Chap 5 不等式与极限定理

1 概率不等式

- **定义(Markov 不等式)** $Y \geq 0, \forall a > 0$, 有

$$P(Y \geq a) \leq \frac{E(Y)}{a}.$$

- **证明** 令示性变量

$$I = \begin{cases} 1, Y \geq a; \\ 0, Y < a. \end{cases}$$

从而有 $I \leq \dfrac{Y}{a}$, 两边取期望, 即得

$$P(Y \geq a) \leq \frac{E(Y)}{a}.$$

- **定义(Chebyshev 不等式)** $Var(Y)$ 存在, $\forall a > 0$, 有

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

- **证明** 注意到

$$\begin{aligned} P(|Y - E(Y)| \geq a) &= P((Y - E(Y))^2 \geq a^2) \\ &\leq \frac{E[(Y - E(Y))^2]}{a^2} \\ &= \frac{Var(Y)}{a^2}. \end{aligned}$$

- **注** 若 $Var(Y) = 0$, 则 $P(Y = E(Y)) = 1$. ($Y = E(Y)$ $a.s.$)
- **定义(Chernoff 不等式)** $\forall a > 0, t > 0$, 有

$$P(Y \geq a) \leq \frac{E(e^{tY})}{e^{ta}}.$$

- **证明** 注意到

$$\begin{aligned} P(Y \geq a) &= P(e^{tY} \geq e^{ta}) \text{ (保证 } e^{tY} > 0) \\ &\leq \frac{E(e^{tY})}{e^{ta}}. \end{aligned}$$

- **例** $X \sim N(0, 1)$, 估计 $P(|X| \geq 3)$.
- **解答** 我们有

$$P(|X|\geq 3)\leq\begin{cases}\dfrac{E(|X|)}{3}=\dfrac{1}{3}\sqrt{\dfrac{2}{\pi}}\approx 0.27; & (Markov)\\ \dfrac{Var(X)}{3^2}=\dfrac{1}{9}\approx 0.11; & (Chebyshev)\\ \dfrac{2E(e^{tX})}{e^{3t}}=2e^{\frac{t^2}{2}-3t}\leq 2e^{-\frac{9}{2}}\approx 0.02. & (Chernoff)\end{cases}$$

2 大数定律 (LLN)

- **定义** $X_1, X_2, \cdots$ **iid** (独立同分布), $E(X_i)=\mu$, $Var(X_i)=\sigma^2>0$. 定义:

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\ E(\overline{X})=\mu,\ Var(\overline{X})=\frac{\sigma^2}{n}\to 0.$$

- **定义(Khinchin 弱大数定律)(WLLN)**

  若 $X_1, X_2, \cdots$ **iid**, $E(X_i)=\mu$, $Var(X_i)=\sigma^2>0$. 则 $\forall\varepsilon>0$, 有

$$\lim_{n\to\infty}P(|\overline{X}-\mu|\geq\varepsilon)=0.$$

- **证明** 我们有

$$P(|\overline{X}-\mu|\geq\varepsilon)\leq\frac{Var(\overline{X})}{\varepsilon^2}=\frac{\sigma^2}{n\varepsilon^2}\to 0, \text{当}\, n\to 0.$$

- **注**
  - $\mu\approx\overline{X}$ (在很大概率意义下可以用作样本均值估计);
  - $\forall\varepsilon>0, \forall\alpha>0, \exists N>0$ 使得当 $n\geq N$ 时

$$P(|\overline{X}-\mu|\geq\varepsilon)\leq\alpha.$$

    其中 ε 体现了精度，α 体现了置信度.
  - **Bernoulli LLN**: $X_i\sim B(p)$, 则特殊地得到 **Bernoulli** 大数定律。
  - 方差有限条件可去掉, 结论依然成立;
  - 可推广至不同的条件:
    - X_i 两两不相关，$Var(X_i)$ 一致有界 (**Chebyshev**);
    - $Var(\overline{X})\to 0$ (**Markov**).
- **定义(依概率收敛)**

$$Y_n\xrightarrow{P}Y\iff\forall\varepsilon>0, \lim_{n\to\infty}P(|Y_n-Y|\geq\varepsilon)=0.$$

- **注** **WLLN** $\Rightarrow\overline{X}\xrightarrow{P}\mu$ (考虑偏差).
- **定义(Kolmogov 强大数定律)(SLLN)**

若 $X_1, X_2, \cdots$ **iid**, $E(X_i)=\mu$. 则有

$$P(\omega\in\Omega \mid \lim_{n\to\infty}\overline{X_n}(\omega)=\mu)=P(\lim_{n\to\infty}\overline{X}=\mu)=1.$$

- **注** 若 $X_i\sim B(p)$ 则 $\overline{X}$ 为频率,从而概率的频率解释是合理的.
- **定义(以概率 1 收敛)**

$$Y_n\xrightarrow{a.s.}Y\Longleftrightarrow P(\lim_{n\to\infty}Y_n=Y)=1.$$

- **注** **SLLN** $\Rightarrow \overline{X}\xrightarrow{a.s.}\mu$ (逐点考虑).
- **例** (**Monte Carlo** 积分)
- **解答** 在 $[a,b]\times[0,c]$ 上取点 (X_i,Y_i) **iid** 在矩形内均匀分布. 定义

$$I_i=\begin{cases}1,(X_i,Y_i)\in D;\\0,(X_i,Y_i)\notin D.\end{cases}$$

则 $I_i\xrightarrow{iid}B(p)$. 我们有

$$P=\frac{1}{(b-a)c}\int_a^b g(x)dx\approx\frac{1}{n}\sum_{i=1}^n I_i.$$

- **例** 两种收敛有什么差别?
- **解答** 考虑 $\Omega=[0,1]$ 均匀分布 (从而有 $(\Omega,\mathscr{F},P)$). 我们构造

$$\begin{gathered}Y_1(\omega)=\omega+I_{[0,1]}(\omega)\\Y_2(\omega)=\omega+I_{[0,\frac{1}{2}]}(\omega)\\Y_3(\omega)=\omega+I_{[\frac{1}{2},1]}(\omega)\\Y_4(\omega)=\omega+I_{[0,\frac{1}{3}]}(\omega)\\Y_5(\omega)=\omega+I_{[\frac{1}{3},\frac{2}{3}]}(\omega)\\Y_6(\omega)=\omega+I_{[\frac{2}{3},1]}(\omega)\\\cdots\\Y(\omega)=\omega.\end{gathered}$$

因此有 $Y_n\xrightarrow{P}Y$, 但是 $Y_n\xrightarrow{a.s.}Y$ 不成立.

这是因为 $\forall\,\omega_0\in(0,1)$, $Y_n(\omega_0)$ 是振荡的, 它的极限不存在.

3 中心极限定理 (CLT)

- **定义(中心极限定理)(CLT)**

若 $X_1,X_2,\cdots$ **iid**, $E(X_i)=\mu$, $Var(X_i)=\sigma^2>0$. 我们有

$$\lim_{n\to\infty} P\left(\frac{X_1+\cdots+X_n-n\mu}{\sqrt{n}\sigma}\leq x\right)=\Phi(x), \forall\, x\in\mathbb{R}.$$

其中 $\Phi(x)$ 为 $N(0,1)$ 的 **CDF**. 也即

$$\lim_{n\to\infty} P\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\leq x\right)=\Phi(x), \forall\, x\in\mathbb{R}.$$

- **证明** 只在 X_i 的 **MGF** 存在情形下证明, 记 $M(t)=M_{X_i}(t)$.

  不失一般性地, 令 $\mu=0$, $\sigma^2=1$. 因此

$$\begin{gathered}M(0)=E(1)=1,\\ M'(0)=E(X_i)=\mu=0,\\ M''(0)=E(X_i^2)=\sigma^2=1.\end{gathered}$$

  我们得到

$$\begin{aligned}E(e^{t\frac{X_1+\cdots+X_n}{\sqrt{n}}})&=M^n(\tfrac{t}{\sqrt{n}})\\ &=\left(1+\tfrac{t^2}{2n}+o(\tfrac{t^2}{n})\right)^n\\ &\to e^{\frac{t^2}{2}}.\end{aligned}$$

- **注**
  - 上述 **CLT** 通常称为 **Lindeberg-Levy CLT**;
  - **CLT** $\Rightarrow X_1+\cdots+X_n\sim N(n\mu,n\sigma^2)$, $\overline{X}\sim N(\mu,\dfrac{\sigma^2}{n})$;
  - **(DeMoivre-Laplace CLT)**

    若 $X_i\sim B(p)$, 则 $\sum\limits_{i=1}^{n}X_i\sim B(n,p)\xrightarrow{CLT}$ 正态分布.

- **定义(二项分布下 CLT 的连续性修正)**

  我们有 $P(t_1\leq\sum\limits_{i=1}^{n}X_i\leq t_2)\approx\Phi(y_2)-\Phi(y_1)$. 其中

$$\begin{cases}y_1=\dfrac{t_1-np-\frac{1}{2}}{\sqrt{np(1-p)}},\\ y_2=\dfrac{t_2-np+\frac{1}{2}}{\sqrt{np(1-p)}}.\end{cases}$$

  修正形式可计算单点 $P(S_n=k)$ 的概率, 对其他离散变量也同样适用.

- **定义(依分布收敛)**

$$Y_n\xrightarrow{d}Y\Longleftrightarrow\lim_{n\to\infty}F_n(x)=F(x).$$

- **注 CLT** $\Rightarrow Z_n = \dfrac{X_1+\cdots+X_n-n\mu}{\sqrt{n}\sigma} = \dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ (标准化).
- **例(选举问题)** 设 p 为选民支持率(未知), 随机调查 n 个人, 支持比例为 $p_n = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$, 其中 $X_i \sim B(p)$. 若 $\varepsilon = 0.03$, $1-\alpha = 0.95$, 求 n 的取值.
- **解答** 有

$$P(|p_n - p| \geq \varepsilon) \leq \alpha.$$

由 **CLT** 可得

$$\begin{aligned} P(|p_n - p| \geq \varepsilon) &= 1 - P\Big(-\frac{\sqrt{n}\varepsilon}{\sigma} \leq \frac{p_n - p}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\sqrt{n}\varepsilon}{\sigma}\Big) \\ &= 1 - \Phi\Big(\frac{\sqrt{n}\varepsilon}{\sqrt{p(1-p)}}\Big) + \Phi\Big(-\frac{\sqrt{n}\varepsilon}{\sqrt{p(1-p)}}\Big) \\ &= 2 - 2\Phi\Big(\frac{\sqrt{n}\varepsilon}{\sqrt{p(1-p)}}\Big) \leq \alpha. \end{aligned}$$

即得

$$\Phi\Big(\frac{\sqrt{n}\varepsilon}{\sqrt{p(1-p)}}\Big) \geq 1 - \frac{\alpha}{2}.$$

为使得对任意 p 成立, 取 $p = \dfrac{1}{2}$, 即有

$$\Phi\Big(2\sqrt{n}\varepsilon\Big) \geq 1 - \frac{\alpha}{2}.$$

注意到 $\Phi(1.96) \approx 0.975 = 1 - \dfrac{\alpha}{2}$, 因此取 $n \geq 1068$ 即可 (与 N 无关).

4 Review

4.1 尾部概率控制

4.2 极限定理

- **LLN**: 弱 or 强
- **CLT**

4.3 三种收敛

4.4 CLT 应用

$$
\begin{cases}
X_1 + \cdots + X_n \sim N(n\mu, n\sigma^2); \\
\overline{X} \sim N(\mu, \dfrac{\sigma^2}{n}).
\end{cases}
$$