

信息学中的概率统计

王若松

前沿计算研究中心
北京大学

尾不等式、大数定律与中心极限定理

1. 尾不等式
2. 大数定律
3. 中心极限定理
4. 应用举例

1. 尾不等式

- ▶ 古典概率模型 \Rightarrow 离散随机变量
- ▶ 几何概率模型 \Rightarrow 连续随机变量

- ▶ 概率和频率？
- ▶ 频率: n 次试验中事件发生的比例: $f_n(A) = \frac{n_A}{n}$
- ▶ n_A : 事件 A 发生的次数, n : 总试验次数

- ▶ 频率 $f_n(A)$ 随着 n 增大逐渐趋于一个稳定值, 也即是事件 A 发生的概率
- ▶ 是否有 $\lim_{n \rightarrow \infty} f_n(A) = P(A)$?

1. 尾不等式

- ▶ 回顾：如果某个随机试验只有两个可能的结果 A 和 \bar{A} ，且 $P(A) = p$ ，将试验独立地重复进行 n 次，称为 **n 重伯努利试验**
- ▶ n_A ：事件 A 发生的次数, n ：总试验次数, $f_n(A) = \frac{n_A}{n}$
- ▶ 令 A_i 表示第 i 次试验中 A 是否发生, $n_A = \sum_{i=1}^n 1_{A_i}$
- ▶ $n_A \sim B(n, p)$, $E(n_A) = np$
- ▶ 对于 $\epsilon > 0$
 - ▶ 给出 $P(|f_n(A) - p| \geq \epsilon) = P(|n_A - E(n_A)| \geq n\epsilon)$ 的上界？**尾不等式/集中不等式**
 - ▶ 证明 $\lim_{n \rightarrow \infty} P(|f_n(A) - p| < \epsilon) = 1$? **大数定律**

1. 尾不等式

- ▶ 尾不等式：给定随机变量 X ，给出 $P(X \geq k)$ 的上界
 - ▶ X_i 表示球与桶模型中第*i*个桶中球的数量，证明 $P(\max\{X_1, X_2, \dots, X_n\} \geq 4\log n) \leq 1/n$
 - ▶ $X \sim N(0, 1)$ ，证明 $P(X \geq x) \leq \frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}}$
- ▶ 集中不等式：给定随机变量 X ，给出 $P(|X - E(X)| \geq k)$ 的上界
 - ▶ $X \sim N(\mu, \sigma^2)$ ，证明 $P(|X - \mu| \geq k\sigma) \leq 1 - \frac{e^{-\frac{k^2}{2}}}{k} \cdot \sqrt{\frac{2}{\pi}}$
 - ▶ $Y = \sum_{i=1}^n X_i^2$ ， $X_i \sim N(0, 1)$ 且相互独立，证明 $P(|Y - E(Y)| \geq \Delta n) \leq 2e^{-n\Delta^2/8}$

1. 尾不等式

- ▶ **马尔可夫不等式**: X 为非负随机变量 $P(X \geq a \cdot E(X)) \leq \frac{1}{a}$
- ▶ **切比雪夫不等式**: $P(|X - E(X)| \geq c \cdot \sigma(X)) \leq 1/c^2$
- ▶ 给出 $P(|f_n(A) - p| \geq \epsilon) = P(|n_A - E(n_A)| \geq n\epsilon)$ 的上界?
- ▶ $n_A \sim B(n, p)$, $\sigma(n_A) = \sqrt{np(1-p)}$
- ▶ $P(|n_A - E(n_A)| \geq n\epsilon) \leq \frac{p(1-p)}{n\epsilon^2}$
- ▶ $\lim_{n \rightarrow \infty} P(|f_n(A) - p| < \epsilon) = 1$

1. 尾不等式

- ▶ $n_A \sim B(n, p), \sigma(n_A) = \sqrt{np(1-p)}$
- ▶ $P(|n_A - E(n_A)| \geq n\epsilon) \leq \frac{p(1-p)}{n\epsilon^2}$
- ▶ 能否对 $P(|n_A - E(n_A)| \geq n\epsilon)$ 给出更好的上界?
- ▶ $\text{Var}\left(\sum_{i=1}^n 1_{A_i}\right) = \sum_{i=1}^n \sum_{j=1}^n E((1_{A_i} - p)(1_{A_j} - p)) = \sum_{i=1}^n \text{Var}(1_{A_i}) = np(1-p)$
- ▶ 仅需 A_i 两两独立
- ▶ 对于 n 重伯努利试验, A_i 相互独立
- ▶ 如何利用到不同试验相互独立?

1. 尾不等式

- ▶ 给定随机变量 X , 对于正整数 k
 - ▶ 定义 $E(X^k)$ 为 X 的 k 阶(原点)矩
 - ▶ 定义 $E((X - E(X))^k)$ 为 X 的 k 阶中心矩
- ▶ 数学期望: 一阶矩
- ▶ 方差: 二阶中心矩
- ▶ $E(X^2)$: 二阶矩

- ▶ 切比雪夫不等式: 对 $(X - E(X))^2$ 使用马尔可夫不等式
- ▶ 对 $(X - E(X))^3$ 使用马尔可夫不等式?
- ▶ 对 $(X - E(X))^4$ 使用马尔可夫不等式?

1. 尾不等式

- ▶ 给定随机变量 $X \sim B(n, p)$, 如何计算 $E((X - E(X))^4)$?
- ▶ $E((X - E(X))^4) = E(X^4) - 4E(X)E(X^3) + 6E(X^2)(E(X))^2 - 3(E(X))^4$
- ▶ 如何计算 $E(X^4), E(X^3)$?
- ▶ $X = \sum_{i=1}^n X_i$, X_i 独立同分布且 X_i 服从参数为 p 的伯努利分布
- ▶ $E((X - E(X))^4) = E((\sum_{i=1}^n (X_i - p))^4)$
- ▶ $= \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} E((X_{i_1} - p)(X_{i_2} - p)(X_{i_3} - p)(X_{i_4} - p))$
- ▶ 如何计算 $\sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} E((X_{i_1} - p)(X_{i_2} - p)(X_{i_3} - p)(X_{i_4} - p))$?

1. 尾不等式

- ▶ 给定随机变量 X , 定义 $M_X(t) = E(e^{tX})$ 为 X 的**矩生成函数**
- ▶ 作业二第二题: $M_X(t) = \sum_{i=0}^{+\infty} \frac{t^i}{i!} E(X^i)$
- ▶ $E(X^4) = \frac{d^4 M_X(t)}{dt^4} |_{t=0}$
- ▶ 令 $Y = X - E(X)$, $E((X - E(X))^4) = E(Y^4) = \frac{d^4 M_Y(t)}{dt^4} |_{t=0}$
- ▶ $M_Y(t) = E(e^{t(X-E(X))}) = M_X(t) \cdot e^{-tE(X)}$
 - ▶ 作业二: $M_X(t) = (1 - p + pe^t)^n$, $e^{-tE(X)} = e^{-t \cdot np}$
- ▶ $E(Y^4) = \frac{d^4 M_Y(t)}{dt^4} |_{t=0} = np(1-p)^4 + n(1-p)p^4 + 3n(n-1)p^2(1-p)^2$

1. 尾不等式

- $E((X - E(X))^4) = np(1-p)^4 + n(1-p)p^4 + 3n(n-1)p^2(1-p)^2$
- 当 $p = 1/2$, $E((X - E(X))^4) = \frac{n(3n-2)}{16} = O(n^2)$
- $P(|X - E(X)| \geq n\epsilon) = P((X - E(X))^4 \geq (n\epsilon)^4) \leq \frac{o(n^2)}{(n\epsilon)^4} = o\left(\frac{1}{n^2\epsilon^4}\right)$
- 对比切比雪夫不等式: $P(|X - E(X)| \geq n\epsilon) \leq o\left(\frac{1}{n\epsilon^2}\right)$

- $E((X - E(X))^4) = \sum_{1 \leq i_1, i_2, i_3, i_4 \leq n} E((X_{i_1} - p)(X_{i_2} - p)(X_{i_3} - p)(X_{i_4} - p))$
- 仍然无法完全利用 n 重伯努利试验不同试验相互独立
- 计算六阶中心矩, 八阶中心矩?

1. 尾不等式

- ▶ Chernoff bound: 对 e^{tX} 使用马尔可夫不等式, 而 $M_X(t) = E(e^{tX})$
- ▶ 给定随机变量 X
 - ▶ 对于任意 $t > 0$, $P(X \geq k) \leq M_X(t) \cdot e^{-tk}$
 - ▶ 对于任意 $t < 0$, $P(X \leq k) \leq M_X(t) \cdot e^{-tk}$
- ▶ 证明
 - ▶ $P(X \geq k) = P(e^{tX} \geq e^{tk}) \leq E(e^{tX}) \cdot e^{-tk}$
 - ▶ $P(X \leq k) = P(e^{tX} \geq e^{tk}) \leq E(e^{tX}) \cdot e^{-tk}$
- ▶ 使用时, 选择最优的 t

1. 尾不等式

- ▶ 例: $X \sim \pi(\lambda)$, 给出 $P(X \geq x)$ 的上界
- ▶ 作业二第六题: $M_X(t) = E(e^{tX}) = e^{\lambda(e^t - 1)}$
- ▶ 对于 $t > 0$, $P(X \geq x) = P(e^{tX} \geq e^{tx}) \leq e^{\lambda(e^t - 1) - tx}$
- ▶ 如何最小化 $e^{\lambda(e^t - 1) - tx}$?
 - ▶ 求导, 得到 $\lambda e^t - x = 0 \Rightarrow t = \ln(x/\lambda)$
 - ▶ 因此, 当 $x > \lambda$, $P(X \geq x) \leq e^{\lambda\left(\frac{x}{\lambda} - 1\right) - x \cdot \ln\left(\frac{x}{\lambda}\right)} = \frac{e^{-\lambda}(e\lambda)^x}{x^x}$

1. 尾不等式

► 例: $X \sim N(\mu, \sigma^2)$, 给出 $P(X - E(X) \geq k\sigma)$ 的上界

$$\blacktriangleright E(e^{tX}) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2} + tx} dx = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

$$\blacktriangleright P(X \geq k\sigma + \mu) \leq e^{\mu t + \frac{\sigma^2 t^2}{2}} \cdot e^{-t(\mu + k\sigma)} = e^{\frac{\sigma^2 t^2}{2} - k\sigma t}$$

$$\blacktriangleright \text{最小化 } e^{\frac{\sigma^2 t^2}{2} - k\sigma t} \Rightarrow t = \frac{k}{\sigma}$$

$$\blacktriangleright P(X - E(X) \geq k\sigma) \leq e^{-\frac{k^2}{2}}$$

$$\blacktriangleright \text{对比作业三第五题: } P(X - E(X) \geq k\sigma) \leq \frac{e^{-\frac{k^2}{2}}}{k\sqrt{2\pi}}$$

1. 尾不等式

- ▶ $X \sim B(n, p)$, $M_X(t) = (1 - p + pe^t)^n$
- ▶ $P(X - E(X) \geq n\epsilon) \leq M_X(t) \cdot e^{-t(E(X)+n\epsilon)}$
- ▶ $= (1 - p + pe^t)^n \cdot e^{-nt(p+\epsilon)} = e^{-tn\epsilon} \cdot \left((1 - p)e^{-tp} + p \cdot e^{t(1-p)} \right)^n$
- ▶ 结论: $(1 - p)e^{-tp} + p \cdot e^{t(1-p)} \leq e^{t^2/8}$
- ▶ $P(X - E(X) \geq n\epsilon) \leq e^{-tn\epsilon + \frac{nt^2}{8}}$
 - ▶ 最小化 $-tn\epsilon + \frac{nt^2}{8} \Rightarrow t = 4\epsilon$
 - ▶ $P(X - E(X) \geq n\epsilon) \leq e^{-2n\epsilon^2}$

1. 尾不等式

- ▶ $X \sim B(n, p)$, $M_X(t) = (1 - p + pe^t)^n$
- ▶ $P(X - E(X) \leq -n\epsilon) \leq M_X(t) \cdot e^{-t(E(X)-n\epsilon)}$
- ▶ $= (1 - p + pe^t)^n \cdot e^{-nt(p-\epsilon)} = e^{tn\epsilon} \cdot \left((1 - p)e^{-tp} + p \cdot e^{t(1-p)} \right)^n$
- ▶ $P(X - E(X) \leq -n\epsilon) \leq e^{tn\epsilon + \frac{nt^2}{8}}$
 - ▶ 最小化 $tn\epsilon + \frac{nt^2}{8} \Rightarrow t = -4\epsilon$
 - ▶ $P(X - E(X) \leq -n\epsilon) \leq e^{-2n\epsilon^2}$
- ▶ $P(|X - E(X)| \geq n\epsilon) \leq 2 \cdot e^{-2n\epsilon^2}$
 - ▶ 对比切比雪夫不等式: $P(|X - E(X)| \geq n\epsilon) \leq O\left(\frac{1}{n\epsilon^2}\right)$
 - ▶ 对比四阶中心矩给出的上界: $P(|X - E(X)| \geq n\epsilon) \leq O\left(\frac{1}{n^2\epsilon^4}\right)$

1. 尾不等式

- ▶ Hoeffding引理：若实数随机变量 $a \leq X \leq b$, 则 $E(e^{t(X-E(X))}) \leq e^{\frac{t^2(b-a)^2}{8}}$
 - ▶ $X \sim B(1, p) \Rightarrow E(e^{t(X-E(X))}) = (1-p)e^{-tp} + p \cdot e^{t(1-p)} \leq e^{t^2/8}$
- ▶ Chernoff-Hoeffding不等式：若 $X = \sum_{i=1}^n X_i$, X_i 相互独立且 $a \leq X_i \leq b$, $k \geq 0$
 - ▶ $P(X \geq E(X) + k) \leq e^{-\frac{2k^2}{n(b-a)^2}}$
 - ▶ $P(X \leq E(X) - k) \leq e^{-\frac{2k^2}{n(b-a)^2}}$
- ▶ 若 $X \sim B(n, p)$, 则有
 - ▶ $P(X \geq n(p + \epsilon)) \leq e^{-2n\epsilon^2}$
 - ▶ $P(X \leq n(p - \epsilon)) \leq e^{-2n\epsilon^2}$

1. 尾不等式

- ▶ Hoeffding引理：若实数随机变量 $a \leq X \leq b$, 则 $E(e^{t(X-E(X))}) \leq e^{\frac{t^2(b-a)^2}{8}}$
- ▶ Chernoff-Hoeffding不等式：若 $X = \sum_{i=1}^n X_i$, X_i 相互独立且 $a \leq X_i \leq b$
 - ▶ $P(X \geq E(X) + t) \leq e^{-\frac{2t^2}{n(b-a)^2}}$
 - ▶ $P(X \leq E(X) - t) \leq e^{-\frac{2t^2}{n(b-a)^2}}$
- ▶ 由 Chernoff bound, $P(X \geq E(X) + k) \leq M_{X-E(X)}(t)e^{-kt}$
- ▶ $M_{X-E(X)}(t) = E\left(e^{t(X-E(X))}\right) = \prod_{i=1}^n E\left(e^{t(X_i-E(X_i))}\right) \leq e^{\frac{nt^2(b-a)^2}{8}}$
- ▶ 最小化 $e^{\frac{nt^2(b-a)^2}{8}} \cdot e^{-kt} \Rightarrow t = \frac{4k}{n(b-a)^2}$
- ▶ $P(X \geq E(X) + k) \leq e^{\frac{nt^2(b-a)^2}{8}} e^{-kt} = e^{-\frac{2k^2}{n(b-a)^2}}$

1. 尾不等式

- ▶ 证明尾不等式的手段
- ▶ 直接对 $P(X \geq k)$ 进行缩放
 - ▶ $X \sim N(0, 1)$, $P(X \geq x) \leq e^{-\frac{x^2}{2}} \cdot \frac{1}{x\sqrt{2\pi}}$
- ▶ 计算偶数阶中心矩，使用马尔可夫不等式
 - ▶ $X \sim B(n, p)$, $P(|X - E(X)| \geq n\epsilon) \leq \frac{p(1-p)}{n\epsilon^2}$
 - ▶ $X \sim B\left(n, \frac{1}{2}\right)$, $P(|X - E(X)| \geq n\epsilon) \leq O\left(\frac{1}{n^2\epsilon^4}\right)$

1. 尾不等式

► 计算矩生成函数，使用Chernoff bound

- $X \sim \pi(\lambda)$, $P(X \geq x) \leq \frac{e^{-\lambda}(e\lambda)^x}{x^x}$
- $X \sim N(\mu, \sigma^2)$, $P(X - E(X) \geq k\sigma) \leq e^{-\frac{k^2}{2}}$
- Chernoff-Hoeffding不等式：若 $X = \sum_{i=1}^n X_i$ ， X_i 相互独立且 $a \leq X_i \leq b$
 - $P(X \geq E(X) + k) \leq e^{-\frac{2k^2}{n(b-a)^2}}$
 - $P(X \leq E(X) - k) \leq e^{-\frac{2k^2}{n(b-a)^2}}$

2. 大数定律

- ▶ **伯努利大数定律**: 在 n 重伯努利试验中, 令 n_A 为事件 A 发生的次数, $P(A) = p$
- ▶ 对于任意 $\epsilon > 0$, $\lim_{n \rightarrow \infty} P\left(\left|\frac{n_A}{n} - p\right| < \epsilon\right) = 1$
- ▶ **大数定律的一般形式**: 对于随机变量 $\{X_n\}$, 对于任意 $\epsilon > 0$,
$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| < \epsilon\right) = 1$$
- ▶ 其他大数定律?

2. 大数定律

- ▶ **马尔可夫大数定律:** 若 $\frac{1}{n^2} \text{Var}(\sum_{i=1}^n X_i) \rightarrow 0$, 则 $\{X_n\}$ 服从大数定律
- ▶ 也即对于任意 $\epsilon > 0$, $\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| < \epsilon\right) = 1$
- ▶ 证明: 由切比雪夫不等式
- ▶ $P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| < \epsilon\right) \geq 1 - \frac{\text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right)}{\epsilon^2}$
- ▶ $\text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \text{Var}(\sum_{i=1}^n X_i) \rightarrow 0$

2. 大数定律

- ▶ **马尔可夫大数定律：**若 $\frac{1}{n^2} \text{Var}(\sum_{i=1}^n X_i) \rightarrow 0$, 则 $\{X_n\}$ 服从大数定律
- ▶ 也即对于任意 $\epsilon > 0$, $\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| < \epsilon\right) = 1$
- ▶ 例1： $\{X_n\}$ 两两不相关，且对于任意 i , 有 $\text{Var}(X_i) \leq c$ 。证明 $\{X_n\}$ 服从大数定律。
- ▶ $\frac{1}{n^2} \text{Var}(\sum_{i=1}^n X_i) = \frac{c}{n} \rightarrow 0$
- ▶ 例2： $\{X_n\}$ 为一列同分布且标准差 $\sigma = \sigma(X_i)$ 存在且有限的随机变量。 X_i 仅与 X_{i-1} 和 X_{i+1} 相关。证明 $\{X_n\}$ 服从大数定律。
 - ▶ $\text{Cov}(X_i, X_{i+1}) \leq \sigma^2$
 - ▶ $\text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) \leq n \cdot \sigma^2 + 2(n-1)\sigma^2$
 - ▶ $\frac{1}{n^2} \text{Var}(\sum_{i=1}^n X_i) \rightarrow 0$

2. 大数定律

- ▶ 辛钦大数定律： $\{X_n\}$ 独立同分布，且数学期望 $\mu = E(X_i)$ 存在，则 $\{X_n\}$ 服从大数定律
- ▶ 也即对于任意 $\epsilon > 0$, $\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| < \epsilon\right) = 1$

- ▶ 对比马尔可夫大数定律
- ▶ 需要独立同分布的假设，不需要对方差进行假设

2. 大数定律

- ▶ 随机变量序列的收敛性
- ▶ 令 $\{Y_n\}$ 为一列随机变量, Y 为随机变量。若对于任意 $\epsilon > 0$, $\lim_{n \rightarrow \infty} P(|Y_n - Y| < \epsilon) = 1$, 则称 $\{Y_n\}$ **依概率收敛于** Y , 记为 $Y_n \xrightarrow{P} Y$
- ▶ 独立同分布情况的大数定律: $Y_n = \frac{1}{n} \sum_{i=1}^n X_i, Y = \mu = E(X_i)$
- ▶ 思考: $\{X_n\}$ 独立同分布, X_i 服从柯西分布, 也即 X_i 的概率密度函数为 $f(x) = \frac{1}{\pi(x^2+1)}$, $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$ 是否依概率收敛?

2. 大数定律

- ▶ 随机变量序列的收敛性
- ▶ 通过分布函数来定义收敛性?
- ▶ 要求函数序列 F_n 点点收敛于 F ?
 - ▶ 对于任意 x , $F_n(x) \rightarrow F(x)$
- ▶ 例: 设 $\{X_n\}$ 为一列随机变量, $P\left(X_n = \frac{1}{n}\right) = 1$ 。 $P(X = 0) = 1$ 。 是否有 F_n 点点收敛于 F_X ?

2. 大数定律

- ▶ $\{X_n\}$ 为一列随机变量，分布函数为 $\{F_n(x)\}$ 。 X 为随机变量，分布函数为 $F(x)$ 。
 对于 $F(x)$ 的任意连续点 x ，均有 $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ ，则称 $\{X_n\}$ 依分布收敛于 X ，
 记为 $X_n \xrightarrow{d} X$ 。
- ▶ 定理：依概率收敛 \Rightarrow 依分布收敛
- ▶ 依分布收敛 \Rightarrow 依概率收敛？
 - ▶ $P(X = +1) = P(X = -1) = \frac{1}{2}$, $Y = -X$
- ▶ X 服从单点分布，则 $X_n \xrightarrow{P} X$ 等价于 $X_n \xrightarrow{d} X$

2. 大数定律

- ▶ 给定随机变量 X , 定义 $\phi_X(t) = E(e^{itX})$ 为 X 的**特征函数**
 - ▶ $\phi_X(-it) = M_X(t) = E(e^{tX})$
- ▶ $P(X = c) = 1, \phi_X(t) = e^{itc}$
- ▶ $X \sim \pi(\lambda), M_X(t) = e^{\lambda(e^t - 1)}, \phi_X(t) = e^{\lambda(e^{it} - 1)}$
- ▶ $X \sim N(\mu, \sigma^2), M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}, \phi_X(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}$
- ▶ $X \sim B(n, p), M_X(t) = (1 - p + pe^t)^n, \phi_X(t) = (1 - p + pe^{it})^n$
- ▶ X 服从柯西分布, $f(x) = \frac{1}{\pi(x^2+1)}, M_X(t) = ?$
 - ▶ $\phi_X(t) = e^{-|t|}$

2. 大数定律

- ▶ 给定随机变量 X , 定义 $\phi_X(t) = E(e^{itX})$ 为 X 的**特征函数**
- ▶ 性质:
- ▶ $E(e^{itX})$ 对于任意实数 t 均存在: $|e^{itx}| \leq 1$
- ▶ $\phi_{aX+b}(t) = E(e^{it(aX+b)}) = \phi_X(at) \cdot e^{itb}$
- ▶ X_1, X_2, \dots, X_n 相互独立, $X = \sum_{i=1}^n X_i$, $\phi_X(t) = \prod_{i=1}^n \phi_{X_i}(t)$
- ▶ $\phi_X^{(k)}(0) = E(X^k) \cdot i^k$

2. 大数定律

- ▶ 给定随机变量 X , 定义 $\phi_X(t) = E(e^{itX})$ 为 X 的**特征函数**
- ▶ 唯一性定理: 随机变量的分布函数由其特征函数唯一决定。
- ▶ 例1: $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$, X, Y 相互独立。证明 $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ 。

$$\phi_{X+Y}(t) = e^{i\mu_1 t - \frac{\sigma_1^2 t^2}{2}} \cdot e^{i\mu_2 t - \frac{\sigma_2^2 t^2}{2}} = e^{i(\mu_1 + \mu_2)t - \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}}$$

2. 大数定律

- ▶ 给定随机变量 X , 定义 $\phi_X(t) = E(e^{itX})$ 为 X 的**特征函数**
- ▶ 唯一性定理: 随机变量的分布函数由其特征函数唯一决定。
- ▶ 例2: X_1, X_2, \dots, X_n 独立同分布, X_i 服从柯西分布。计算 $X = \frac{1}{n} \sum_{i=1}^n X_i$ 的分布函数。
 - ▶ $\phi_{X_i}(t) = e^{-|t|}$
 - ▶ 令 $Y = \sum_{i=1}^n X_i$, $\phi_Y(t) = e^{-n|t|}$, $\phi_X(t) = \phi_Y\left(\frac{t}{n}\right) = e^{-|t|/n}$
- ▶ 推广: X_1, X_2, \dots, X_n 独立同分布, X_i 服从柯西分布。 $\sum_{i=1}^n a_i \cdot X_i \sim |a|_1 \cdot X$, X 服从柯西分布, $|a|_1 = \sum_{i=1}^n |a_i|$ 。

2. 大数定律

- ▶ 给定随机变量 X , 定义 $\phi_X(t) = E(e^{itX})$ 为 X 的**特征函数**
- ▶ 连续性定理: $X_n \xrightarrow{d} X$ 等价于 $\phi_{X_n}(t) \rightarrow \phi_X(t)$
- ▶ 例: $X_n \sim \pi(n), Y_n = \frac{X_n - n}{\sqrt{n}}$, 证明 $Y_n \xrightarrow{d} X \sim N(0,1)$
- ▶ $\phi_{X_n}(t) = e^{n(e^{it}-1)}, \phi_{Y_n}(t) = \phi_{X_n}\left(\frac{t}{\sqrt{n}}\right) \cdot e^{-it\sqrt{n}} = e^{n\left(e^{it/\sqrt{n}}-1\right)-it\sqrt{n}}$
- ▶ $e^{it/\sqrt{n}} = 1 + \frac{it}{\sqrt{n}} - \frac{t^2}{2n} + o(1/n)$
- ▶ $\phi_{Y_n}(t) = e^{n\left(\frac{it}{\sqrt{n}}-\frac{t^2}{2n}+o(1/n)\right)-it\sqrt{n}} \rightarrow e^{-\frac{t^2}{2}}$

2. 大数定律

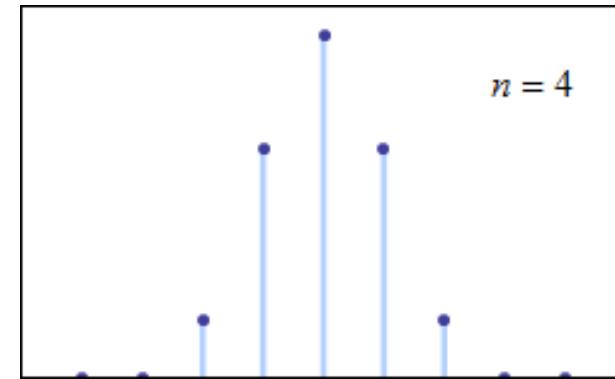
- ▶ 辛钦大数定律： $\{X_n\}$ 独立同分布，且数学期望 $\mu = E(X_i)$ 存在，则 $\{X_n\}$ 服从大数定律
- ▶ 也即对于任意 $\epsilon > 0$, $\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| < \epsilon\right) = 1$
- ▶ 令 $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$ 。 $\phi_{X_i}(t) = 1 + i\mu t + o(t)$
- ▶ $\phi_{Y_n}(t) = \left(1 + \frac{i\mu t}{n} + o\left(\frac{t}{n}\right)\right)^n \rightarrow e^{i\mu t}$

3. 中心极限定理

- ▶ 考虑独立同分布随机变量 $\{X_n\}$, $E(X_n) = \mu$, $\text{Var}(X_n) = \sigma^2$
- ▶ 大数定律: $\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{P} \mu$, 也即 $\frac{\sum_{i=1}^n (X_i - \mu)}{n} \xrightarrow{P} 0$
- ▶ 令 $Y_n = \sum_{i=1}^n X_i$, Y_n 的极限分布是什么?
- ▶ 令 $\tilde{Y}_n = \frac{Y_n - E(Y_n)}{\sigma(Y_n)} = \frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n}\sigma}$ 为 Y_n 的标准化
- ▶ 是否也有 $\tilde{Y}_n \xrightarrow{P} 0$?

3. 中心极限定理

- De Moivre-Laplace 定理：
- $\{X_n\}$ 为独立同分布，服从参数为 p 的伯努利分布的随机变量
- $Y_n = \sum_{i=1}^n X_i \sim B(n, p)$, $\tilde{Y}_n = \frac{Y_n - E(Y_n)}{\sigma(Y_n)} = \frac{\sum_{i=1}^n (X_i - p)}{\sqrt{np(1-p)}}$ 为 Y_n 的标准化
- $\tilde{Y}_n \xrightarrow{d} Z \sim N(0, 1)$
- Y_n 近似服从 $\pi(\lambda)$, $\lambda = np$
- 若 $Y_n \sim \pi(\lambda)$, $\tilde{Y}_n = \frac{Y_n - \lambda}{\sqrt{\lambda}} \xrightarrow{d} Z \sim N(0, 1)$



3. 中心极限定理

- ▶ Lindeberg-Lévy 定理:
- ▶ $\{X_n\}$ 独立同分布, $E(X_n) = \mu$, $\text{Var}(X_n) = \sigma^2$
- ▶ $Y_n = \sum_{i=1}^n X_i$, $\tilde{Y}_n = \frac{Y_n - E(Y_n)}{\sigma(Y_n)} = \frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n}\sigma}$ 为 Y_n 的标准化
- ▶ $\tilde{Y}_n \xrightarrow{d} Z \sim N(0, 1)$

- ▶ $\phi_{X_n - \mu}(t) = 1 - \frac{\sigma^2}{2} t^2 + o(t^2)$
- ▶ $\phi_{\tilde{Y}_n}(t) = \left(\phi_{X_n - \mu} \left(\frac{t}{\sqrt{n}\sigma} \right) \right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \right)^n \rightarrow e^{-\frac{t^2}{2}}$

3. 中心极限定理

- ▶ Berry-Esseen定理：
- ▶ $\{X_n\}$ 独立同分布, $E(X_n) = \mu$, $\text{Var}(X_n) = \sigma^2$, $E(|X_n - \mu|^3)$ 有限
- ▶ $Y_n = \sum_{i=1}^n X_i$, $\tilde{Y}_n = \frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n}\sigma}$ 为 Y_n 的标准化, $Z \sim N(0, 1)$
- ▶ 对于任意 x , $|P(\tilde{Y}_n \leq x) - P(Z \leq x)| \leq O(1) \cdot \frac{E(|X_n - \mu|^3)}{\sigma^3 \sqrt{n}}$
- ▶ 例: $X_n \sim B(1, p)$, $Y_n \sim B(n, p)$, $\tilde{Y}_n = \frac{\sum_{i=1}^n (X_i - p)}{\sqrt{np(1-p)}}$, $Z \sim N(0, 1)$
- ▶ $E(|X_n - \mu|^3) = p(1-p)(p^2 + (1-p)^2)$
- ▶ 对于任意 x , $|P(\tilde{Y}_n \leq x) - P(Z \leq x)| \leq O(1) \cdot \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}}$

4. 应用举例

- ▶ 思考：使用随机数生成器的计算机程序的样本空间
 - ▶ 无限长的0/1随机序列
- ▶ 问题一：某计算机程序有 $1/3$ 的概率崩溃，有 $2/3$ 的概率返回正确的结果
- ▶ 如何通过重复运行提高得到正确结果的概率？
- ▶ 独立地重复运行 T 次，成功概率为 $1 - \frac{1}{3^T}$
- ▶ $T = O(\log(1/\delta)) \Rightarrow$ 成功概率至少为 $1 - \delta$

4. 应用举例

- ▶ 问题二：某计算机程序有 $1/3$ 的概率返回错误的结果，有 $2/3$ 的概率返回正确的结果。假设只有一种正确的结果，错误的结果可能有多种。
- ▶ 如何通过重复运行来提高得到正确结果的概率？

- ▶ 独立地重复运行 T 次，设返回的结果为 $\omega_1, \omega_2, \dots, \omega_T$
- ▶ 输出 $\omega_1, \omega_2, \dots, \omega_T$ 中**出现频率最高的结果**

- ▶ 事件 A_i 表示第 i 次运行返回错误的结果， $P(A_i) = 1/3$ ， $E(1_{A_i}) = 1/3$
- ▶ 出现频率最高的结果为错误结果 $\Rightarrow X = \sum_{i=1}^n 1_{A_i} \geq \frac{T}{2}$

4. 应用举例

- ▶ 问题二：某计算机程序有 $1/3$ 的概率返回错误的结果，有 $2/3$ 的概率返回正确的结果。假设只有一种正确的结果，错误的结果有多种。
- ▶ 如何通过重复运行来提高得到正确结果的概率？
- ▶ 事件 A_i 表示第 i 次运行返回错误的结果， $P(A_i) = 1/3$, $E(1_{A_i}) = 1/3$
- ▶ 出现频率最高的结果为错误结果 $\Rightarrow X = \sum_{i=1}^n 1_{A_i} \geq \frac{T}{2}$
- ▶ $X \sim B(T, 1/3)$, $E(X) = T/3$
- ▶ Chernoff bound: $P(X - E(X) \geq T\epsilon) \leq e^{-2T\epsilon^2}$
- ▶ $\epsilon = 1/6 \Rightarrow P(X \geq T/2) \leq e^{-\frac{T}{18}}$
- ▶ $T = O(\log(1/\delta))$, 成功概率至少为 $1 - \delta$

4. 应用举例

- ▶ 随机快速排序
- ▶ 假设：输入是 $(1, 2, \dots, n)$ 的排列
- ▶ 算法
 - ▶ 在当前数组中随机选择一个主元（pivot） x
 - ▶ 根据元素与 x 大小的关系，将其余元素分为 L 和 R
 - ▶ 递归处理 L 和 R
- ▶ 令随机变量 T 为算法的时间复杂度。如何计算 $E(T)$ ？
- ▶ $T = O\left(\sum_{i < j} 1_{C_{i,j}}\right)$, 其中事件 $C_{i,j}$ 为是否比较过 i 与 j

4. 应用举例

► 算法

- ▶ 在当前数组中随机选择一个主元 (pivot) x
- ▶ 根据元素与 x 大小的关系，将其余元素分为 L 和 R
- ▶ 递归处理 L 和 R
- ▶ $T = O\left(\sum_{i < j} 1_{C_{i,j}}\right)$, 其中事件 $C_{i,j}$ 为是否比较过 i 与 j
- ▶ 算法比较过 i 与 j , 当且仅当 i 或 j 是 $\{i, i + 1, \dots, j\}$ 中首个被选为主元的元素
- ▶ $P(C_{i,j}) = \frac{2}{j-i+1}$
- ▶ $E(T) = \sum_{i < j} \frac{2}{j-i+1} = O(n \log n)$
- ▶ 马尔可夫不等式: $P(T \geq c \cdot E(T)) \leq 1/c$

4. 应用举例

▶ 算法

- ▶ 在当前数组中随机选择一个主元 (pivot) x
- ▶ 根据元素与 x 大小的关系，将其余元素分为 L 和 R
- ▶ 递归处理 L 和 R

▶ 令 D_i 为元素 i 被比较的次数。如何给出 D_i 的尾不等式？

▶ 观察：

- ▶ 若某次选择的主元落在 $[n/4, 3n/4]$ ，则元素 i 所在数组大小至少减少 $1/4$ 。
- ▶ 主元落在 $[n/4, 3n/4]$ 之间的概率为 $1/2$
- ▶ 若选择的主元落在 $[n/4, 3n/4]$ 的次数至少有 $3 \log n$ ，则 i 完成排序
 - $(3/4)^{3 \log n} \leq 1/n$

▶ $P(D_i > 20 \log n) \leq P(X_i < 3 \log n)$, 其中 $X_i \sim B(20 \log n, 1/2)$

4. 应用举例

▶ 算法

- ▶ 在当前数组中随机选择一个主元 (pivot) x
- ▶ 根据元素与 x 大小的关系，将其余元素分为 L 和 R
- ▶ 递归处理 L 和 R
- ▶ 令 D_i 为元素 i 被比较的次数
- ▶ $P(D_i > 20 \log n) \leq P(X_i < 3 \log n)$, 其中 $X_i \sim B(20 \log n, 1/2)$
- ▶ Chernoff-Hoeffding 不等式：若 $X = \sum_{i=1}^n X_i$, X_i 相互独立且 $a \leq X_i \leq b$
 - ▶ $P(X \geq E(X) + k) \leq e^{-\frac{2k^2}{n(b-a)^2}}$
 - ▶ $P(X \leq E(X) - k) \leq e^{-\frac{2k^2}{n(b-a)^2}}$
- ▶ $P(X_i \leq 3 \log n) \leq e^{-4 \log n} \leq 1/n^4$

4. 应用举例

► 算法

- 在当前数组中随机选择一个主元 (pivot) x
- 根据元素与 x 大小的关系，将其余元素分为 L 和 R
- 递归处理 L 和 R
- 令 D_i 为元素 i 被比较的次数
- $P(D_i > 20 \log n) \leq P(X_i \leq 3 \log n)$, 其中 $X_i \sim B(20 \log n, 1/2)$
- $P(X_i \leq 3 \log n) \leq e^{-4 \log n} \leq 1/n^4$
- $P(D_i > 20 \log n) \leq 1/n^4$
- Union Bound: $P\left(\max_i D_i > 20 \log n\right) \leq 1/n^3$
- $P(T > 20 \cdot n \log n) \leq 1/n^3$

4. 应用举例

- ▶ 有 n 个学生，每次选出一些学生进行拔河比赛，共进行 m 次比赛
- ▶ 将全部 n 个学生分为固定的两组，使得 m 次比赛尽量公平

- ▶ 给定 $S_1, S_2, \dots, S_m \subseteq \{1, 2, \dots, n\}$
- ▶ 对于 $\chi: \{1, 2, \dots, n\} \rightarrow \{-1, 1\}$ ，定义 $\text{disc}_\chi(S_i) = |\sum_{j \in S_i} \chi(j)|$
- ▶ 找到 χ 使得 $\max\{\text{disc}_\chi(S_1), \text{disc}_\chi(S_2), \dots, \text{disc}_\chi(S_m)\}$ 尽量小

4. 应用举例

- ▶ 给定 $S_1, S_2, \dots, S_m \subseteq \{1, 2, \dots, n\}$
- ▶ 对于 $\chi: \{1, 2, \dots, n\} \rightarrow \{-1, +1\}$, 定义 $\text{disc}_\chi(S_i) = |\sum_{j \in S_i} \chi(j)|$
- ▶ 找到 χ 使得 $\max\{\text{disc}_\chi(S_1), \text{disc}_\chi(S_2), \dots, \text{disc}_\chi(S_m)\}$ 尽量小
- ▶ 将 $\chi(j)$ 独立等概率设为 -1 或 $+1$
- ▶ 如何给出 $\max\{\text{disc}_\chi(S_1), \text{disc}_\chi(S_2), \dots, \text{disc}_\chi(S_m)\}$ 的上界?
- ▶ 考虑固定的 $i \in \{1, 2, \dots, m\}$, 给出 $\text{disc}_\chi(S_i)$ 的上界
- ▶ 对 $i \in \{1, 2, \dots, m\}$ 使用 Union bound

4. 应用举例

- ▶ 对于 $\chi: \{1, 2, \dots, n\} \rightarrow \{-1, +1\}$, 定义 $\text{disc}_\chi(S_i) = |\sum_{j \in S_i} \chi(j)|$
- ▶ 对于固定的 $i \in \{1, 2, \dots, m\}$
- ▶ $E(\sum_{j \in S_i} \chi(j)) = \sum_{j \in S_i} E(\chi(j)) = 0$
- ▶ Chernoff-Hoeffding不等式: 若 $X = \sum_{i=1}^n X_i$, X_i 相互独立且 $a \leq X_i \leq b$
 - ▶ $P(X \geq E(X) + k) \leq e^{-\frac{2k^2}{n(b-a)^2}}$
 - ▶ $P(X \leq E(X) - k) \leq e^{-\frac{2k^2}{n(b-a)^2}}$
- ▶ $P(\sum_{j \in S_i} \chi(j) \geq k) \leq e^{-\frac{k^2}{2n}}, \quad P(\sum_{j \in S_i} \chi(j) \leq -k) \leq e^{-\frac{k^2}{2n}}$
- ▶ $P(\text{disc}_\chi(S_i) \geq k) \leq 2 \cdot e^{-\frac{k^2}{2n}}$

4. 应用举例

- ▶ 给定 $S_1, S_2, \dots, S_m \subseteq \{1, 2, \dots, n\}$
- ▶ 对于 $\chi: \{1, 2, \dots, n\} \rightarrow \{-1, +1\}$, 定义 $\text{disc}_\chi(S_i) = |\sum_{j \in S_i} \chi(j)|$
- ▶ 找到 χ 使得 $\max\{\text{disc}_\chi(S_1), \text{disc}_\chi(S_2), \dots, \text{disc}_\chi(S_m)\}$ 尽量小
- ▶ 将 $\chi(j)$ 独立等概率设置为 -1 或 $+1$
- ▶ 对于固定的 $i \in \{1, 2, \dots, m\}$, $P(\text{disc}_\chi(S_i) \geq k) \leq 2e^{-\frac{k^2}{2n}}$
- ▶ $P(\max\{\text{disc}_\chi(S_1), \text{disc}_\chi(S_2), \dots, \text{disc}_\chi(S_m)\} \geq k) \leq 2m \cdot e^{-\frac{k^2}{2n}}$
- ▶ $P\left(\max\{\text{disc}_\chi(S_1), \text{disc}_\chi(S_2), \dots, \text{disc}_\chi(S_m)\} \geq \sqrt{n \log m}\right) \leq \frac{1}{2}$
- ▶ 如何设计成功概率至少为 $1 - \delta$ 的算法?

4. 应用举例

- ▶ 给定数据 $x_1, x_2, \dots, x_n \in \mathbb{R}^d$
- ▶ 设计映射 $F: \mathbb{R}^d \rightarrow \mathbb{R}^k$, 使得对于任意 $1 \leq i, j \leq n$,
 - ▶ $(1 - \epsilon)|x_i - x_j|_2^2 \leq |F(x_i) - F(x_j)|_2^2 \leq (1 + \epsilon)|x_i - x_j|_2^2$
- ▶ 应用: 压缩高维数据到低维 (k 尽量小), 保留距离信息
- ▶ 构造随机映射 $F: \mathbb{R}^d \rightarrow \mathbb{R}^k$, 对固定的 i, j , 证明 $|F(x_i) - F(x_j)|_2^2 \in (1 \pm \epsilon)|x_i - x_j|_2^2$
- ▶ 对全部 i, j 使用 Union bound
- ▶ 如何构造随机映射 $F: \mathbb{R}^d \rightarrow \mathbb{R}^k$?
- ▶ 回顾: 若 X_i 独立同分布, 且 $X_i \sim N(0, 1)$, 则 $\sum_{i=1}^d a_i X_i \sim N(0, |a|_2^2)$

4. 应用举例

- ▶ 设计映射 $F: \mathbb{R}^d \rightarrow \mathbb{R}^k$, 使得对于任意 $1 \leq i, j \leq n$,
 - ▶ $(1 - \epsilon)|x_i - x_j|_2^2 \leq |F(x_i) - F(x_j)|_2^2 \leq (1 + \epsilon)|x_i - x_j|_2^2$
- ▶ 构造随机映射 $F: \mathbb{R}^d \rightarrow \mathbb{R}^k$, 对固定的 i, j , 证明 $|F(x_i) - F(x_j)|_2^2 \in (1 \pm \epsilon)|x_i - x_j|_2^2$
- ▶ 回顾: 若 X_i 独立同分布, 且 $X_i \sim N(0, 1)$, 则 $\sum_{i=1}^d a_i X_i \sim N(0, |a|_2^2)$
- ▶ 随机向量 A , 独立同分布, 且 $A_i \sim N(0, 1)$ 。 $F(x) = A^T x$
- ▶ $F(x_i) - F(x_j) = A^T(x_i - x_j)$
- ▶ $A^T(x_i - x_j) \sim N(0, |x_i - x_j|_2^2)$
- ▶ $E\left(\left(F(x_i) - F(x_j)\right)^2\right) = |x_i - x_j|_2^2$ 。
- ▶ 如何证明尾不等式?

4. 应用举例

- ▶ A 为 $k \times d$ 矩阵， A 每个元素均服从 $N(0,1)$ ，且不同元素相互独立
- ▶ 思路：设 $F(x) = Ax$ ，则有 $F(x_i) - F(x_j) = A(x_i - x_j)$
- ▶ 考虑固定的 $\Delta x = x_i - x_j$ ，
- ▶ 证明以较高概率有 $|\Delta y|_2^2 \in (1 \pm \epsilon) |\Delta x|_2^2$ ， $\Delta y = F(x_i) - F(x_j) = A \cdot \Delta x$
- ▶ 再对全部的 $\frac{n(n-1)}{2}$ 对 x_i 和 x_j 使用Union bound
- ▶ 对于固定向量 $\Delta x \in \mathbb{R}^d$ ，向量 $\Delta y = A \cdot \Delta x$ 服从何种分布？
- ▶ $(\Delta y)_i \sim N(0, |\Delta x|_2^2)$ ，且 Δy 不同元素相互独立。 $|\Delta y|_2^2$ 的分布？
- ▶ $E(|\Delta y|_2^2) = k \cdot |\Delta x|_2^2$ 。 $|\Delta y|_2^2 / |\Delta x|_2^2 \sim \chi^2(k)$
- ▶ 定义 $F(x) = \frac{1}{\sqrt{k}} Ax$

4. 应用举例

- ▶ 定义 $F(x) = \frac{1}{\sqrt{k}} Ax$
- ▶ A 为 $k \times d$ 矩阵， A 每个元素均服从 $N(0,1)$ ，且不同元素相互独立
- ▶ 对于固定向量 $\Delta x \in \mathbb{R}^d$ ，向量 $\Delta y = A \cdot \Delta x$ 服从何种分布？
 - ▶ $(\Delta y)_i \sim N(0, |\Delta x|_2^2)$ ，且 Δy 不同元素相互独立
- ▶ 令 $z = \frac{\Delta y}{|\Delta x|_2}$ ，则 $z_i \sim N(0, 1)$ ，且 z 的不同元素相互独立
- ▶ $|z|_2^2 = z_1^2 + z_2^2 + \dots + z_k^2 = \frac{|\Delta y|_2^2}{|\Delta x|_2^2} \sim \chi^2(k)$
- ▶ $E(|z|_2^2) = k$ ， $P(|z|_2^2 \geq (1 + \epsilon) \cdot k) \leq e^{-k\epsilon^2/8}$ ， $P(|z|_2^2 \leq (1 - \epsilon) \cdot k) \leq e^{-k\epsilon^2/8}$
- ▶ $P(|z|_2^2 \notin (1 \pm \epsilon)k) \leq 2e^{-k\epsilon^2/8} \Rightarrow P(|\Delta y|_2^2 / |\Delta x|_2^2 \notin (1 \pm \epsilon)k) \leq 2e^{-k\epsilon^2/8}$
- ▶ $P(|\Delta y|_2^2 \notin (1 \pm \epsilon) \cdot k \cdot |\Delta x|_2^2) \leq 2e^{-k\epsilon^2/8} \Rightarrow P(|A \cdot \Delta x|_2^2 \notin (1 \pm \epsilon) \cdot k \cdot |\Delta x|_2^2) \leq 2e^{-k\epsilon^2/8}$
- ▶ $P\left(\left|\frac{1}{\sqrt{k}} A \cdot \Delta x\right|_2^2 \notin (1 \pm \epsilon) |\Delta x|_2^2\right) \leq 2e^{-k\epsilon^2/8}$

4. 应用举例

- ▶ 定义 $F(x) = \frac{1}{\sqrt{k}}Ax$, $F(x_i) - F(x_j) = \frac{1}{\sqrt{k}}A(x_i - x_j)$
- ▶ 对固定的 $1 \leq i, j \leq n$, 令 $\Delta x = x_i - x_j$,
 - ▶ $P\left(\left|\frac{1}{\sqrt{k}}A \cdot \Delta x\right|_2^2 \notin (1 \pm \epsilon)|\Delta x|_2^2\right) \leq 2e^{-k\epsilon^2/8}$
 - ▶ $P\left(\left|\frac{1}{\sqrt{k}}A(x_i - x_j)\right|_2^2 \notin (1 \pm \epsilon)|x_i - x_j|_2^2\right) \leq 2e^{-k\epsilon^2/8}$
 - ▶ $P\left(|F(x_i) - F(x_j)|_2^2 \notin (1 \pm \epsilon)|x_i - x_j|_2^2\right) \leq 2e^{-k\epsilon^2/8}$
- ▶ 事件 E 表示: $|F(x_i) - F(x_j)|_2^2 \in (1 \pm \epsilon)|x_i - x_j|_2^2$ 对全部 $1 \leq i, j \leq n$ 成立
- ▶ $P(\overline{E}) \leq 2e^{-\frac{k\epsilon^2}{8}} \cdot \frac{n(n-1)}{2} \leq e^{-\frac{k\epsilon^2}{8}} \cdot n^2$
- ▶ $k = O\left(\frac{\log n}{\epsilon^2}\right) \Rightarrow P(E) \geq \frac{1}{2}$

4. 应用举例

- ▶ Johnson-Lindenstrauss Lemma
- ▶ 给定 $x_1, x_2, \dots, x_n \in \mathbb{R}^d$, 存在 $F: \mathbb{R}^d \rightarrow \mathbb{R}^k$, $k = O\left(\frac{\log n}{\epsilon^2}\right)$
- ▶ 对于任意 $1 \leq i, j \leq n$, $(1 - \epsilon)|x_i - x_j|_2^2 \leq |F(x_i) - F(x_j)|_2^2 \leq (1 + \epsilon)|x_i - x_j|_2^2$
- ▶ $F = \frac{1}{\sqrt{k}}Ax$, A 每个元素均服从 $N(0,1)$, 且不同元素相互独立
- ▶ 等价形式: $F = Ax$, A 每个元素均服从 $N(0,1/k)$, 且不同元素相互独立
- ▶ F 与数据 x_1, x_2, \dots, x_n 无关, 可被高效构造
- ▶ F 是一个线性变换
- ▶ 最终维度与初始维度 d 无关, 与数据数量 n 仅为对数关系