1_tutorial_solution

anomymous

2019 年 3 月 13 日

1. Let X be a r.v., $\mu = E(X)$ and $\sigma^2 = E[(X-\mu)^2]$. If $X^* = \frac{X-\mu}{\sigma}$, please prove $P[|X^*| \geq c] \leq \frac{1}{c^2}$.

proof:

$$\begin{aligned} P[|X^*| \geq c] &= P(|\frac{x-\mu}{\sigma}| \geq c) \\ &= P(|x-\mu| \geq c\sigma) \end{aligned}$$

根据马尔可夫不等式，

$$P(|x-\mu| \geq c\sigma) \leq \frac{\sigma^2}{c^2\sigma^2} = \frac{1}{c^2} \tag{1}$$

2. Let X_i $(i = 1, 2, \cdots, n)$ be i.i.d., $\mu = E(X_i)$ and $\sigma^2 = E[(X_i-\mu)^2]$. If $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, please prove $P[|\overline{X}-\mu| \geq \epsilon] \leq \frac{\sigma^2}{n\epsilon^2}$.

proof:

$$\begin{aligned} var(\bar{x}) &= var(\frac{1}{n}\sum_{i=1}^{n} X_i) & \qquad E(\bar{x}) &= E(\frac{1}{n}\sum_{i=1}^{n} X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} var(X_i) & &= \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\ &= \frac{n}{n^2} var(X_i) & &= \frac{n}{n} E(X_i) \\ &= \frac{\sigma^2}{n} & &= \mu \end{aligned}$$

应用切比雪夫不等式，

$$P(|\bar{x}-\mu| \geq \sigma) \leq \frac{\sigma^2}{n\epsilon^2} \tag{2}$$

3. In n tosses of a fair coin, let X be the number of heads, what's the probability of $X < \frac{n}{4}$ heads?

proof: 可知$X \in Ber(p), p = \frac{1}{2}$.所以 $E(X) = \frac{1}{2}n$. 通过切诺夫不等式，我们可得

$$P(X < \frac{n}{4}) = P(X < (1 - \frac{1}{2})E(X)) < exp(-\frac{n}{2} \times \frac{\frac{1}{4}}{2}) = e^{-\frac{n}{16}} \tag{3}$$

4. Let X_i be a sequence of independent r.v.s with $P(X_i = 1) = p_i$ and $P(X_i = 0) = 1 - p_i$. r.v. $X = \sum_{i=1}^{n} X_i$ and $\mu = \sum_{i=1}^{n} p_i$. Please prove the following conclusions.

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

proof: 设 $Y_i = e^{tX_i}$, $Y = e^{tX}$, $t > 0$. 则有 $Y = Y_1 Y_2 \cdots Y_n$.

$$\begin{aligned}
E[Y] &= E[Y_1 Y_2 \cdots Y_n] \\
&= \prod_{i=1}^{n} E[Y_i] \ (X_i \ are \ independ.so \ Y_i \ are \ independent) \\
&= \prod_{i=1}^{n} E[e^{tX_i}] \\
&= \prod_{i=1}^{n} (p_i e^t + (1 - p_i)e^0) \ (Law \ of \ total \ probability) \\
&= \prod_{i=1}^{n} (1 - p_i + p_i e^t)
\end{aligned}$$

则

$$
\begin{aligned}
P(X>\mu(1+\delta)) &= P[Y \geq e^{t\mu(1+\delta)}] \\
&\leq \frac{E[Y]}{e^{t\mu(1+\delta)}} \\
&= \frac{\prod_{i=1}^{n}(1-p_i+p_ie^t)}{e^{t\mu(1+\delta)}} \\
&< \frac{\prod_{i=1}^{n}(e^{-p_i(1-e^t)})}{e^{t\mu(1+\delta)}} \quad (bacause\ 1+x<e^x\ (x>0)) \\
&= \frac{e^{-\sum_{i=1}^{n}p_i(1-e^t})}{e^{t\mu(1+\delta)}} \\
&= \frac{e^{-\mu(1-e^t)}}{e^{t\mu(1+\delta)}} \\
&= e^{-\mu[1-e^t+t(1+\delta)]}
\end{aligned}
$$

综上所述呢，就是$P(X>\mu(1+\delta))<e^{-\mu[1-e^t+t(1+\delta)]}$. 取$t=ln(1+\delta)$, 使得$-\mu[1-e^t+t(1+\delta)]$最小，此时有

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

接下来证明不等式2，这里有一个附加条件，$delta$ 在0的极小邻域内。
现在证明$(\frac{e^\delta}{(1+\delta)^(1+\delta)})^\mu < \exp{(-\mu\delta^2/4)}$ 在$\delta\in(0,1]$时成立。
也就是证明$\delta-(1+\delta)ln(1+\delta)\leq\frac{\delta^2}{4}$. 对$f(\delta)=\delta-(1+\delta)ln(1+\delta)+\frac{\delta^2}{4}$求导得：

- $f'(\delta)=-ln(1+\delta)+\frac{\delta}{2}$
- $f"(\delta)=\frac{\delta-1}{2(1+\delta)}$

在$0<\delta\leq 1$时，$f"(\delta)\leq 0$,所以$f"(\delta)$在$(0,1]$递减。
同理可得，$f(x)$ 在$(0,1]$递减。又因为$f(0)=0$.
所以$f(x)<0$在$(0,1]$成立。
所以$(\frac{e^\delta}{(1+\delta)^(1+\delta)})^\mu < \exp{(-\mu\delta^2/4)}$ 在$\delta\in(0,1]$时成立。

5. For the situation of our running example (8 billion bits, 1 billion members of the set S), calculate the false-positive rate if we use three hash functions? What if we use four hash functions?

solution:
- $(1-e^{-\frac{3}{8}})^3, k=3$
- $(1-e^{-\frac{3}{8}})^4, k=4$

6. Suppose we have n bits of memory available, and our set S has m members. Instead of using k hash functions, we could divide the n bits into k arrays, and hash once to each array. As a function of n, m, and k, what is the probability of a false positive? How does it compare with using k hash functions into a single array?

solution:

$$[1-(1-\frac{k}{n})]^k = [1-e^{-\frac{k}{n}n}]^k \tag{4}$$