温兆和 10205501432. 数据科学算法作业2

1. 证: $E(X^*)=0$. $Var(X^*)=\frac{1}{\sigma^2}Var(X)=1$.

由切比雪夫不等式: 当 $c>0$.

$$P(|X^*-E(X^*)|\geq c)\leq\frac{Var(X^*)}{c^2}$$

即 $P(|X^*|\geq c)\leq\frac{1}{c^2}$.

2. 证: $E(\bar{X})=0$. $Var(\bar{X})=\frac{\sigma^2}{n}$

由切比雪夫不等式: 当 $\varepsilon>0$

$$P(|\bar{X}-E(\bar{X})|\geq\varepsilon)\leq\frac{Var(\bar{X})}{\varepsilon^2}$$

即: $P(|\bar{X}-\mu|\geq\varepsilon)\leq\frac{\sigma^2}{n\varepsilon^2}$.

3. 解: (1). $P(X<\frac{n}{4})=P(X>\frac{3}{4}n)$

$=P(X-\frac{n}{2}>\frac{n}{4})=\frac{1}{2}P(|X-\frac{n}{2}|>\frac{n}{4})$

$\leq\frac{1}{2}\times\frac{n}{4}\times\frac{16}{n^2}=\frac{2}{n}$.

(2) $P(X<\frac{n}{4})=P(X<\frac{n}{2}(1-\frac{1}{2}))$

$<\left(\frac{e^{-0.5}}{\sqrt{0.5}}\right)^{\frac{n}{2}}=\left(\frac{2}{e}\right)^{\frac{n}{4}}$

4. 证: (1). 对 $\forall t>0$. 有:

$X>(1+\delta)\mu \Rightarrow tX>t(1+\delta)\mu$

$\Rightarrow e^{tX}>e^{t(1+\delta)\mu}$.

故 $P(X>(1+\delta)\mu)=P(e^{tX}>e^{t(1+\delta)\mu})$

由马尔可夫不等式: $P(e^{tX}>e^{t(1+\delta)\mu})<\frac{\prod_{i=1}^{n}E(e^{tX_i})}{e^{t(1+\delta)\mu}}$

(由于 $e^{tX}=e^{t\sum X_i}=\prod_{i=1}^{n}e^{tX_i}$ 且诸 X_i 独立).

又: $E(e^{tX_i})=p_ie^t+(1-p_i)e^0=1+p_i(e^t-1)$

由于 $x>0$ 时 $1+x<e^x$ 且 $p_i(e^t-1)$ 当 $t>0$ 时恒有正：

$$E(e^{tX_i}) < e(p_i(e^{t-1}))$$

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

故 $P(X>(1+\delta)\mu) \leq \dfrac{\exp(\mu(e^t-1))}{\exp(t(1+\delta)\mu)}$

$$= \exp(\mu(e^t-1-t-t\delta))$$

对 $f(t)=e^t-1-t-t\delta$ 求导，求其最小值：

令 $f'(t)=e^t-1-\delta=0 \Rightarrow t=\ln(1+\delta)$

故 $P(X>(1+\delta)\mu) \leq \exp(\mu(\delta-\ln(1+\delta)-\delta\ln(1+\delta)))$

$$= \left(\frac{e^\delta}{(e^{\ln(1+\delta)})^{(1+\delta)}}\right)^\mu = \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$$

(2). ~~由泰勒公式：~~ $\cancel{\ln(1+\delta) \geq \delta - \frac{\delta^2}{2} = \delta(1-\frac{\delta}{2}) \geq \frac{\delta}{1+\frac{\delta}{2}}}$

考虑函数 $f(x)=\ln(1+x)-\dfrac{2x}{2+x}$，

$f'(x)=\dfrac{x(x+4)}{(1+x)(2+x)^2}$ 当 $x\in(0,1)$ $f'(x)>0$ $f(x)$ 递增。

又 $f(0)=0$ 故 $x\in(0,1)$ 时 $f(x)>0$。

故 $\ln(1+\delta) \geq \dfrac{\delta}{1+\delta/2}$。

故 $P(X>(1+\delta)\mu) < \left(\dfrac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$

$$< \left(\frac{e^{-\delta}}{\exp\left(\frac{\delta(1+\delta)}{1+\frac{\delta}{2}}\right)}\right)^\mu = \left(\exp\left(-\delta+2\delta-\frac{2\delta}{2+\delta}\right)\right)^\mu$$

$$= \exp\left(-\frac{\mu\delta^2}{2+\delta}\right) < \exp\left(-\frac{\mu\delta^2}{3}\right).$$

5. 证：由切诺夫不等式：$P(|X-\mu|>\delta\mu)$

$=P(X-\mu>\delta\mu)+P(X-\mu<-\delta\mu)$

$=P(X>(1+\delta)\mu)+P(X<(1-\delta)\mu)$

$<\exp\left(-\frac{\mu\delta^2}{2}\right)+\exp\left(-\frac{\mu\delta^2}{3}\right)$

$<2\exp\left(-\frac{\mu\delta^2}{3}\right)$

6. 解：$P(|\bar{X}-p|\leq\varepsilon p)\geq 1-\delta \Leftrightarrow P(|\bar{X}-p|>\varepsilon p)\leq\delta$

$\Leftrightarrow P(\bar{X}>(1+\varepsilon)p)+P(\bar{X}<(1-\varepsilon)p)\leq\delta$

$\Leftrightarrow P(X>(1+\varepsilon)\mu)+P(X<(1-\varepsilon)\mu)\leq\delta$

其中 $X=\sum_{i=1}^{n}X_i$. $\mu=\sum_{i=1}^{n}p_i=np$.

由第5题结论：$P(X>(1+\varepsilon)\mu)+P(X<(1-\varepsilon)\mu)<2\exp\left(-\frac{\mu\varepsilon^2}{3}\right)$

故当 $2\exp\left(-\frac{\mu\varepsilon^2}{3}\right)\leq\delta$，题设条件成立.

$2\exp\left(-\frac{\mu\varepsilon^2}{3}\right)\leq\delta$

$\Leftrightarrow -\frac{np\varepsilon^2}{3}\leq\ln\frac{\delta}{2}$

$\Leftrightarrow n\geq\frac{-3\ln\frac{\delta}{2}}{p\varepsilon^2}$

故当 $n\geq -3\ln\frac{\delta}{2}/p\varepsilon^2$ 时题设条件成立.