第三章：1，2，3，4，5，6

1. 随机变量 X 的期望 $\mu = E(X)$ 和方差 $\sigma^2 = E[(X-\mu)^2]$。如果定义随机变量 $X^* = \frac{X-\mu}{\sigma}$，证明

$$P[|X^*| \geq c] \leq \frac{1}{c^2}.$$

解

$$P[|X^*| \geq c] = P[|\frac{X-\mu}{\delta}| \geq c] = P[|X-\mu| \geq c\delta] \leq \frac{\delta^2}{c^2\delta^2} = \frac{1}{c^2}$$

2. 令 X_i $(i = 1, 2, \cdots, n)$ 为一组独立同分布的随机变量，期望 $\mu = E(X_i)$ 和方差 $\sigma^2 = E[(X_i-\mu)^2]$。如果定义 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$，证明

$$P[|\overline{X} - \mu| \geq \epsilon] \leq \frac{\sigma^2}{n\epsilon^2}.$$

解　由于

$$Var(\overline{X}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

$$E(\overline{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}E(\sum_{i=1}^{n} X_i) = \frac{1}{n} \cdot nE(X_i) = \mu$$

所以

$$P[|\overline{X} - \mu| \geq \epsilon] \leq \frac{Var(\overline{X})}{\epsilon} = \frac{\sigma^2}{n\epsilon^2}$$

3. 假设抛一枚均匀的硬币 n 次，随机变量 X 定义为正面朝上的次数，

a. 请运用 Chebyshev 不等式给出事件 $X < \frac{n}{4}$ 的概率上界；

b. 请运用 Chernoff 不等式给出事件 $X < \frac{n}{4}$ 的概率上界；

a. **解**

$$P(X < \frac{n}{4}) = P(X > \frac{3n}{4}) < P(|X - \frac{n}{2}| > \frac{n}{4}) < \frac{Var(X)}{(n/4)^2} = \frac{n}{4}$$

b. **解**

$$P(X < \frac{n}{4}) = P(X > \frac{3n}{4}) = P(X > (1+\frac{1}{2})\frac{n}{2})) < \exp(-\frac{n}{2}(\frac{1}{2})^2/4) = \exp(-n/32)$$

4. 假设 X_i 独立随机变量序列，满足

$$P(X_i = 1) = p_i \text{ 和} P(X_i = 0) = 1 - p_i.$$

令 $\mu = \sum_{i=1}^{n} p_i$，定义随机变量 $X = \sum_{i=1}^{n} X_i$。证明以下结论：

a. $P(X > (1+\delta)\mu) < \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$；

b. $P(X > (1+\delta)\mu) < \exp(-\mu\delta^2/3)$。

a. **解**　对于 t>0

$$P(X > (1+\delta)\mu) = P(\exp(tX) > \exp(t(1+\delta)\mu)) < \frac{\prod_{i=1}^{n} E(\exp(tX_i))}{\exp(t(1+\delta)\mu)}$$

因为 $1 - x < e^{-x}$，所以

$$E(\exp(tX_i)) = p_i e^t + (1-p_i) = 1 - p_i(1-e^t) < \exp(p_i(e^t - 1))$$

$$\prod_{i=1}^{n} E(\exp(tX_i)) < \prod_{i=1}^{n} \exp(p_i(e^t-1)) = \exp(\mu(e^t - 1))$$

$$P(X > (1+\delta)\mu) < \frac{\exp(\mu(e^t-1))}{\exp(t(1+\delta)\mu)} = \exp(\mu(e^t - 1 - t - t\delta))$$

对 t 求偏导，得到 $\mu(e^t - 1 - \delta) = 0$，所以 $t = \ln(1+\delta)$

$$P(X > (1+\delta)\mu) < \exp(\mu(e^t - 1 - t - t\delta)) = (\frac{e^\delta}{(1+\delta)^(1+\delta)})^\mu.$$

b. **解**

$$(1+\delta)^{1+\delta} = e^{\ln(1+\delta)^{(1+\delta)}} = e^{(1+\delta)\ln(1+\delta)}$$

$$(1+\delta)\ln(1+\delta) = (1+\delta)(\delta - \frac{\delta^2}{2} + \frac{\delta^3}{3} \cdots) > \delta + \frac{\delta^2}{3}$$

所以

$$P(X > (1+\delta)\mu) < (\frac{e^\delta}{\exp(\delta + \frac{\delta^2}{4})})^\mu = \exp(-\mu\delta^2/3)$$

5. 假设 X_i 独立随机变量序列，满足

$$P(X_i = 1) = p_i \text{ 和} P(X_i = 0) = 1 - p_i.$$

令 $\mu = \sum_{i=1}^{n} p_i$，定义随机变量 $X = \sum_{i=1}^{n} X_i$。证明

$$P(|X - \mu| > \delta\mu) < 2\exp(-\mu\delta^2/3).$$

解　由于第四题结论

$$P(X > (1+\delta)\mu) < \exp(-\mu\delta^2/3)$$

又因为

$$P(X < (1-\delta)\mu) < \exp(-\mu\delta^2/2) < \exp(-\mu\delta^2/3)$$

所以

$$P(|X - \mu| > \delta\mu) = P(X > (1+\delta)\mu) + P(X < (1-\delta)\mu) < 2\exp(-\mu\delta^2/3).$$

6. 设一组独立随机变量 $X_i \sim Bernoulli(p)$, 其中 $i = 1, \cdots, n$。令 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$。请问当 n 取值为多少时使得

$$P(|\overline{X} - p| \le \epsilon p) \ge 1 - \delta.$$

解　由于 $Var(\overline{X}) = \frac{1}{n}p(1-p)$，$E(\overline{X}) = p$

所以有

$$1 - P(|\overline{X} - p| \le \epsilon p) = P(|\overline{X} - p| > \epsilon p) < \frac{\frac{1}{n}p(1-p)}{\epsilon^2 p^2} \le \delta$$

所以当 $n \ge \frac{1-p}{\delta\epsilon^2 p}$ 使得 $P(|\overline{X} - p| \le \epsilon p) \ge 1 - \delta$