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

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

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

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

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

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

解　由于第四题结论

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

又因为

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

所以

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

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

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

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

所以有

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

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

7. 假设某平台拥有 k 台服务器，需要对外响应 n 个服务（每项服务都在很短的时间内完成）。当平台收到一个服务请求时，将这个服务请求随机分发到一个服务器上。假设 X_i 表示第 i 台服务器上被分发的服务请求数量。

a. 随机变量 X_i 服从什么分布?

b. $X_1,X_2,\cdots,X_k$ 是不是独立随机变量？为什么?

c. 在考虑负载均衡问题时，如果仅关心随机变量 $M\geq(1+\epsilon)\frac{n}{k}$ 的可能性大小，其中随机变量 $M=\max\{X_1,X_2,\cdots,X_k\}$。请给出以下概率的上界

$$Pr(M\geq(1+\epsilon)\frac{n}{k}).$$

a. **解** 二项分布，$X \sim b(n, \frac{1}{k})$

b. **解** 不是独立随机变量, 这是由于 $\sum_{i=1}^{k} X_i = n$ 举例来说, 如果 $k = 2$, $P(X_1 = m, X_2 = n - m) = C_n^m (\frac{1}{2})^m (\frac{1}{2})^{n-m} \neq P(X_1 = m) \cdot P(X_2 = n - m) = C_n^m (\frac{1}{2})^m (\frac{1}{2})^{n-m} \cdot C_n^{n-m} (\frac{1}{2})^{n-m} (\frac{1}{2})^m$

c. **解**

$$
\begin{aligned}
P(M \geq (1+\epsilon)\frac{n}{k}) &= P(max\{X_1, X_2, \cdots, X_k\} \geq (1+\epsilon)\frac{n}{k}) \\
&= P(\bigcup_{i=1}^{k} X_i \geq (1+\epsilon)\frac{n}{k}) \\
&\leq \sum_{i=1}^{k} P(X_i \geq (1+\epsilon)\frac{n}{k}) \\
&\leq \sum_{i=1}^{k} \exp{(-\frac{n\epsilon^2}{4k})} \\
&= k \cdot \exp(-\frac{n\epsilon^2}{4k})
\end{aligned}
$$