概率统计第十二讲：
极限定理

史灵生　清华数学系

1、本讲提要

1. 大数定律
   - 弱大数定律：Chebyshev、Bernoulli和Khinchin
   - 强大数定律：Borel和Kolmogorov
   - 应用
2. 中心极限定理
   - 独立同分布：Lindeberg-Lévy和De Moivre-Laplace极限定理
   - 独立不同分布：Lindeberg和Liapunov中心极限定理
   - 应用

2、Chebyshev大数定律

Chebyshev大数定律

设$\{X_n\}$为一列独立同分布的随机变量序列，存在数学期望和有限的方差。记$\mu = EX_n$，则对任意$\varepsilon > 0$，

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| < \varepsilon\right) = 1。$$

定律表明：

用一个随机变量的一组独立观测值的算术平均值作为它的数学期望（概率意义上的平均值）是合理的，即使我们并不确切地知道这个随机变量的概率分布。这样的结论是符合我们的日常经验的，实际上人类证明这个结论之前（甚至在得到这个结论的确切表述之前）就一直信奉并广泛地使用着这个经验，所以这个严格的数学定理还是被叫做“定律”，相比于物理、化学这样的实验自然科学，这种命名方式在数学中是极其少见的。

3、大数定律

证明：（利用Chebyshev不等式）

记$\sigma^2 = Var(X_n)$，则

$$P\left(\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| \geq \varepsilon\right) = P\left[\left|\frac{X_1+\cdots+X_n}{n} - E\left(\frac{X_1+\cdots+X_n}{n}\right)\right| \geq \varepsilon\right]$$
$$\leq \frac{1}{\varepsilon^2} Var\left(\frac{X_1+\cdots+X_n}{n}\right) = \frac{\sigma^2}{n\varepsilon^2} \to 0,\quad n \to \infty.$$

推论（Bernoulli大数定律）

设事件A发生的概率为p，记S_n为n次独立试验中A发生的次数，则事件A出现的频率$\frac{S_n}{n}$依概率收敛于p。

证明：设$X_k = \begin{cases} 1, & \text{若第}k\text{次试验中事件}A\text{发生；} \\ 0, & \text{否则。} \end{cases}$

则$X_k \overset{\text{i.i.d.}}{\sim} b(1,p)$，$k = 1,2,\ldots,n$，$EX_k = p$，$Var(X_k) = p(1-p)$，而$S_n = X_1 + \cdots + X_n$。

4、大数定律

注：

1. 上述推论从理论上证实了频率的“相对稳定性”，即概率是频率的极限，这是我们从日常经验出发对概率的最自然的理解方式。
2. Jakob Bernoulli最早证明了上述形式的大数定律，但是“大数定律”这个称呼是Poisson引进的，他将Bernoulli的结论推广为：如果在一列独立试验的第k次试验中事件A发生的概率为p_k，则$\frac{S_n}{n}-\frac{p_1+\cdots+p_n}{n}\xrightarrow{P}0, \quad n\to\infty$。
3. 如果所有X_n两两不相关，存在数学期望，并且方差一致有界，则大数定律依然成立。（Chebyshev定律一般形式）
4. 甚至当所有X_n渐近不相关$\left(\lim\limits_{|k-l|\to\infty} Cov(X_k,X_l)=0\right)$，存在数学期望，并且方差一致有界时，大数定律也成立。（Bernstein大数定律）

5、Khinchin大数定律

Khinchin大数定律（无需对方差提出要求，方差可不存在）

当$\{X_n\}$独立同分布，存在数学期望μ时，大数定律也成立。

证明：

- $$\varphi_{\frac{X_1+\cdots+X_n}{n}}(t) = \prod_{k=1}^{n} \varphi_{X_k}\left(\frac{t}{n}\right) = \left[\varphi_{X_1}\left(\frac{t}{n}\right)\right]^n$$
  $$= \left[1 + i\mu\frac{t}{n} + o\left(\frac{t}{n}\right)\right]^n \to e^{i\mu t}, \qquad n \to \infty.$$
- 因此对任意$x \neq \mu$，$F_{\frac{X_1+\cdots+X_n}{n}}(x) \to F_\mu(x) = I_{[\mu,\infty)}(x)$，
  $$P\left(\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| < \varepsilon\right) = P\left(\mu - \varepsilon < \frac{X_1+\cdots+X_n}{n} < \mu + \varepsilon\right)$$
  $$= \lim_{x \nearrow \mu+\varepsilon} F_{\frac{X_1+\cdots+X_n}{n}}(x) - F_{\frac{X_1+\cdots+X_n}{n}}(\mu - \varepsilon)$$
  $$\to \lim_{x \nearrow \mu+\varepsilon} F_\mu(x) - F_\mu(\mu - \varepsilon) = 1 - 0 = 1, \quad n \to \infty。$$

6、强大数定律

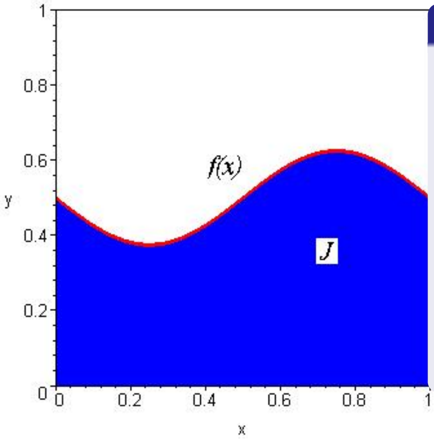
上述用“依概率收敛”表述的大数定律，被统称为“弱大数定律”。另外，我们还有“强大数定律”。

- Borel强大数定律：设$X_1, \ldots, X_n, \ldots$独立同分布，$EX_1 = \mu$，$Var(X_1) < \infty$。则
$$P\left\{\omega : \lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu\right\} = 1.$$
- 设事件A发生的概率为p，记S_n为n次独立试验中A发生的次数，则相对频率满足$P\left\{\omega : \lim_{n\to\infty} \frac{S_n}{n} = p\right\} = 1$。
- Kolmogorov强大数定律：设$X_1, \ldots, X_n, \ldots$独立，满足
$$\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < +\infty,$$
则$P\left\{\omega : \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n}(X_k - EX_k) = 0\right\} = 1$。

7、Monte-Carlo随机模拟

例（随机投点法）

设$0 \le f(x) \le 1$，求积分$J = \int_0^1 f(x)\mathrm{d}x$的近似值。



解：

- 设$X, Y \sim U[0,1]$相互独立，
- $P[Y \le f(X)] = \int_0^1 \int_0^{f(x)} \mathrm{d}y\mathrm{d}x = \int_0^1 f(x)\mathrm{d}x = J$。
- 可用频率来近似$P[Y \le f(X)]$：

1、产生$[0,1]$上的$2n$个均匀随机数：x_i，y_i，$i = 1, 2, \ldots, n$。

2、记录满足$y_i \le f(x_i)$的次数S_n，则$J \approx S_n/n$。

8、Monte-Carlo随机模拟

例（平均值法）

求积分$J=\int_0^{\frac{\pi}{2}}\cos\sqrt{x}\mathrm{d}x$的近似值。

解：

- 注意到$J=\int_0^{\frac{\pi}{2}}\cos\sqrt{x}\mathrm{d}x=\int_0^{\frac{\pi}{2}}\frac{\pi}{2}\cos\sqrt{x}\cdot\frac{2}{\pi}\mathrm{d}x$，
- 因此$J=E\left(\frac{\pi}{2}\cos\sqrt{X}\right)$，其中$X\sim U(0,\frac{\pi}{2})$。
- 设$\{X_n\}_{n=1}^{\infty}$是一列独立随机变量，分别服从$(0,\frac{\pi}{2})$上均匀分布。则$\{\frac{\pi}{2}\cos\sqrt{X_n}\}_{n=1}^{\infty}$独立同分布。
- 又$\left|\frac{\pi}{2}\cos\sqrt{X_n}\right|<2$，故$\frac{\pi}{2}\cos\sqrt{X_n}$有数学期望和方差。
- 由Chebyshev大数定律，
$$\frac{\frac{\pi}{2}\cos\sqrt{X_1}+\cdots+\frac{\pi}{2}\cos\sqrt{X_n}}{n}\xrightarrow{P}E\left(\frac{\pi}{2}\cos\sqrt{X_1}\right)=J。$$

9、点估计

习题4.1第13题

设随机变量序列$\{X_n\}$独立同分布，分别服从$(0,\beta)$上的均匀分布. 则$Y_n = \max\{X_1, X_2, \ldots, X_n\}$依概率收敛于$\beta$。

证明：

- 对$0 < \varepsilon < \beta$，
$$\begin{aligned} P(|Y_n - \beta| \geq \varepsilon) &= P(Y_n \leq \beta - \varepsilon) \\ &= P(X_1 \leq \beta - \varepsilon, \ldots, X_n \leq \beta - \varepsilon) \\ &= P(X_1 \leq \beta - \varepsilon) \cdots P(X_n \leq \beta - \varepsilon) \\ &= \left(\frac{\beta - \varepsilon}{\beta}\right)^n = \left(1 - \frac{\varepsilon}{\beta}\right)^n \to 0, \quad n \to \infty。\end{aligned}$$
- 因此$Y_n \xrightarrow{P} \beta$。

10、点估计

- 在实际问题中，我们可能并不知道β的准确值，
- 因此我们需要从一些观测结果估计β的值，
- 所以$Y_n = \max\{X_1, X_2, \ldots, X_n\}$可以作为$\beta$的一个估计值。
- 而根据大数定律$\bar{X} = (X_1 + \cdots + X_n)/n \xrightarrow{P} EX_1 = \beta/2$，
- 所以$2\bar{X} \xrightarrow{P} \beta$，即$2\bar{X}$也可以作为$\beta$的一个估计值。
- 作为β的估计值，$\max\{X_1, X_2, \ldots, X_n\}$和$2\bar{X}$哪个更好呢？
- 注意到：$E(2\bar{X}) = 2\frac{EX_1+\cdots+EX_n}{n} = \beta$，$\forall n \in \mathbb{N}$，
- $F_{Y_n}(x) = \prod_{i=1}^{n} F_{X_i}(x) = (x/\beta)^n$，
- $p_{Y_n}(x) = nx^{n-1}/\beta^n$，
- $E\max\{X_1, X_2, \ldots, X_n\} = EY_n = n\int_0^\beta (x/\beta)^n \mathrm{d}x = \frac{n}{n+1}\beta$。

11、无偏估计

- 所以$E\max\{X_1, X_2, \ldots, X_n\}$作为$\beta$的估计值，有系统误差（称为“偏”，*bias*）

$$E\max\{X_1, X_2, \ldots, X_n\} - \beta = -\frac{\beta}{n+1},$$

- 而$2\bar{X}$是没有系统误差的（称为无偏的，*unbiased*）。
- 易见$\frac{n+1}{n}\max\{X_1, X_2, \ldots, X_n\}$和$2\bar{X}$都是$\beta$的无偏估计，那么它们中到底哪个更好呢？
- 我们比较

$$E\left(\frac{n+1}{n}\max\{X_1, X_2, \ldots, X_n\} - \beta\right)^2, \quad E(2\bar{X} - \beta)^2$$

即$\frac{n+1}{n}\max\{X_1, X_2, \ldots, X_n\}$和$2\bar{X}$的方差，方差小者为优。

12、无偏估计

- 与上述计算$E\max\{X_1, X_2, \ldots, X_n\}$的方法类似，可得
- $E(\max\{X_1, X_2, \ldots, X_n\})^2 = \frac{n}{n+2}\beta^2$，
- $Var(\max\{X_1, \ldots, X_n\}) = \frac{n}{n+2}\beta^2 - \frac{n^2}{(n+1)^2}\beta^2 = \frac{n}{(n+1)^2(n+2)}\beta^2$，
- $$Var\left(\frac{n+1}{n}\max\{X_1, \ldots, X_n\}\right) = \frac{(n+1)^2}{n^2} Var(\max\{X_1, \ldots, X_n\}) = \frac{\beta^2}{n(n+2)}。$$
- $Var(2\bar{X}) = 4Var(\bar{X}) = 4Var\left(\sum_{i=1}^{n} X_i/n\right) = \frac{\beta^2}{3n}$。
- 故$n \geq 1$时，$Var\left(\frac{n+1}{n}\max\{X_1, X_2, \ldots, X_n\}\right) \leq Var(2\bar{X})$，
- $Var\left(\frac{n+1}{n}\max\{X_1, \ldots, X_n\}\right)/Var(2\bar{X}) = \frac{3}{n+2} \to 0$，$n \to \infty$，
- 即$\frac{n+1}{n}\max\{X_1, \ldots, X_n\}$作为$\beta$的无偏估计比$2\bar{X}$有更小方差。
- 如果能找到所有无偏估计的方差下界，甚至找到最小方差无偏估计（*minimum variance unbiased estimate*，*MVUE*），那当然是最理想的。

13、Lindeberg-Lévy中心极限定理

大学物理中有一个Galton [1] 钉板的演示实验[2]，说明大量随机因素可以呈现出宏观规律性，而正态分布就是这种宏观规律性的一个重要表现。

Lindeberg-Lévy中心极限定理

设$\{X_n\}$是独立同分布的随机变量序列，存在数学期望μ和有限方差σ^2，则当$n \to \infty$时，$\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$的分布函数收敛到标准正态的分布函数$\Phi$，而且这收敛是一致的，即

$$\lim_{n\to\infty}\sup_{x\in\mathbb{R}}\left|P\left(\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}\le x\right)-\Phi(x)\right|=0.$$

[1] Francis Galton（1822年2月16日-1911年1月17日）爵士，英国探险家、统计学家、人类学家，现代优生学（eugenics，源于希腊语$\epsilon\mu\gamma\epsilon\nu\eta\zeta$）的创立者

[2] 参看http://en.wikipedia.org/wiki/Bean_machine

14、Lindeberg-Lévy中心极限定理

- $E\left(\frac{X_1+\cdots+X_n}{n}\right)=\mu,\quad Var\left(\frac{X_1+\cdots+X_n}{n}\right)=\sigma^2/n,$
- $\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}=\left(\frac{X_1+\cdots+X_n}{n}-\mu\right)\Big/\frac{\sigma}{\sqrt{n}}$ 的期望为0，方差为1，
- 其特征函数为

$$\begin{aligned}\varphi_n(t) &= \varphi_{X_1+\cdots+X_n-n\mu}\left(\frac{t}{\sigma\sqrt{n}}\right)=\left[\varphi_{X_1-\mu}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n\\ &= \left[1+iE(X_1-\mu)\frac{t}{\sigma\sqrt{n}}+\frac{i^2E(X_1-\mu)^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2+o\left(\frac{t^2}{n\sigma^2}\right)\right]^n\\ &= \left[1-\frac{t^2}{2n}+o\left(\frac{t^2}{n\sigma^2}\right)\right]^n\to e^{-\frac{t^2}{2}},\qquad n\to\infty。\end{aligned}$$

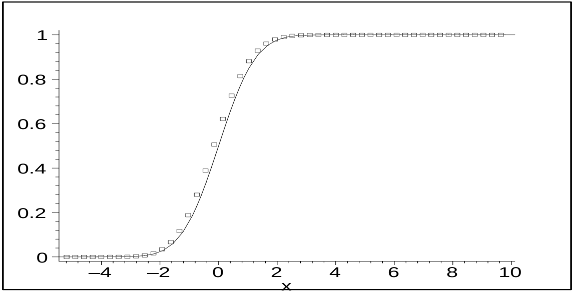
- 故当$n\to\infty$时，$\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$的分布函数收敛到正态$N(0,1)$ 的分布函数。这样我们就得到分布函数的弱收敛(即在极限函数连续点上逐点收敛)，而$\Phi(x)$是连续函数，所以这就是在整个$\mathbb{R}$上逐点收敛。（一致收敛性见书中习题4.3第7题）

15、De Moivre-Laplace中心极限定理

推论（De Moivre 1733-Laplace 1812）

设$X_n \sim b(n,p)$。则 $\lim\limits_{n\to\infty} P\left(\frac{X_n-np}{\sqrt{npq}} \leq x\right) = \Phi(x),\quad \forall x \in \mathbb{R}$，而且这里的收敛对$x$是一致的。

下图中我们可以看到$X \sim b(60, 0.35)$的标准化$(X-np)/\sqrt{npq}$的概率分布函数（方格）和$N(0,1)$的概率分布$\Phi(x)$函数曲线。



16、Lindeberg中心极限定理

Lindeberg中心极限定理

设$X_1, X_2, \ldots, X_n, \ldots$独立，密度分别为$p_1(x), p_2(x), ..., p_n(x), ...$，

记$B_n = \left[\sum_{i=1}^{n} Var(X_i)\right]^{1/2}$，若对任意$\tau > 0$，有

$$\lim_{n\to\infty} \frac{1}{\tau^2 B_n^2} \sum_{i=1}^{n} \int_{|x-EX_i|>\tau B_n} (x - EX_i)^2 p_i(x)\mathrm{d}x = 0.$$

则$\sum_{i=1}^{n}(X_i - EX_i)/B_n$按分布收敛于$N(0,1)$。

17、Liapunov中心极限定理

Liapunov中心极限定理

设$X_1, X_2, \ldots, X_n, \ldots$独立，满足对某个$\delta > 0$，

$$\lim_{n\to\infty}\left[\sum_{j=1}^{n} Var(X_j)\right]^{-1-\delta/2}\sum_{i=1}^{n} E|X_i - EX_i|^{2+\delta} = 0,$$

则$\sum_{i=1}^{n}(X_i - EX_i)\Big/\sqrt{\sum_{j=1}^{n} Var(X_j)}$按分布收敛于$N(0,1)$。

正态$N(\mu,\sigma^2)$随机数的产生：

1. 由中心极限定理知，若$x_1, x_2, ..., x_{12}$是独立的(0,1)上均匀随机数，则可用$y = \sum_{i=1}^{12} x_i - 6$近似地作为标准正态随机数。
2. $z = \sigma y + \mu$可看成服从正态$N(\mu,\sigma^2)$的一个随机数。
3. 重复（1）和（2）n次可得服从正态$N(\mu,\sigma^2)$的n个随机数。

18、骰子投掷

例

将一颗骰子连掷100次，则点数之和不少于300的概率是多少？

解：

- 设X_k为第k次掷出的点数，k=1，2，…，100，则$X_1, X_2, \ldots, X_{100}$独立同分布。
- $E(X_1) = 7/2$，$Var(X_1) = \dfrac{1}{6}\sum\limits_{k=1}^{6} k^2 - 49/4 = 35/12$。
- 由Lindeberg-Lévy中心极限定理得：$\dfrac{\sum_{i=1}^{100} X_i - 100\times 7/2}{10\sqrt{35/12}}$ 近似服从标准正态分布。
- $P\left(\sum\limits_{i=1}^{100} X_i \geq 300\right) \approx 1 - \Phi\left(\dfrac{300-100\times 7/2}{10\sqrt{35/12}}\right) = 0.9983\cdots$。

19、保险

例

在一家保险公司里有10000个人参加寿命保险，每人每年付12元保险费。在一年内一个人死亡的概率为0.6%，死亡时其家属可向保险公司领得1000元，问：

1. 保险公司亏本的概率有多大？
2. 其他条件不变，为使保险公司一年的利润不少于60000元的概率不低于90%，赔偿金至多可设为多少？

解：设X表示一年内死亡的人数，则$X \sim b(10000, 0.6\%)$。

- 由De Moivre-Laplace中心极限定理得：$\dfrac{X-10000\times0.6\%}{\sqrt{10000\times0.6\%\times99.4\%}}$ 近似服从标准正态分布。

1. 设Y表示保险公司一年的利润，则$Y = 120000 - 1000X$，

$$P(Y<0) = P(120000-1000X<0) = 1-P(X\le 120) \approx 1-\Phi(8) \approx 0.$$

20、保险

例

在一家保险公司里有10000个人参加寿命保险，每人每年付12元保险费。在一年内一个人死亡的概率为0.6%，死亡时其家属可向保险公司领得1000元，问：
（2）为使保险公司一年的利润不少于6万元的概率不低于90%，赔偿金至多可设为多少？

解：（2）

- 设赔偿金为a元，则令$P(Y \geq 60000) \geq 0.9$。
- $$\begin{aligned} P(Y \geq 60000) &= P(120000 - aX \geq 60000) \\ &= P(X \leq 60000/a) \geq 0.9 \text{。} \end{aligned}$$
- 由De Moivre-Laplace中心极限定理，上式近似于
$$\Phi\left(\frac{60000/a - 10000 \times 0.6\%}{\sqrt{10000 \times 0.6\% \times 99.4\%}}\right) \geq 0.9 \Rightarrow a \leq 858\text{。}$$