# 数据科学与工程算法基础 习题3

Author：GONGGONGJOHN

## 1

由于 $\mu=E(X),\sigma^2=E\left[(X-\mu)^2\right]$

故由Chebyshev不等式可知

$$\begin{aligned} P\left(|X^*|\geq c\right) &= P\left(\left|\frac{X-\mu}{\sigma}\right|\geq c\right) \\ &= P(|X-\mu|\geq c|\sigma|) \\ &\leq \frac{\sigma^2}{c^2\sigma^2} \\ &= \frac{1}{c^2} \end{aligned}$$

## 2

由 $X_i$ 独立同分布可知

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}E(X_i) \\ &= \mu \\ Var(\bar{X}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}Var(X_i) \\ &= \frac{\sigma^2}{n} \end{aligned}$$

故由Chebyshev不等式得

$$P\left(|\bar{X}-\mu|\geq\varepsilon\right)\leq\frac{\frac{\sigma^2}{n}}{\varepsilon^2}=\frac{\sigma^2}{n\varepsilon^2}$$

## 3

（1）易见 $X\sim b(n,\frac{1}{2})$

于是 $E(X)=\frac{n}{2},Var(X)=\frac{n}{4}$

现取 $\varepsilon=\frac{n}{4}$，则由Chebyshev不等式可知

$$P\left(\left|X-\frac{n}{2}\right|\geq\frac{n}{4}\right)\leq\frac{\frac{n}{4}}{\frac{n^2}{16}}=\frac{4}{n}$$

由二项分布的概率密度函数可知 $P(X\geq\frac{3}{4}n)=P(X\leq\frac{n}{4})$

故

$$\begin{aligned}P\left(\left|X-\frac{n}{2}\right|\geq\frac{n}{4}\right)&=P\left(\{X\geq\frac{3}{4}n\}\cap\{X\leq\frac{n}{4}\}\right)\\&=P(X\geq\frac{3}{4}n)+P(X\leq\frac{n}{4})\\&=2P(X\leq\frac{n}{4})\leq\frac{4}{n}\end{aligned}$$

也即 $P(X\leq\frac{n}{4})\leq\frac{2}{n}$

当 $\frac{n}{4}\notin\mathbb{Z}^+$ 时,

$$P\left(X<\frac{n}{4}\right)=P\left(X\leq\frac{n}{4}\right)\leq\frac{2}{n}$$

当 $\frac{n}{4}\in\mathbb{Z}^+$ 时,

由于

$$P\left(X=\frac{n}{4}\right)=\binom{n}{\frac{n}{4}}\left(\frac{1}{2}\right)^{\frac{n}{4}}\left(\frac{1}{2}\right)^{n-\frac{n}{4}}=\binom{n}{\frac{n}{4}}\left(\frac{1}{2}\right)^{n}$$

故

$$\begin{aligned}P\left(X<\frac{n}{4}\right)&=P\left(X\leq\frac{n}{4}\right)-P\left(X=\frac{n}{4}\right)\\&\leq\frac{2}{n}-\binom{n}{\frac{n}{4}}\left(\frac{1}{2}\right)^{n}\end{aligned}$$

**(2)** 取 $\delta=\frac{1}{2}$

故由Chernoff不等式可知

$$\begin{aligned}P\left(X<\frac{n}{4}\right)&=P\left(X<(1-\frac{1}{2})\frac{n}{2}\right)\\&<\exp\left\{-\frac{1}{2}\cdot\frac{n}{2}\cdot\frac{1}{4}\right\}\\&=\exp\left\{-\frac{n}{16}\right\}\end{aligned}$$

# 4

**(1)** 对任意 $t>0$，有

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

而由 $1-x<e^{-x}$ 可知

$$
\begin{aligned}
E\left(\exp\{tX_i\}\right) &= p_i e^t + (1-p_i) \\
&= 1 - p_i(1-e^t) \\
&< \exp\left\{p_i(e^t-1)\right\}
\end{aligned}
$$

故

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

于是

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

要使得该式对任意 $t > 0$ 成立，则左式应当小于右式的最小值。

令 $f(t) = e^t - 1 - t - t\delta$，由 $e^x$ 的单调性及 $\mu > 0$ 可知

$$
\mathbf{minimize}_t \exp\left\{\mu(e^t-1-t-t\delta)\right\} \leftrightarrow \mathbf{minimize}_t\ f(t)
$$

求导易知当 $t = \ln(1+\delta)$ 时，$f(t)$ 取到最小值

因此

$$
\begin{aligned}
P(X > (1+\delta)\mu) &< \exp\{\mu(\exp\{\ln(1+\delta)\} - 1 - \ln(1+\delta) - \delta\ln(1+\delta))\} \\
&= \exp\left\{\mu \ln \frac{e^\delta}{(1+\delta)^{1+\delta}}\right\} \\
&= \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu
\end{aligned}
$$

**（2）**

> **Lemma:** 当 $x > 0$ 时，$\ln(1+x) \geq x/(1+\frac{x}{2})$
>
> **Proof:** 令 $f(x) = \ln(1+x) - \frac{2x}{2+x}$
>
> 易见 $f(x) = \ln(1+x) - 2 + \frac{4}{2+x}$
>
> 故 $f'(x) = \frac{1}{1+x} - \frac{4}{(2+x)^2}$
>
> 由此可得唯一驻点 $x = 0$
>
> 而 $f(0) = 0, f(2) = \ln 3 - 1 > 0$
>
> 故对任意 $x > 0$，$\ln(1+x) \geq x/(1+\frac{x}{2})$

由（1）可知对任意 $\delta \in (0,1)$，有

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

而由引理可得 $(1+\delta)\ln(1+\delta) \geq \frac{2\delta(1+\delta)}{2+\delta}$

因此

$$
\begin{aligned}
P(X > (1+\delta)\mu) &< \exp\left\{\mu\left(\delta - \frac{2\delta(1+\delta)}{2+\delta}\right)\right\} \\
&= \exp\left\{-\frac{\mu\delta^2}{2+\delta}\right\} \\
&< \exp\left\{-\frac{\mu\delta^2}{3}\right\}
\end{aligned}
$$

# 6

由 $X_i \overset{i.i.d}{\sim} b(1,p)$ 可知

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

故由Chebyshev定理可得

$$P(|\bar{X} - p| \geq \varepsilon p) \leq \frac{1-p}{n\varepsilon^2 p}$$

因此

$$
\begin{aligned}
P(|\bar{X} - p| \leq \varepsilon p) &= 1 - P(|\bar{X} - p| \geq \varepsilon p) \\
&\geq 1 - \frac{1-p}{n\varepsilon^2 p}
\end{aligned}
$$

要使得 $P(|\bar{X} - p| \leq \varepsilon p) \geq 1 - \delta$ 成立，则有

$$1 - \frac{1-p}{n\varepsilon^2 p} \geq 1 - \delta$$

也即

$$n \geq \frac{1-p}{\varepsilon^2 p \delta}$$