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

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

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

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

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: $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$.

(2) 样本方差: $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n-1}\sum_{i=1}^{n}X_i^2-\frac{n}{n-1}\bar{X}^2$. 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: $B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$, $k=1,2,\cdots$.

3. 顺序统计量

将样本 $X_1,\cdots,X_n$ 的 n 个观测值按其值从小到大排列成：$X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$. 则称$(X_{(1)},\cdots,X_{(n)})$为**顺序统计量**. $X_{(k)}$称为第 k 个顺序统计量，其中 $X_{(1)}$和 $X_{(n)}$分别称为最小和最大顺序统计量，即

$$X_{(1)}=\min\{X_1,\cdots,X_n\},\ X_{(n)}=\max\{X_1,\cdots,X_n\}.$$

4. 经验分布函数

设 $X_1,\cdots,X_n$ 为来自总体 X 的简单随机样本. 则对于任意的实数 x，称

$$F_n(x)=\frac{\{i\mid X_i\leqslant x\}}{n}=\frac{1}{n}\sum_{i=1}^{n}I(X_i\leqslant x)$$

为总体 X 的经验分布函数. 其中 $I(\cdot)$为示性函数.

(三) 常用的抽样分布

常用的抽样分布有 χ^2 分布、t 分布和 F 分布，要了解这些分布的典型模式以及会查相应的分位数表.

1. χ^2 分布

(1) 定义：设 $X_1,\cdots,X_n$ 为 n 个独立的标准正态随机变量，则称

$$X=X_1^2+\cdots+X_n^2$$

服从自由度为 n 的 χ^2 分布，记为 $X\sim\chi^2(n)$，其概率密度为

$$f(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}, & x>0,\\ 0, & x\leqslant 0.\end{cases}$$

(2) χ^2 分布的性质：如果 $X\sim\chi^2(m)$，$Y\sim\chi^2(n)$，且 X 与 Y 相互独立，则 $X+Y\sim\chi^2(m+n)$. 此性质说明 χ^2 分布具有可加性.

设 $X\sim\chi^2(n)$，则 $EX=n$，$DX=2n$.

(3) χ^2 分布的上侧分位数：设随机变量 X 服从自由度为 n 的 χ^2 分布. 对于给定的 $\alpha\ (0<\alpha<1)$，称满足条件

$$P\{X>\chi_\alpha^2(n)\}=\alpha$$

的 $\chi_\alpha^2(n)$为分布 $\chi^2(n)$的上侧 α 分位数.

2. t 分布

(1) 定义：设随机变量 $X\sim N(0,1)$，$Y\sim\chi^2(n)$，且 X 与 Y 相互独立，则称随机变量

$$T=\frac{X}{\sqrt{Y/n}}$$

服从自由度为 n 的 t 分布，记作 $T\sim t(n)$，其概率密度函数为

$$f(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\cdot\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},\ x\in\mathbf{R}.$$

(2) 渐近正态性：当 $t(n)$的自由度 n 非常大时，$t(n)$分布与标准正态分布非常接近，即

$$\lim_{n\to\infty}f(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}=\varphi(x).$$

(3) t 分布的上侧分位数：设随机变量 T 服从自由度为 n 的 t 分布，对于给定的 $\alpha\ (0<\alpha<1)$，称满足条件

$$P\{T>t_\alpha(n)\}=\alpha$$

的 $t_\alpha(n)$为分布 $t(n)$的上侧 α 分位数.

由于 t 分布的概率密度是偶函数，故

$$t_{1-\alpha}(n)=-t_\alpha(n).$$

3. F 分布

(1) 定义：设随机变量 $X\sim\chi^2(m)$，$Y\sim\chi^2(n)$，且二者相互独立，则称随机变量

$$F=\frac{X/m}{Y/n}$$

服从自由度为(m,n)的 F 分布，记作 $F\sim F(m,n)$，其概率密度为

$$f(x)=\begin{cases}\dfrac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)}m^{\frac{m}{2}}n^{\frac{n}{2}}\dfrac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}}, & x>0,\\ 0, & x\leqslant 0.\end{cases}$$

(2) F 分布的性质：如果 $F\sim F(m,n)$，则$\dfrac{1}{F}\sim F(n,m)$.

(3) F 分布的上侧分位数：设随机变量 F 服从自由度为(m,n)的 F 分布，对于给定的 $\alpha\ (0<\alpha<1)$，称满足条件

$$P\{F>F_\alpha(m,n)\}=\alpha$$

的 $F_\alpha(m,n)$为 $F(m,n)$分布的上侧 α 分位数.

由 F 分布的性质，有

$$F_{1-\alpha}(m,n)=\frac{1}{F_\alpha(n,m)}.$$

（四）正态总体的抽样分布

1. 一个正态总体

设总体 $X\sim N(\mu,\sigma^2)$，$X_1,\cdots,X_n$ 是来自 X 的简单随机样本，$\overline{X}$ 和 S^2 分别为样本均值和样本方差.

(1) 样本均值的分布：$\overline{X}\sim N\left(\mu,\dfrac{\sigma^2}{n}\right)$，即$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$.

(2) 样本方差的分布：$\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$.

(3) 样本均值 $\overline{X}$ 和样本方差 S^2 相互独立.

基于上面三点，有

$$T=\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1).$$

(4) $\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(X_i-\mu)^2\sim\chi^2(n)$.

2. 两个正态总体

设总体 $X\sim N(\mu_1,\sigma_1^2)$，总体 $Y\sim N(\mu_2,\sigma_2^2)$，$X_1,\cdots,X_m$ 和 $Y_1,\cdots,Y_n$ 是分别来自总体 X 和总体 Y 的简单随机样本，$\overline{X}$，S_X^2 和 $\overline{Y}$，S_Y^2 是相应的样本均值和样本方差.

(1) 样本均值差的分布：$\overline{X}-\overline{Y}\sim N\left(\mu_1-\mu_2,\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}\right)$.

(2) 样本方差比的分布：$F=\dfrac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2}\sim F(m-1,n-1)$.

(3) 如果 $\sigma_1^2=\sigma_2^2$，则

$$T=\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}}}\cdot\sqrt{\frac{mn}{m+n}}\sim t(m+n-2).$$

(4) $F=\dfrac{n\sigma_2^2}{m\sigma_1^2}\cdot\dfrac{\sum\limits_{i=1}^{m}(X_i-\mu_1)^2}{\sum\limits_{i=1}^{n}(Y_i-\mu_2)^2}\sim F(m,n)$.

·例 题 详 解·

例 3.6.1 设总体 X 服从正态分布 $N(\mu_1,\sigma^2)$，总体 Y 服从正态分布 $N(\mu_2,\sigma^2)$，$X_1,X_2,\cdots,X_{n_1}$ 和 Y_1，$Y_2,\cdots,Y_{n_2}$ 分别是来自总体 X 和 Y 的简单随机样本，则

$$E\left[\frac{\sum_{i=1}^{n_1}(X_i-\bar{X})^2+\sum_{j=1}^{n_2}(Y_j-\bar{Y})^2}{n_1+n_2-2}\right]=\underline{\qquad\qquad}.$$

[答案] σ^2.

[提示] 本题主要考查样本方差的无偏性、随机变量和的数学期望.

[解] 记

$$S_1^2=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\bar{X})^2,S_2^2=\frac{1}{n_2-1}\sum_{j=1}^{n_2}(Y_j-\bar{Y})^2,$$

则

$$ES_1^2=ES_2^2=\sigma^2,$$

故

$$\text{原式}=E\left[\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\right]=\frac{(n_1-1)ES_1^2+(n_2-1)ES_2^2}{n_1+n_2-2}=\sigma^2.$$

[典型错误] 许多考生不能把本题和样本方差的无偏性联系起来，而直接求数学期望，从而造成计算错误或无从下手.

例 3.6.2 设总体 X 服从正态分布 $N(0,2^2)$，而 $X_1,X_2,\cdots,X_{15}$ 是来自总体 X 的简单随机样本，则随机变量

$$Y=\frac{X_1^2+\cdots+X_{10}^2}{2(X_{11}^2+\cdots+X_{15}^2)}$$

服从________分布，参数为________.

[答案] F，$(10,5)$.

[提示] 本题主要考查抽样分布.

[解] 由于 $X_i\sim N(0,2^2)$，则 $\frac{1}{2}X_i\sim N(0,1)$. 故

$$U=\frac{1}{4}X_1^2+\cdots+\frac{1}{4}X_{10}^2\sim\chi^2(10),$$

$$V=\frac{1}{4}X_{11}^2+\cdots+\frac{1}{4}X_{15}^2\sim\chi^2(5),$$

$$Y=\frac{U/10}{V/5}\sim F(10,5).$$

[典型错误] 不熟悉 F 分布的定义.

例 3.6.3 设随机变量 X 和 Y 相互独立且都服从正态分布 $N(0,3^2)$，而 $X_1,\cdots,X_9$ 和 $Y_1,\cdots,Y_9$ 分别是来自总体 X 和 Y 的简单随机样本，则统计量 $U=\frac{X_1+\cdots+X_9}{\sqrt{Y_1^2+\cdots+Y_9^2}}$ 服从________分布，参数为________.

[答案] t，9.

[提示] 本题主要考查 t 分布.

[解] 令

$$X_i'=\frac{X_i}{3},\quad Y_i'=\frac{Y_i}{3},\quad i=1,2,\cdots,9.$$

则

$$X_i'\sim N(0,1),\quad Y_i'\sim N(0,1),\quad i=1,2,\cdots,9,$$

$$X'=X_1'+\cdots+X_9'\sim N(0,3^2),$$

$$Y'=Y_1'^2+\cdots+Y_9'^2\sim\chi^2(9),$$

因此

$$U=\frac{X_1+\cdots+X_9}{\sqrt{Y_1^2+\cdots+Y_9^2}}=\frac{X_1'+\cdots+X_9'}{\sqrt{Y_1'^2+\cdots+Y_9'^2}}=\frac{X'}{\sqrt{Y'}}=\frac{X'/3}{\sqrt{Y'/9}}.$$

由于 $X'/3\sim N(0,1)$，$Y'\sim\chi^2(9)$，故 $U\sim t(9)$．

[典型错误] 不熟悉 t 分布的定义．

例 3.6.4 设 X_1，X_2，X_3，X_4 是来自正态总体 $N(0,2^2)$的简单随机样本．$X=a(X_1-2X_2)^2+b(3X_3-4X_4)^2$．则当 $a=$________，$b=$________时，统计量 X 服从 χ^2分布，其自由度为________．

[答案] $\frac{1}{20}$，$\frac{1}{100}$，2．

[提示] 本题主要考查χ^2分布．

[解] 根据χ^2分布的定义，若$Y_1,Y_2,\cdots,Y_m$ 服从标准正态分布，则 $Y=Y_1^2+Y_2^2+\cdots+Y_m^2$ 服从自由度为 m 的χ^2分布．对于本题，若 X 服从χ^2分布，则 $m=2$．且需

$$\sqrt{a}(X_1-2X_2)\sim N(0,1),\ \sqrt{b}(3X_3-4X_4)\sim N(0,1),$$

于是 $D[\sqrt{a}(X_1-2X_2)]=(a+4a)DX_1=5a\times 2^2=1$，即 $a=\frac{1}{20}$；

$D[\sqrt{b}(3X_3-4X_4)]=(9b+16b)DX_1=25b\times 2^2=1$，即 $b=\frac{1}{100}$．

[典型错误] 错误地认为 $D(X_1-2X_2)=DX_1-4DX_2$．

例 3.6.5 在天平上重复称量一重为 a 的物品，假设各次称量结果相互独立且同服从正态分布 $N(a,0.2^2)$．若以 $\bar{X}_n$ 表示 n 次称量结果的算术平均值，则为使 $P\{|\bar{X}_n-a|<0.1\}\geqslant 0.95$，$n$ 的最小值应不小于正整数________．

[答案] 16．

[提示] 本题主要考查样本均值的抽样分布、标准正态分布的分位数．

[解] 由于
$$\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i\sim N\left(a,\frac{1}{n}\cdot 0.2^2\right),$$

故
$$z=\frac{\bar{X}_n-a}{\frac{0.2}{\sqrt{n}}}\sim N(0,1),\ P\{|z|<1.96\}\geqslant 0.95,$$

因而使
$$P\{|\bar{X}_n-a|<0.1\}=P\left\{\frac{\sqrt{n}\,|\bar{X}_n-a|}{0.2}<\frac{\sqrt{n}}{2}\right\}\geqslant 0.95.$$

令 $\frac{\sqrt{n}}{2}\geqslant 1.96$，即得 n 的最小值应不小于 16．

[典型错误] 不清楚样本均值的抽样分布，不知道标准正态分布的分位数．

例 3.6.6 设随机变量 X 服从正态分布 $N(0,1)$．对给定的 α $(0<\alpha<1)$，数 u_α 满足 $P\{X>u_\alpha\}=\alpha$．若$P\{|X|<x\}=\alpha$，则 x 等于(　　)．

(A) $u_{\frac{\alpha}{2}}$　　(B) $u_{1-\frac{\alpha}{2}}$　　(C) $u_{\frac{1-\alpha}{2}}$　　(D) $u_{1-\alpha}$

[答案] (C)．

[提示] 本题主要考查概率运算、标准正态分布．

[解] 由标准正态分布概率密度的对称性，可知 $P\{X\leqslant -x\}=P\{X\geqslant x\}$，则

$$\alpha=P\{|X|<x\}=1-P\{|X|\geqslant x\}=1-(P\{X\geqslant x\}+P\{X\leqslant -x\})=1-2P\{X\geqslant x\},$$

故
$$P\{X\geqslant x\}=P\{X>x\}=\frac{1-\alpha}{2}.$$

再由 u_α 的定义知

$$x=u_{\frac{1-\alpha}{2}}.$$

[典型错误] 没能正确理解 u_α 的定义，从而无从下手解决本题．

例 3.6.7 设随机变量 $X\sim t(n)$ $(n>1)$，$Y=\frac{1}{X^2}$，则(　　)．

(A) $Y\sim\chi^2(n)$　　(B) $Y\sim\chi^2(n-1)$　　(C) $Y\sim F(n,1)$　　(D) $Y\sim F(1,n)$

[答案] (C).

[提示] 本题主要考查 t 分布、χ^2分布及 F 分布的定义.

[解] 设 $W\sim N(0,1)$，$Z\sim\chi^2(n)$，且 W，Z 独立，则 $X=\dfrac{W}{\sqrt{Z/n}}\sim t(n)$. 又 $W\sim N(0,1)$. 所以 $W^2\sim\chi^2(1)$，所以$\dfrac{1}{X^2}=\dfrac{Z/n}{W^2/1}\sim F(n,1)$.

[典型错误] 因为有相当一些考生只是在辅导班上学了点数理统计，理解不深，所以无法做此题.

例 3.6.8 设随机变量 X 和 Y 都服从标准正态分布，则(　　).

(A) $X+Y$ 服从正态分布　　(B) X^2+Y^2 服从χ^2分布

(C) X^2 和 Y^2 都服从χ^2分布　　(D) X^2/Y^2 服从 F 分布

[答案] (C).

[提示] 本题主要考查正态分布、χ^2分布、F 分布的定义和性质.

[解] 当随机变量 X 和 Y 都服从正态分布，且二者相互独立时，(A)、(B)、(C)、(D)四选项均成立. 当未给出 X、Y 相互独立这一条件时，(A)、(B)、(D)均不一定成立.

[典型错误] 对二元正态分布的函数的分布不清楚，而错误地选择(A).

例 3.6.9 设$X_1,X_2,\cdots,X_n$ $(n\geqslant2)$为来自总体 $N(0,1)$的简单随机样本，$\bar X$ 为样本均值，S^2 为样本方差，则(　　).

(A) $n\bar X\sim N(0,1)$　　(B) $nS^2\sim\chi^2(n)$

(C) $\dfrac{(n-1)\bar X}{S}\sim t(n-1)$　　(D) $\dfrac{(n-1)X_1^2}{\sum\limits_{i=2}^{n}X_i^2}\sim F(1,n-1)$

[答案] (D).

[提示] 本题主要考查抽样分布，如正态分布、χ^2分布、t 分布和 F 分布. 考查样本方差的定义和分布. 注意到大纲中定义的样本方差为 $S^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar X)^2$,而不是 $S^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\bar X)^2$.另外,在本题中总体均值与方差均是已知的.

[解] 由均值及方差的定义和性质可知. $\bar X\sim N\left(0,\dfrac{1}{n}\right)$. $(n-1)S^2\sim\chi^2(n-1)$，故(A)、(B)、(C)选项不正确. 正确的为(D). 事实上，$X_1^2\sim\chi^2(1)$，且与 $\sum\limits_{i=2}^{n}X_i^2$ 独立,另外$\sum\limits_{i=2}^{n}X_i^2\sim\chi^2(n-1)$,故$\dfrac{X_1^2/1}{\sum\limits_{i=2}^{n}X_i^2/(n-1)}\sim F(1,n-1)$.

[典型错误] 不清楚 S^2 的定义而错选(B).

例 3.6.10 设总体 X 服从正态分布$N(\mu,\sigma^2)$ $(\sigma>0)$. 从该总体中抽取简单随机样本$X_1,X_2,\cdots,X_{2n}$ $(n\geqslant2)$，其样本均值为 $\bar X=\dfrac{1}{2n}\sum\limits_{i=1}^{2n}X_i$,求统计量 $Y=\sum\limits_{i=1}^{n}(X_i+X_{n+i}-2\bar X)^2$ 的数学期望 EY.

[提示] 本题主要考查样本的独立性和同分布性质、数学期望的计算. 如果将X_1+X_{n+1}、X_2+X_{n+2}、$\cdots$，X_n+X_{2n}视作正态总体$N(2\mu,2\sigma^2)$的 n 个简单随机样本，容易计算出 EY. 也可以将$(X_i+X_{n+i}-2\bar X)^2$分成$[(X_i-\bar X)+(X_{n+i}-\bar X)]^2$. 分别考虑$X_1,X_2,\cdots,X_n$ 以及$X_{n+1},X_{n+2},\cdots,X_{2n}$两个样本的性质，注意$X_1,X_2,\cdots,X_{2n}$是相互独立的.

[解法 1] 考虑X_1+X_{n+1}，X_2+X_{n+2}，$\cdots$，X_n+X_{2n}. 将其视为取自总体 $N(2\mu,2\sigma^2)$的简单随机样本，则其样本均值为 $\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i+X_{n+i})=\dfrac{1}{n}\sum\limits_{i=1}^{2n}X_i=2\bar X$,样本方差为$\dfrac{1}{n-1}Y$.

由于 $E\left(\frac{1}{n-1}Y\right)=2\sigma^2$，所以 $EY=(n-1)(2\sigma^2)=2(n-1)\sigma^2$.

[解法 2] `记

$$\bar{X}'=\frac{1}{n}\sum_{i=1}^{n}X_i,\ \bar{X}''=\frac{1}{n}\sum_{i=1}^{n}X_{n+i}.$$

显然有 $2\bar{X}=\bar{X}'+\bar{X}''$. 因此

$$\begin{aligned}EY&=E\left[\sum_{i=1}^{n}(X_i+X_{n+i}-2\bar{X})^2\right]\\&=E\left\{\sum_{i=1}^{n}[(X_i-\bar{X}')+(X_{n+i}-\bar{X}'')]^2\right\}\\&=E\left\{\sum_{i=1}^{n}[(X_i-\bar{X}')^2+2(X_i-\bar{X}')(X_{n+i}-\bar{X}'')+(X_{n+i}-\bar{X}'')^2]\right\}\\&=E\left[\sum_{i=1}^{n}(X_i-\bar{X}')^2\right]+0+E\left[\sum_{i=1}^{n}(X_{n+i}-\bar{X}'')^2\right]\\&=(n-1)\sigma^2+(n-1)\sigma^2\\&=2(n-1)\sigma^2.\end{aligned}$$

[典型错误]

① 相当多的考生将 $\bar{X}=\frac{1}{2n}\sum_{i=1}^{2n}X_i$ 写成 $\frac{1}{n}\sum_{i=1}^{n}X_i$.

② 认为 $E(X_i\bar{X})$ 等于 $EX_iE\bar{X}$，其实 X_i 与 $\bar{X}$ 并不独立. 所以并不等.

③ 认为 $D(X_i+\bar{X})$ 等于 $DX_i+D\bar{X}$，错误性质同②；认为 $D(X_i-2\bar{X})=DX_i-4D\bar{X}$，这里比前面多一层错误.

例 3.6.11 从正态总体 $N(3.4,6^2)$ 中抽取容量为 n 的样本，如果要求其样本均值位于区间 $(1.4,5.4)$ 内的概率不小于 0.95. 问样本容量 n 至少应取多大?

附表：标准正态分布表

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

z	1.28	1.645	1.96	2.33
$\Phi(z)$	0.900	0.950	0.975	0.990

[提示] 本题主要考查抽样分布、区间估计、正态概率表. 题目要求满足条件 $P\{1.4<\bar{X}<5.4\}\geqslant 0.95$ 的 n，只需知道 $\bar{X}$ 的抽样分布即可.

[解] 以 $\bar{X}$ 表示该样本均值，则

$$\frac{\bar{X}-3.4}{6}\sqrt{n}\sim N(0,1).$$

从而有

$$\begin{aligned}P\{1.4<\bar{X}<5.4\}&=P\{-2<\bar{X}-3.4<2\}\\&=P\{|\bar{X}-3.4|<2\}=P\left\{\frac{|\bar{X}-3.4|}{6}\sqrt{n}<\frac{2\sqrt{n}}{6}\right\}\\&=2\Phi\left(\frac{\sqrt{n}}{3}\right)-1\geqslant 0.95,\end{aligned}$$

故
$$\Phi\left(\frac{\sqrt{n}}{3}\right)\geqslant 0.975,$$

由此得
$$\frac{\sqrt{n}}{3}\geqslant 1.96.$$

即 $$n \geqslant (1.96\times 3)^2 \approx 34.57,$$
所以 n 至少应取 35.

[典型错误] 不清楚样本均值的分布，不清楚正态分布的对称性.

例 3.6.12 设 $X_1, X_2, \cdots, X_9$ 是来自正态总体 X 的简单随机样本，

$$Y_1=\frac{1}{6}(X_1+\cdots+X_6),\quad Y_2=\frac{1}{3}(X_7+X_8+X_9),$$

$$S^2=\frac{1}{2}\sum_{i=7}^{9}(X_i-Y_2)^2, Z=\frac{\sqrt{2}(Y_1-Y_2)}{S},$$

证明统计量 Z 服从自由度为 2 的 t 分布.

[提示] 本题主要考查 t 分布、χ^2分布、随机变量和的分布. 要证明统计量 Z 服从自由度为 2 的 t 分布，必须证明 $Z=\dfrac{U}{\sqrt{\chi^2/2}}$，其中 $U\sim N(0,1)$，χ^2服从自由度为 2 的χ^2分布.

[证] 记 $DX=\sigma^2$（未知），易见

$$EY_1=EY_2,\ DY_1=\frac{\sigma^2}{6},\ DY_2=\frac{\sigma^2}{3}.$$

由于 Y_1 和 Y_2 独立，可见 $E(Y_1-Y_2)=0$，

$$D(Y_1-Y_2)=\frac{\sigma^2}{6}+\frac{\sigma^2}{3}=\frac{\sigma^2}{2},$$

从而

$$U=\frac{Y_1-Y_2}{\sigma/\sqrt{2}}\sim N(0,1).$$

由正态总体样本方差的性质，知

$$\chi^2=\frac{2S^2}{\sigma^2}$$

服从自由度为 2 的χ^2分布.

由于 Y_1 与 Y_2、Y_1 与 S^2 以及 Y_2 与 S^2 独立，可见 Y_1-Y_2 与 S^2 独立.

于是，由服从 t 分布随机变量的结构，知

$$Z=\frac{\sqrt{2}(Y_1-Y_2)}{S}=\frac{U}{\sqrt{\chi^2/2}}$$

服从自由度为 2 的 t 分布.

[典型错误]

① 没有说明 Y_1-Y_2 与 S^2 独立.

② 不知$\dfrac{2S^2}{\sigma^2}\sim\chi^2(2)$.

七、参数估计(数学三)

统计推断，即是用样本推断总体，是数理统计学的核心内容. 估计与检验是其中的两大类问题. 参数估计就是根据样本对总体中的未知参数或数字特征进行估计，常分为点估计与区间估计两种. 本节的重点是理解参数的点估计、估计量与估计值的概念，了解估计量的无偏性、有效性和相合性的概念及其证明方法，掌握求估计量的矩估计法和最大似然估计法，了解区间估计的概念，掌握建立未知参数置信区间的一般方法，掌握正态总体参数及其相关特征的置信区间的求法.

• 考试内容与要求 •

考试内容