第五章 大数定律与中心极限定理

【内容预览】

$$\begin{cases} \text{切比雪夫不等式：} P\{|X-EX| \geqslant \varepsilon\} \leqslant \dfrac{DX}{\varepsilon^2} \\ \text{大数定理} \begin{cases} \text{切比雪夫大数定律} \\ \text{独立同分布的大数定理} \\ \text{辛钦大数定理} \end{cases} \\ \text{中心极限定理} \begin{cases} \text{列维林德伯格中心极限定理} \\ \text{拉普拉斯中心极限定理} \end{cases} \end{cases}$$

【知识清单】

5.1、切比雪夫不等式

设X为方差有限的随机变量，则对任意的$\varepsilon > 0$，有$P\{|X-EX| \geqslant \varepsilon\} \leqslant \dfrac{DX}{\varepsilon^2}$

技巧：切比雪夫不等式中有一个猫嘴的表情包：$\geq \varepsilon \leq$，有很多同学将这个公式记成了：$P\{|X-EX| \leq \varepsilon\} \geq \dfrac{DX}{\varepsilon^2}$，这是错误的.

另外，切比雪夫不等式有一个变形：$P\{|X-EX| < \varepsilon\} \geqslant 1-\dfrac{DX}{\varepsilon^2}$

注意：公式中的随机变量X可以替换为任何随机变量.

例 5-1：设X是随机变量，$D(X)=2$，则由切比雪夫不等式，有$P(|X-EX| \geq 2) \leq$__________.

解：$P(|X-EX| \geq \varepsilon) \leq \dfrac{\sigma^2}{\varepsilon^2} \Rightarrow P(|X-EX| \geq 2) \leq \dfrac{2}{2^2} = \dfrac{1}{2}$

5.2、大数定理

1.依概率收敛

设$Y_1,Y_2,\cdots,Y_n,\cdots$是一个随机变量序列，$a$是一个常数. 若对于任意正数$\varepsilon$有$\lim\limits_{n\to\infty} P\{|Y_n-a| < \varepsilon\} = 1$，则称序列$Y_1,Y_2,\cdots,Y_n,\cdots$**依概率收敛于$\boldsymbol{a}$**，记为$Y_n \xrightarrow{P} a$.

性质：设$X_n \xrightarrow{P} a$，$Y_n \xrightarrow{P} b$，又设函数$g(X_n,Y_n)$存在，则$g(X_n,Y_n) \xrightarrow{P} g(a,b)$.

定理 1 切比雪夫大数定理 设$X_1,X_2,\cdots,X_n,\cdots$相互独立且存在$c>0$使得$DX_i\leq c(i=1,2,\cdots)$，则对任意的$\varepsilon>0$，

有$\lim\limits_{n\to\infty}P\left\{\left|\frac{1}{n}\sum\limits_{i=1}^{n}X_i-\frac{1}{n}\sum\limits_{i=1}^{n}EX_i\right|<\varepsilon\right\}=1.$

定理 2 弱大数定理（辛钦大数定理） 设$X_1,X_2,\cdots,X_n,\cdots$相互独立同分布且$EX_i=\mu(i=1,2,\cdots)$，则对任意的

$\varepsilon>0$，有$\lim\limits_{n\to\infty}P\left\{\left|\frac{1}{n}\sum\limits_{i=1}^{n}X_i-\mu\right|<\varepsilon\right\}=1$，也可记为$\frac{1}{n}\sum\limits_{i=1}^{n}X_i\xrightarrow{P}\mu.$

证明：这里只在随机变量的方差$D(X_k)=\sigma^2\ (k=1,2,\cdots)$存在这一条件下证明上述结果. 因为

$E\left(\frac{1}{n}\sum\limits_{k=1}^{n}X_k\right)=\frac{1}{n}\sum\limits_{k=1}^{n}E(X_k)=\frac{1}{n}(n\mu)=\mu$，又由独立性得：$D\left(\frac{1}{n}\sum\limits_{k=1}^{n}X_k\right)=\frac{1}{n^2}\sum\limits_{k=1}^{n}D(X_k)=\frac{1}{n^2}(n\sigma^2)=\frac{\sigma^2}{n}$，

由切比雪夫不等式得：$1\geq P\left\{\left|\frac{1}{n}\sum\limits_{k=1}^{n}X_k-\mu\right|<\varepsilon\right\}\geq1-\frac{\sigma^2/n}{\varepsilon^2}$，当$n\to\infty$，即得$\lim\limits_{n\to\infty}P\left\{\left|\frac{1}{n}\sum\limits_{k=1}^{n}X_k-\mu\right|<\varepsilon\right\}=1.$

定理 3 伯努利大数定理 设f_A是n次独立重复试验中事件A发生的次数，p是事件A在每次试验中发生的概率，

则对于任意正数$\varepsilon>0$，有

$$\lim_{n\to\infty}P\left\{\left|\frac{f_A}{n}-p\right|<\varepsilon\right\}=1\text{或}\lim_{n\to\infty}P\left\{\left|\frac{f_A}{n}-p\right|\geq\varepsilon\right\}=0$$

在第一章提到的对古典概型的另一种定义正是利用了伯努利大数定理.

☞**技巧：**上述大数定律皆指随机变量组在独立的条件下，若干个随机变量的平均值依概率收敛到其数学期望的平均值，即$\frac{1}{n}\sum\limits_{i=1}^{n}X_i\xrightarrow{P}E\left(\frac{1}{n}\sum\limits_{i=1}^{n}X_i\right)=\mu.$

更进一步，$\frac{1}{n}\sum\limits_{i=1}^{n}X_i^k\xrightarrow{P}E\left(\frac{1}{n}\sum\limits_{i=1}^{n}X_i^k\right)$（要求$k$阶矩存在）

5.3、中心极限定理

定理 1（列维林德伯格中心极限定理）设$X_1,X_2,\cdots,X_n,\cdots$独立同分布，且数学期望$EX_i=\mu$和方差$DX_i=\sigma^2$都

存在$(i=1,2,\cdots)$，则对任意实数x，有$\lim\limits_{n\to\infty}P\left\{\dfrac{\sum\limits_{i=1}^{n}X_i-n\mu}{\sqrt{n}\,\sigma}\leq x\right\}=\Phi(x).$

定理 2（棣莫弗-拉普拉斯中心极限定理）设$X_n\sim B(n,p)$，且$X_1,X_2,\cdots,X_n,\cdots$相互独立，对任意实数$x$，有

$$\lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \le x\right\} = \Phi(x)$$

.

☞技巧：定理二是定理一的特例，注意到：在n趋于∞的情况下，我们构造的这个随机变量的极限分布是标准正态分布.

本章在考试中出现的形式较为固定，大多数情况下会考察一道应用大数定理或中心极限定理求解生活中问题的应用题，部分学校喜欢在填空题中考察大数定理或中心极限定理的定义以及切比雪夫不等式的应用.

在本章的重要题型部分，我们将通过例题主要讲解这两类题型的一般做题思路.

【重要题型】

题型1：切比雪夫不等式

题目一般以选择填空的方式来考察，需要首选表达成具体的切比雪夫不等式的形式，然后计算出不等式中未知的参数，代入不等式得到答案. 其中尤其需要注意不等式中出现的究竟是“≥”还是“≤”.

例 5-2：设随机变量 X 的概率密度为 $f(x)=\begin{cases}\frac{1}{2}x^2e^{-x}, & x>0\\ 0, & x\le 0\end{cases}$，利用切比雪夫不等式估计概率 $P(1<X<5)\ge$________.

解：$EX=\int_0^{\infty}\frac{1}{2}x^3e^{-x}dx=3, EX^2=\int_0^{\infty}\frac{1}{2}x^4e^{-x}dx=12, DX=12-9=3$

注：这里利用公式 $\int_0^{+\infty}e^{-x}x^k dx=k!$ 简化计算

$$P(1<X<5)=P(|X-3|<2)\ge 1-\frac{3}{2^2}=\frac{1}{4}.$$

本题中没有明显给出切比雪夫的标准形式，这时需要计算一下期望与方差来完善其最初的形式. 另外题目中的符号为“≥”，因此最终的结果需要进一步的处理.

题型2：中心极限定理

该类题目主要考察对中心极限定理的利用，包括选择填空、应用题等题型，对于应用题关键在于要找到对应的均值与方差，得到正确的近似正态分布.

例 5-3：设随机变量 $X_1,X_2,\cdots X_{100}$ 相互独立,并且均服从参数为 $\frac{1}{2}$ 的指数分布，利用中心极限定理计算 $P\left(\sum_{i=1}^{100}X_i<240\right)=$________.

解：$\mu=\frac{1}{\lambda}=2, \sigma^2=\frac{1}{\lambda^2}=4$,由中心极限定理，$\sum_{i=1}^{n}X_i\sim N(n\mu,n\sigma^2)$ ，$\sum_{i=1}^{100}X_i\sim N(200,400)$

因此 $\dfrac{\sum\limits_{i=1}^{100} X_i - 200}{\sqrt{400}} \sim N(0,1)$，那么 $P\left(\dfrac{\sum\limits_{i=1}^{100} X_i - 200}{\sqrt{400}} \le x\right) = \Phi(x)$ 故

$$P\left(\sum_{i=1}^{100} X_i < 240\right) = P\left(\frac{\sum\limits_{i=1}^{100} X_i - 200}{20} < \frac{240-200}{20}\right) = \Phi(2)$$

例 5-4：设某种元件的寿命 X(小时)服从指数分布,其概率密度函数为 $f(x) = \begin{cases} \frac{1}{100} e^{-\frac{x}{100}}, & x > 0 \\ 0, & x \le 0 \end{cases}$，各个元件的寿命相互独立，随机地取 36 个元件，求寿命总和大于 4200 小时的概率近似值.

解： 由概率密度函数可知 $\lambda = \dfrac{1}{100}$.

设第 i 只元件的寿命为 $X_i (i=1,2,\cdots,36)$，且各个元件的寿命相互独立.

那么 $EX = \dfrac{1}{\lambda} = 100$，$DX = \dfrac{1}{\lambda^2} = 10000$.

则令 $Y = \sum\limits_{i=1}^{36} X_i$，$EY = 3600$，$DY = 36 \times 100^2$，

所以 $\dfrac{\sum\limits_{i=1}^{36} X_i - 3600}{\sqrt{36 \times 100^2}} = \dfrac{\sum\limits_{i=1}^{36} X_i - 3600}{600} \sim N(0,1)$

则 $P(Y > 4200) = P\left(\sum\limits_{i=1}^{36} X_i > 4200\right) = 1 - P\left(\sum\limits_{i=1}^{36} X_i \le 4200\right)$.

$$1 - P\left(\sum_{i=1}^{36} X_i \le 4200\right) = 1 - P\left(\frac{\sum\limits_{i=1}^{36} X_i - 3600}{600} \le \frac{4200-3600}{600} = 1\right) = 1 - \Phi(1) = 0.1587.$$

【精选习题】

基础篇

1.设随机变量 X 的概率密度为 $f(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$，则 $P\left(|X - EX| \ge 2\sqrt{DX}\right)$ 等于 (　　)

A、$\dfrac{9-8\sqrt{2}}{9}$　　B、$\dfrac{6+4\sqrt{2}}{9}$　　C、$\dfrac{6-8\sqrt{2}}{9}$　　D、$\dfrac{6-4\sqrt{2}}{9}$

2.设随机变量 $X \sim U(-1,2)$,由切比雪夫不等式得 $P(|X-0.5| < 1) >$ ＿＿＿＿＿.

3.设 $EX = -2, EY = 2, DX = 1, DY = 4, \rho(X,Y) = 0.5$,则 $D(X+Y) =$ ＿＿＿＿＿，根据切比雪夫不等式估计概

率$P\{|X+Y|\geq 6\}$__________.

4. 设 $X_1, X_2 \cdots X_n$ 相互独立，且均服从泊松分布 $P(2)$，则由林德伯格—列维中心极限定理

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - 2n}{\sqrt{n}} < 2\right) = (\quad).$$

A、$\Phi(\sqrt{2})$　　B、$\Phi(2)$　　C、$\Phi(1)$　　D、1

5.从废品率为 0.03 的大量产品中随机抽取 1000 个，根据棣莫弗-拉普拉斯中心极限定理的结论，废品数X近似服从的分布为__________.（要求同时写出分布的参数）

6.河北某中学一宿舍楼按 400 名学生的规模建造，该校学生每天洗漱时间是有规定的，按每人在规定的时间内大约有 10%的时间占用一个水龙头计算，为能以 95%的概率保证用水需要，该宿舍至少要做安装多少个水龙头？（假设每人用水情况相互独立）（Φ（1.645）=0.95）

7.假设一条自动生产线生产的产品的合格率是 0.8.要使一批产品的合格率达到 76% 与 84% 之间的概率不小于 90%，问: 这批产品至少要生产多少件？$\Phi(1.64)=0.95$

8.从长沙望城前往长沙火车南站有2条路线可走，第一条路线穿过市区，路线较短，但交通拥挤，所需时间（分钟）服从正态分布$N(50,100)$，第二条路沿环城公路，但路线较长，且交通阻塞较多，所需时间（分钟）服从正态分布$N(60,16)$.若某人只有 65 分钟可用，应该走哪条路较好?

$(\Phi(1.5)=0.9332, \Phi(1.25)=0.8944, \Phi(2.33)=0.99, \Phi(2.58)=0.995)$

提高篇

9.在一次集体登山活动中，假设每个人意外受伤的概率为 1%，每个人是否意外受伤是相互独立的。

（1）为保证没有人意外受伤的概率大于 0.90，问：应当如何控制参加登山活动的人数?

（2）如果有 100 人参加这次登山活动，求意外受伤的人数小于等于 2 人的概率的近似值。(要求用中心极限定理解题，结果可用正态分布的分布函数表示）

10.某地区某种疾病的患病率为 0.1，现通过血检对该地区的 10000 个人进行普查. 若血样呈阳性，则有此种疾病；呈阴性则无此种疾病；且各人患此种疾病相互独立.

（1）试利用中心极限定理估计患病人数在$[970, 1030]$内的概率（用标准正态分布函数表示）;

（2）现血检有两套方案：方案一是逐一化验，共需要 10000 次化验；方案二是分组检查，将 10000 个人分组，每组 k 个人（假设$m=\dfrac{10000}{k}$为整数），每组 k 个人的血液提取一半混合一起检查，若为阴性，说明这 k 个人全为阴性，这时只需要 1 次化验；若为阳性，则再将这k个人的另外一半血液逐一检查，这时需要$k+1$次化验. 问 k 满足什么条件时，方案二的平均检查次数小于方案一；当$k=4$时，方案二的平均检查次数?