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

$$\Pr[X \geqslant E[X]+\epsilon] = \Pr[e^{tX} \geqslant e^{tE[X]+t\epsilon}] \leqslant e^{-t\epsilon-tE[X]}E[e^{tX}].$$

特别地, 有

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

类似地, 对任意 $\epsilon>0$ 和 $t<0$ 有

$$\Pr[X \leqslant E[X]-\epsilon] = \Pr[e^{tX} \geqslant e^{tE[X]-t\epsilon}] \leqslant e^{t\epsilon-tE[X]}E[e^{tX}].$$

同理有

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

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.

6.2.1 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

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

对任意 $\epsilon \in (0,1)$ 有

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

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.

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

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

利用随机变量的独立性以及 $1+x \leqslant e^x$ 有

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

$$\leqslant \quad \exp\left(\sum_{i=1}^{n} p_i(e^t-1)\right) = \exp(\mu(e^t-1)).$$

由此可得

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

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

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

对第二个不等式, 只需证明当 $\epsilon \in (0,1)$ 有

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

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 $f'(\epsilon) \leqslant 0$, 即函数 $f(\epsilon)$ 在 $[0,1]$ 上单调递减. 当 $\epsilon \geqslant 0$ 时有 $f(\epsilon) \leqslant f(0) = 0$, 所以 $\exp(f(\epsilon)) \leqslant 1$.

下面的定理给出了 $\Pr[\sum_{i=1}^{n} X_i \leqslant (1-\epsilon)\mu]$ 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

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

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

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

证明 根据 Taylor 展开式有

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

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

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

对任意 $t>0$, 根据 Chernoff 方法有

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

通过对上式右边求最小值解得 $t=\epsilon$, 带入上式得到

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

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

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

6.2.2 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

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

证明 由凸函数的性质可知

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

两边再同时取期望有

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

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

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

根据泰勒中值定理有

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

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

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

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

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

证明 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

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

根据 Chernoff 引理, 对任意 $X_i \in [a,b]$ 有

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

由此得到

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

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:

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

从而完成证明.

6.2.3 Gaussian 和 Sub-Gaussian 随机变量不等式

首先考虑独立同分布的 Gaussian 随机变量:

定理 6.9 设随机变量 $X_1, \ldots, X_n$ 相互独立、且服从 $X_i \sim \mathcal{N}(\mu, \sigma)$, 对任意 $\epsilon > 0$ 有

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

证明 对随机变量 $X_i \sim \mathcal{N}(\mu, \sigma)$, 根据正太分布的性质有

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) \sim \mathcal{N}(0, \sigma^2/n) \quad \Rightarrow \quad \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}(X_i - \mu) \sim \mathcal{N}(0, 1).$$

若 $X' \sim \mathcal{N}(0,1)$, 对任意 $\epsilon > 0$, 根据以前的定理有

$$P(X' \geqslant \epsilon) \leqslant \frac{1}{2}e^{-\epsilon^2/2}.$$

因此得到

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) \geqslant \epsilon\right] = \Pr\left[\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}(X_i - \mu) \geqslant \epsilon\sqrt{n}/\sigma\right] \leqslant \frac{1}{2}\exp(-n\epsilon^2/2\sigma^2),$$

定理得证.

下面定义 Sub-Gaussian 随机变量, 将有界随机变量和 Gaussian 随机变量统一起来:

定义 6.3 对任意 $t \in (-\infty, +\infty)$, 若随机变量 X 满足

$$E[e^{(X-E[X])t}] \leqslant \exp(bt^2/2),$$