 1$ 有

$$E[X^n] = M_X^{(n)}(0),$$

这里 $M_X^{(n)}(t)$ 表示矩生成函数在 $t=0$ 的 n 阶导数, 而 $E[X^n]$ 被称为随机变量 X 的 n 阶矩 *(moment)*.

Proof. 由 Tayler 公式有

$$e^{tX} = \sum_{i=1}^{\infty} \frac{(tX)^i}{i!}.$$

两边同时取期望有

$$E[e^{tX}] = \sum_{i=1}^{\infty} \frac{t^i}{i!} E[X^i].$$

对上式两边分别对 t 求 n 阶导数并取 $t=0$ 有 $M_X^{(n)}(t) = E[X^n]$. □

定理6.4. 对随机变量 X 和 Y, 如果存在常数 $\delta>0$, 使得当 $t \in (-\delta,\delta)$ 时有 $M_X(t)=M_Y(t)$ 成立, 那么 X 与 Y 有相同的分布.

上述定理表明随机变量的矩生成函数可唯一确定随机变量的分布, 其证明超出了本书的范围. 若随机变量 X 与 Y 独立, 则有

$$M_{X+Y}(t)=E[e^{(X+Y)t}]=E[e^{tX}e^{tY}]=E[e^{tX}]\cdot E[e^{tY}]=M_X(t)M_Y(t).$$

于是得到

推论6.3. 对任意独立的随机变量 X 和 Y 有 $M_{X+Y}(t)=M_X(t)M_Y(t)$.

下面将利用矩生成函数来证明一系列不等式. 给定任意随机变量 X 和任意 $t>0$ 和 $\epsilon>0$, 利用 Markov 不等式有

$$\Pr[X\geq\epsilon]=\Pr[e^{tX}\geq e^{t\epsilon}]\leq e^{-t\epsilon}E[e^{tX}].$$

特别地, 有

$$\Pr[X\geq\epsilon]\leq\min_{t>0}\left\{e^{-t\epsilon}E[e^{-tX}]\right\}.$$

类似地, 对任意 $\epsilon>0$ 和 $t<0$ 有

$$\Pr[X\leq\epsilon]=\Pr[tX\geq t\epsilon]\leq e^{-t\epsilon}E[e^{tX}].$$

同理有

$$\Pr[X\leq\epsilon]\leq\min_{t<0}\left\{e^{-t\epsilon}E[e^{-tX}]\right\}.$$

上述方法称为 '**Chernoff 方法**', 是证明集中不等式一种最根本最重要的方法. 下面将针对特定的分布或特定的条件, 先求解矩生成函数 $E[e^{tX}]$, 然后求解最小值 t 的取值.

6.3.1 二值随机变量的 Chernoff 不等式

定理6.5. 设随机变量 $X_1,X_2,\cdots,X_n$ 相互独立且满足 $X_i\sim Ber(p_i)$, 令 $\mu=\sum_{i=1}^n E[X_i]=\sum_{i=1}^n p_i$. 对任意 $\epsilon>0$ 有

$$\Pr\left[\sum_{i=1}^n X_i\geq(1+\epsilon)\mu\right]\leq\left(\frac{e^\epsilon}{(1+\epsilon)^{(1+\epsilon)}}\right)^\mu;$$

对任意 $\epsilon\in(0,1)$ 有

$$\Pr\left[\sum_{i=1}^n X_i\geq(1+\epsilon)\mu\right]\leq e^{-\mu\epsilon^2/3}.$$

上述第一个不等式给出了最紧的不等式上界, 第二个不等式是第一个不等式的适当放松.

Proof. 令 $\bar{X}=\sum_{i=1}^n X_i$. 对任意 $t>0$, 根据 Chernoff 方法有

$$\Pr[\bar{X}\geq(1+\epsilon)\mu]=\Pr[e^{t\bar{X}}\geq e^{t(1+\epsilon)\mu}]\leq e^{-t(1+\epsilon)\mu}E[e^{t\bar{X}}].$$

利用随机变量的独立性以及 $1+x \le e^x$ 有

$$\begin{aligned}
E[e^{t\bar{X}}] &= E[e^{\sum_{i=1}^n tX_i}] = \prod_{i=1}^n E[e^{tX_i}] \\
&= \prod_{i=1}^n [(1-p_i)+p_i e^t] = \prod_{i=1}^n [1+p_i(e^t-1)] \\
&\le \exp\left(\sum_{i=1}^n p_i(e^t-1)\right) = \exp(\mu(e^t-1)).
\end{aligned}$$

由此可得

$$\Pr[\bar{X} \ge (1+\epsilon)\mu] \le \exp\left(-t(1+\epsilon)\mu + \mu(e^t-1)\right).$$

对上式求最小值解得 $t_{\min} = \ln(1+\epsilon)$, 代入可得

$$\Pr[\bar{X} \ge (1+\epsilon)\mu] \le \left(\frac{e^\epsilon}{(1+\epsilon)^{(1+\epsilon)}}\right)^\mu.$$

对第二个不等式, 只需证明当 $\epsilon \in (0,1)$ 有

$$f(\epsilon) = \ln\left(\frac{e^\epsilon}{(1+\epsilon)^{(1+\epsilon)}}\right) + \frac{\epsilon^2}{3} = \epsilon - (1+\epsilon)\ln(1+\epsilon) + \frac{\epsilon^2}{3} \le 0.$$

易知 $f(0)=0$ 和 $f(1)<0$. 当 $\epsilon \in (0,1)$,

$$f'(\epsilon) = -\ln(1+\epsilon) + 2\epsilon/3, \quad f''(\epsilon) = -\frac{1}{1+\epsilon} + \frac{2}{3}.$$

于是得到 $f'(0)=0$, $f'(1) = -0.0265 < 0$ 和 $f'(1/2) = -0.0721 < 0$, 由连续函数性质有 $f'(\epsilon) \le 0$, 即函数 $f(\epsilon)$ 在 $[0,1]$ 上单调递减. 当 $\epsilon \ge 0$ 时有 $f(\epsilon) \le f(0) = 0$, 所以 $\exp(f(\epsilon)) \le 1$. □

下面的定理给出了 $\Pr[\sum_{i=1}^n X_i \ge (1-\epsilon)\mu]$ 的估计, 证明作为练习题留给大家完成.

定理6.6. 设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立且满足 $X_i \sim Ber(p_i)$, 令 $\mu = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n p_i$. 对任意 $\epsilon \in (0,1)$ 有

$$\Pr\left[\sum_{i=1}^n X_i \ge (1-\epsilon)\mu\right] \le \left(\frac{e^{-\epsilon}}{(1-\epsilon)^{(1-\epsilon)}}\right)^\mu \le \exp(-\mu\epsilon^2/2).$$

定义6.2. 若随机变量 $X \in \{+1,-1\}$ 满足

$$\Pr(X=+1) = \Pr(X=-1) = 1/2,$$

则称 X 为 *Rademacher* 随机变量.

我们有如下定理:

定理6.7. 对 n 个独立的 *Rademacher* 随机变量 $X_1, X_2, \cdots, X_n$, 有

$$\Pr\left(\frac{1}{n}\sum_{i=1}^n X_i \ge \epsilon\right) \le \exp(-n\epsilon^2/2) \quad \text{和} \quad \Pr\left(\frac{1}{n}\sum_{i=1}^n X_i \le -\epsilon\right) \le \exp(-n\epsilon^2/2).$$

Proof. 根据 Taylor 展开式有

$$\frac{1}{2}\exp(t)+\frac{1}{2}\exp(-t)=\sum_{i\geq 0}\frac{t^{2i}}{(2i)!}\leq\sum_{i\geq 0}\frac{(t^2/2)^i}{i!}=\exp(t^2/2).$$

若随机变量 $X\in\{+1,-1\}$ 且满足 $\Pr(X=1)=\Pr(X=-1)=1/2$, 则有

$$E[e^{tX}]=\frac{1}{2}e^t+\frac{1}{2}e^{-t}\leq\exp(t^2/2).$$

对任意 $t>0$, 根据 Chernoff 方法有

$$\begin{aligned}\Pr\left(\frac{1}{n}\sum_{i=1}^n X_i\geq\epsilon\right) &\leq \exp(-nt\epsilon)E\left[\exp\left(\sum_{i=1}^n tX_i\right)\right]\\ &= \exp(-nt\epsilon)\prod_{i=1}^n E\left[\exp(tX_i)\right]\leq\exp(-nt\epsilon+nt^2/2).\end{aligned}$$

通过对上式右边求最小值解得 $t=\epsilon$, 带入上式得到

$$\Pr\left(\frac{1}{n}\sum_{i=1}^n X_i\geq\epsilon\right)\leq\exp(-n\epsilon^2/2).$$

同理证明另一个不等式. □

推论6.4. 对独立同分布的随机变量 $X_1,X_2,\cdots,X_n$ 满足 $P(X_1=0)=P(X_1=1)=1/2$, 有

$$\Pr\left(\frac{1}{n}\sum_{i=1}^n X_i-\frac{1}{2}\geq\epsilon\right)\leq\exp(-2n\epsilon^2)\quad 和 \quad \Pr\left(\frac{1}{n}\sum_{i=1}^n X_i-\frac{1}{2}\leq-\epsilon\right)\leq\exp(-2n\epsilon^2).$$

6.3.2 有界随机变量的 Chernoff 不等式

本节研究有界的随机变量 $X_i\in[a,b]$ 的 Chernoff 不等式. 首先介绍著名的 Chernoff 引理.

引理6.4. 设随机变量 $X\in[0,1]$ 的期望 $\mu=E[X]$. 对任意 $t>0$ 有

$$E[e^{tX}]\leq\exp(t\mu+t^2/8).$$

Proof. 由凸函数的性质可知

$$e^{tX}=e^{tX+(1-X)0}\leq Xe^t+(1-X)e^0,$$

两边再同时取期望有

$$E(e^{tX})\leq 1-\mu+\mu e^t=\exp(\ln(1-\mu+\mu e^t)).$$

令 $f(t)=\ln(1-\mu+\mu e^t)$, 我们有 $f(0)=0$ 以及

$$f'(t)=\frac{\mu e^t}{1-\mu+\mu e^t}\Rightarrow f'(0)=\mu.$$

进一步有

$$f''(t)=\frac{\mu e^t}{1-\mu+\mu e^t}-\frac{\mu^2e^{2t}}{(1-\mu+\mu e^t)^2}\leq 1/4.$$

根据泰勒中值定理有

$$f(t) = f(0) + tf'(0) + f''(\xi)t^2/2 \leq t\mu + t^2/8.$$

引理得证. □

由上面的 Chernoff 引理进一步推导出

推论6.5. 设随机变量 $X \in [a,b]$ 的期望 $\mu = E[x]$. 对任意 $t > 0$ 有

$$E(e^{tX}) \leq \exp(\mu t + t^2(b-a)^2/8).$$

根据上述推论, 我们得到有界随机变量的 Chernoff 不等式:

定理6.8. 假设 $X_1, \ldots, X_n$ 是 n 独立的随机变量且满足 $X_i \in [a,b]$. 对任意 $\epsilon > 0$ 有

$$\begin{aligned}
\Pr\left[\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n E[X_i] \geq \epsilon\right] &\leq \exp(-2n\epsilon^2/(b-a)^2), \\
\Pr\left[\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n E[X_i] \leq -\epsilon\right] &\leq \exp(-2n\epsilon^2/(b-a)^2).
\end{aligned}$$

Proof. 这里给出第一个不等式的证明, 第二个不等式证明作为习题. 对任意 $t > 0$, 根据 Chernoff 方法有

$$\begin{aligned}
&\Pr\left[\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n E[X_i] \geq \epsilon\right] \\
&= \Pr\left[\sum_{i=1}^n t(X_i - E[X_i]) \geq nt\epsilon\right] \\
&\leq \exp(-nt\epsilon)E\left[\exp\left(\sum_{i=1}^n t(X_i - E[X_i])\right)\right] \\
&= \exp(-nt\epsilon)\prod_{i=1}^n E\left[\exp(t(X_i - E[X_i]))\right].
\end{aligned}$$

根据 Chernoff 引理, 对任意 $X_i \in [a,b]$ 有

$$E\left[\exp(t(X_i - E[X_i]))\right] \leq \exp((b-a)^2t^2/8).$$

由此得到

$$\Pr\left[\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n E[X_i] \geq \epsilon\right] \leq \exp(-nt\epsilon + nt^2(b-a)^2/8).$$

对上式右边取最小值求解 $t = 4\epsilon/(b-a)^2$, 然后带入上式可得:

$$\Pr\left[\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n E[X_i] \geq \epsilon\right] \leq \exp(-2n\epsilon^2/(b-a)^2).$$

从而完成证明. □