2019 Lec 3

2019年3月28日 13:46

Chebshev's inequality.

$$\Pr(|X-\mu_x| \geq t\sigma_x)$$

$$= \Pr\left((X-\mu_x)^2 \geq (t\sigma_x)^2\right)$$

$$Y \triangleq (X-\mu_x)^2 \geq 0$$

$$\Pr(Y \geq (t\sigma)^2) \leq \frac{E(Y)}{(t\sigma_x)^2}$$

$$E(Y) = E\left((X-\mu_x)^2\right) = \sigma_x^2$$

---

chernoff bounds

RV: $x_1, x_2, \ldots, x_n$

$$\begin{cases} \Pr(x_i=1) = p_i \\ \Pr(x_i=0) = 1-p_i \end{cases}$$

$$\mu = \sum_{i=1}^{n} p_i \qquad X = \sum_{i=1}^{n} x_i \qquad E(X) = \mu$$

$$\Pr(X > (1+\delta)\mu) \leq \frac{E(X)=\mu}{(1+\delta)\mu} = \frac{1}{1+\delta}$$

$$= \Pr\left(e^{\lambda X} > e^{\lambda(1+\delta)\mu}\right)$$

$$\leq \frac{E(e^{\lambda X})}{e^{\lambda(1+\delta)\mu}}$$

$$\leq \frac{e^{\mu(e^{\lambda}-1)}}{e^{\lambda(1+\delta)\mu}}$$

set $\lambda = \ln(1+\delta)$

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

$$E(e^{\lambda X})$$

$$= E\left(e^{\lambda(X_1+\cdots+X_n)}\right)$$

$$= E\left(\prod_{i=1}^{n} e^{\lambda X_i}\right)$$

independence

$$= \prod_{i=1}^{n} E(e^{\lambda X_i})$$

$$= \prod_{i=1}^{n} \left((1-p_i)\cdot 1 + p_i \cdot e^{\lambda}\right)$$

$$= \prod_{i=1}^{n} \left(1 + p_i(e^{\lambda}-1)\right) \leq \prod_{i=1}^{n} e^{p_i(e^{\lambda}-1)}$$

$$\leq f(\Sigma p_i, \lambda)$$

$$= e^{\mu(e^{\lambda}-1)}$$

$1+x \le e^x$.

balls & bins.

m balls. n bins.

$m=1$, $\Pr(k=1)=1$

$m=2$. $\begin{cases} \Pr(k=1) = 1-\frac{1}{n} \\ \Pr(k=2) = \frac{1}{n} \end{cases}$

$m=?$ $\Pr(k=1) = 1-o(1)$

$m=o(\sqrt{n})$ $\Pr(\max(X_1, \cdots, X_n) \ge 2) = o(1)$.

$$\Pr(X_1 \ge 2 \text{ OR } X_2 \ge 2 \cdots \text{ OR } X_n \ge 2) \le \sum_{i=1}^{n} \Pr(X_i \ge 2)$$

$$= n \cdot \Pr(X_1 \ge 2) = \Theta\left(\frac{m^2}{n}\right) = o(1)$$

$m = o(\sqrt{n})$.

$$\Pr(X_1 \ge 2) \le \boxed{\binom{m}{2}\left(\frac{1}{n}\right)^2} = \Theta\left(\frac{m^2}{n^2}\right)$$

$$\Pr(X_1 \ge 2) = \sum_{k=2}^{m} \Pr(X_1 = k)$$

$$= \sum_{k=2}^{m} \binom{m}{k} \cdot \left(\frac{1}{n}\right)^k \left(1-\frac{1}{n}\right)^{m-k}$$

$$= 1 - \Pr(X_1 = 0) - \Pr(X_1 = 1)$$

$$= 1 - \left(1-\frac{1}{n}\right)^m - m \cdot \frac{1}{n} \cdot \left(1-\frac{1}{n}\right)^{m-1}$$

Case 2. $m = \Theta(\sqrt{n}) = c\sqrt{n}$.

$$\Pr(X_1 \ge 2) \le \binom{m}{2}\left(\frac{1}{n}\right)^2 \approx \frac{c^2}{2n}$$

$$\Pr(k > 1) \le n \cdot \Pr(X_1 \ge 2) \le \frac{c^2}{2}$$

$$\Pr(k=1) = \underbrace{\frac{n-1}{n} \cdot \frac{n-2}{n} \cdot \frac{n-3}{n} \cdots \frac{n-m+1}{n}}_{= \Pr(E_1 \cdots E_m)}$$

$\bar{E}_i$: $= \Pr(\bar{U}_1)\,\Pr(\bar{U}_2 \mid \bar{U}_1)\,\Pr(\bar{U}_3 \mid \bar{U}_1 \bar{U}_2)$

$\qquad\quad 1 \qquad\quad \frac{n-1}{n} \qquad\quad \frac{n-2}{n}$

$1-x \le e^{-x}$

$$= \left(1-\frac{1}{n}\right)\cdot\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{m-1}{n}\right) \qquad \ge \left(1-\frac{m-1}{n}\right)^{m-1}$$

$$\le e^{-\frac{1}{n}}\cdot e^{-\frac{2}{n}}\cdots e^{-\frac{m-1}{n}}$$

$$\approx e^{-\frac{m^2}{2n}} = e^{-\frac{c^2}{2}} < 1$$

$$\Pr(k \ge 2) \ge 1 - e^{-\frac{c^2}{2}}$$

$\left(1-\frac{1}{n}\right)^n \sim \frac{1}{e}$

$\left(1-\frac{m-1}{n}\right)^{m-1} \sim \left(1-\frac{m-1}{n}\right)^{\frac{n}{m-1}\cdot\frac{(m-1)^2}{n}}$

$\sim \left(\frac{1}{e}\right)^{\frac{m^2}{n}}$

$$\Pr(k \ge 3) \le n \cdot \Pr(X_1 \ge 3)$$

$$\le n \cdot \binom{m}{3}\cdot\left(\frac{1}{n}\right)^3 \qquad m = \Theta(n^{\frac{1}{2}})$$

$$= \Theta\left(n \cdot \frac{n^{1.5}}{n^3}\right) = \Theta\left(\frac{1}{n^{0.5}}\right)$$

$$m \approx n^{\frac{2}{3}}$$

$$\Pr(X_1 \ge 2) \le \binom{m}{2}\left(\frac{1}{n}\right)^2$$

$$\Pr(X_1 \ge 2) \ge \Pr(X_1 = 2) = \binom{m}{2}\left(\frac{1}{n}\right)^2\left(1-\frac{1}{n}\right)^{m-2}$$

$$\left(1-\frac{1}{n}\right)^{m-2} = \left(1-\frac{1}{n}\right)^{n\cdot\frac{m-2}{n}} \sim \frac{1}{e^{\frac{m-2}{n}}} \sim \frac{1}{e^{\frac{1}{\sqrt{n}}}}$$

Case 3: $m = n$. Prove $k = \frac{\ln n}{\ln\ln n}$ w.h.p.

1°. $\Pr\left(k \le \frac{\ln n}{\ln\ln n}\right) = 1 - o(1)$, $\quad \frac{\ln n}{\ln\ln n} = x$

$\Pr(k \ge x) = o(1)$

$$\le n \cdot \Pr(X_1 \ge x) \le n \cdot \binom{m}{x}\cdot\left(\frac{1}{n}\right)^x = o(1)$$

$$= n \cdot \binom{n}{x}\left(\frac{1}{n}\right)^x \le n \cdot \left(\frac{en}{x}\right)^x \cdot \left(\frac{1}{n}\right)^x$$

Stirling's formula: $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ $\qquad = n\cdot\left(\frac{e}{x}\right)^x = o(1)$

$$\left(\frac{n}{x}\right)^x \le \binom{n}{x} \le \left(\frac{en}{x}\right)^x$$

$$\left(\frac{n}{x}\right)^x \le \binom{n}{x} \le \left(\frac{en}{x}\right)^x$$

2^0 $\Pr(k \ge c \cdot x) = (-o(1))$.

$\Pr(k \le c \cdot x) = \Pr(\bar{E}_1 \wedge \cdots \wedge \bar{E}_n)$

E_i: $x_i \le c \cdot x$. $\quad Y_i = \begin{cases} 1 & E_i \text{ 没发生} \\ 0. & \bar{E}_i \text{ 发生} \end{cases}$

$$\Pr(k \le c \cdot x) = \Pr(\forall i, Y_i = 0) = \Pr\left(\sum_{i=1}^n Y_i = 0\right)$$

$$\le \Pr\left(\left|\sum_{i=1}^n Y_i - E\left(\sum_{i=1}^n Y_i\right)\right| \ge E\left(\sum_{i=1}^n Y_i\right)\right) \le \frac{\sigma^2\left(\sum_{i=1}^n Y_i\right)}{\left(E\left(\sum_{i=1}^n Y_i\right)\right)^2} \le \cdots$$

① $\sum_{i=1}^n E(Y_i) = n \cdot \Pr(X_1 > cx) \ge n \cdot \Pr(X_1 = cx)$

$$= n \cdot \binom{n}{cx} \cdot \left(\frac{1}{n}\right)^{cx} \cdot \left(1 - \frac{1}{n}\right)^{n \cdot \frac{n-cx}{n}}$$

$$\ge n \cdot \left(\frac{n}{cx}\right)^{cx} \cdot \left(\frac{1}{n}\right)^{cx} \cdot \left(\frac{1}{e}\right)^{\frac{n-cx}{n}} \ge n\left(\frac{1}{cx}\right)^{cx} \cdot \frac{1}{e}$$

$$= \Theta\left(n^{1-c}\right)$$

$$\left(\frac{\ln\ln n}{c\ln n}\right)^{c\frac{\ln n}{\ln\ln n}}$$

$$e^{c \cdot \frac{\ln n}{\ln\ln n} \cdot (\ln\ln\ln n - \ln\ln n - \ln c)}$$

$$= e^{-c\ln n} = \frac{1}{n^c}$$

② $\sigma^2\left(\sum_{i=1}^n Y_i\right)$.

$\mathrm{Var}(X+Y)$

$= \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Cov}(X,Y)$

$\mathrm{Var}\left(\sum_{i=1}^n Y_i\right)$

$$= \sum_{i=1}^n \mathrm{Var}(Y_i) + \sum_{i,j} \underbrace{\mathrm{Cov}(Y_i, Y_j)}_{\le 0}$$

$\le \sum_{i=1}^n (P_i - P_i^2) + 0$

$\le n + 0.$

$\le n.$

$$\Pr(k < cx) = \Pr(Y_1 + \cdots + Y_n = 0)$$

$$\Pr(k < cX) = \Pr(Y_1 + \cdots + Y_n = 0)$$

$$\leq \frac{\mathrm{Var}\left(\sum_{i=1}^{n} Y_i\right)}{E^2\left(\sum_{i=1}^{n} Y_i\right)} = O\left(\frac{n}{(n^{1-c})^2}\right) \sim \frac{1}{n^{\frac{1}{3}}}$$

$$c = \frac{1}{3}$$

$$\frac{\ln n}{3 \ln\ln n} < k < \frac{\ln n}{\ln\ln n}$$ with high probability.