$$F_Y(y)=P\{-\sqrt{y}\leqslant X\leqslant\sqrt{y}\}=\int_{-\sqrt{y}}^{\sqrt{y}}f_X(x)\mathrm{d}x$$

$$=\int_{-\sqrt{y}}^{0}\frac{1}{2}\mathrm{d}x+\int_{0}^{\sqrt{y}}\frac{1}{4}\mathrm{d}x=\frac{3}{4}\sqrt{y}.$$

没有分成 $0<y<1$，$1\leqslant y<4$ 来讨论.

③ 误认为

$$F\left(-\frac{1}{2},\ 4\right)=P\left\{X\leqslant-\frac{1}{2},\ Y\leqslant4\right\}=P\left\{X\leqslant-\frac{1}{2}\right\}P\{Y\leqslant4\}$$

或

$$F\left(-\frac{1}{2},\ 4\right)=\int_{-\infty}^{-\frac{1}{2}}\int_{-\infty}^{4}f(x,y)\mathrm{d}x\mathrm{d}y=\int_{-\infty}^{-\frac{1}{2}}f_X(x)\mathrm{d}x\cdot\int_{-\infty}^{4}f_Y(y)\mathrm{d}y.$$

五、大数定律和中心极限定理

大数定律是说明大量随机现象的平均值稳定性的一系列数学定理的总称，它有多种形式，常用的有伯努利大数定律和辛钦大数定律，切比雪夫大数定律是大数定律比较一般的形式.

中心极限定理揭示了“随机变量之和在一定条件下的极限分布是正态分布”，有着广泛的应用，也有着丰富的内容.

本节的一个重点是了解关于独立重复观测结果平均水平的极限分布是正态分布的列维－林德伯格中心极限定理和关于二项分布收敛于正态分布的棣莫弗－拉普拉斯中心极限定理；并会应用这两个定理近似计算有关事件的概率.

·考试内容与要求·

考试内容

切比雪夫大数定律　伯努利大数定律　辛钦(Khinchine)大数定律　棣莫弗－拉普拉斯(De Moivre－Laplace)定理　列维－林德伯格(Levy－Lindberg)定理

考试要求

1. 了解切比雪夫大数定律、伯努利大数定律和辛钦大数定律(独立同分布随机变量序列的大数定律).

2. 了解棣莫弗－拉普拉斯中心极限定理(二项分布以正态分布为极限分布)、列维－林德伯格中心极限定理(独立同分布随机变量序列的中心极限定理)，并会用相关定理近似计算有关随机事件的概率.

·考试内容解析·

(一) 依概率收敛

对于随机变量序列 $X_1,X_2,\cdots,X_n,\cdots$，如果存在一个常数 A，对于任意给定的 $\varepsilon>0$，总有

$$\lim_{n\to\infty}P\{|X_n-A|<\varepsilon\}=1,$$

则称随机变量列 $X_1,X_2,\cdots,X_n,\cdots$依概率收敛于 A，记作

$$X_n\xrightarrow{P}A\ 或\ P-\lim_{n\to\infty}X_n=A.$$

(二) 大数定律

1. 切比雪夫大数定律

设随机变量序列 $X_1,X_2,\cdots,X_n,\cdots$两两独立或两两不相关，其期望与方差 EX_i，DX_i 均存在，且存在常数 $M>0$，使得 $DX_i\leqslant M$，$i=1,2,\cdots$，则对于任意的正数 ε，有

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

特别地，对于独立同分布的随机变量序列 $X_1,X_2,\cdots$(即对于任意的正整数 $k>1$,随机变量 $X_1,X_2,\cdots,X_k$ 相互独立且具有相同的概率分布)，若其期望 $EX_i=\mu$ 和方差 $DX_i=\sigma^2$ 存在，则对于任意的 $\varepsilon>0$，有

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

即

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} 1.$$

2. 伯努利大数定律

设 $p = P(A)$是事件 A 在每次试验中事件A 发生的概率，$f_n(A)$为事件 A 在n 次独立重复试验中事件A 发生的频数，则 $f_n(A)$依概率收敛于 p，即

$$f_n(A) \xrightarrow{P} p.$$

伯努利大数定律表明，在相同条件下进行 n 次独立重复试验，当 n 充分大时，随机事件 A 发生的频率稳定在事件A 发生的概率$P(A)$附近，即伯努利大数定律是反映频率稳定性的数学定理.

3. 辛钦大数定律

设 $X_1, X_2, \cdots, X_n, \cdots$是独立同分布的随机变量列，只要数学期望 $EX_i = \mu$ ($i = 1, 2, \cdots$)存在，则对于任意的正数 ε，有

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

即

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} 1.$$

辛钦大数定律比独立同分布条件下的切比雪夫大数定律更一般.

（三）中心极限定理

1. 棣莫弗－拉普拉斯中心极限定理

设随机变量 X_n 服从参数为n 和p 的二项分布，即 $X_n \sim B(n, p)$ $(0<p<1)$，则对于任意实数 x，有

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

(1) 局部定理：对于任意 p $(0<p<1)$和 k $(0\leqslant k\leqslant n)$，当 n 充分大时，有

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}.$$

(2) 积分定理：对于任意 p $(0<p<1)$和 k_1，k_2 $(0\leqslant k_1<k_2\leqslant n)$，当 n 充分大时，有

$$\sum_{k=k_1}^{k_2} \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi}}\int_{u_1}^{u_2} e^{-\frac{t^2}{2}} dt,$$

其中 $u_i = (k_i - np)/\sqrt{npq}$，$i = 1, 2$.

2. 列维－林德伯格中心极限定理

设 $X_1, X_2, \cdots, X_n, \cdots$是独立同分布的随机变量列，其数学期望 $\mu = EX_i$ 和方差 $\sigma^2 = DX_i$ 均存在，$i = 1, 2, \cdots$，则当 n 充分大时，$S_n = X_1 + X_2 + \cdots + X_n$ 近似地服从正态分布，即对于任意的实数 x，有

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

也可以说，对于任意实数 $a<b$，当 n 充分大时，有

$$P\left\{a < \sum_{i=1}^{n} X_i < b\right\} \approx \frac{1}{\sqrt{2\pi}}\int_{u_1}^{u_2} e^{-\frac{t^2}{2}} dt,$$

其中 $u_1 = (a - n\mu)/(\sqrt{n}\sigma)$，$u_2 = (b - n\mu)/(\sqrt{n}\sigma)$.

·例 题 详 解·

例 3.5.1 设随机变量 X 的方差为 2，则根据切比雪夫不等式有估计 $P\{|X - EX| \geqslant 2\} \leqslant$ ________.

[答案] $\frac{1}{2}$.

[提示] 本题主要考查切比雪夫不等式.

[解] 切比雪夫不等式为：对于任意 $\varepsilon>0$,

$$P\{|X-EX|\geqslant\varepsilon\}\leqslant\frac{DX}{\varepsilon^2}.$$

令 $\varepsilon=2$，$DX=2$，则 $P\{|X-EX|\geqslant 2\}\leqslant\frac{DX}{2^2}=\frac{1}{2}$.

[典型错误] 不知切比雪夫不等式，故束手无策.

例 3.5.2 设总体 X 服从参数为 2 的指数分布，$X_1,X_2,\cdots,X_n$ 为来自总体 X 的简单随机样本．则当 $n\to\infty$ 时，$Y_n=\frac{1}{n}\sum\limits_{i=1}^{n}X_i^2$ 依概率收敛于________.

[答案] $\frac{1}{2}$.

[提示] 本题主要考查独立同分布大数定律的概念及运用、指数分布的数字特征.

[解] 由于 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的简单随机样本．因而 $X_1,X_2,\cdots,X_n$ 相互独立，并可以推出 $X_1^2,X_2^2,\cdots,X_n^2$ 也相互独立并且同分布.

又 X 服从参数为 2 的指数分布，所以

$$EX_i=\frac{1}{2},\ DX_i=\frac{1}{4}\quad(i=1,2,\cdots,n),$$

则
$$EX_i^2=DX_i+(EX_i)^2=\frac{1}{4}+\left(\frac{1}{2}\right)^2=\frac{1}{2}\quad(i=1,2,\cdots,n).$$

由独立同分布大数定律知 $\frac{1}{n}\sum\limits_{i=1}^{n}X_i^2$ 依概率收敛于 $\frac{1}{2}$.

[典型错误] 对大数定律不熟，空白者多.

例 3.5.3 设随机变量 X 与 Y 的数学期望都是 2．方差分别为 1 和 4．而相关系数为 0.5．则根据切比雪夫不等式．知 $P\{|X-Y|\geqslant 6\}\leqslant$________.

[答案] $\frac{1}{12}$.

[提示] 本题主要考查切比雪夫不等式、随机变量和的方差．由题意知 $E(X-Y)=0$．故只要把 $X-Y$ 看成一个随机变量 Z，则可以用切比雪夫不等式，而

$$D(X-Y)=DX+DY-2\rho\cdot\sqrt{DX}\cdot\sqrt{DY}.$$

[解] 由于 $E(X-Y)=2-2=0$，故由切比雪夫不等式有

$$\begin{aligned}P\{|X-Y|\geqslant 6\}&=P\{|X-Y-E(X-Y)|\geqslant 6\}\\&\leqslant\frac{D(X-Y)}{6^2}=\frac{1}{36}[DX+DY-2\mathrm{cov}(X,Y)]\\&=\frac{1}{36}(DX+DY-2\rho\cdot\sqrt{DX}\cdot\sqrt{DY})\\&=\frac{1}{12}.\end{aligned}$$

[典型错误] 本题的难点在于把 $X-Y$ 看成一个新的随机变量，由于 X 与 Y 是相关的，故 $D(X-Y)\neq DX+DY$.

例 3.5.4 设随机变量 $X_1,X_2,\cdots,X_n$ 相互独立，$S_n=X_1+X_2+\cdots+X_n$．则根据列维－林德伯格中心极限定理，当 n 充分大时，S_n 近似服从正态分布，只要 $X_1,X_2,\cdots,X_n$ (　　).

(A) 有相同的数学期望　　(B) 有相同的方差

(C) 服从同一指数分布　　(D) 服从同一离散分布

[答案] (C).

[提示] 本题主要考查列维－林德伯格中心极限定理．列维－林德伯格中心极限定理：如果 $X_1,X_2,\cdots,$

X_n 是独立同分布的随机变量．$EX_i=\mu$，$DX_i=\sigma^2$，$i=1,2,\cdots,n$，则 S_n 近似服从正态分布$N(n\mu,n\sigma^2)$．

［解］ 对照列维－林德伯格中心极限定理，四个选项中只有(C)符合条件．这是因为 $X_1,\cdots,X_n$ 服从同一指数分布时，EX_i 和 DX_i 均存在．

［典型错误］ 不熟悉列维－林德伯格中心极限定理要求独立同分布且有有限的期望与方差．

例 3.5.5 设$X_1,X_2,\cdots,X_n,\cdots$为独立同分布的随机变量列，且均服从参数为$\lambda$ $(\lambda>1)$的指数分布，记$\Phi(x)$为标准正态分布函数，则(　　)．

(A) $\lim\limits_{n\to\infty}P\left\{\dfrac{\sum\limits_{i=1}^{n}X_i-n\lambda}{\lambda\sqrt{n}}\leqslant x\right\}=\Phi(x)$　　(B) $\lim\limits_{n\to\infty}P\left\{\dfrac{\sum\limits_{i=1}^{n}X_i-n\lambda}{\sqrt{n\lambda}}\leqslant x\right\}=\Phi(x)$

(C) $\lim\limits_{n\to\infty}P\left\{\dfrac{\lambda\sum\limits_{i=1}^{n}X_i-n}{\sqrt{n}}\leqslant x\right\}=\Phi(x)$　　(D) $\lim\limits_{n\to\infty}P\left\{\dfrac{\sum\limits_{i=1}^{n}X_i-\lambda}{\sqrt{n\lambda}}\leqslant x\right\}=\Phi(x)$

［答案］ (C)．

［提示］ 本题主要考查列维－林德伯格中心极限定理．

［解］ 因为 $X_1,X_2,\cdots$是独立同分布的指数分布随机变量列，则 $EX_i=\dfrac{1}{\lambda}$，$DX_i=\dfrac{1}{\lambda^2}$，且由列维－林德伯格中心极限定理知

$$\zeta_n=\frac{\lambda}{\sqrt{n}}\sum_{i=1}^{n}\left(X_i-\frac{1}{X}\right)=\frac{\lambda\sum\limits_{i=1}^{n}X_i-n}{\sqrt{n}}$$

的极限分布为标准正态分布，故知选项(C)正确．

［典型错误］ 不熟悉列维－林德伯格中心极限定理的表达方式及指数分布的期望与方差．

例 3.5.6 一生产线生产的产品成箱包装．每箱的重量是随机的．假设每箱平均重 50 千克、标准差为 5 千克．若用最大载重量为 5 吨的汽车承运，试利用中心极限定理说明每辆车最多可以装多少箱，才能保障不超载的概率大于 0.977．($\Phi(2)=0.977$，其中$\Phi(x)$是标准正态分布函数．)

［提示］ 本题主要考查列维－林德伯格中心极限定理及概率计算．中心极限定理是针对 n 个独立同分布的随机变量和而言的，根据题意构造这样的一个独立随机变量和．最后求出 n．

［解］ 设X_i $(i=1,2,\cdots,n)$是装运的第 i 箱的重量(单位:千克)，n 是所求箱数．由条件可以把X_1，$X_2,\cdots,X_n$ 视为独立同分布的随机变量，而 n 箱的总重量

$$T_n=X_1+X_2+\cdots+X_n$$

是独立同分布的随机变量之和．

由条件知 $EX_i=50$，$\sqrt{DX_i}=5$；$ET_n=50n$，$\sqrt{DT_n}=5\sqrt{n}$(单位:千克)．

根据列维－林德伯格中心极限定理，T_n 近似服从正态分布$N(50n,25n)$．

箱数 n 决定于条件

$$P\{T_n\leqslant 5\,000\}=P\left\{\frac{T_n-50n}{5\sqrt{n}}\leqslant\frac{5\,000-50n}{5\sqrt{n}}\right\}$$
$$\approx\Phi\left(\frac{1\,000-10n}{\sqrt{n}}\right)>0.977=\Phi(2)．$$

由此可见

$$\frac{1\,000-10n}{\sqrt{n}}>2，$$

从而 $n<98.019\,9$，即最多可以装 98 箱．

［典型错误］ 不会建立相应的概率模型及不会应用列维－林德伯格中心极限定理．

六、数理统计的基本概念(数学三)

本节是数理统计的基础知识，“数学四”不要求．

本节内容主要包括总体、样本、统计量等概念以及几个很有用的抽样分布．大纲要求：

（1）理解总体、简单随机样本、统计量及抽样分布的概念．掌握常用的统计量．

（2）了解常用的 χ^2 分布、t 分布和 F 分布的概念、性质以及服从这些分布的随机变量的典型模式，了解分位数的概念并会查表计算．

（3）了解正态总体的常用抽样分布：正态分布、χ^2 分布、t 分布和 F 分布．

• 考试内容与要求 •

考试内容

总体　个体　简单随机样本　统计量　经验分布函数　样本均值　样本方差和样本矩　χ^2 分布　t 分布　F 分布　分位数　正态总体的常用抽样分布

考试要求

1．理解总体、简单随机样本、统计量、样本均值、样本方差及样本矩的概念．其中样本方差定义为

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 .$$

2．了解产生 χ^2 变量、t 变量和 F 变量的典型模式；理解标准正态分布、χ^2 分布、t 分布和 F 分布的分位数，会查相应的数值表．

3．掌握正态总体的抽样分布：样本均值、样本方差、样本矩、样本均值差、样本方差比的抽样分布．

4．理解经验分布函数的概念和性质，会根据样本值求经验分布函数．

• 考试内容解析 •

（一）总体和样本

1．总体

一般而言，总体是指与所研究的问题有关的对象（个体）的全体所构成的集合．但在数理统计中，总体就是一个服从某概率分布的随机变量 X，其概率分布称为总体分布，其数字特征称为总体数字特征．

2．样本与简单随机抽样

样本是按一定规定从总体中抽出的一部分个体，所谓"按一定规定"是指总体中的每一个个体均有同等的被抽出的机会．也可以说 n 个独立且与总体 X 同分布的随机变量 $(X_1,\cdots,X_n)$ 为来自总体 X 的一个简单随机样本，简称为样本或一组样本，$X_i\ (i=1,\cdots,n)$ 称为第 i 个样本，n 称为样本容量．样本的具体观测值 $(x_1,\cdots,x_n)$ 称为样本值．

对于总体 X 的 n 次独立重复观测，称做来自总体 X 的 n 次简单随机抽样．

（二）统计量和样本矩

1．统计量

完全由样本决定的量，叫做统计量．统计量只依赖于样本，而不能依赖于任何其他未知的量．特别是它不能依赖于总体分布中所包含的未知参数．

统计量是样本的函数．它是一个随机变量，其分布称为抽样分布．

2．样本矩

设 $(X_1,\cdots,X_n)$ 是来自总体 X 的简单随机样本．

（1）样本均值：$\overline{X} = \frac{1}{n}\sum_{i=1}^{n}X_i$ ．

（2）样本方差：$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}X_i^2 - \frac{n}{n-1}\overline{X}^2$．$S$ 称为样本标准差．

（3）样本 k 阶原点矩：$A_k = \frac{1}{n}\sum_{i=1}^{n}X_i^k,\ k = 1,2,\cdots$．

（4）样本 k 阶中心矩：$B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k,\ k = 1,2,\cdots$．