概率论与数理统计

庄玮玮
weizh@ustc.edu.cn

安徽 合肥

2020 年 4 月

第 四 章　　大数律和中心极限定理

§4.1 强大数率

在讨论数学期望的性质时，我们已经预料到对于独立同分布的随机变量$X_1, X_2, \cdots$，其样本均值

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

收敛到数学期望$\mathrm{E}\, X_1 = \mu$，即

$$\lim_{n\to\infty} \overline{X}_n = \mu.$$

这就是将要介绍的强大数律.

用 A a.s. 表示事件 A 发生的概率是1. 也就是说 $P(A)=1$ 和 A a.s. 是等价的. 按照这一记号,

$$\lim_{n\to\infty}\overline{X}_n=\mu \text{ a.s.}$$

和

$$P\big(\lim_{n\to\infty}\overline{X}_n=\mu\big)=1$$

是等价的, 都说明事件 "$\overline{X}_n$ 收敛到μ" 的概率是1.

定理 4.1.1 (强大数律) 如果 $X_1, X_2, \cdots$ 是独立同分布的随机变量, $\mu = \mathrm{E}\, X_1$, 则

$$\lim_{n\to\infty} \overline{X}_n = \mu \text{ a.s..}$$

因为概率等于1的事件在实际中必然发生, 所以在强大数律中, 如果用x_n表示X_n的观测值, 则有

$$\lim_{n\to\infty} \frac{x_1 + x_2 + \cdots + x_n}{n} = \mu.$$

因为强大数律的数学证明并不需要概率的频率定义, 所以它从理论上保证了概率的频率定义是正确的.

例4.1.1 在赌对子时，甲每次下注100元．用 S_n 表示他下注 n 次后的盈利，则 $\lim_{n\to\infty} S_n = -\infty$ a.s..

解 用 X_i 表示甲第 i 次下注后的盈利，则 $X_1, X_2, \cdots, X_n$ 独立同分布，$\mu = \mathrm{E}\,X_i = -18.6$．用强大数律得到

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n} \to -18.6 \text{ a.s.},$$

于是有 $\lim_{n\to\infty} S_n = -\infty$ a.s.．也就是说，如果甲一直赌下去，必然输光．

例4.1.1 在赌对子时，甲每次下注100元．用 S_n 表示他下注 n 次后的盈利，则 $\lim_{n\to\infty} S_n = -\infty$ a.s..

解 用 X_i 表示甲第 i 次下注后的盈利，则 $X_1, X_2, \cdots, X_n$ 独立同分布，$\mu = \mathrm{E}\, X_i = -18.6$．用强大数律得到

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n} \to -18.6 \text{ a.s.},$$

于是有 $\lim_{n\to\infty} S_n = -\infty$ a.s.．也就是说，如果甲一直赌下去，必然输光．

例4.1.2 在敏感问题调查中，已经推导了服用过兴奋剂的运动员在全体运动员中所占的比例 p 满足公式

$$p = 2p_1 - 1,$$

其中p_1是回答“是”的概率. 实际问题中，p_1是未知的，需要经过调查得到. 如果调查了 n 个运动员，则用回答“是”的比例 $\hat{p}_1$ 估计 p_1. 于是自然用$\hat{p} = 2\hat{p}_1 - 1$估计p. 当$n \to \infty$时，证明 $\hat{p} \to p$ a.s..

证明 对 $j=1,2,\cdots$ 引入随机变量

$$X_j=\begin{cases}1, & \text{第}j\text{个人回答“是”},\\ 0, & \text{第}j\text{个人回答“否”},\end{cases}$$

则 $X_1,X_2,\cdots$ 独立同分布，满足

$$P(X_j=1)=p_1,\quad \mathrm{E}\,X_j=p_1,\quad \hat{p}_1=\frac{1}{n}\sum_{j=1}^{n}X_j.$$

根据强大数律得到，当被调查的人数 $n\to\infty$ 时， $\hat{p}_1\to p_1$ a.s.，所以当 $n\to\infty$ 时，

$$\hat{p}=(2\hat{p}_1-1)\to(2p_1-1)=p \text{ a.s.}.$$

§4.2 切比雪夫不等式

有强大数律, 自然就有弱大数律. 为了介绍弱大数律, 先介绍随机变量的依概率收敛和切比雪夫不等式.

定义4.2.1 设 $U, U_1, U_2, \cdots$ 是随机变量. 如果对任何 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P(|U_n - U| \geqslant \varepsilon) = 0,$$

则称 U_n 依概率收敛到 U, 记做 $U_n \xrightarrow{p} U$.

引理4.2.1 (切比雪夫不等式) 设随机变量 X 有数学期望 μ 和方差 $\mathrm{Var}(X)$, 则对常数 $\varepsilon > 0$, 有

$$P(|X-\mu| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^2}\mathrm{Var}(X).$$

证明　用 $\mathrm{I}[A]$ 表示事件 A 的示性函数. 定义 $Y = |X - \mu|$, 则无论 $\{Y \geqslant \varepsilon\}$ 是否发生, 总有

$$\mathrm{I}[Y \geqslant \varepsilon] \leqslant \frac{1}{\varepsilon^2} Y^2.$$

因为示性函数$\mathrm{I}[Y \geqslant \varepsilon]$服从伯努利分布, 所以

$$\begin{aligned} P(|X - \mu| \geqslant \varepsilon) = & \quad P(Y \geqslant \varepsilon) = \mathrm{E}\,\mathrm{I}[Y \geqslant \varepsilon] \\ \leqslant & \quad \frac{1}{\varepsilon^2}\mathrm{E}\,Y^2 = \frac{1}{\varepsilon^2}\mathrm{Var}\,(X). \end{aligned}$$

切比雪夫不等式是概率论中最重要和最基本的不等式.

推论4.2.2(弱大数律) 设随机变量 $X_1, X_2, \cdots$ 独立同分布, $\mu = \mathrm{E}\, X_1$, 则对任何 $\varepsilon > 0$, 有

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

从理论上讲, 从强大数律可以推出弱大数律. 但是 $\sigma^2 = \mathrm{Var}\,(X_1) < \infty$ 时, 可以用切比雪夫不等式给出简单的证明如下: 由

$$\mathrm{Var}\,(\overline{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right) = \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{Var}\,(X_j) = \frac{\sigma^2}{n}$$

和切比雪夫不等式得到

$$P\big(|\overline{X}_n - \mu| \geqslant \varepsilon\big) \leqslant \frac{1}{\varepsilon^2}\mathrm{Var}\,(\overline{X}_n) = \frac{1}{n\varepsilon^2}\sigma^2 \to 0, \text{ 当 } n \to \infty.$$

要理解弱大数律的确弱于强大数律, 只要理解依概率收敛的确弱于a.s.收敛. 看下面的例子.

用随机变量U表示一位专职司机在一个工作日内因交通事故造成的损失, U 取值越大说明损失越大, $U=0$ 表示没有造成实际损失. 对于正数 ε, 用 $U \geqslant \varepsilon$ 表示造成较大的损失. 对于一位优秀的老司机来讲, 假设他的 U 已经很小. 为方便,假设他的$U=0$.

设甲某是一位新的专职司机，用 U_n 表示他在第 n 个工作日的交通事故造成的损失. 因为他的开车经验在不断提高，所以随着时间的推移，他的 U_n 会向老司机的 $U=0$ 收敛. 如果 $U_n \xrightarrow{p} U$，则 $n \to \infty$ 时，我们只能得到

$$P(U_n \geqslant \varepsilon) = \ P(|U_n - U| \geqslant \varepsilon) \to 0.$$

所以对任意大的n，都不能保证 $P(U_n \geqslant \varepsilon) = 0$. 也就是说，无论有多长的开车经验，这位新司机因交通事故造成较大损失的概率都是正数，从而都有可能造成较大的损失.

用 u_n 表示 U_n 的观测值. 如果 $U_n \to U$ a.s.，则实际中有 $u_n \to 0$. 说明存在n_0，使得$n \geqslant n_0$时，$u_n < \varepsilon$. 也就是说，从某天开始，这位新司机就再也不会发生有较大损失的交通事故了.

§4.3 中心极限定理

强大数律和弱大数律分别讨论了随机变量的样本均值的几乎处处收敛和依概率收敛. 中心极限定理研究当 n 较大时, 随机变量的部分和

$$S_n = \sum_{j=1}^{n} X_j$$

的概率分布问题. 先看几个随机变量和的分布的例子.

例4.3.1 设 $\{X_j\}$ 独立同分布都服从 $\mathcal{B}(1,p)$ 分布, 则部分和

$$S_n = \sum_{j=1}^{n} X_j \sim \mathcal{B}(n,p).$$

随着n的增加, $\mathcal{B}(n,p)$的概率分布的折线图越来越接近正态概率密度的形状.

例4.3.1 设 $\{X_j\}$ 独立同分布都服从 $\mathcal{B}(1,p)$ 分布，则部分和

$$S_n = \sum_{j=1}^{n} X_j \sim \mathcal{B}(n,p).$$

随着n的增加，$\mathcal{B}(n,p)$的概率分布的折线图越来越接近正态概率密度的形状.

例4.3.2 设$\{X_j\}$独立同分布且都服从泊松分布 $\mathcal{P}(\lambda)$, 部分和

$$S_n = \sum_{j=1}^{n} X_j \sim \mathcal{P}(\backslash\lambda).$$

随着 n 的增加, 概率分布的折线图越来越接近正态概率密度的形状.

例4.3.3 设 $\{X_j\}$ 独立同分布且都服从几何分布 $P(X=k)=pq^{k-1}$, $k=1,2,\cdots$, $p+q=1$, 则部分和 $S_n=\sum_{j=1}^{n}X_j$ 服从 帕斯卡分布

$$P(S_n=k)=\mathrm{C}_{k-1}^{n-1}p^nq^{k-n},\quad k=n,n+1,\cdots.$$

这是因为可以将 S_n 视为第 n 次击中目标时的射击次数.

取$p = 0.6$时，S_n的概率分布折线图见图4.3.1. 随着n的增加，概率分布折线图也越来越接近正态概率密度的形状.

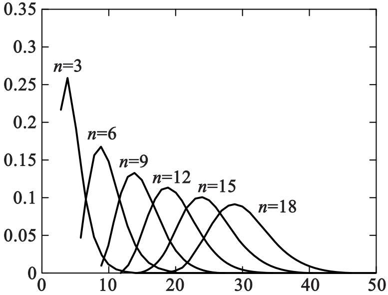

图4.3.1 帕斯卡分布的概率分布折线图，$p = 0.6$，$n = 3, 6, \cdots, 18$

例4.3.4 设$\{X_j\}$独立同分布都服从指数分布$\mathcal{E}(\lambda)$, 则部分和 $S_n=\sum_{j=1}^{n}X_j$ 服从 $\Gamma(n,\lambda)$ 分布(略去推导), 概率密度是

$$f_n(x)=\frac{\lambda^n}{\Gamma(n)}x^{n-1}\mathrm{e}^{-\lambda x},\ \ x\geqslant 0.$$

取$\lambda = \pi$，$n = 3m$，$m = 1, 2, \cdots, 7$时，S_n的概率密度见图4.3.2，横坐标是x，纵坐标是$f_n(x)$.随着n的增加，概率密度的图形也越来越接近正态概率密度曲线.

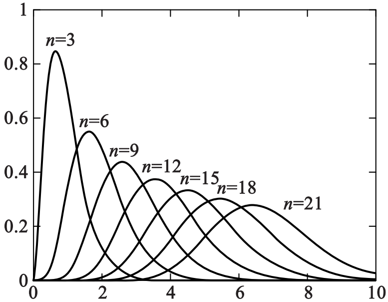

图4.3.2 $\Gamma(n, \lambda)$分布的概率密度曲线图，$\lambda = \pi$，$n = 3, 6, \cdots, 21$

例4.3.5 设 X_1, X_2, X_3 相互独立且都在 $(0,1)$ 上均匀分布，$S_3 = X_1 + X_2 + X_3$，则 $ES_3 = 3/2$，$\mathrm{Var}(X_1) = 1/12$，$\mathrm{Var}(S_3) = 1/4$. 用 $g(x)$ 表示 S_3 的标准化

$$U = \frac{S_3 - 3/2}{\sqrt{1/4}} = 2S_3 - 3$$

的概率密度. 图4.3.3 是 $g(x)$ 和标准正态概率密度 $\varphi(x)$ 的比较，二者已经大体相同.

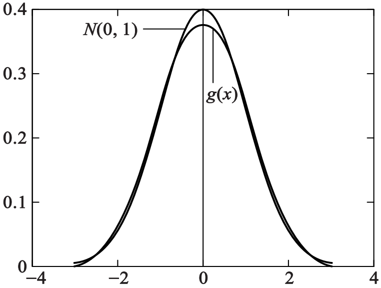

图4.3.3 $g(x)$ 和$\varphi(x)$的图形

以上的例子都显示，独立同分布随机变量和的分布近似于正态分布. 这就是将要介绍的中心极限定理.

设随机变量 $X_1, X_2, \cdots$ 独立同分布，$\mathrm{E}\, X_1 = \mu$, $\mathrm{Var}\,(X_1) = \sigma^2 > 0$.

用 $S_n = \sum_{j=1}^{n} X_j$ 表示部分和，用

$$Z_n = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$$

表示 S_n 的标准化，用 $\Phi(x)$ 表示服从 $N(0,1)$的分布函数.

定理4.3.1(中心极限定理) 在上述条件下，当$n \to \infty$时，有

$$P\left(Z_n \leqslant x\right) \to \Phi(x), \ x \in (-\infty, \infty),$$

称Z_n依分布收敛到$N(0,1)$，记做

$$Z_n \xrightarrow{d} N(0,1).$$

定理4.3.2 设 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布，X_i 的分布是

$$P(X_i = 1) = p, \qquad P(X_i = 0) = 1 - p \quad (0 < p < 1).$$

则对任何实数 x，有

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{np(1-p)}}(X_1 + \cdots + X_n - np \le x\right) = \Phi(x).$$

例4.3.6 某研究所有50位科研人员. 由于研究兴趣的不同, 每次的学术报告会平均只有30位参加. 假设每个人是否参加报告会是独立同分布的, 估算下次报告会参加人数不多于25人的概率.

解 用$X_i = 1$或0分别表示第i个人参加或不参加, 则$X_1, X_2, \cdots, X_{50}$ 独立同分布, $S_n = X_1 + X_2 + \cdots + X_{50}$是参加报告会的总人数. 由$X_i \sim \mathcal{B}(1, 3/5)$ 得到

$$\mathrm{E}\, S_n = 50 \times 0.6 = 30, \ \mathrm{Var}\,(S_n) = 50 \times 0.6 \times 0.4 = 12.$$

例4.3.6 某研究所有50位科研人员. 由于研究兴趣的不同, 每次的学术报告会平均只有30位参加. 假设每个人是否参加报告会是独立同分布的, 估算下次报告会参加人数不多于25人的概率.

解 用$X_i = 1$或0分别表示第i个人参加或不参加, 则$X_1, X_2, \cdots, X_{50}$ 独立同分布, $S_n = X_1 + X_2 + \cdots + X_{50}$是参加报告会的总人数. 由$X_i \sim \mathcal{B}(1, 3/5)$ 得到

$$\mathrm{E}\, S_n = 50 \times 0.6 = 30,\ \ \mathrm{Var}\,(S_n) = 50 \times 0.6 \times 0.4 = 12.$$

用中心极限定理得到

$$\begin{aligned} P(S_n \leqslant 25) &= P\left(\frac{S_n - 30}{\sqrt{12}} \leqslant \frac{25-30}{\sqrt{12}}\right) \\ &\approx \Phi(-5/\sqrt{12}) \\ &= 1 - \Phi(1.4434) \approx 0.075. \end{aligned}$$

例4.3.6蕴涵了如下结论：如果$S_n \sim \mathcal{B}(n,p)$，则$n$较大时，有

$$P(S_n \leqslant s) \approx \Phi\left(\frac{s-np}{\sqrt{npq}}\right).$$

这里较大的n，指起码要求n使得$n\min\{p, 1-p\} \geqslant 5$.

例4.3.7 设某地区原有一家小型电影院，因不敷需要，拟筹建一家较大型的。根据分析，该地区每日平均看电影者约有 $n = 1600$ 人，且预计新电影院建成开业后，平均约有3/4的观众将去这家电影院。现该电影院在计划座位时，要求座位数尽可能的多，但“空座达到200或更多”的概率又不能超过0.1，问设多少座位为好？

因为样本均值$\overline{X}_n$的标准化等于S_n的标准化:

$$\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} = Z_n,$$

所以有如下的推论.

推论4.3.2 在定理4.3.1的条件下, 对较大的 n, 有

$$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leqslant x\right) \approx \Phi(x),$$

$$P\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leqslant x\right) \approx \Phi(x).$$

中心极限定理是概率论中最重要的基本定理，在本书的统计部分将多次使用中心极限定理. 在一些实际问题中，随机变量的方差σ^2是未知的，这时可以用

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2 \quad 或 \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2$$

代替推论4.3.2中的σ^2，得到下面的定理4.3.3.

定理4.3.3 (中心极限定理) 在定理4.3.1的条件下，当n较大时近似地有

$$Z_n = \frac{\overline{X}_n - \mu}{\hat{\sigma}/\sqrt{n}} \sim N(0,1).$$