《概率论与数理统计》期末速通

5. 大数定律与中心极限定理

5.1 大数定律

[**定义5.1.1**] 设 $\{X_n\}\ \ (n=1,2,\cdots)$ 是一列随机变量序列. 对常数 a , 若对 $\forall \varepsilon>0$, 有 $\lim\limits_{n\to+\infty}P\{|X_n-a|<\varepsilon\}=1$, 即事件 $|X_n-a|<\varepsilon$ 几乎必然发生, 则称 $\{X_n\}$ **依概率收敛**于 a , 记作 $X_n\xrightarrow{P}a$.

[**定理5.1.1**] 对随机变量序列 $\{X_n\},\{Y_n\}$ 和常数 a,b , 若 $X_n\xrightarrow{P}a,Y_n\xrightarrow{P}b$, 则对 $\forall$ 二元连续函数, 随机变量序列 $g(X_n,Y_n)\xrightarrow{P}g(a,b)$.

[**注**] 本定理将用于讨论未知参数估计量的一致性(或相合性).

[**定理5.1.2**] [**弱大数定律**, **Khinchin大数定理**] 设随机变量 $X_1,X_2,\cdots$ 独立同分布, 且存在数学期望 $E(X_k)=\mu\ \ (k=1,2,\cdots)$. 对 n 个变量的算术平均值 $\dfrac{1}{n}\sum\limits_{i=1}^n X_i$, 对 $\forall\varepsilon>0$, 有 $\lim\limits_{n\to+\infty}P\left\{\left|\dfrac{1}{n}\sum\limits_{i=1}^n X_i-\mu\right|<\varepsilon\right\}=1$ 或 $\lim\limits_{n\to+\infty}P\left\{\left|\dfrac{1}{n}\sum\limits_{i=1}^n X_i-\mu\right|\geq\varepsilon\right\}=0$, 或记作 $\dfrac{1}{n}\sum\limits_{k=1}^n X_k\xrightarrow{P}\mu=E(X_k)\ \ (n\to+\infty)$.

[**证**] 下面在这些随机变量的方差 $D(X_k)=\sigma^2\ \ (k=1,2,\cdots)$ 存在的条件下证明.

因 $E\left(\dfrac{1}{n}\sum\limits_{k=1}^n X_k\right)=\dfrac{1}{n}E\left(\sum\limits_{k=1}^n X_k\right)=\dfrac{1}{n}\sum\limits_{i=1}^n E(X_i)=\mu$,

$$D\left(\frac{1}{n}\sum_{k=1}^n X_k\right)=\frac{1}{n^2}D\left(\sum_{k=1}^n X_k\right)\xlongequal{\text{独立性}}\frac{1}{n^2}\cdot n\sigma^2=\frac{\sigma^2}{n},$$

由Chebyshev不等式 $P\{|X-E(X)|<\varepsilon\}\geq 1-\dfrac{D(X)}{\varepsilon^2}$:

$$1-\frac{\frac{\sigma^2}{n}}{\varepsilon^2}\leq P\left\{\left|\frac{1}{n}\sum_{k=1}^n X_k-\mu\right|<\varepsilon\right\}\leq 1.$$ 令 $n\to+\infty$ 即证.

[**定理5.1.3**] [**Bernoulli大数定理**, **弱大数定律的推论**] 设 f_A 为 n 次独立重复试验中事件 A 发生的次数, p 为 A 在每次试验中发生的概率, 则对 $\forall\varepsilon>0$, 有 $\lim\limits_{n\to+\infty}P\left\{\left|\dfrac{f_A}{n}-p\right|<\varepsilon\right\}=1$ 或 $\lim\limits_{n\to+\infty}P\left\{\left|\dfrac{f_A}{n}-p\right|\geq\varepsilon\right\}=0$, 或记作 $\dfrac{f_A}{n}\xrightarrow{P}p$.

[**证**] 设随机变量 $X_k=[\text{第}k\text{次试验中}A\text{发生}]\ \ (k=1,\cdots,n)$, 则 $X_1,\cdots,X_n$ 相互独立, 且都服从(0-1)分布.

因 $f_A\sim b(n,p)$, 则 $f_A=X_1+\cdots+X_n$.

由**弱大数定律**: $\dfrac{f_A}{n}=\dfrac{1}{n}\sum\limits_{k=1}^{n}X_k\xrightarrow{P}E(X_k)=p\ \ (k=1,\cdots,n)$.

[**注1**] 频率的稳定性: 对 $\forall\varepsilon>0$, 当试验次数 n 充分大时, 事件"频率 $\dfrac{f_A}{n}$ 与概率 p 的偏差 $<\varepsilon$"几乎必然发生.

[**注2**] 由实际推断原理: 实际应用中, 试验次数很大时, 可用事件的频率代替其概率.

5.2 中心极限定理

[**定理5.2.1**] [**独立同分布的中心极限定理**] 若随机变量 $X_1,\cdots,X_n,\cdots$ 独立同分布, 且期望 $E(X_k)=\mu\ \ (k=1,2,\cdots)$, 方差 $D(X_k)=\sigma^2\ \ (k=1,2,\cdots)$, 则随机变量 $\sum\limits_{k=1}^{n}X_k$ 的标准化变量

$Y_n=\dfrac{\sum\limits_{k=1}^{n}X_k-E\left(\sum\limits_{k=1}^{n}X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{n}X_k\right)}}=\dfrac{\sum\limits_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}$ 的分布函数 $F_n(x)$ 对 $\forall x$ 满足

$\lim\limits_{n\to+\infty}F_n(x)=\lim\limits_{n\to+\infty}P\left\{\dfrac{\sum\limits_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}\le x\right\}=\displaystyle\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t=\varPhi(x)$, 即 $Y_n\overset{近似}{\sim}N(0,1)$.

[**注1**] 实际应用中有很多随机变量 X 由大量的相互独立的随机因素综合影响而成, 其中每个因素在总的影响中所起的作用很小, 这样的 X 往往服从正态分布. 本定理表明: 独立的随机变量的个数不断增大时, 其和的分布趋于正态分布.

[**注2**] 一般 $\sum\limits_{k=1}^{n}X_k$ 的分布函数不易求, n 充分大时, 可用 $\varPhi(x)$ 给出其近似分布.

[**注3**] 本定理的另一形式: $\dfrac{\sum\limits_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}=\dfrac{\frac{1}{n}\sum\limits_{k=1}^{n}X_k-\mu}{\frac{\sigma}{\sqrt{n}}}=\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\overset{近似}{\sim}N(0,1)$, 其中 $\overline{X}=\dfrac{1}{n}\sum\limits_{k=1}^{n}X_k$ 为 $X_1,\cdots,X_n$ 的算术平均值.

因 $Z=\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\overset{近似}{\sim}N(0,1)$, 则 $\overline{X}=\dfrac{\sigma}{\sqrt{n}}Z+\mu\overset{近似}{\sim}N\left(\mu,\dfrac{\sigma^2}{n}\right)$.

[**定理5.2.2**] [**De Moivre-Laplace定理**] 若随机变量 $\eta_n\sim b(n,p)\ \ (n=1,2,\cdots;0<p<1)$, 则对 $\forall x$, 有 $\lim\limits_{n\to+\infty}P\left\{\dfrac{\eta_n-np}{\sqrt{np(1-p)}}\le x\right\}=\displaystyle\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t=\varPhi(x)$.

[**证**] 设随机变量 $X_k=[第k次试验中A发生]\ \ (k=1,\cdots,n)$, 则 $X_1,\cdots,X_n$ 相互独立, 且都服从(0-1)分布.

因 $E(X_k)=p,D(X_k)=p(1-p)\ \ (k=1,\cdots,n)$, 且 $\eta_n=\sum\limits_{k=1}^{n}X_k$,

由**独立同分布的中心极限定理**: $\dfrac{\eta_n-np}{\sqrt{np(1-p)}}=\dfrac{\sum\limits_{k=1}^{n}X_k-np}{\sqrt{n}\cdot\sqrt{p(1-p)}}\overset{近似}{\sim}N(0,1)$.

[**注1**] 本定理是**独立同分布的中心极限定理**的特例.

[**注2**] 本定理表明: 二项分布的极限分布是正态分布.

随机变量 $X \sim b(n,p)$ 的概率的计算:

① $n \leq 5$ 时, 直接计算: $P\{X=k\} = C_n^k \cdot p^k \cdot (1-p)^{n-k} \ \ (k=0,1,\cdots,n)$.

② n 很大, p 很小, 且 $\lambda = np < 10$ 时, 由Poisson定理: $P\{X=k\} \approx \dfrac{\lambda^k \mathrm{e}^{-\lambda}}{k!} \ \ (k=0,1,\cdots,n)$.

③ $n \geq 50$, p 不是很小时, 由**De Moivre-Laplace定理**: $\dfrac{\eta_n - np}{\sqrt{np(1-p)}} \overset{\text{近似}}{\sim} N(0,1)$.

[**例5.2.1**] 设 $V_1,\cdots,V_{20}$ 是相互独立的随机变量, 且都在区间 $(0,10)$ 上服从均匀分布. 设随机变量 $V = \sum\limits_{k=1}^{20} V_k$, 估计概率 $P\{V > 105\}$.

[**解**] $E(V_k) = \dfrac{0+10}{2} = 5, D(V_k) = \dfrac{(10-0)^2}{12} = \dfrac{100}{12}$.

由**独立同分布的中心极限定理**: $\dfrac{\sum\limits_{k=1}^{20} V_k - 20 \cdot 5}{\sqrt{20} \cdot \sqrt{\dfrac{100}{12}}} \overset{\text{近似}}{\sim} N(0,1)$.

$$P\{V > 105\} = P\left\{\sum_{k=1}^{20} V_k > 105\right\} = P\left\{\frac{\sum\limits_{k=1}^{20} V_k - 100}{\sqrt{20} \cdot \sqrt{\dfrac{100}{12}}} > \frac{105-100}{\sqrt{20} \cdot \sqrt{\dfrac{100}{12}}}\right\}$$

$$= 1 - P\left\{\frac{\sum\limits_{k=1}^{20} V_k - 100}{\sqrt{20} \cdot \sqrt{\dfrac{100}{12}}} \leq \frac{5}{\sqrt{\dfrac{2000}{12}}}\right\} = 1 - \varPhi\left(\frac{5}{\sqrt{\dfrac{2000}{12}}}\right).$$

[**例5.2.2**] 某船每遭受一次冲击时, 角度有 $p = \dfrac{1}{3}$ 的概率偏移. 若该船遭受了 90000 次冲击, 估计其中角度有 $29500 \sim 30500$ 次偏移的概率.

[**解**] 将每次冲击视为一次试验, 且各试验相互独立.

设 90000 次冲击中有 X 次角度偏移, 则随机变量 $X \sim b\left(90000, \dfrac{1}{3}\right)$.

由**De Moivre-Laplace定理**:

$$P\{29500 \leq X \leq 30500\} = P\left\{\frac{29500 - 90000 \cdot \dfrac{1}{3}}{\sqrt{90000 \cdot \dfrac{1}{3} \cdot \dfrac{2}{3}}} \leq \frac{X - 90000 \cdot \dfrac{1}{3}}{\sqrt{90000 \cdot \dfrac{1}{3} \cdot \dfrac{2}{3}}} \leq \frac{30500 - 90000 \cdot \dfrac{1}{3}}{\sqrt{90000 \cdot \dfrac{1}{3} \cdot \dfrac{2}{3}}}\right\}$$

$$= P\left\{-\frac{5}{\sqrt{2}} \leq \frac{X - 30000}{100\sqrt{2}} \leq \frac{5}{\sqrt{2}}\right\} \approx \varPhi\left(\frac{5}{\sqrt{2}}\right) - \varPhi\left(-\frac{5}{\sqrt{2}}\right) = 2\varPhi\left(\frac{5}{\sqrt{2}}\right) - 1.$$

[**例5.2.3**] 设每个学生来参加家长会的人数是随机变量, 其中无家长、1 名家长、2 名家长参会的概率分别为 $0.05, 0.8, 0.15$, 各学生参会的家长数独立. 若有 400 名学生, 估计:

(1) 参加会议的家长总数 $X > 450$ 的概率.

(2) 有 1 名家长参会的学生人数 ≤ 340 的概率.

[**解**]

(1) 设第 $k\ \ (k = 1, \cdots, 400)$ 个学生参会的家长数为 X_k , 则 $X_1, \cdots, X_{400}$ 独立同分布, 其分布律为:

X_k	0	1	2
p	0.05	0.8	0.15

则 $E(X_k) = 1.1, D(X_k) = 0.19$, 且 $X = \sum_{k=1}^{400} X_k$.

由**独立同分布的中心极限定理**:

$$P\{X > 450\} = P\left\{\sum_{k=1}^{400} X_k > 450\right\} = P\left\{\frac{\sum\limits_{k=1}^{400} X_k - 440}{\sqrt{400} \cdot \sqrt{0.19}} > \frac{450 - 440}{\sqrt{400} \cdot \sqrt{0.19}}\right\}$$

$$= 1 - P\left\{\frac{\sum\limits_{k=1}^{400} X_k - 440}{\sqrt{400} \cdot \sqrt{0.19}} \leq \frac{1}{2 \cdot \sqrt{0.19}}\right\} \approx 1 - \varPhi\left(\frac{1}{2 \cdot \sqrt{0.19}}\right).$$

(2) 设有 1 名家长参会的学生人数为 Y , 则 $Y \sim b(400, 0.8)$.

由**De Moivre-Laplace定理**:

$$P\{Y \leq 340\} = P\left\{\frac{Y - 400 \cdot 0.8}{\sqrt{400 \cdot 0.8 \cdot 0.2}} \leq \frac{340 - 400 \cdot 0.8}{\sqrt{400 \cdot 0.8 \cdot 0.2}}\right\} = P\left\{\frac{Y - 400 \cdot 0.8}{\sqrt{400 \cdot 0.8 \cdot 0.2}} \leq 2.5\right\} \approx \varPhi(2.5)$$

.