定理6.6. 设$X_1, X_2, \cdots, X_n$是n个独立的*Bernoulli*随机变量，且满足$X_i \sim Ber(p_i)$. 设$X = \sum_{i=1}^n X_i$, $\mu = \sum_{i=1}^n p_i$, 对$\forall \epsilon \in (0,1)$, 有

$$\Pr[X \geq (1-\epsilon)\mu] < \left(\frac{e^{-\epsilon}}{(1-\epsilon)^{(1-\epsilon)}}\right)^{\mu} \leq \exp(-\mu\epsilon^2/2).$$

对于更为特殊的随机变量$X \in \{+1, -1\}$, 且满足

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

我们有如下定理：

定理6.7. 设$X_1, X_2, \cdots, X_n$是n个独立同分布随机变量，满足$\Pr(X_i = 1) = \Pr(X_i = -1) = 1/2$, 则有

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

Proof. 根据$\exp(t)$和$\exp(-t)$的Taylor展开式有

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

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

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

通过对上式右边求最小值解得$t = \epsilon$, 带入上式得到

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

同理证明另外一个不等式. □

6.3.4 有界随机变量和Chernoff不等式

本小节研究随机变量$X_i \in [a,b]$, 其对应的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

$$e^{tX} \leq Xe^t + (1-X)e^0 \Rightarrow E(e^{tX}) \leq 1 - \mu + \mu e^t = \exp(\ln(1-\mu+\mu e^t)) \tag{2}$$

令$f(t) = \ln(1-\mu+\mu e^t)$, 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

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

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6.3. 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$\epsilon > 0$, 我们有

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

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

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

根据Chernoff引理以及简单整理可得, 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. □

6.4 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$X_1, \ldots, X_n$是n独立同分布的随机变量, 且服从$X_i \sim \mathcal{N}(\mu, \sigma)$. 对任意$\epsilon > 0$, 有*

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

Proof. 如果$X_i \sim \mathcal{N}(\mu, \sigma)$, 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

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

因此得到

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

定理得证. □

下面研究Sub-Gaussian随机变量, 可以将有界随机变量和Gaussian随机变量统一起来:

定义6.2. *对任意$t \in \mathcal{R}$, 如果随机变量X满足*

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

则称随机变量X是服从参数为b的亚高斯(Sub-Gaussian)随机变量.

直观而言：亚高斯分布随机变量表示随机变量的尾部分布不会比一个高斯分布更严重.

例6.2. *对任意有界的随机变量 $X \in [a, b]$, 根据 Chernoff 引理有*

$$E[e^{(X-\mu)t}] \leq \exp(t^2(b-a)^2/8),$$

因此有界随机变量是服从参数为$(b-a)^2/4$的亚高斯型随机变量.

例6.3. 如果随机变量X服从高斯分布$\mathcal{N}(\mu,\sigma^2)$, 我们有

$$E[e^{(X-\mu)t}]=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{xt}e^{-x^2/2\sigma^2}dx=e^{\sigma^2t^2/2}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-(t\sigma-x/\sigma)^2/2}d(x/\sigma)=e^{\sigma^2t^2/2}.$$

因此$Gaussian$型随机变量是服从参数为σ^2的亚高斯型随机变量.

由前面的两个例子可知高斯随机变量和有界随机变量都是亚高斯随机变量. 对于亚高斯型随机变量, 我们有如下不等式:

定理6.10. 假设$X_1,\ldots,X_n$是独立地、服从参数为b的亚高斯随机变量, 对任意$\epsilon>0$, 有

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

Proof. 对任意$t>0$, 根据Chernoff方法有

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

通过求解上式最小值可得$t_{\min}=\epsilon/b$, 从而完成证明. □

对于亚高斯型随机变量, 我们还可以估计最大值期望的不等式:

定理6.11. 假设$X_1,\ldots,X_n$是相互独立地、服从参数为b的亚高斯随机变量, 且满足$E[X_i]=0$. 我们有

$$E\left[\max_{i\in[n]}X_i\right]\leq\sqrt{2b\ln n}.$$

Proof. 对任意$t>0$, 根据Jensen不等式有

$$\exp\left(tE\left[\max_{i\in[n]}X_i\right]\right)\leq E\left[\exp\left(t\max_{i\in[n]}X_i\right)\right]=E\left[\max_{i\in[n]}\exp\left(tX_i\right)\right]\leq\sum_{i=1}^n E[\exp(tX_i)]\leq n\exp(t^2b/2).$$

对上式两边同时取对数和除以t有

$$E\left[\max_{i\in[n]}X_i\right]\leq\frac{\ln n}{t}+\frac{bt}{2}.$$

通过求解上式最小值可得$t_{\min}=\sqrt{2\ln n/b}$, 从而完成证明. □

前面所讲的概率不等式, 可以以另外一种表达形式给出, 这里以定理 6.10为例, 假设$X_1,\ldots,X_n$是独立地、服从参数为b的亚高斯随机变量, 对任意$\epsilon>0$, 有

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

令$\delta=\exp(-n\epsilon^2/2b)$, 求解出

$$\epsilon=\sqrt{2b\ln(1/\delta)/n},$$

带入得到下面的不等式至少以$1-\delta$的概率成立：

$$\frac{1}{n}\sum_{i=1}^n(X_i-\mu)\leq\sqrt{\frac{2b}{n}\ln\frac{1}{\delta}}.$$

所有不等式都可以采用$1-\delta$的形式描述.