جلسه دوم

خلاصه‌سازی برای مه‌داده

برخی کران‌های احتمالاتی

متغیر تصادفی، امید ریاضی

- X: متغیر تصادفی
- مقادیر X از مجموعه S
- تعریف

$$\mathbb{E}X = \sum_{j \in S} j \cdot \mathbb{P}(X = j).$$

- لم:
- $$\mathbb{E}(aX + bY) = a \cdot \mathbb{E}X + b \cdot \mathbb{E}Y.$$

متغیرهای مستقل

- **Definition** (**Independency** of random variables):
  - P[X=x,Y=y] = P[X=x] . P[Y=y]

- **Lemma**: E[X Y] = E[X] E[Y]
  - اثبات:
  - LHS = $\sum_{x,y} P[X=x, Y=y] \cdot x \cdot y = (\sum_{x} x \cdot P[X=x])(\sum_{y} y \cdot P[Y=y])$

- **Lemma**: Var[X+Y] = Var[X] + Var[Y]
  - اثبات:
  - Var[X+Y] = E[(X+Y)^2] – E[X+Y]^2
    - = E[X^2] + E[Y^2] + 2E[XY] – E[X]^2 – E[Y]^2 – 2E[X]E[Y]
    - = E[X^2]–E[X]^2 + E[Y^2]–E[Y]^2

کران مارکوف

Lemma 2 (Markov)**.** *If X is a nonnegative random variable, then for all $\lambda > 0$,*

$$\mathbb{P}(X > \lambda) < \frac{\mathbb{E}X}{\lambda}$$

اثبات

کران مارکوف

Lemma 2 (Markov)**.** *If X is a nonnegative random variable, then for all $\lambda > 0$,*

$$\mathbb{P}(X > \lambda) < \frac{\mathbb{E}X}{\lambda}$$

اثبات

$$P(X > \lambda) \cdot \lambda$$

$$\wedge$$

$$EX$$

Lemma 3 (Chebyshev). ~~*Under the same conditions as Markov*~~,

$$\mathbb{P}(|X - \mathbb{E}X| > \lambda) < \frac{\mathbb{E}(X - \mathbb{E}X)^2}{\lambda^2} = \frac{\mathrm{Var}[X]}{\lambda^2} \tag{2.3}$$

اثبات

$$\mathbb{P}(|X - \mathbb{E}X| > \lambda) = \mathbb{P}((X - \mathbb{E}X)^2 > \lambda^2).$$

Lemma 2 (Markov). *If X is a nonnegative random variable, then for all $\lambda > 0$,*

$$\mathbb{P}(X > \lambda) < \frac{\mathbb{E}X}{\lambda}$$

Lemma 3 (Chebyshev). ~~*Under the same conditions as Markov*~~,

$$\mathbb{P}(|X - \mathbb{E}X| > \lambda) < \frac{\mathbb{E}(X - \mathbb{E}X)^2}{\lambda^2} = \frac{\mathrm{Var}[X]}{\lambda^2} \tag{2.3}$$

$$\text{ص} = P(\underbrace{(x-EX)^2}_{Y} > \lambda^2)$$

$$< \frac{E[(x-EX)^2]}{\lambda^2}$$

کران چرنوف

Lemma 4 (Chernoff)**.** *Suppose* $X_1, X_2, \ldots, X_n$ *are independent random variables with* $X_i \in [0,1]$*. Let* $X = \sum_{i=1}^{n} X_i$ *with* $\mu = \mathbb{E}X$*. If* $0 < \epsilon < 1$*, then*

$$\mathbb{P}(|X - \mathbb{E}X| > \varepsilon\mu) \leq 2 \cdot e^{-\varepsilon^2 \mu/3}. \tag{2.5}$$

اثبات

- فرض: Xiها برنولی، با احتمال pi
- به جای ۱/۳، هر عدد ثابتی

کران چرنوف

Lemma 4 (Chernoff)**.** *Suppose $X_1, X_2, \ldots, X_n$ are independent random variables with $X_i \in [0,1]$. Let $X = \sum_{i=1}^{n} X_i$ with $\mu = \mathbb{E}X$. If $0 < \epsilon < 1$, then*

$$\mathbb{P}(|X - \mathbb{E}X| > \varepsilon\mu) \leq 2 \cdot e^{-\varepsilon^2 \mu/3}. \tag{2.5}$$

$$Y = \frac{1}{n} X$$

$$\mu_Y = \frac{1}{n} \mu_X$$

$$P(\cancel{n}|Y - EY| > \varepsilon \cancel{n} \mu_Y) \leq 2 \cdot e^{-\varepsilon^2 n \mu_Y / 3}$$

کران چرنوف

Lemma 4 (Chernoff)**.** *Suppose $X_1, X_2, \ldots, X_n$ are independent random variables with $X_i \in [0,1]$. Let $X = \sum_{i=1}^{n} X_i$ with $\mu = \mathbb{E}X$. If $0 < \epsilon < 1$, then*

$$\mathbb{P}(|X - \mathbb{E}X| > \varepsilon\mu) \leq 2 \cdot e^{-\varepsilon^2 \mu/3}. \tag{2.5}$$

اثبات

$\mu = \sum p_i$

- فرض: Xiها برنولی، با احتمال pi
- به جای ۱/۳، هر عدد ثابتی

$$\mathbb{P}(|X - \mathbb{E}X| > \varepsilon\mu) \leq 2 \cdot e^{-\varepsilon^2\mu/3}. \tag{2.5}$$

اثبات یک طرف

$$\mathbb{P}(X > (1+\varepsilon)\mu) =$$

$$P(X - EX > \varepsilon\mu) + (X - EX < -\varepsilon\mu)$$

$$= \mathbb{P}(|X - \mathbb{E}X| > \varepsilon\mu) \leq 2 \cdot e^{-\varepsilon^2\mu/3}. \tag{2.5}$$

اثبات یک طرف

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

$$< E e^{tX} \cdot e^{-t(1+\varepsilon)\mu}$$

$$= E \prod_{x_i} e^{tx_i} \cdot \quad \text{"}$$

$$= \prod E e^{tx_i} \cdot \quad \text{"} \qquad 1 + x \leq e^{x}$$

$$= \prod_i \left(p_i \cdot e^{t} + (1-p_i)\right) \cdot \quad \text{"}$$

$$p_i(e^{t} - 1) + 1$$

$$\leq e^{\sum_i p_i (e^{t} - 1)} \cdot \quad \text{"} = e^{(e^{t}-1)\cdot\mu - t(1+\varepsilon)\mu}$$

$$= e^{\mu(\underbrace{e^{t} - 1}_{\varepsilon} - t - \varepsilon t)}$$

$$= \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{\mu} \simeq e^{-\Theta(\varepsilon^2\mu)} \qquad \left(\ln(1+\epsilon) \geq \frac{2\epsilon}{2+\epsilon}\right)$$

$$e^{\mu(e^t-1-t-\varepsilon t)}$$

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

$$t = \ln(1+\varepsilon)$$

$$= \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{\mu} \simeq e^{-\Theta(\varepsilon^2\mu)} \qquad \left(\ln(1+\epsilon) \geq \frac{2\epsilon}{2+\epsilon}\right)$$

اثبات طرف دیگر

$$\mathbb{P}(X < (1-\varepsilon)\mu) \leq (e^{-\varepsilon}/(1-\varepsilon)^{1-\varepsilon})^{\mu} \simeq e^{-\Theta(\varepsilon^2\mu)}$$

سه کران مهم برای ادامه

Lemma 2 (Markov). *If X is a nonnegative random variable, then for all $\lambda > 0$,*

$$\mathbb{P}(X > \lambda) < \frac{\mathbb{E}X}{\lambda}$$

Lemma 3 (Chebyshev). ~~*Under the same conditions as Markov*~~,

$$\mathbb{P}(|X - \mathbb{E}X| > \lambda) < \frac{\mathbb{E}(X - \mathbb{E}X)^2}{\lambda^2} = \frac{\mathrm{Var}[X]}{\lambda^2} \tag{2.3}$$

Lemma 4 (Chernoff). *Suppose $X_1, X_2, \ldots, X_n$ are independent random variables with $X_i \in [0,1]$. Let $X = \sum_{i=1}^n X_i$ with $\mu = \mathbb{E}X$. If $0 < \epsilon < 1$, then*

$$\mathbb{P}(|X - \mathbb{E}X| > \varepsilon\mu) \leq 2 \cdot e^{-\varepsilon^2 \mu/3}. \tag{2.5}$$

شمارش تقریبی

شمارش تقریبی

- صورت مساله:

  - `init();` sets $n \leftarrow 0$
  - `update();` increments $n \leftarrow n + 1$
  - `query();` outputs n or an estimate of n

- حافظه؟

- کران پایین برای حافظه

- هدف:

  - تقریب n با حافظه کم‌تر

  - و $\mathbb{P}(|\tilde{n} - n| > \varepsilon n) < \delta,$