第四章 大数定律与中心极限定理

Instructor: 郝壮

haozhuang@buaa.edu.cn
School of Economics and Management
Beihang University

December 27, 2021

期末考试

时间: 2022-01-12 第19周星期三 15:40-17:40 (一)201

答疑

1. 答疑时间： **1月11日**

14:00-15:00：李卫国老师助教答疑(许岷);

15：00-16:00: 李明全老师助教答疑(田嘉琳);

16：00-17:00: 郝壮老师助教答疑(田曼);

17:00-18:00: 张军欢老师助教答疑(李雨萌).

2. 答疑地点： **A940**

3. 上述所有时段各班级共享, 均可答疑.

第四章大数定律与中心极限定理

- §4.1 随机变量序列的两种收敛性
- §4.2 特征函数
- §4.3 大数定律
- §4.4 中心极限定理

随机变量序列: 一组有序排列的随机变量.

随机变量序列既可以是同分布, 也可以是不同分布. 如:

- 一列伯努利实验中每次实验的结果记为$X_1, X_2, \ldots, X_n, \ldots$, 可记为$\{X_n\}$ (分布相同)
- n重伯努利实验中随n增大试验成功的次数记为$S_1, S_2, \ldots, S_n, \ldots$, 可记为$\{S_n\}$ (分布不同)

随机变量序列的两种收敛性:

- i) 依概率收敛 (大数定律讨论的是依概率收敛);
- ii) 依分布收敛 (中心极限定理讨论的是依分布收敛).

定义4.1.1 (依概率收敛) 设$\{X_n\}$为一组随机变量序列, X为一随机变量, 若对任意的$\epsilon > 0$, 有

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$$

则称随机变量序列$\{X_n\}$**依概率收敛于**X, 记为

$$X_n \xrightarrow{P} X$$

$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$ 等价于
$\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$

当X为退化分布(degenerate distribution)时,
即$P(X = c) = 1$时, 则称序列$\{X_n\}$依概率收敛于常数c,
$X_n \xrightarrow{P} c$

依概率收敛的性质

定理4.1.1 设 $\{X_n\}, \{Y_n\}$ 是两个随机变量序列, 若$X_n \xrightarrow{P} a$, $Y_n \xrightarrow{P} b$, 则$\{X_n\}$与$\{Y_n\}$的加, 减, 乘, 除, 依概率收敛到 a 与 b 的加, 减, 乘, 除:

$$
\begin{aligned}
&(1)\ X_n \pm Y_n \xrightarrow{P} a \pm b \\
&(2)\ X_n \times Y_n \xrightarrow{P} a \times b \\
&(3)\ X_n \div Y_n \xrightarrow{P} a \div b (b \neq 0)
\end{aligned}
$$

下面证明(1), 乘除运算的收敛证明见教材.

证明(1)： $X_n \pm Y_n \xrightarrow{P} a \pm b$

因为:

$$\left\{|(X_n+Y_n)-(a+b)| \geqslant \varepsilon\right\} \subset \left\{\left(|X_n-a| \geqslant \frac{\varepsilon}{2}\right) \cup \left(|Y_n-b| \geqslant \frac{\varepsilon}{2}\right)\right\}$$

所以

$$\begin{aligned} 0 &\leqslant P\left(|(X_n+Y_n)-(a+b)| \geqslant \varepsilon\right) \\ &\leqslant P\left(|X_n-a| \geqslant \frac{\varepsilon}{2}\right) + P\left(|Y_n-b| \geqslant \frac{\varepsilon}{2}\right) \to 0 \quad (n \to \infty) \end{aligned}$$

即

$$P\left(|(X_n+Y_n)-(a+b)| < \varepsilon\right) \to 1 \quad (n \to \infty)$$

由此得, $X_n + Y_n \xrightarrow{P} a+b$, 同理可证$X_n - Y_n \xrightarrow{P} a-b$

定义4.1.2 设随机变量$X, X_1, X_2, \ldots$的分布函数分别为$F(x), F_1(x), F_2(x), \ldots$. 若在$F(x)$的**连续点上都有**

$$\lim_{n \to +\infty} F_n(x) = F(x)$$

则称分布函数列$\{F_n(x)\}$ **弱收敛** (weak convergence)于$F(x)$, 记为

$$F_n(x) \xrightarrow{W} F(x)$$

相应记

$$X_n \xrightarrow{L} X$$

$\{X_n\}$**按分布收敛**(convergence in distribution)于X.

($\mathcal{L}$: law of probability distribution of X, 概率分布律, 简称概率分布). 按分布收敛也可记为 $X_n \xrightarrow{D} X$ 或 $X_n \xrightarrow{d} X$.

注意

- 弱收敛是对分布函数列$\{F_n(x)\}$而言的.
- 依分布收敛是对具有分布函数列$\{F_n(x)\}$的随机变量序列$\{X_n\}$而言的.
- 弱收敛和依分布收敛不考虑非连续点上分布函数列$\{F_n(x)\}$的性质.

定理4.1.2

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{L} X$$

逆命题不成立.

证明: 设随机变量$X, X_1, X_2, \ldots$的分布函数分别为$F(x), F_1(x), F_2(x), \ldots$. 为证$X_n \xrightarrow{L} X$, 相当于证$F_n(x) \xrightarrow{W} F(x)$, 所以只需证对所有的$x$, 有

$$F(x-0) \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant F(x+0)$$

$\varliminf_{n\to\infty} F_n(x)$表示下极限, $\varlimsup_{n\to\infty} F_n(x)$表示上极限.

若上式成立, 则当x是$F(x)$的连续点时, 有$F(x-0) = F(x+0)$, 由此可得$F_n(x) \xrightarrow{W} F(x)$.

为证

$$F(x-0) \leqslant \varliminf_{n\to\infty} F_n(x) \leqslant \varlimsup_{n\to\infty} F_n(x) \leqslant F(x+0)$$

先令$x' < x$, 则

$$\begin{aligned}\{X \leqslant x'\} &= \{X_n \leqslant x, X \leqslant x'\} \cup \{X_n > x, X \leqslant x'\} \\ &\subset \{X_n \leqslant x\} \cup \{|X_n - X| \geqslant x - x'\}\end{aligned}$$

所以$F(x') \leqslant F_{\mathrm{n}}(x) + P(|X_{\mathrm{n}} - X| \geqslant x - x')$

由$X_n \xrightarrow{P} X$, 得$P(|X_n - X| \geqslant x - x') \to 0 (n \to \infty)$ 所以有

$$F(x') \leqslant \varliminf_{n\to\infty} F_n(x)$$

再令$x' \to x$, 得

$$F(x-0) \leqslant \varliminf_{n\to\infty} F_n(x)$$

同理可证, 当$x'' > x$时, 有

$$\overline{\lim_{n\to\infty}} F_n(x) \leqslant F\left(x''\right)$$

令$x'' \to x$, 得

$$\overline{\lim_{n\to\infty}} F_n(x) \leqslant F(x+0)$$

$Q.E.D.$

依概率收敛与按分布收敛的关系

定理4.1.2

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{L} X$$

逆命题不成立. 反例如下:

设随机变量X的分布列为

$$P(X=-1)=\frac{1}{2}, \quad P(X=1)=\frac{1}{2}$$

令$X_n=-X$, 则X_n与X同分布, 即X_n与X有相同的分布函数, 故$X_n \xrightarrow{L} X$.

但对任意$0<\epsilon<2$, 有

$$P\left(|X_n-X| \geqslant \varepsilon\right)=P(2|X| \geqslant \varepsilon)=1 \neq 0$$

所以X_n依分布收敛于X, 但不是依概率收敛于X.

依概率收敛与按分布收敛的关系

一般按分布收敛和依概率收敛不等价, 但当随机变量退化为常数时, 二者等价.

定理4.1.3 若c为常数, 则

$$X_n \xrightarrow{L} c \Leftrightarrow X_n \xrightarrow{P} c$$

必要性已证明, 下面证明充分性.

证明： 因为c为常数, 故

$$F(x)=\begin{cases}0, & x<c\\ 1, & x\geqslant c\end{cases}$$

对于任意的$\varepsilon>0$,

$$\begin{aligned}P\left(|X_n-c|\geqslant\varepsilon\right)&=P\left(X_n\geqslant c+\varepsilon\right)+P\left(X_n\leqslant c-\varepsilon\right)\\&\leqslant P\left(X_n>c+\varepsilon/2\right)+P\left(X_n\leqslant c-\varepsilon\right)\\&=1-F_n(c+\varepsilon/2)+F_n(c-\varepsilon)\end{aligned}$$

由于 $x=c+\varepsilon/2$和$x=c-\varepsilon$均为$F(x)$的连续点,
且$F_n(x)\longrightarrow F(x)$, 所以当$n\to\infty$, 有

$$F_{\mathrm{n}}(c+\varepsilon/2)\to F(c+\varepsilon/2)=1,\quad F_n(c-\varepsilon)\to F(c-\varepsilon)=0$$

所以

$$P\left(|X_n-c|\geqslant\varepsilon\right)\to 0$$

即$X_n\xrightarrow{P}c$, $Q.E.D.$

第二版教材错误(第三版已改)

pp. 213 习题4.1 第一题

1. 如果 $X_n \xrightarrow{P} X$, 且 $Y_n \xrightarrow{P} Y$. 试证 $: P(X = Y) = 1$.

应为

1. 如果 $X_n \xrightarrow{P} X$, 且 $X_n \xrightarrow{P} Y$. 试证 $: P(X = Y) = 1$.

第14次作业

习题4.1中题目1, 18

习题4.2中题目1, 2, 5, 6, 13, 15

§4.2 特征函数(characteristic function)

特征函数是处理概率论问题的有力工具, 引入特征函数的概念提供了一种研究随机变量的方法. 其作用在于:

- 可将卷积运算化成乘法运算
- 可将求各阶矩的积分运算化成微分运算
- 可将求随机变量序列的极限分布化成一般的函数极限问题
- **证明大数定律**
- **证明中心极限定理**

§4.2 特征函数(characteristic function)

特征函数是处理概率论问题的有力工具, 引入特征函数的概念提供了一种研究随机变量的方法. 其作用在于:

- 可将卷积运算化成乘法运算
- 可将求各阶矩的积分运算化成微分运算
- 可将求随机变量序列的极限分布化成一般的函数极限问题
- **证明大数定律**
- **证明中心极限定理**

在某些情况下, 利用分布函数求解随机变量问题较困难. 比如求多个独立随机变量和的分布时, 用分布函数求解的话, 涉及到多重卷积, 非常困难. 而转换成特征函数就相对简单些. 所以特征函数在大数定律和中心极限定律分析中应用广泛, 因为这些定律都涉及到多个独立随机变量和.

前导知识1: 复随机变量

定义: 复随机变量 设 $X(w), Y(w)$ 是定义在Ω上的**实随机变量**, 则称$Z = Z(w) = X(w) + iY(w)$为**复随机变量**. $i = \sqrt{-1}$ 是虚数单位.

- 称$\overline{Z} = X(w) - iY(w)$为$Z(w)$的**复共轭(conjugate)随机变量**.
- 复随机变量$Z = X(w) + iY(w)$的**模(modulus)**$|Z|$定义为$\sqrt{X^2 + Y^2}$, 或$|Z|^2 = X^2 + Y^2$
- $Z\overline{Z} = X^2 + Y^2 = |Z|^2$

前导知识2: 复随机变量

- 复随机变量的**期望**: $E(Z) = E(X) + iE(Y)$
- 复随机变量的**独立性**: $Z_1 = X_1 + iY_1$和$Z_2 = X_2 + iY_2$相互独立**当且仅当**(X_1, Y_1)与(X_2, Y_2)相互独立
  - (X_1, Y_1)与(X_2, Y_2)相互独立：即联合分布=边际分布乘积 $F(x_1, y_1, x_2, y_2) = F(x_1, y_1)F(x_2, y_2)$
- **欧拉公式**: $e^{iX} = \cos X + i \sin X$ (两侧同时泰勒展开即可证明成立), 其模$|e^{iX}| = \sqrt{\cos^2 X + \sin^2 X} = 1$
- 若X与Y独立, 则e^{iX}和e^{iY}也独立

定义4.2.1 设 X 是一随机变量, 称$\varphi(t)=E(e^{itX})$, $-\infty<t<\infty$ 为 X**的特征函数**.

- 当离散随机变量X的分布列为

$$p_k=P\left(X=x_k\right), k=1,2,\cdots$$

时, 则X的特征函数为

$$\varphi(t)=\sum_{k=1}^{\infty} e^{itx_k}p_k, \quad -\infty<t<\infty$$

- 当连续随机变量X的密度函数为$p(x)$, 则X的特征函数为

$$\varphi(t)=\int_{-\infty}^{\infty} \mathrm{e}^{itx}p(x)\mathrm{d}x, \quad -\infty<t<\infty$$

上述变换是$p(x)$ **的傅里叶变换**

注意

- 特征函数不是随机变量, 是一般函数, 具有一般函数的性质. 如讨论特征函数收敛性时不存在依概率收敛或依分布收敛的概念!
- 因为$|e^{itX}| = \sqrt{cos^2(tX) + sin^2(tX)} = 1$, 小于无穷大, 所以$E(e^{itX})$**必定存在**, 即任一随机变量的特征函数总是存在的.
- 随机变量的分布相同, 特征函数也相同 (定理4.2.4), 所以也常称某**分布的特征函数**.
- 无论随机变量是连续的还是离散的, 其特征函数均为定义在$-\infty < t < \infty$的连续函数(性质4.2.1).

注意

特征函数的计算中用到复变函数, 为此注意:

(1) 欧拉公式: $e^{ix} = cos(x) + i \cdot sin(x)$;
$e^{-ix} = cos(-x) + i \cdot sin(-x) = cos(x) - i \cdot sin(x)$

(2) 复数的共轭(conjugate): $\overline{a+bi} = a - bi$

(3) 复数的模(modulus): $|a+bi| = \sqrt{a^2+b^2}$

常用分布的特征函数

- **单点分布** $P(X=a)=1$, 其特征函数为

$$\varphi(t)=\sum_{k=1}^{\infty} e^{itx_k}p_k=\mathrm{e}^{ita}$$

- **0-1 分布** $P(X=x)=p^x(1-p)^{1-x}$, $x=0,1$, 其特征函数为

$$\begin{aligned}\varphi(t)&=\sum_{k=1}^{\infty} e^{itx_k}p_k=(1-p)e^{it0}+pe^{it1}\\&=p\mathrm{e}^{it}+q, \text{其中} q=1-p\end{aligned}$$

常用分布的特征函数

- **泊松分布** $P(\lambda)$, $P(X=k)=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$, $k=0,1,\cdots$, 其特征函数为

$$\varphi(t)=\sum_{k=0}^{\infty}\mathrm{e}^{\mathrm{i}kt}\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}=\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda e^{it}}=\mathrm{e}^{\lambda\left(\mathrm{e}^{it}-1\right)}$$

(中间推导为$\mathrm{e}^{\lambda e^{it}}$在$\lambda=0$点的泰勒展开逆推)

- **均匀分布** 因为密度函数为

$$p(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & \text{其他}\end{cases}$$

所以, 其特征函数为

$$\varphi(t)=\int_a^b\frac{e^{itx}}{b-a}dx=\frac{e^{ibt}-e^{iat}}{it(b-a)}$$

标准正态分布 $N(0,1)$, 因为密度函数为

$$p(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}, \quad -\infty < x < \infty$$

所以, 其特征函数为

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}tx} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} \frac{(\mathrm{i}tx)^n}{n!} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x$$
$$= \sum_{n=0}^{\infty} \frac{(\mathrm{i}t)^n}{n!} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^n \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x \right]$$

上式中括号内是标准正态分布n阶矩$E(X^n)$, 当n为奇数时$E(X^n)=0$, 当n为偶数时, 如$n=2m$,

$$E\left(X^n\right)=E\left(X^{2m}\right)=(2m-1)!!=\frac{(2m)!}{2^m\cdot m!}\text{(教材例2.7.1)}$$

(! !: double factorial,双阶乘, 相同奇偶性的所有正整数乘积)

带入原式,

$$\varphi(t)=\sum_{m=0}^{\infty}\frac{(\mathrm{i}t)^{2m}}{(2m)!}\cdot\frac{(2m)!}{2^m\cdot m!}=\sum_{m=0}^{\infty}\left(-\frac{t^2}{2}\right)^m\frac{1}{m!}=\mathrm{e}^{-\frac{t^2}{2}}$$

常用分布的特征函数

- **指数分布** $Exp(\lambda)$ 因为密度函数为

$$p(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x}, & x > 0 \\ 0, & x \leqslant 0 \end{cases}$$

所以, 其特征函数为 (推导方法与教材不同)

$$\begin{aligned} \varphi(t) &= \int_0^{\infty} \mathrm{e}^{\mathrm{itx}} \lambda \mathrm{e}^{-\lambda \mathrm{x}} \mathrm{d}x = \lambda \int_0^{\infty} \mathrm{e}^{(\mathrm{it}-\lambda)\mathrm{x}} \mathrm{d}x \\ &= \frac{\lambda}{it-\lambda} \cdot e^{(it-\lambda)x} \Big|_0^{+\infty} = \frac{\lambda}{it-\lambda} (\cos(tx) + i\sin(tx)) e^{-\lambda x} \Big|_0^{+\infty} \\ &= \frac{\lambda}{it-\lambda}(0-1) = \left(1 - \frac{\mathrm{i}t}{\lambda}\right)^{-1} \end{aligned}$$

上式利用了欧拉公式: $\mathrm{e}^{\mathrm{i}tx} = \cos(tx) + \mathrm{i}\sin(tx)$

性质4.2.1 (特征函数$\varphi(t)=E(\mathrm{e}^{\mathrm{i}tX})$的模) $|\varphi(t)|\leq\varphi(0)=1$

证明：(与教材中证明方法不同)

$$
\begin{aligned}
&|\varphi(t)|=\left|E(\mathrm{e}^{\mathrm{i}tX})\right|=|E[cos(tX)+\mathrm{i}sin(tX)]| \\
&=|E[cos(tX)]+\mathrm{i}E[sin(tX)]|=\sqrt{E^2[cos(tX)]+E^2[sin(tX)]} \\
&\leqslant\sqrt{E[cos^2(tX)]+E[sin^2(tX)]}=1=\varphi(0)
\end{aligned}
$$

性质4.2.2 $\varphi(-t)=\overline{\varphi(t)}$ **(共轭)**

证明:

$$
\begin{aligned}
\varphi(-t)&=E\left[e^{-itX}\right]=E[\cos(tx)+i\sin(-tx)] \\
&=E[\cos(tx)-i\sin(tx)]=\overline{\varphi(t)}
\end{aligned}
$$

性质4.2.3 若$Y=aX+b$, 则$\varphi_Y(t)=e^{ibt}\varphi_X(at)$

证明:

$$\varphi_Y(t)=E\left(\mathrm{e}^{it(aX+b)}\right)=\mathrm{e}^{ibt}E\left(\mathrm{e}^{\mathrm{iat}X}\right)=\mathrm{e}^{ibt}\varphi_X(at)$$

性质4.2.4 (独立随机变量之和的特征函数) 若 X与 Y 独立, 则 $\varphi_{X+Y}(t)=\varphi_X(t)\varphi_Y(t)$

证明: 因为 X与 Y 独立, 所以e^{itX}和e^{itY}也独立, 所以

$$E\left(\mathrm{e}^{it(X+Y)}\right)=E\left(\mathrm{e}^{itX}\mathrm{e}^{itY}\right)=E\left(\mathrm{e}^{itX}\right)E\left(\mathrm{e}^{itY}\right)=\varphi_X(t)\cdot\varphi_Y(t)$$

性质4.2.5 若$E(X^l)$存在, 则X的特征函数可l次求导, 且对$1 \leqslant k \leqslant l$, 有$\varphi^{(k)}(0) = i^k E(X^k)$.

证明: 因为

$$\varphi(t) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}tx} p(x) \mathrm{d}x, \quad -\infty < t < \infty$$

所以对$\varphi(t)$求k次导得

$$\varphi^{(k)}(t) = \int_{-\infty}^{\infty} \mathrm{i}^k x^k \mathrm{e}^{\mathrm{i}tx} p(x) \mathrm{d}x = \mathrm{i}^k E\left(X^k \mathrm{e}^{\mathrm{i}tX}\right)$$

令$t = 0$,

$$\varphi^{(k)}(0) = \mathrm{i}^k E\left(X^k\right)$$

- **可以用上式求随机变量的各阶矩.** 特别地, 期望和方差

$$E(X) = \frac{\varphi'(0)}{\mathrm{i}}, \quad \mathrm{Var}(X) = -\varphi''(0) + (\varphi'(0))^2$$

常用分布的特征函数2

- **二项分布**$b(n,p)$
  若Y服从二项分布$Y \sim b(n,p)$, 则Y是n个服从伯努利分布$b(1,p)$的独立同分布的$X_1, ...X_n$的和,
  $Y = X_1 + ... + X_n$. 因为$X_i, i = 1, ...n$的特征函数为

$$\varphi_{X_i}(t) = p\mathrm{e}^{it} + q$$

由性质4.2.4, Y特征函数为

$$\varphi_Y(t) = \left(p\mathrm{e}^{\mathrm{i}t} + q\right)^n$$

常用分布的特征函数2

- **一般正态分布**$N(\mu, \sigma^2)$
  若Y服从一般正态分布$Y \sim N(\mu, \sigma^2)$, X服从标准正态分布$X \sim N(0, 1)$, 则$Y = \sigma X + \mu$. 因为X的特征函数为

$$\varphi_X(t) = \mathrm{e}^{-\frac{t^2}{2}}$$

由性质**4.2.3**, Y特征函数为

$$\varphi_Y(t) = \varphi_{\sigma X+\mu}(t) = \mathrm{e}^{\mathrm{i}\mu t}\varphi_X(\sigma t) = \exp\left\{\mathrm{i}\mu t - \frac{\sigma^2 t^2}{2}\right\}$$

常用分布的特征函数2

由上述性质亦可推出 (证明见教材)

- **伽玛分布**$Ga(n,\lambda)$**的特征函数为**

$$\varphi_Y(t) = \left(1 - \frac{\mathrm{i}t}{\lambda}\right)^{-n}$$

(独立同分布的指数分布之和, 应用性质4.2.4可推出)

- **卡方分布**$\chi^2(n)$**的特征函数为**

$$\varphi(t) = (1 - 2it)^{-n/2}$$

(Gamma分布的特殊形式:$\chi^2(n) = Ga(n/2, 1/2)$)

教材错误

p.p. 196页几何分布

$$p_k = pq^{k-1}, k = 1, 2, \cdots$$

的特征函数应为

$$\frac{pe^{it}}{1 - qe^{it}}$$

而非

$$\frac{p}{1 - qe^{it}}$$

请利用几何分布的分布列推导几何分布的特征函数.

特征函数的定理

定理4.2.1(略) (一致连续性) $\varphi(t)$在$(-\infty,\infty)$上任意点连续(即一致连续 uniformly continuous). 证明略.

定理4.2.2(略) (非负定性) 随机变量X的特征函数$\varphi(t)$是非负定的, 即对任意n个实数$t_1,t_2,\ldots,t_n$和n个复数$z_1,z_2,\ldots,z_n$有

$$\sum_{k=1}^{n}\sum_{j=1}^{n}\varphi\left(t_k-t_j\right)z_k\bar{z}_j\geqslant 0$$

定理4.2.3(略) (逆转公式) 设$F(x)$和$\varphi(t)$分别为随机变量X的分布函数和特征函数, 则对$F(x)$的任意两个连续点$x_1 < x_2$, 有

$$F(x_2) - F(x_1) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-itx_1} - e^{-itx_2}}{it} \varphi(t)dt$$

定理4.2.4 (唯一性定理) 随机变量的分布函数由其特征函数唯一决定.

证明: 对$F(x)$的每一个连续点x, 当y沿着$F(x)$的连续点趋向于$-\infty$时, 由逆转公式得

$$F(x) = \lim_{y\to-\infty} \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ity} - e^{-itx}}{it} \varphi(t)dt$$

而分布函数由其连续点上的值唯一决定, 故结论成立.

定理4.2.5 当X为连续随机变量, 其密度函数为$p(x)$, 特征函数为$\varphi(t)$. 如果$\int_{-\infty}^{\infty} |\varphi(t)|dt < \infty$, 则

$$p(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-itx}\varphi(t)\mathrm{d}t$$

证明见教材.

所以, 特征函数是密度函数的傅里叶变换, 密度函数是特征函数的傅里叶逆变换:

$$\begin{aligned} \varphi(t) &= \int_{-\infty}^{\infty} e^{itx}p(x)dx \\ p(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi(t)dt \end{aligned}$$

4.2.3 特征函数唯一决定分布函数

定理4.2.6 分布函数序列$\{F_n(x)\}$弱收敛于分布函数$F(x)$的充要条件是$\{F_n(x)\}$的特征函数序列$\{\varphi_n(x)\}$收敛于$F(x)$的特征函数$\varphi(t)$.

该定理说明分布函数序列的弱收敛性与其相应的特征函数序列的点点收敛性是等价的.

证明略.

第14次作业

习题4.1中题目1, 18

习题4.2中题目1, 2, 5, 6, 13, 15

§4.3 大数定律(Law of Large Numbers)

- 大数定律是统计学的基础.
- 讨论 "概率是频率的稳定值" 的确切含义.
- 给出几种重要的大数定律: 伯努利大数定律, 切比雪夫大数定律, 马尔可夫大数定律, 辛钦大数定律.

- 利用切比雪夫不等式证明伯努利大数定律.
- 利用切比雪夫不等式证明切比雪夫大数定律.
- 利用切比雪夫不等式证明(特定情况下)马尔可夫大数定律.
- 利用特征函数证明辛钦大数定律.

4.3.1 伯努利大数定律

记S_n 是n重伯努利试验中事件A出现的次数, 称$\dfrac{S_n}{n}$为事件A出现的**频率(relative frequency)**.

记每次试验中A发生的概率为p, $P(A) = p$, 则S_n服从二项分布$b(n,p)$, $E(S_n) = np$, $Var(S_n) = np(1-p)$. 因此频率$\dfrac{S_n}{n}$的期望和方差分别为

$$E\left(\frac{S_n}{n}\right) = p, \quad \mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{p(1-p)}{n}$$

即**频率的期望**就是**概率**. (另外, 频率的方差随着n增大趋向于0.)

那么当$n \to \infty$时, $\left|\dfrac{S_n}{n} - p\right|$的极限性质是什么?

定理4.3.1 (伯努利大数定律)

设S_n 是n重伯努利试验中事件A出现的次数, 每次试验中 $P(A)=p$, 则对任意的 $\varepsilon>0$, 有

$$\lim_{n\to+\infty} P\left(\left|\frac{S_n}{n}-p\right|<\varepsilon\right)=1$$

伯努利大数定律说明, 随着n的增大, 事件的频率和概率的偏差大于预先给定精度的可能性越来越小, 频率稳定于概率.

大数定律是统计学的根本基础. 提供了用频率确定概率的理论依据.

证明用到切比雪夫不等式.

回顾2.3.3 切比雪夫不等式

定理2.3.1 设随机变量X的期望和方差存在, 则对任意正数ε, 有下面不等式成立

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

$$P(|X - E(X)| < \varepsilon) \geq 1 - \frac{Var(X)}{\varepsilon^2}$$

切比雪夫不等式给出了大偏差发生概率的上界(upper bound), 方差越大, 上界越大.

证明伯努利大数定律:

因为$S_n \sim b(n,p)$, 且

$$E\left(\frac{S_n}{n}\right) = p, \quad \mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{p(1-p)}{n}$$

由切比雪夫不等式(2.3.3)

$$1 \geqslant P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) \geqslant 1 - \frac{\mathrm{Var}\left(\frac{S_n}{n}\right)}{\varepsilon^2} = 1 - \frac{p(1-p)}{n\varepsilon^2}$$

当$n \to \infty$时, $1 - \frac{p(1-p)}{n\varepsilon^2} \to 1$, 因此

$$\lim_{n\to+\infty} P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) = 1$$

Q.E.D.

注意: 不加概率P的说法 "$\left|\frac{S_n}{n}-p\right|<\varepsilon$总成立" 不成立.

因为对$0<p<1$,

$$P\left(\frac{S_n}{n}=1\right)=P\left(X_1=1,\cdots,X_n=1\right)=p^n$$
$$P\left(\frac{S_n}{n}=0\right)=P\left(X_1=0,\cdots,X_n=0\right)=(1-p)^n$$

即使n很大, 上式虽极小但不为0.

4.3.2 常用的几个大数定律

利用伯努利大数定律给出**大数定律一般化形式**, 并介绍常用的大数定律.

伯努利大数定律讨论的是一个独立同分布(IID)的随机变量序列$\{X_n\}$, 其共同分布为两点分布$b(1,p)$. 记

$$X_i = \begin{cases} 1, \text{第}i\text{次试验}A\text{发生} \\ 0, \text{第}i\text{次试验}A\text{不发生} \end{cases}$$

则$\{X_n\}$是独立的二点分布随机变量序列. 该序列前n个随机变量之和$S_n = \sum_{i=1}^{n} X_i$有如下变换

$$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad p = E(\frac{S_n}{n}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right)$$

大数定律一般化形式

所以伯努利大数定律又可写成

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right)\right| < \varepsilon\right) = 1$$

上式具有一般化意义(频率和概率的关系 to 随机变量序列(的观察)的算数平均和期望的算术平均的关系)

在独立同分布情况下, 上式可以简化为?

大数定律一般化形式

所以伯努利大数定律又可写成

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right)\right| < \varepsilon\right) = 1$$

上式具有一般化意义(频率和概率的关系 to 随机变量序列(的观察)的算数平均和期望的算术平均的关系)

在独立同分布情况下, 上式可以简化为?

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E\left(X_i\right)\right| < \varepsilon\right) = 1$$

为什么上式重要?

大数定律一般化形式

所以伯努利大数定律又可写成

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right)\right| < \varepsilon\right) = 1$$

上式具有一般化意义(频率和概率的关系 to 随机变量序列(的观察)的算数平均和期望的算术平均的关系)

在独立同分布情况下, 上式可以简化为?

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E\left(X_i\right)\right| < \varepsilon\right) = 1$$

为什么上式重要? 现实中大量时候不知道随机变量分布因而不知道期望, 不过可以通过n个随机变量的观察的均值近似求得期望. (当然, 需要在大数定理成立的条件下才成立.)

大数定律一般化形式

定义4.3.1 若随机变量序列$\{X_n\}$满足:

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right)\right| < \varepsilon\right) = 1$$

则称该随机变量序列$\{X_n\}$**服从大数定律**.

对不同性质的随机变量序列$\{X_n\}$(独立, 相关, 同分布, 不同分布), 下面给出不同的大数定律.

定理4.3.2 设$\{X_n\}$为一列两两不相关的随机变量序列, 若每个X_i方差存在, 且每个X_i方差存在并有共同的上界, 即$Var(X_i) \leqslant c, i = 1, 2, ...,$ 则 $\{X_n\}$服从大数定律, 即对任意$\varepsilon > 0$, 有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right)\right| < \varepsilon\right) = 1$$

证明同样用到切比雪夫不等式.

证明: 因为$\{X_n\}$为一列两两不相关的随机变量序列, 故

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left(X_i\right)\leqslant\frac{c}{n}$$

由切比雪夫不等式, 对任意 $\varepsilon>0$, 有

$$\begin{aligned}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E\left(X_i\right)\right|<\varepsilon\right)&\geqslant 1-\frac{\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)}{\varepsilon^2}\\&\geqslant 1-\frac{c}{n\varepsilon^2}\end{aligned}$$

所以当$n\to\infty$时有

$$\lim_{n\to\infty}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E\left(X_i\right)\right|<\varepsilon\right)=1$$

Q.E.D.

注意

- 切比雪夫大数定律只要求$\{X_n\}$互不相关, 并不要求同分布.
- **切比雪夫大数定律推论: 如果随机变量序列独立同分布且方差有限, 则必定服从大数定律.**
- 伯努利大数定律是切比雪夫大数定律的特例.

例4.3.2 设$\{X_n\}$是独立同分布的随机变量序列, $E\left(X_n^4\right) < \infty$. 若令$E\left(X_n\right) = \mu$, $\operatorname{Var}\left(X_n\right) = \sigma^2$, 考察

$$Y_n = \left(X_n - \mu\right)^2, \quad n = 1, 2, \cdots$$

则随机变量序列 $\{Y_n\}$ 服从大数定律, 即对任意的 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)^2 - \sigma^2\right| \geqslant \varepsilon\right) = 0$$

例4.3.2

证明： 显然 $\{Y_n\}$ 是独立同分布随机变量序列, 其方差

$$\operatorname{Var}(Y_n) = \operatorname{Var}(X_n - \mu)^2 = E(X_n - \mu)^4 - \sigma^4$$

由于 $E\left(X_n^4\right)$ 存在, 故 $E\left(X_n^3\right), E\left(X_n^2\right), E(X_n - \mu)^4$ 也都存在.

所以$\operatorname{Var}(Y_n)$有共同上界.

由切比雪夫大数定律知$\{Y_n\}$ 服从大数定律. $Q.E.D.$

马尔可夫大数定律 (Markov's LLN)

定理4.3.3 若随机变量序列$\{X_n\}$满足, 在$n \to \infty$时有:

$$\frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \to 0(\text{马尔可夫条件})$$

则 $\{X_n\}$服从大数定律, 即对任意 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right)\right| < \varepsilon\right) = 1$$

- 马尔可夫大数定律对随机变量序列$\{X_n\}$已经没有任何同分布, 独立性, 不相关的假设
- 马尔可夫大数定律可推出切比雪夫大数定律
- 教材中说“利用切比雪夫不等式即可得证”不准确. 在随机变量独立的条件下, 证明可直接由切比雪夫不等式得出; 但在不独立情况下证明较难(略).

辛钦大数定律 (Khinchin's LLN)

定理4.2.4 若随机变量序列$\{X_n\}$独立同分布, 且X_i的数学期望存在, 则$\{X_n\}$服从大数定律, 即对任意 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right)\right| < \varepsilon\right) = 1$$

- 辛钦大数定律要求独立同分布, 不要求方差存在(可为无穷大).

辛钦大数推论1:

- 对随机变量X独立观察n次, 记每次观察值为X_i, 当n足够大时, 可以用$\frac{1}{n}\sum_{i=1}^{n} X_i$近似$E(X)$.

辛钦大数推论2:

- 若$E(X^k)$存在, 对随机变量X独立观察n次,记每次观察值为X_i, 当n足够大时, 可以用$\frac{1}{n}\sum_{i=1}^{n} X_i^k$近似$E(X^k)$.
(因为若随机变量序列$\{X_n\}$独立同分布, 则$\{X_n^k\}$独立同分布)

辛钦大数定律的证明

证明: 随机变量序列$\{X_n\}$独立同分布,
且$E(X_i) = a, i = 1, 2, \ldots$. 要证明

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{p} a, \quad n \to \infty$$

为此记

$$Y_{\text{n}} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

由定理4.1.3 ($X_n \xrightarrow{P} a \Leftrightarrow X_n \xrightarrow{L} a$), 只需证$Y_n \xrightarrow{L} a$, 又由定理4.2.6 (特征函数唯一决定分布函数), 只需证$\varphi_{Y_n}(t) \to \varphi_a(t) = \mathrm{e}^{iat}$

辛钦大数定律的证明

随机变量序列$\{X_n\}$独立同分布，所以有相同的特征函数$\varphi(t)$. 又因为性质4.2.5 $\varphi'(0)/i = E(X_i) = a$，从而$\varphi(t)$在0点的泰勒展开为

$$\varphi(t) = \varphi(0) + \varphi'(0)t + o(t) = 1 + \mathrm{i}at + o(t)$$

再由$\{X_n\}$的独立性知道Y_n的特征函数为(性质4.2.4)

$$\varphi_{Y_n}(t) = \left[\varphi\left(\frac{t}{n}\right)\right]^n = \left[1 + \mathrm{i}a\frac{t}{n} + o\left(\frac{1}{n}\right)\right]^n$$

对任意的t有

$$\lim_{n\to\infty}\left[\varphi\left(\frac{t}{n}\right)\right]^n = \lim_{n\to\infty}\left[1 + ia\frac{t}{n} + o\left(\frac{1}{n}\right)\right]^n = \mathrm{e}^{iat}$$

故$\varphi_{Y_n}(t) \to \phi_a(t) = \mathrm{e}^{iat}$, $Q.E.D.$

注意点

- (1) 伯努利大数定律是切比雪夫大数定律的特例.
- (2) 切比雪夫大数定律是马尔可夫大数定律的特例.
- (3) 伯努利大数定律是辛钦大数定律的特例.

第15次作业

习题4.3中第5, 11, 12, 13题

§4.4 中心极限定理 (Central Limit Theorem, CLT)

大数定律讨论的是随机变量序列的**算术平均**的**概率极限**问题 (特例是讨论频率的概率极限问题).

中心极限定理讨论的是**随机变量和**的**分布极限**问题.

为何学习中心极限定理

第一：

- 独立随机变量$\{X_n\}$ 的和$Y_n = \sum_{i=1}^{n} X_i$是一类重要的随机变量, 可描述大量现实现象.
- 但在n很大时, 其精确分布很难求得(例4.4.2 多重卷积).
- 通过中心极限定理, 可描述其渐进分布(asymptotic distribution).

第二：

- 中心极限定理在统计学和计量经济学中有极其广范的应用.

4.4.1 独立随机变量和

设 $\{X_n\}$ 为独立随机变量序列, 记其和为

$$Y_n = \sum_{i=1}^{n} X_i$$

对Y_n标准化,

$$Y_n^* = \frac{Y_n - E(Y_n)}{\sqrt{\mathrm{Var}(Y_n)}}$$

$E(Y_n^*) = 0, Var(Y_n^*) = 1.$

讨论标准化的独立随机变量和Y_n^*的极限分布, 中心极限定理指出(在较宽松条件下)该极限分布为正态分布.

通常地, 运用中心极限定理可推出$Y_n = \sum_{i=1}^{n} X_i$的近似分布.

4.4.2 独立同分布下的中心极限定理

定理4.4.1 林德贝格-勒维中心极限定理 (Lindeberg-Levy CLT): 设$\{X_n\}$为独立同分布随机变量序列, X_i的数学期望为μ, $E(X_i)=\mu$, 方差为$\sigma^2>0$, $Var(X_i)=\sigma^2$, 记

$$Y_n^* = \frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$$

则对任意实数y, Y_n^* 的分布弱收敛于标准正态分布, 即

$$\lim_{n\to\infty} P\left(Y_n^*\leqslant y\right) = \lim_{n\to\infty} P\left\{\frac{\sum\limits_{i=1}^{n}X_i-n\mu}{\sigma\sqrt{n}}\leq y\right\}$$

$$=\Phi(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}e^{-\frac{t^2}{2}}dt$$

即$\{Y_n^*\}$依分布收敛于标准正态分布的随机变量:

$Y_n^* \xrightarrow{L} X\sim N(0,1)$.

证明： 为证上式，只需证$\{Y_n^*\}$的分布函数列弱收敛于标准正态分布.

又由特征函数唯一性(定理4.2.6)，只需证特征函数列收敛于标准正态分布的特征函数.

为此设$X_n-\mu$的特征函数为$\varphi_{X_n-\mu}(t)$，则$\{Y_n^*\}$的特征函数为(性质4.2.3, 4.2.4)

$$\varphi_{Y_n^*}(t)=\left[\varphi_{X_n-\mu}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

又因为$E(X_n-\mu)=0, Var(X_n-\mu)=\sigma^2$，所以有

$$\varphi'_{X_n-\mu}(0)=0,\quad \varphi''_{X_n-\mu}(0)=-\sigma^2$$

于是特征函数$\varphi(t)$有展开式

$$\begin{aligned}\varphi_{X_n-\mu}(t) &= \varphi_{X_n-\mu}(0)+\varphi'_{X_n-\mu}(0)t+\varphi''_{X_n-\mu}(0)\frac{t^2}{2}+o\left(t^2\right)\\ &= 1-\frac{1}{2}\sigma^2t^2+o\left(t^2\right)\end{aligned}$$

从而

$$\lim_{n\to\infty}\varphi_{Y_n^*}(t)=\lim_{n\to\infty}\left[1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)\right]^n=e^{-t^2/2}$$

而$e^{-t^2/2}$是$N(0,1)$的特征函数. $Q.E.D.$

- 无论$\{X_i\}$独立同何种分布, 只要期望和方差存在, 其和的分布都可以用正态分布逼近.
- 应用举例 (见教材)：正态随机数的产生; 误差分析; 计算置信水平.

拓展知识

- 如果将Y_n^*进行变换写为

$$
\begin{aligned}
Y_n^* &= \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \\
&= \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \\
&= \sqrt{n}(\bar{X} - \mu)/\sigma \xrightarrow{L} N(0, 1)
\end{aligned}
$$

  其中$\bar{X}$为均值, 则n较大时, $\bar{X}$ 的渐进分布为$N\left(\mu, \sigma^2/n\right)$.
- 上式说明对一个具有一般分布的总体做简单随机抽样抽取到容量为n的"样本" (下学期介绍), 其均值(称为样本均值)的渐进分布为正态分布.

拓展知识

- 如果将Y_n^*进行变换写为

$$\begin{aligned} Y_n^* &= \frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \\ &= \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \\ &= \sqrt{n}(\bar{X}-\mu)/\sigma \xrightarrow{L} N(0,1) \end{aligned}$$

  其中$\bar{X}$为均值, 则n较大时, $\bar{X}$ 的渐进分布为$N\left(\mu,\sigma^2/n\right)$.
- 上式说明对一个具有一般分布的总体做简单随机抽样抽取到容量为n的"样本" (下学期介绍), 其均值(称为样本均值)的渐进分布为正态分布.
- 欢迎下学期选修我的**应用统计学**课程!

一个重要变换

有了

$$\lim_{n\to\infty} P\left\{\frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \leq y\right\} = \Phi(y)$$

则

$$\lim_{n\to\infty} P\left\{\sum_{i=1}^{n} X_i \leq y\sigma\sqrt{n} + n\mu\right\} = \Phi(y)$$

设$y\sigma\sqrt{n} + n\mu = K$, 则$y = \dfrac{K - n\mu}{\sigma\sqrt{n}}$, 则

$$\lim_{n\to\infty} P\left\{\sum_{i=1}^{n} X_i \leq K\right\} = \Phi(\frac{K - n\mu}{\sigma\sqrt{n}})$$

例

设 X 为一次射击中命中的环数, 其分布列为

X	10	9	8	7	6
P	0.8	0.1	0.05	0.02	0.03

求100次射击中命中环数在900环到930环之间的概率.

解: 设 X_i 为第 i 次射击命中的环数, 则X_i 独立同分布, 且 $E(X_i) = 9.62$, $Var(X_i) = 0.82$, 则

$$
\begin{aligned}
P\left\{900 \le \sum_{i=1}^{100} X_i \le 930\right\} &\approx \Phi\left(\frac{930 - 100 \times 9.62}{\sqrt{100 \times 0.82}}\right) \\
&-\Phi\left(\frac{900 - 100 \times 9.62}{\sqrt{100 \times 0.82}}\right) \\
&= \Phi(-3.53) - \Phi(-6.85)
\end{aligned}
$$

课堂练习(5')

每袋味精的净重为随机变量, 平均重量为 100克, 标准差为10克. 一箱内装200袋味精, 求一箱味精的净重大于20, 500克的概率?

解:

课堂练习(5')

每袋味精的净重为随机变量, 平均重量为 100克, 标准差为10克. 一箱内装200袋味精, 求一箱味精的净重大于20,500克的概率?

解: 设箱中第i 袋味精的净重为 X_i, 则X_i 独立同分布, 且 $E(X_i) = 100$, $Var(X_i) = 100$,

由中心极限定理得, 所求概率为:

$$
\begin{aligned}
P\left\{\sum_{i=1}^{n} X_i > 20500\right\} &\approx 1 - \Phi\left(\frac{20500 - 200 \times 100}{\sqrt{200 \times 100}}\right) \\
&= 1 - \Phi(3.54) = 0.0002
\end{aligned}
$$

故一箱味精的净重大于20,500克的概率为0.0002. (很小)

4.4.3 二项分布的正态近似

定理4.4.2 棣莫弗—拉普拉斯(de Moivre-Laplace)中心极限定理 设n重伯努利试验中，事件A在每次试验中出现的概率为$p(0 < p < 1)$，记S_n为n次试验中事件A出现的次数，则S_n为服从二项分布$b(n, p)$的随机变量，且记

$$Y_n^* = \frac{S_n - np}{\sqrt{npq}}$$

其中$q = 1 - p$. 对任意实数y，有

$$\lim_{n\to\infty} P\left(Y_n^* \leqslant y\right) = \Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{t^2}{2}} \mathrm{d}t$$

棣莫弗—拉普拉斯中心极限定理是林德贝格—勒维中心极限定理的特例 (S_n是N个独立同伯努利分布随机变量的和). 证明可直接应用林德贝格—勒维中心极限定理即可推出.

注意点

二项分布是离散分布, 而正态分布是连续分布, 所以用正态分布作为二项分布的近似时, 可作如下修正(设$k_1 < k_2$均为整数):

$$
\begin{aligned}
P\left(k_1 \leqslant S_n \leqslant k_2\right) =& P\left(k_1 - 0.5 < S_n < k_2 + 0.5\right) \\
\approx & \Phi\left(\frac{k_2 + 0.5 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - 0.5 - np}{\sqrt{npq}}\right)
\end{aligned}
$$

课堂练习(5')

问题： 设每颗炮弹命中目标的概率为0.01, 求500发炮弹中命中 5 发的概率.

解:

课堂练习(5')

问题： 设每颗炮弹命中目标的概率为0.01, 求500发炮弹中命中 5 发的概率.

解： 设 X 表示命中的炮弹数, 则$X \sim b(500, 0.01)$

(1) $P(X=5) = C_{500}^{5} \times 0.01^{5} \times 0.99^{495} = 0.17635$

(2) 应用正态逼近:

$$
\begin{aligned}
P(X=5) =& P(4.5 < X < 5.5) \\
\approx& \Phi\left(\frac{5.5-5}{\sqrt{4.95}}\right) - \Phi\left(\frac{4.5-5}{\sqrt{4.95}}\right) = 0.1742
\end{aligned}
$$

棣莫弗-拉普拉斯CLT的三类应用问题

现实中有大量可以用N重伯努利实验(两种结果)描述的现象，下面介绍3类棣莫弗-拉普拉斯CLT的应用问题：

- 已知n和y，求概率$P\left(Y_n^* \leqslant y\right) \approx \Phi(y) = \beta$;
- 已知n和概率，求y (分位数);
- 已知y和概率，求n (样本量).

例4.4.5 100个独立工作 (工作的概率为0.9) 的部件组成一个系统, 求系统中至少有85个部件工作的概率.

解:

例4.4.5 100个独立工作 (工作的概率为0.9) 的部件组成一个系统, 求系统中至少有85个部件工作的概率.

解: 用$X_i = 1$表示第i个部件正常工作, 反之记为$X_i = 0$. 又记$Y = X_1 + X_2 + \cdots + X_{100}$, 则 $E(Y) = 90$, $Var(Y) = npq = 9$. 由此得:

$$P\{Y \geq 85\} \approx 1 - \Phi(\frac{85 - 0.5 - 90}{\sqrt{9}}) = 0.966$$

0.5为正态近似修正.

一, 给定n和y, 求概率

例4.4.5 100个独立工作 (工作的概率为0.9) 的部件组成一个系统, 求系统中至少有85个部件工作的概率.

解: 用$X_i = 1$表示第i个部件正常工作, 反之记为$X_i = 0$. 又记$Y = X_1 + X_2 + \cdots + X_{100}$, 则 $E(Y) = 90$, $Var(Y) = npq = 9$. 由此得:

$$P\{Y \geq 85\} \approx 1 - \Phi(\frac{85 - 0.5 - 90}{\sqrt{9}}) = 0.966$$

0.5为正态近似修正.

如果不做修正 $1 - \Phi(\frac{85 - 90}{\sqrt{9}}) = 0.962$ (相差不大).

有关修正的进一步解释

关于

$$P\{Y \geq 85\} \approx 1 - \Phi(\frac{85 - 0.5 - 90}{\sqrt{9}}) = 0.966$$

思考：为何上式的正态修正是-0.5 **而非**$+0.5$**?**

有关修正的进一步解释

关于

$$P\{Y \geq 85\} \approx 1 - \Phi(\frac{85 - 0.5 - 90}{\sqrt{9}}) = 0.966$$

思考：为何上式的正态修正是-0.5 而非$+0.5$?

因为$P\{Y \geq 85\}$中包括85, 故用$P\{Y > 85 - 0.5\}$做近似, 而非用$P\{Y > 85 + 0.5\}$做近似.

以下的-0.5或$+0.5$修正同理.

例4.4.7 有200台独立工作 (工作的概率为0.7) 的机床, 每台机床工作时需$15kw$电力. 问共需多少电力, 才可有95%的可能性保证正常生产?

解:

例4.4.7 有200台独立工作 (工作的概率为0.7) 的机床, 每台机床工作时需15kw电力. 问共需多少电力, 才可有95%的可能性保证正常生产?

解: 用$X_i = 1$表示第i台机床正常工作, 反之记为$X_i = 0$. 又记$Y = X_1 + X_2 + \cdots + X_{200}$, 则
$E(Y) = 140, Var(Y) = npq = 42.$

设供电量为y, 则从中解得

$$P\{15Y \leq y\} \approx \Phi(\frac{y/15 + 0.5 - 140}{\sqrt{42}}) \geq 0.95$$

故$y \geqslant 2252.$

例4.4.7 有200台独立工作 (工作的概率为0.7) 的机床, 每台机床工作时需15kw电力. 问共需多少电力, 才可有95%的可能性保证正常生产?

解: 用$X_i = 1$表示第i台机床正常工作, 反之记为$X_i = 0$. 又记$Y = X_1 + X_2 + \cdots + X_{200}$, 则
$E(Y) = 140, Var(Y) = npq = 42.$

设供电量为y, 则从中解得

$$P\{15Y \leq y\} \approx \Phi(\frac{y/15 + 0.5 - 140}{\sqrt{42}}) \geq 0.95$$

故$y \geqslant 2252.$

如果不做修正 $\Phi(\frac{y/15 - 140}{\sqrt{42}}) \geq 0.95$ 可得$y \geqslant 2260$ (相差不大).

例4.4.8 用调查对象中的收看比例k/n作为某电视节目的收视率p的估计. 要有 90%的把握，使k/n与p的差异不大于0.05，问至少要调查多少对象？

解:

例4.4.8 用调查对象中的收看比例k/n作为某电视节目的收视率p的估计. 要有 90%的把握, 使k/n与p的差异不大于0.05, 问至少要调查多少对象?

解: 用 Y_n表示n 个调查对象中收看此节目的人数, 则Y_n服从$b(n,p)$分布, k 为Y_n的实际取值 (某一个Y_n的实现). 由题意

$$
\begin{aligned}
&P(|Y_n/n-p|<0.05)\\
&=P\left(\frac{n(-0.05+p)-np}{\sqrt{npq}}<\frac{Y_n-np}{\sqrt{npq}}<\frac{n(0.05+p)-np}{\sqrt{npq}}\right)\\
&=P\left(\frac{-0.05n}{\sqrt{npq}}<\frac{Y_n-np}{\sqrt{npq}}<\frac{0.05n}{\sqrt{npq}}\right)\\
&\approx 2\Phi\left(0.05\sqrt{n/p(1-p)}\right)-1\geq 0.90
\end{aligned}
$$

从中解得$0.05\sqrt{n/p(1-p)}\geq 1.645$, 又由$p(1-p)\leq 0.25$, 可解得 $n\geq 270.6\Rightarrow n=271$

拓展思考

如果调查100人, 有10人观看, 未知的收视率p应该如何估算?

拓展思考

如果调查100人, 有10人观看, 未知的收视率p应该如何估算?

统计学中介绍点估计可以回答估算值的问题 (如矩估计).

定理4.4.3 林德贝格中心极限定理 设$\{X_n\}$为独立随机变量序列, 记 $E(X_i)=\mu_i$, $Var(X_i)=\sigma_i^2$, 要讨论随机变量的和$Y_n=\sum\limits_{i=1}^{n}X_i$的分布, 若任对$\tau>0$, 有

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

其中$B_n=\sqrt{Var(Y_n)}=\sqrt{\sigma_1^2+\sigma_2^2+\ldots+\sigma_n^2}$ (林德贝格条件), 则

$$\lim_{n\to\infty}P\left\{\frac{1}{B_n}\sum_{i=1}^{n}(X_i-\mu_i)\leq y\right\}=\Phi(y)$$

证明略.(林德贝格条件较难验证.)

定理4.4.4 李雅普诺夫中心极限定理 设$\{X_n\}$为独立随机变量序列, 记 $E(X_i)=\mu_i$, $Var(X_i)=\sigma_i^2$, 要讨论随机变量的和$Y_n=\sum\limits_{i=1}^{n}X_i$的分布, 若存在$\delta>0$, 满足:

$$\lim_{n\to\infty}\frac{1}{B_n^{2+\delta}}\sum_{i=1}^{n}\left(|X_i-\mu_i|^{2+\delta}\right)=0$$

其中$B_n=\sqrt{Var(Y_n)}=\sqrt{\sigma_1^2+\sigma_2^2+\ldots+\sigma_n^2}$ (李雅普诺夫条件),

则

$$\lim_{n\to\infty}P\left\{\frac{1}{B_n}\sum_{i=1}^{n}(X_i-\mu_i)\le y\right\}=\Phi(y)$$

证明略.

例4.4.9 设 $X_1, X_2, ..., X_{99}$相互独立, 且服从不同的$0-1$分布 $P(X_i=1)=p_i=1-\frac{i}{100}$. 试求$P\left(\sum_{i=1}^{99} X_i \geq 60\right)$

解: 设$X_{100}, X_{101}, \ldots$相互独立, 且与X_{99}同分布, 则可以验证$\{X_n\}$满足$\delta=1$的李雅普诺夫条件, 且$E\left(\sum_{i=1}^{99} X_i\right)=49.5$, $Var\left(\sum_{i=1}^{99} X_i\right)=16.665$. 由此得

$$P\left(\sum_{i=1}^{99} X_i \geq 60\right) = P\left\{\frac{\sum_{i=1}^{99} X_i - 49.5}{\sqrt{16.665}} \geq \frac{60-49.5}{\sqrt{16.665}}\right\}$$

$$\approx 1-\phi(2.5735)=0.005$$

第16次作业 (不交)

习题4.4中第1, 3, 4, 5, 8, 13, 14, 18题

期末考试

题型: 填空, 计算, 证明

范围: 课上习题, PPT, 作业, 教材, 教材中衍生题目

说明：请携带不具有储存、编程、查询能力的计算器

期末考试

时间: 2022-01-12 第19周星期三第8，9，10节 15:40-17:40 (一)201

第四十二条 主、监考教师应在考试前提醒考生除准备考试所必需的文具（如钢笔、圆珠笔、铅笔、橡皮、绘图仪器和无字典存储和编程功能的电子计算器）外，不得携带手机、智能手表、文曲星、商务通或带有存储、编程、查询功能的高档计算器进入座位，否则按考试违纪论处，如在考试过程中发现考生查看或使用上述工具，视为作弊。考试中不准使用自备的答题纸和草稿纸。书籍（开卷考试除外）、书包及上述工具等物品一律按要求放到指定地点。

答疑

1. 答疑时间： **1月11日**

14:00-15:00：李卫国老师助教答疑(许岷);

15：00-16：00: 李明全老师助教答疑(田嘉琳);

16：00-17：00: 郝壮老师助教答疑(田曼);

17：00-18：00: 张军欢老师助教答疑(李雨萌).

2. 答疑地点： **A940**

3. 上述所有时段各班级共享, 均可答疑.

祝大家考试顺利, 寒假愉快!