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 $\lim_{n\to\infty} f_n(A) = P(A)$?

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$\overline{A}$，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 $\lim_{n\to\infty} P(|f_n(A) - p| < \epsilon) = 1$? **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$P(\max\{X_1, X_2, \dots, X_n\} \geq 4\log n) \leq 1/n$
  - $X \sim \pi(\lambda)$，证明$P(X \geq x) \leq \frac{e^{-\lambda}(e\lambda)^x}{x^x}$
  - $X \sim N(0, 1)$，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
  - $X \sim \pi(\lambda)$，证明$P(|X - \lambda| \geq 0.2\lambda) \leq 2 \cdot e^{-0.01\lambda}$
  - $X \sim N(\mu, \sigma^2)$，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

$E(X) \geq P(X \geq aE(X)) \cdot aE(X) + P(X < aE(X)) \cdot X$

$P(X \geq aE(X)) \leq \frac{1}{a}$.

- **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**：$P(|X - E(X)| \geq c \cdot \sigma(X)) \leq 1/c^2$ 平方即证

- 给出 $P(|f_n(A) - p| \geq \epsilon) = P(|n_A - E(n_A)| \geq n\epsilon)$的上界?
- $n_A \sim B(n,p)$， $\sigma(n_A) = \sqrt{np(1-p)}$
- $P(|n_A - E(n_A)| \geq n\epsilon) \leq \frac{p(1-p)}{n\epsilon^2}$ ✓
- $\lim_{n\to\infty} P(|f_n(A) - p| < \epsilon) = 1$ ✓

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?

- $\mathrm{Var}\left(\sum_{i=1}^n 1_{A_i}\right) = \sum_{i=1}^n \sum_{j=1}^n E((1_{A_i} - p)(1_{A_j} - p)) = \sum_{i=1}^n \mathrm{Var}\left(1_{A_i}\right) = np(1-p)$ ✓
- 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$\left(X-E(X)\right)^2$使用马尔可夫不等式
- 对$\left(X-E(X)\right)^3$使用马尔可夫不等式?
- 对$\left(X-E(X)\right)^4$使用马尔可夫不等式?

北京大学 PEKING UNIVERSITY

1. 尾不等式

- 给定随机变量$X \sim B(n,p)$，如何计算$E\left(\left(X-E(X)\right)^4\right)$?

对称系数

- $E\left(\left(X-E(X)\right)^4\right)=E(X^4)-4E(X)E(X^3)+6E(X^2)\left(E(X)\right)^2-3\left(E(X)\right)^4$
- 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
- $E\left(\left(X-E(X)\right)^4\right)=E\left(\left(\sum_{i=1}^{n}(X_i-p)\right)^4\right)$
- $=\sum_{1\le i_1,i_2,i_3,i_4\le n}E\left((X_{i_1}-p)(X_{i_2}-p)(X_{i_3}-p)(X_{i_4}-p)\right)$
- 如何计算 $\sum_{1\le i_1,i_2,i_3,i_4\le n}E\left(\left(X_{i_1}-p\right)\left(X_{i_2}-p\right)\left(X_{i_3}-p\right)\left(X_{i_4}-p\right)\right)$?

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： $M_X(t)=\sum_{i=0}^{+\infty}\frac{t^i}{i!}E(X^i)$

proof: $e^{tX}=1+tX+\frac{1}{2!}(tX)^2+\cdots$

$E(e^{tX})=1+tE(X)+\frac{1}{2!}t^2E(X^2)+\cdots$

$=\sum\frac{t^i}{i!}\cdot E(X^i)$.

- $E(X^4)=\frac{d^4M_X(t)}{dt^4}|_{t=0}$ 消$t^5,t^6,\cdots$
- 令 $Y=X-E(X)$，$E\left(\left(X-E(X)\right)^4\right)=E(Y^4)=\frac{d^4M_Y(t)}{dt^4}|_{t=0}$
- $M_Y(t)=E\left(e^{t(X-E(X))}\right)=M_X(t)\cdot e^{-tE(X)}$
  - 作业二：$M_X(t)=(1-p+pe^t)^n$， $e^{-tE(X)}=e^{-t\cdot np}$

- $E(Y^4)=\frac{d^4M_Y(t)}{dt^4}|_{t=0}=np(1-p)^4+n(1-p)p^4+3n(n-1)p^2(1-p)^2$

1. 尾不等式

- $E\left((X-E(X))^4\right) = np(1-p)^4 + n(1-p)p^4 + 3n(n-1)p^2(1-p)^2$
- 当$p = 1/2$， $E\left((X-E(X))^4\right) = \frac{n(3n-2)}{16} = O(n^2)$
- $P(|X-E(X)| \ge n\epsilon) \overset{=}{\cancel{\le}} P\left((X-E(X))^4 \ge (n\epsilon)^4\right) \overset{\text{Markov}}{\le} \frac{O(n^2)}{(n\epsilon)^4} = O\left(\frac{1}{n^2\epsilon^4}\right)$ ✓
  - 对比切比雪夫不等式： $P(|X-E(X)| \ge n\epsilon) \le O\left(\frac{1}{n\epsilon^2}\right)$

- $E\left((X-E(X))^4\right) = \sum_{1\le i_1,i_2,i_3,i_4\le n} E((X_{i_1}-p)(X_{i_2}-p)(X_{i_3}-p)(X_{i_4}-p))$
- 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?

北京大学 PEKING UNIVERSITY

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

  Markov
  - $P(X \geq k) = P\left(e^{tX} \geq e^{tk}\right) \leq E(e^{tX}) \cdot e^{-tk}$ $(t>0)$
  - $P(X \leq k) = P\left(e^{tX} \geq e^{tk}\right) \leq E(e^{tX}) \cdot e^{-tk}$ $(t<0)$

- 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：$E(e^{tX}) = e^{\lambda(e^t-1)}$
- 对于$t > 0$，$P(X \geq x) = P(e^{tX} \geq e^{tx}) \leq \underbrace{e^{\lambda(e^t-1)-tx}}_{M_X(t)}$

- 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$x > \lambda$，$P(X \geq x) \leq e^{\lambda\left(\frac{x}{\lambda}-1\right)-x\cdot\ln\left(\frac{x}{\lambda}\right)} = \dfrac{e^{-\lambda}(e\lambda)^x}{x^x}$

北京大学 PEKING UNIVERSITY

1. 尾不等式

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

$$= -\frac{1}{2\sigma^2}\left[x^2 - 2\mu x + \mu^2 - 2t\sigma^2 x\right]$$

$$= -\frac{1}{2\sigma^2}\left\{\underbrace{\left[x-(\mu+t\sigma^2)\right]^2}_{\text{正则性}} \underbrace{- t^2\sigma^4 - 2\mu t\sigma^2}_{\text{外提}}\right\}$$

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

$$= e^{\frac{t^2\sigma^2}{2} + \mu t}.$$

- $P(X \geq k\sigma + \mu) \leq e^{\mu t + \frac{\sigma^2 t^2}{2}} \cdot e^{-t(\mu + k\sigma)} = e^{\frac{\sigma^2 t^2}{2} - k\sigma t}$
- 最小化 $e^{\frac{\sigma^2 t^2}{2} - k\sigma t} \Rightarrow t = \frac{k}{\sigma}$
- $P(X - E(X) \geq k\sigma) \leq e^{-\frac{k^2}{2}}$
  - 对比作业三第三题：$P(|X - E(X)| \geq k\sigma) \leq \frac{e^{-\frac{k^2}{2}}}{k\sqrt{2\pi}}$

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

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

1. 尾不等式

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

1. 尾不等式

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

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
  - $X \sim N(0,1)$，$P(X \geq x) \leq e^{-\frac{x^2}{2}} \cdot \frac{1}{x\sqrt{2\pi}}$ 作业 §3.3

- 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 $a \leq X_i \leq b$
    - $P(X \geq E(X) + k) \leq e^{-\frac{2k^2}{n(b-a)^2}}$
    - $P(X \geq E(X) - k) \leq e^{-\frac{2k^2}{n(b-a)^2}}$

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$\epsilon > 0$, $\lim_{n\to\infty} P\left(\left|\frac{n_A}{n} - p\right| < \epsilon\right) = 1$

- 大数定律的一般形式：对于随机变量$\{X_n\}$，对于任意$\epsilon > 0$,
$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \epsilon\right) = 1$

- 其他大数定律?

北京大学 PEKING UNIVERSITY

2. 大数定律

- **马尔可夫大数定律**：若$\frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^{n} X_i) \to 0$，则$\{X_n\}$服从大数定律
- 也即对于任意$\epsilon > 0$, $\lim_{n\to\infty} P\left(\left|\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{Y} - \underbrace{\frac{1}{n}\sum_{i=1}^{n} E(X_i)}_{E(Y)}\right| < \epsilon\right) = 1$

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

$= \frac{\frac{1}{n^2} Var(\sum X_i)}{\varepsilon^2}$

- 证明： 由切比雪夫不等式
- $P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \epsilon\right) \geq 1 - \frac{\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)}{\epsilon^2}$
- $\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^{n} X_i) \to 0$

2. 大数定律

- **马尔可夫大数定律**：若$\frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^{n} X_i) \to 0$，则$\{X_n\}$服从大数定律
- 也即对于任意$\epsilon > 0,\ \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \epsilon\right) = 1$
- 例1：$\{X_n\}$两两不相关，且对于任意i，有$\mathrm{Var}(X_i) \le c$。证明$\{X_n\}$服从大数定律。
- $\frac{1}{n^2}\mathrm{Var}(\sum_{i=1}^{n} X_i) = \frac{c}{n} \to 0$ ✓

- 例2：$\{X_n\}$为一列同分布且标准差 $\sigma = \sigma(X_i)$存在的随机变量。 X_i仅与 X_{i-1}和 X_{i+1}相关。证明$\{X_n\}$服从大数定律。
- $\mathrm{Cov}(X_i, X_{i+1}) \le \sigma^2$ $|\mathrm{Corr}| \le 1$，送根.
- $\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) \le n \cdot \sigma^2 + 2(n-1)\sigma^2$
- $\frac{1}{n^2}\mathrm{Var}(\sum_{i=}^{n} X_i) \to 0$

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$\epsilon > 0,\ \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \epsilon\right) = 1$

- 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$\epsilon > 0$, $\lim_{n\to\infty} P(|Y_n - Y| < \epsilon) = 1$，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？

柯西分布期望不存在，不能用大数定律.

PEKING UNIVERSITY

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， $F_n(x) \to F(x)$

- 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$\lim_{n\to\infty} F_n(x) = F(x)$，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 ⇒ 依分布收敛
- 依分布收敛 ⇒ 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$\phi_X(t)=E\left(e^{itX}\right)$为 X的**特征函数**
  - $\phi_X(-it)=M_X(t)=E(e^{tX})$

- $P(X=c)=1$， $\phi_X(t)=e^{itc}$
- $X\sim\pi(\lambda)$， $M_X(t)=e^{\lambda(e^t-1)}$， $\phi_X(t)=e^{\lambda(e^{it}-1)}$
- $X\sim N(\mu,\sigma)$， $M_X(t)=e^{\mu t+\frac{\sigma^2t^2}{2}}$， $\phi_X(t)=e^{i\mu t-\frac{\sigma^2t^2}{2}}$
- $X\sim B(n,p)$， $M_X(t)=(1-p+pe^t)^n$， $\phi_X(t)=\left(1-p+pe^{it}\right)^n$
- X服从柯西分布， $f(x)=\frac{1}{\pi(x^2+1)}$， $M_X(t)=?$
  - $\phi_X(t)=e^{-|t|}$

北京大学 PEKING UNIVERSITY

2. 大数定律

- 给定随机变量X，定义$\phi_X(t) = E\left(e^{itX}\right)$为 X的**特征函数**

- 性质：
- $E\left(e^{itX}\right)$对于任意实数t均存在： $\left|e^{itx}\right| \leq 1$
- $\phi_{aX+b}(t) = E\left(e^{it(aX+b)}\right) = \phi_X(at) \cdot e^{itb}$
- $X_1, X_2, \ldots X_n$相互独立， $X = \sum_{i=1}^{n} X_i$， $\phi_X(t) = \prod_{i=1}^{n} \phi_{X_i}(t)$
- $\phi_X^{(k)}(0) = E\left(X^k\right) \cdot i^k$

PEKING UNIVERSITY 北京大学

2. 大数定律

- 给定随机变量X，定义$\phi_X(t) = E\left(e^{itX}\right)$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：$X \sim N(\mu_1, \sigma_1), Y \sim N(\mu_2, \sigma_2)$，$X, Y$相互独立。证明$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$。
- $\phi_{X+Y}(t) = e^{i\mu_1 t - \frac{\sigma_1^2 t^2}{2}} \cdot e^{i\mu_2 t - \frac{\sigma_2^2 t^2}{2}} = e^{i(\mu_1+\mu_2)t - \frac{\left(\sigma_1^2+\sigma_2^2\right)t^2}{2}}$
- 唯一性定理： $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

PEKING UNIVERSITY

2. 大数定律

- 给定随机变量X，定义$\phi_X(t) = E\left(e^{itX}\right)$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$Y = \sum_{i=1}^{n} X_i$， $\phi_Y(t) = e^{-n|t|}$， $\phi_X(t) = \phi_Y\left(\frac{t}{n}\right) = e^{-|t|}$

- 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$\phi_X(t) = E\left(e^{itX}\right)$为 X的**特征函数**
- 连续性定理： $X_n \xrightarrow{d} X$等价于 $\phi_{X_n}(t) \to \phi_X(t)$

- 例： $X_n \sim \pi(n), Y_n = \frac{X_n - n}{\sqrt{n}}$，证明$X_n \xrightarrow{d} X \sim N(0,1)$

- $\phi_{X_n}(t) = e^{n(e^{it}-1)}, \phi_{Y_n}(t) = \phi_{X_n}\left(\frac{t}{\sqrt{n}}\right) \cdot e^{-it\sqrt{n}} = e^{n\left(e^{it/\sqrt{n}}-1\right)-it\sqrt{n}}$
- $e^{it/\sqrt{n}} = 1 + \frac{it}{\sqrt{n}} - \frac{t^2}{2n} + o(1/n)$
- $\phi_{X_n}(t) = e^{n\left(\frac{it}{\sqrt{n}}-\frac{t^2}{2n}+o(1/n)\right)-it\sqrt{n}} \to e^{-\frac{t^2}{2}}$

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$\epsilon > 0,\ \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \epsilon\right) = 1$

- 令 $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$。 $\phi_{X_i}(t) = 1 + i\mu t + o(t)$
- $\phi_{Y_n}(t) = \left(1 + \frac{i\mu t}{n} + o\left(\frac{t}{n}\right)\right)^n \to e^{i\mu t}$

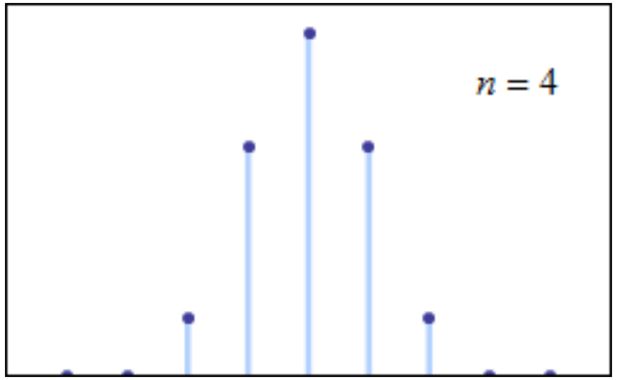
3. 中心极限定理

- 考虑独立同分布随机变量$\{X_n\}$， $E(X_n)=\mu$， $\mathrm{Var}(X_n)=\sigma^2$
- 大数定律：$\frac{\sum_{i=1}^{n} X_n}{n} \xrightarrow{P} \mu$，也即 $\frac{\sum_{i=1}^{n}(X_n-\mu)}{n} \xrightarrow{P} 0$

- 令 $Y_n=\sum_{i=1}^{n} X_n$， Y_n的极限分布是什么?
- 令$\tilde{Y}_n=\frac{Y_n-E(Y_n)}{\sigma(Y_n)}=\frac{\sum_{i=1}^{n}(X_n-\mu)}{\sqrt{n}\sigma}$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
- $Y_n = \sum_{i=1}^{n} X_n \sim B(n,p), \tilde{Y}_n = \frac{Y_n - E(Y_n)}{\sigma(Y_n)} = \frac{\sum_{i=1}^{n}(X_n - p)}{\sqrt{np(1-p)}}$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，$E(X_n) = \mu$， $\mathrm{Var}(X_n) = \sigma^2$
- $Y_n = \sum_{i=1}^{n} X_n$, $\tilde{Y}_n = \frac{Y_n - E(Y_n)}{\sigma(Y_n)} = \frac{\sum_{i=1}^{n}(X_n-\mu)}{\sqrt{n}\sigma}$ 为 Y_n 的标准化
- $\tilde{Y}_n \xrightarrow{d} Z \sim N(0,1)$
- $\phi_{X_n-\mu}(t) = 1 - \frac{\sigma^2}{2}t^2 + o(t^2)$
- $\phi_{\tilde{Y}_n}(t) = \left(\phi_{X_n-\mu}\left(\frac{t}{\sqrt{n}\sigma}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \to e^{-\frac{t^2}{2}}$

3. 中心极限定理

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

- 例： $X_n\sim B(1,p)$， $Y_n\sim B(n,p)$， $\tilde{Y}_n=\frac{\sum_{i=1}^{n}(X_n-p)}{\sqrt{np(1-p)}}$， $Z\sim N(0,1)$
- $E(|X_n-\mu|^3)=p(1-p)(p^2+(1-p)^2)$
- 对于任意x， $\left|P\left(\tilde{Y}_n\le x\right)-P(Z\le x)\right|\le O(1)\cdot\frac{p^2+(1-p)^2}{\sqrt{np(1-p)}}$

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$1-\frac{1}{3^T}$
- $T=O\left(\log(1/\delta)\right)\Rightarrow$成功概率至少为$1-\delta$

北京大学 PEKING UNIVERSITY

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，$P(A_i) = 1/3$， $E\left(1_{A_i}\right) = 1/3$
- 出现频率最高的结果为错误结果⇒ $X = \sum_{i=1}^{n} 1_{A_i} \geq \frac{T}{2}$ 变成众数

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，$P(A_i) = 1/3$，$E\left(1_{A_i}\right) = 1/3$
- 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：$P(X - E(X) \geq T\epsilon) \leq e^{-2T\epsilon^2}$
- $\epsilon = 1/6 \Rightarrow P(X \geq T/2) \leq e^{-\frac{T}{18}}$
- $T = O\left(\log(1/\delta)\right)$，成功概率至少为$1 - \delta$

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

4. 应用举例

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

- 将$\chi(j)$独立等概率设为-1或 $+1$
- 如何给出 $\max\{\mathrm{disc}_\chi(S_1), \mathrm{disc}_\chi(S_2), \dots, \mathrm{disc}_\chi(S_m)\}$的上界?

- 考虑固定的 $i \in \{1,2,\dots,m\}$，给出$\mathrm{disc}_\chi(S_i)$的上界
- 对 $i \in \{1,2,\dots,m\}$使用Union bound

4. 应用举例

- 对于 $\chi: \{1,2,\dots,n\} \to \{-1,+1\}$，定义$\text{disc}_\chi(S_i) = |\sum_{j\in S_i} \chi(j)|$

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

北京大学 PEKING UNIVERSITY

4. 应用举例

- 给定 $S_1, S_2, \ldots, S_m \subseteq \{1,2,\ldots,n\}$
- 对于 $\chi: \{1,2,\ldots,n\} \to \{-1,+1\}$，定义$\mathrm{disc}_\chi(S_i) = \left|\sum_{j\in S_i} \chi(j)\right|$
- 找到 χ使得 $\max\{\mathrm{disc}_\chi(S_1), \mathrm{disc}_\chi(S_2), \ldots, \mathrm{disc}_\chi(S_m)\}$尽量小

- 将$\chi(j)$独立等概率设置为-1或 $+1$
- 对于固定的 $i \in \{1,2,\ldots,m\}$，$P\left(\mathrm{disc}_\chi(S_i) \ge k\right) \le 2e^{-\frac{k^2}{2n}}$
- $P\left(\max\{\mathrm{disc}_\chi(S_1), \mathrm{disc}_\chi(S_2), \ldots, \mathrm{disc}_\chi(S_m)\} \ge k\right) \le 2m \cdot e^{-\frac{k^2}{2n}}$
- $P\left(\max\{\mathrm{disc}_\chi(S_1), \mathrm{disc}_\chi(S_2), \ldots, \mathrm{disc}_\chi(S_m)\} \ge \sqrt{n \log m}\right) \le \frac{1}{2}$
- 如何设计成功概率至少为$1-\delta$的算法?

少写个大O

上界

4. 应用举例

- 给定数据$x_1, x_2, \dots, x_n \in \mathbb{R}^d$
- 设计映射 $F: \mathbb{R}^d \to \mathbb{R}^k$，使得对于任意$1 \le i, j \le n$，
  - $(1-\epsilon)|x_i - x_j|_2^2 \le |F(x_i) - F(x_j)|_2^2 \le (1+\epsilon)|x_i - x_j|_2^2$
- 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$F: \mathbb{R}^d \to \mathbb{R}^k$，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

- 回顾：若X_i独立同分布，且$X_i \sim N(0,1)$，则$\sum_{i=1}^d a_i X_i \sim N(0, |a|_2^2)$

- 考虑$k \times d$矩阵A，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$x \in \mathbb{R}^d$，向量Ax服从何种分布?

4. 应用举例

- 考虑$k \times d$矩阵A，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$x \in \mathbb{R}^d$，向量$y = Ax$服从何种分布?
  - $y_i \sim N(0, |x|_2^2)$，且y不同元素相互独立
- 令$z = \frac{y}{|x|_2}$，则$z_i \sim N(0,1)$，且z的不同元素相互独立
- $|z|_2^2 = z_1^2 + z_2^2 + \cdots + z_k^2 = |y|_2^2/|x|_2^2$

  亦服从正态分布
- $E(|z|_2^2) = k$，$P(|z|_2^2 \geq (1+\epsilon) \cdot k) \leq e^{-k\epsilon^2/8}$，$P(|z|_2^2 \leq (1-\epsilon) \cdot k) \leq e^{-k\epsilon^2/8}$
- $P(|z|_2^2 \notin (1 \pm \epsilon)k) \leq 2e^{-k\epsilon^2/8} \Rightarrow P(|y|_2^2/|x|_2^2 \notin (1 \pm \epsilon)k) \leq 2e^{-k\epsilon^2/8}$
- $P(|y|_2^2 \notin (1 \pm \epsilon) \cdot k \cdot |x|_2^2) \leq 2e^{-k\epsilon^2/8} \Rightarrow P(|Ax|_2^2 \notin (1 \pm \epsilon) \cdot k \cdot |x|_2^2) \leq 2e^{-k\epsilon^2/8}$
- $P\left(|\frac{1}{\sqrt{k}}Ax|_2^2 \notin (1 \pm \epsilon)|x|_2^2\right) \leq 2e^{-k\epsilon^2/8}$
- 定义$F(x) = \frac{1}{\sqrt{k}}Ax$

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$1\le i,j\le n$，令$x=x_i-x_j$，
  - $P\left(|\frac{1}{\sqrt{k}}Ax|_2^2\notin(1\pm\epsilon)|x|_2^2\right)\le 2e^{-k\epsilon^2/8}$
  - $P\left(|\frac{1}{\sqrt{k}}A(x_i-x_j)|_2^2\notin(1\pm\epsilon)|x_i-x_j|_2^2\right)\le 2e^{-k\epsilon^2/8}$
- 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$1\le i,j\le n$成立
- $P(\overline{E})\le 2e^{-\frac{k\epsilon^2}{8}}\cdot\frac{n(n-1)}{2}\le e^{-\frac{k\epsilon^2}{8}}\cdot n^2$
- $k=O\left(\frac{\log n}{\epsilon^2}\right)\Rightarrow P(E)\ge\frac{1}{2}$

4. 应用举例

- Johnson-Lindenstrauss Lemma 概率证法
- 给定$x_1, x_2, \dots, x_n \in \mathbb{R}^d$，存在$F: \mathbb{R}^d \to \mathbb{R}^k$， $k = O\left(\frac{\log n}{\epsilon^2}\right)$
- 对于任意$1 \le i, j \le n$， $(1-\epsilon)|x_i - x_j|_2^2 \le |F(x_i) - F(x_j)|_2^2 \le (1+\epsilon)|x_i - x_j|_2^2$

- $F = \frac{1}{\sqrt{k}} Ax$， 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