期末考试复习

数学科学学院

电子科技大学
UNIVERSITY OF ELECTRONIC SCIENCE
AND TECHNOLOGY OF CHINA

期末考试概述：

1、 设随机事件 A，B 及和事件 $A\cup B$ 的概率分别为 0.4，0.3 和 0.6，若 $\bar{B}$ 表示 B 的对立事件，那么积事件 $A\bar{B}$ 的概率 $P(A\bar{B})=$（　　）.

（A）0.2;　　（B）0.3;　　（C）0.4;　　（D）0.6.

2、 袋中装有 5 个小球，分别标有号码 1、2、3、4、5；从中任取三个小球，用 X 和 Y 分别表示最小号码和最大号码，则 X 的边缘分布律为（　　）.

（A）

X	3	4	5
P	0.4	0.4	0.2

（B）

X	3	4	5
P	0.6	0.3	0.1

（C）

X	1	2	3
P	0.4	0.4	0.2

（D）

X	1	2	3
P	0.6	0.3	0.1

3、 设 $X\sim U(-1,1)$，则 $Y=\dfrac{1}{|X|}$ 的概率密度为（　　）.

（A）$f(y)=\begin{cases}1/y^2, & y>1\\ 0, & 其它\end{cases}$;　　（B）$f(y)=\begin{cases}1/2y^2, & |y|>1\\ 0, & 其它\end{cases}$;

（C）$f(y)=\begin{cases}1/2y^2, & y>1\\ 0, & 其它\end{cases}$;　　（D）$f(y)=\begin{cases}1/y^2, & |y|>1\\ 0, & 其它\end{cases}$.

4、 设 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，且 $P\{|X-\mu_1|<1\}>P\{|Y-\mu_2|<1\}$，则必有（　　）.

（A）$\sigma_1<\sigma_2$;　　（B）$\sigma_1>\sigma_2$;　　（C）$\mu_1<\mu_2$;　　（D）$\mu_1>\mu_2$.

期末考试会涉及到只和前三章有关的题目

1、（10 分）常数 a 与任意随机变量独立吗？给出数学证明过程.

期末考试概述：

去年考了的大题类型：

1.计算随机变量的期望

2、（10 分）设连续型随机变量 X 的概率密度函数为

$$f(x)=\begin{cases} ax, & 0<x<2 \\ cx+b, & 2\le x<4 \\ 0, & \text{其他} \end{cases},$$

已知，$E(X)=2$，$P\{1<X<3\}=\frac{3}{4}$，求：（1）a,b,c 的值；（2）随机变量 $Y=e^X$ 的期望.

2.抽样分布定理、正态分布的概率计算

3、（10 分）设总体 $X\sim N(20,9)$，从总体中抽取一组样本容量为 n 的样本，$\bar{X}$ 是样本均值，试求使得

$$P\{19.9\le\bar{X}\le 20.1\}\ge 0.95$$

成立的最小的样本容量值．已知：$\Phi(1.96)=0.975,\Phi(1.645)=0.95$.

计算概率时对随机变量和a同时标准化

3.两类点估计和优良性准则判断

三、计算题（共 30 分，每题 15 分）

1、（15 分）设总体 $X \sim U(0,\theta)$，参数 $\theta(>0)$ 未知，$X_1,X_2,\cdots,X_n$ 为其样本，求：

（1）θ 的矩估计量 $\hat{\theta}_1$；

（2）θ 的极大似然估计量 $\hat{\theta}_2$；

（3）判断 $\hat{\theta}_1$ 与 $\hat{\theta}_2$ 是否是参数 θ 的无偏估计量.

最大似然估计：若似然方程（组）无解，或似然函数不可导，此法失效，改用其它方法。

注1：需要同学们理解最大似然估计法的本质。

注2：本题的解法需要自己做一遍才能体会。

提醒：看书时多动笔

建议：自己动笔算一下书上的例7.1.8

例 设样本$X_1, X_2, \ldots, X_n$是来自在区间$[0, \theta]$上均匀分布的总体，θ未知，求θ的最大似然估计。

解： 设$x_1, x_2, \ldots, x_n$是样本观测值。

总体的概率密度函数为

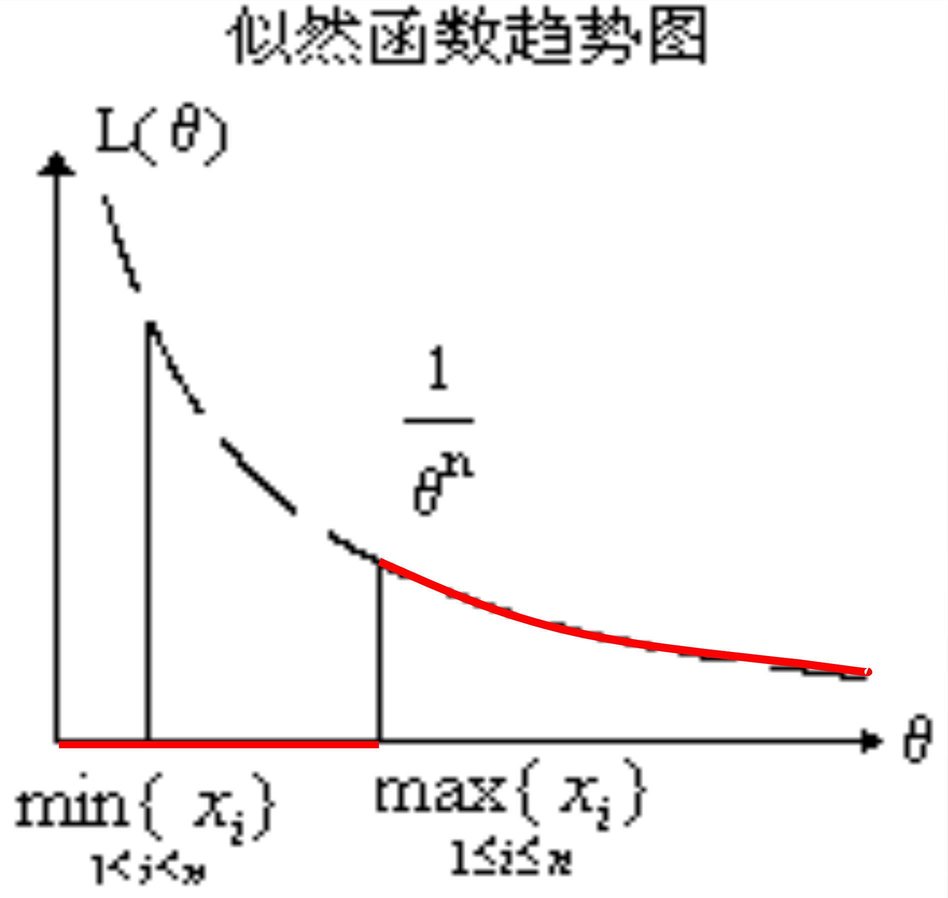
$$f(x)=\begin{cases}\dfrac{1}{\theta} & ,0\le x\le\theta\\ 0 & ,else\end{cases}$$

从而可得似然函数

$$L(x_1,\ldots,x_n;\theta)=\begin{cases}\dfrac{1}{\theta^n} & ,0\le x_i\le\theta, i=1,\cdots,n\\ 0 & ,else\end{cases}$$

注意： 该似然函数不能用一般方法--通过求导构造似然方程。尝试用其他方法求解！

$$L=\begin{cases}\dfrac{1}{\theta^n}, & 0\le x_i\le\theta \\ 0, & 其他\end{cases}=\begin{cases}\dfrac{1}{\theta^n}, & 0\le \min\limits_i x_i, \max\limits_i x_i\le\theta \\ 0, & 其他\end{cases}$$



如图所示，似然函数L在

$\tilde{\theta}=\max\limits_{1\le i\le n}\{x_i\}$处达到极大值

故θ的极大似然估计量为：

$$\hat{\theta}=\max_{1\le i\le n}\{X_i\}$$

另：由$E(X)=\dfrac{\theta}{2}=\bar{X}$ 得矩估计量为 $\hat{\theta}=2\bar{X}$

注1：此例中矩估计量不等于极大似然估计量

注2：此例表明了把密度函数写全的重要性

4.双样本t假设检验

2、（15 分）学生可以在不包含实验课程的 3 学分物理课和包含实验课程的 4 学分物理课之间进行选择，且两个物理课的期末考试是相同的．其中，抽样调查了上 3 学分物理课的 18 名学生，算得平均考试成绩 $\bar{x}=77$，样本标准差 $s_1=6$；抽样调查了上 4 学分物理课的 12 名学生，算得平均考试成绩 $\bar{y}=84$，样本标准差 $s_2=4$；设 3 学分物理课学生成绩总体 X 的均值为 μ_1，方差为 σ_1^2；4 学分物理课学生成绩总体 Y 的均值为 μ_2，方差为 σ_2^2；假设两个总体 X 与 Y 均服从正态分布，相互独立且方差相等 $(\sigma_1^2=\sigma_2^2=\sigma^2)$．

（1）计算两门物理课平均考试成绩差值 $\mu_1-\mu_2$ 的置信水平为 0.99 的置信区间．

（2）是否有证据表明 3 学分课程的学生成绩相较 4 学分的会更低（即 $\mu_1\le\mu_2$）？使用 0.05 的显著性水平．

已知：$t_{0.05}(28)=1.701, t_{0.025}(28)=2.048, t_{0.005}(28)=2.763,$

$t_{0.05}(30)=1.697, t_{0.025}(30)=2.042, t_{0.005}(30)=2.750$

问题：零假设怎么给?

思考：是用证据说明一个事实能成立，这个证据需要强有力。

单侧检验

比较下列两个检验的否定条件：

$$H_0: \mu \le \mu_0 \quad H_1: \mu > \mu_0 \qquad (a)$$

$$H_0: \mu = \mu_0 \quad H_1: \mu > \mu_0 \qquad (a')$$

分析：假设检验是用反证法，检测H_0下的结果是否有利于备择假设。

注：确定H_0的拒绝域（否定H_0的条件）遵循备择有利原则：

选择有利于H_1成立的检验统计量的取值区域作为原假设H_0的拒绝域，也即H_1的接受域。

因此(a')的拒绝域也即（a）的拒绝域。

注：假设检验是单侧还是双侧的，由H_1的形式决定

(双正态总体的抽样分布定理)

设$X_1, X_2, \ldots, X_{n_1}$和$Y_1, Y_2, \ldots, Y_{n_2}$分别是来自正态总体$X \sim N(\mu_1, \sigma_1^2)$和$Y \sim N(\mu_2, \sigma_2^2)$的样本且相互独立，$\overline{X}, S_1^2$及$\overline{Y}, S_2^2$为各自的样本均值和样本方差，则

$$(1)\ F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

(2)当$\sigma_1^2 = \sigma_2^2$时，

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

其中，$S_w = \sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$

联合样本均方差

去年没考的大题类型：

1.多维正态分布的计算

例 设二维随机变量 $(X, Y) \sim N(1, 3^2; 0, 4^2; -0.5)$，并且

$Z = \frac{X}{3} + \frac{Y}{2}$，试求

(1) Z 的数学期望和方差；

(2) X 与 Z 的相关系数 ρ_{XZ}；

(3) X 与 Z 是否相互独立？

定理 设 $X=(X_1, X_2, \ldots, X_n)$ 服从 n 维正态分布 $N(\mu, C)$，$B=(B_{jk})_{m\times n}$ 是任意矩阵，则 BX^T 服从 m 维正态分布 $N(B\mu, BCB^T)$.

例 建议大家用定理和线性变换来做题

去年没考的大题类型：

2.计算统计量的抽样分布

例　统计量的分布

设 $X_1, X_2, \ldots, X_{n+m}$ 是来自正态总体 $X \sim N(0,\sigma^2)$ 的样本，求下列统计量的概率分布：

1. $Y = \dfrac{1}{\sigma^2}\sum_{i=1}^{n+m} X_i^2$　　2. $Z = \sqrt{\dfrac{m}{n}}\sum_{i=1}^{n} X_i \Big/ \sqrt{\sum_{i=n+1}^{n+m} X_i^2}$　　3. $\dfrac{1}{Z^2}$

注意三大分布构造时所需的**独立性**！！

如同时涉及样本均值、样本方差，可能需要**抽样分布定理**。

9、设随机变量 X 和 Y 都服从标准正态分布，则必有（　　）.

（A）$X+Y$ 服从正态分布 $N(0,2)$；　（B）$\dfrac{X}{|Y|}$ 服从 $t(1)$ 分布；

（C）X^2 和 Y^2 都服从 $\chi^2(1)$ 分布；　（D）$\dfrac{X^2}{Y^2}$ 服从 $F(1,1)$ 分布.

定义 设 随机变量X，Y 相互独立，$X\sim N(0,1)$，$Y\sim\chi^2(n)$，记

$$\boxed{T=\frac{X}{\sqrt{Y/n}}\sim t(n)}$$

称随机变量 T 服从自由度为 n 的 t 分布。

定义 设 随机变量X，Y 相互独立，$X\sim\chi^2(n_1)$，$Y\sim\chi^2(n_2)$，记

$$\boxed{F=\frac{X/n_1}{Y/n_2}\sim F(n_1,n_2)}$$

称随机变量F 服从第一自由度为n_1，第二自由度为n_2的F分布。

7. 设总体 $X \sim N(\mu,\sigma^2)$,样本 $X_1,X_2,\cdots,X_n,X_{n+1}$, 记 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$, $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$,

确定统计量 $\frac{X_{n+1}-\overline{X}}{S}\sqrt{\frac{n}{n+1}}$ 的抽样分布。

设$X_1,X_2,\ldots,X_9$是标准正态总体的简单随机样本，$\overline{X}$，S^2是样本方差，证明$9\overline{X}^2+8S^2\sim\chi^2(9)$

8. 设总体 $X \sim N(0,\sigma^2)$, X_1,X_2,X_3,X_4 为其样本，试确定 $\frac{(X_1-X_2)^2}{(X_3+X_4)^2}$ 的分布。

注意到$\overline{X}$和S^2独立，熟记抽样分布定理的所有内容

3.线性回归

(3 分)

六、(共 10 分)

在考察硝酸钠的可溶程度时，对一系列不同温度观察它在 100 毫升的水中溶解的硝酸钠的重量，得到一组观测值 $(x_i,y_i), i=1,2,\cdots,9$，计算得

$$\sum_{i=1}^{9}x_i=234,\ \sum_{i=1}^{9}y_i=811.3,\ \sum_{i=1}^{9}x_i^2=10144,\ \sum_{i=1}^{9}y_i^2=76218.1339,\ \sum_{i=1}^{9}x_iy_i\ \ 4628.6$$

从理论上推测，温度 x_i 与溶解的硝酸钠的重量 Y_i 之间有关系式：

$$Y_i=a+bx_i+\varepsilon_i, i=1,2,\cdots,9$$

式中 $\varepsilon_i, i=1,2,\cdots,9$ 相互独立，均服从正态分布 $N(0,\sigma^2)$

(1) 求未知参数 a,b 的最小二乘估计值和 σ^2 的无偏估计值；(2) 检验线性回归是否显著 ($\alpha=0.01$)？(结果保留四位有效数字)

α \ 自由度	6	7
0.01	0.834	0.798
0.005	0.870	0.836

10、在分析数据 $(x_i,y_i)(i=1,2,\cdots,n)$ 的关系时，某人用线性回归 $y=a+bx$ 来对参数 a,b 估计，计算得到 $\bar{x}=3$，$\bar{y}=10$，请问下列结论正确的是（　　）.

（A）经验回归直线必过点 $(3,10)$；　（B）点 $(3,10)$ 在经验回归直线上方；

（C）经验回归直线一定不过点 $(3,10)$；　（D）点 $(3,10)$ 在经验回归直线下方.

$$l_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-n\bar{x}^2$$

$$l_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}x_iy_i-n\bar{x}\,\bar{y}$$

$$l_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}y_i^2-n\bar{y}^2$$

$$\hat{b}=\frac{l_{xy}}{l_{xx}},\quad \hat{a}=\bar{y}-\hat{b}\bar{x},$$

$$\hat{\sigma}^2=\frac{1}{n-2}(l_{yy}-\hat{b}^2 l_{xx})$$

给定显著性水平α

当 $|R|\le R_\alpha(n-2)$，接受原假设

认为X与Y之间的线性相关关系不显著。

当 $|R|>R_\alpha(n-2)$，拒绝原假设

认为X与Y之间的线性相关关系显著。

期中考试的启发：

一.不重视课件和课本；

计算边缘分布$f_X(x)$注意点：

1. 边缘密度函数$f_X(x)$的非零区间；
2. 完整写下$f_X(x)$，包括为0的区间。

计算条件分布$f_{X|Y}(x|y)$注意点：

1. 先要计算边缘密度函数$f_Y(y)$的非零区间；
2. 当$f_Y(y) > 0$时才有条件分布（概率）存在，需写明；
3. 在条件分布存在时，完整写出$f_{X|Y}(x|y)$。

判断$f_X(x)$和$f_Y(y)$的独立性的注意点：

若要说明不独立，需要 $f(x,y) \neq f_X(x)f_Y(y)$

在平面上一个“面积”非0的区域内成立。

期中考试的启发：

二.经典题型不注重细节；

例：1.边缘密度写的不完整；

2.不讨论条件分布的存在性；

3.全概率公式不写划分；

4.说明不独立时不明确面积非0的区域。

建议：标准题型严格按照课本例题的过程去做。

三.计算能力差；

建议：多动笔，不要抄答案。

期中考试的启发：

四.常见题型严格对照课本用标准的解法。
注：对后半部分章节更加重要。

五.考试的注意事项。
例：1.答题区域要写全步骤、用到的定理、性质(重要)；
2.未出现在书上，而出现在课件上的例子或者结果，不能直接使用。
3.题目看清楚，如有疑问可找巡考。

如：设二维随机变量(X , Y)的联合概率密度为：

$$f(x,y)=\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}(1+\sin x\sin y),\qquad (x,y)\in R^2$$

则可计算得

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},-\infty<x<+\infty$$

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx=\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}},-\infty<y<+\infty$$

即 $X\sim N(0,1)$，$Y\sim N(0,1)$ 类似反例，如要使用需证明！

但显然(X , Y)并不服从二维联合正态分布！

前三章回顾：

重视大量的基础概念：

例：分布函数(包括联合)、概率、(事件和随机变量)两两独立与相互独立、离散(连续)型随机变量等等；

注：以课本和课件为主！重视期中复习课件。

常见题型练习：

例：事件概率的计算；
边缘分布和条件分布的计算；
随机变量的函数的分布。

第四章：

一.数学期望、方差、矩的定义和存在性问题。

1.数学期望不存在的反例（条件收敛，柯西分布）；

二.数学期望和(协)方差的计算

1.常见分布的期望和方差(包括几何分布);

2.期望和方差的性质，如：$D(aX \pm bY)$;

3.随机变量的函数的期望和方差的计算;

4.协方差和相关系数的计算、性质、区别;

5.独立性(讨论联合分布和边缘分布)和相关性(相关系数是否为0)的关系。(反例？)

注：对照书和课件查上面提到的性质等。

第四章：

经典题型：

13.将 n 只球（1 到 n 号）随机地放进 n 个盒子（1 到 n 号）中去，一个盒子装一只球。若一只球装入与球同号的盒子中，称为一个配对。记 X 为总的配对数，计算 $E(X)$ 和 $D(X)$。

三.多元正态分布

要求掌握：

1.多元正态的概率密度函数(向量形式)；
2.多元正态的线性变换的分布；
3.从分布提取相关性信息；
4.独立和不相关的关系。

第五章：

一.大数定律的定义及其证明

1.依概率收敛与大数定律的关系；

2.两类条件(独立同分布，独立且方差有统一上界)。

注：若让证明大数定律，最好给以证明。

二.中心极限定理及相关题型。

1.依分布收敛与中心极限定理的关系；

2.条件(独立同分布)。

技巧：1.计算概率时对随机变量和a同时标准化；

2.注意$\Phi(-3)\approx 0$.

第六章：

一.基本概念：

1.简单随机样本的定义、性质和它的本质(随机变量)；
2.统计量和统计值的概念及大小写问题；
3.常见统计量(样本均值，样本方差，样本k阶中心矩等)；
4.四大分布的定义(注意独立性要求)和性质(可加性，大样本，对称性等)；
5.上侧分位数的定义，以及分布对称时的性质。

二.抽样分布定理：

1.本质：样本均值和样本方差的分布；
2.注意样本均值和样本方差的独立性；
3.单样本和双样本抽样分布定理都要重视。

第六章：

抽样分布定理和相关的证明题：

设 X_1，X_2，…，X_{n+m} 是来自正态总体 $X \sim N(0, \sigma^2)$ 的样本，求下列统计量的概率分布：

1. $Y = \dfrac{1}{\sigma^2} \sum_{i=1}^{n+m} X_i^2$

2. $Z = \sqrt{\dfrac{m}{n}} \sum_{i=1}^{n} X_i \Big/ \sqrt{\sum_{i=n+1}^{n+m} X_i^2}$

3. $\dfrac{1}{Z^2}$

注：通过统计量的形式先确定是哪一个分布

第七章：统计部分的重难点！

一.基本题型：矩法和最大似然估计法。

矩法：从估计量到估计值；
最大似然估计法：从估计值到估计量。

注1：看清是值还是量。
注2：注意似然方程无解时的处理方法。

二.三个准则的基本概念及相关题型。

三.本章大量小的知识点，回归课本，梳理细节。
例：对方差的矩估计不是样本方差，有偏等等；
正态分布的最小方差无偏估计是什么?
相合估计量可能有偏。

例 设总体的分布律为

X	0	1	2	3
P	θ^2	$2\theta(1-\theta)$	θ^2	1-2θ

样本：3,1,3,0,3,1,2,3

求：（1）$\boldsymbol{\theta}$的矩估计值；(2）$\boldsymbol{\theta}$的最大似然估计值

解：（1）$E(X)=2\theta(1-\theta)+2\theta^2+3(1-2\theta)=\bar{X}=2$

（2）$\boldsymbol{L(\theta)=[P(X_i=3)]^4[P(X_i=1)]^2P(X_i=0)P(X_i=2)}$

似然函数通常由边缘分布列或者边缘密度的连乘得到

抽象的样本联合分布函数可能写不出来，直接根据样本写似然函数

区间估计、八、九章：

理清区间估计和假设检验的关系(看8.1课件)。

两类标准题型：假设检验；线性关系是否显著。

注1：自己提取文字信息给定零假设和备择假设。

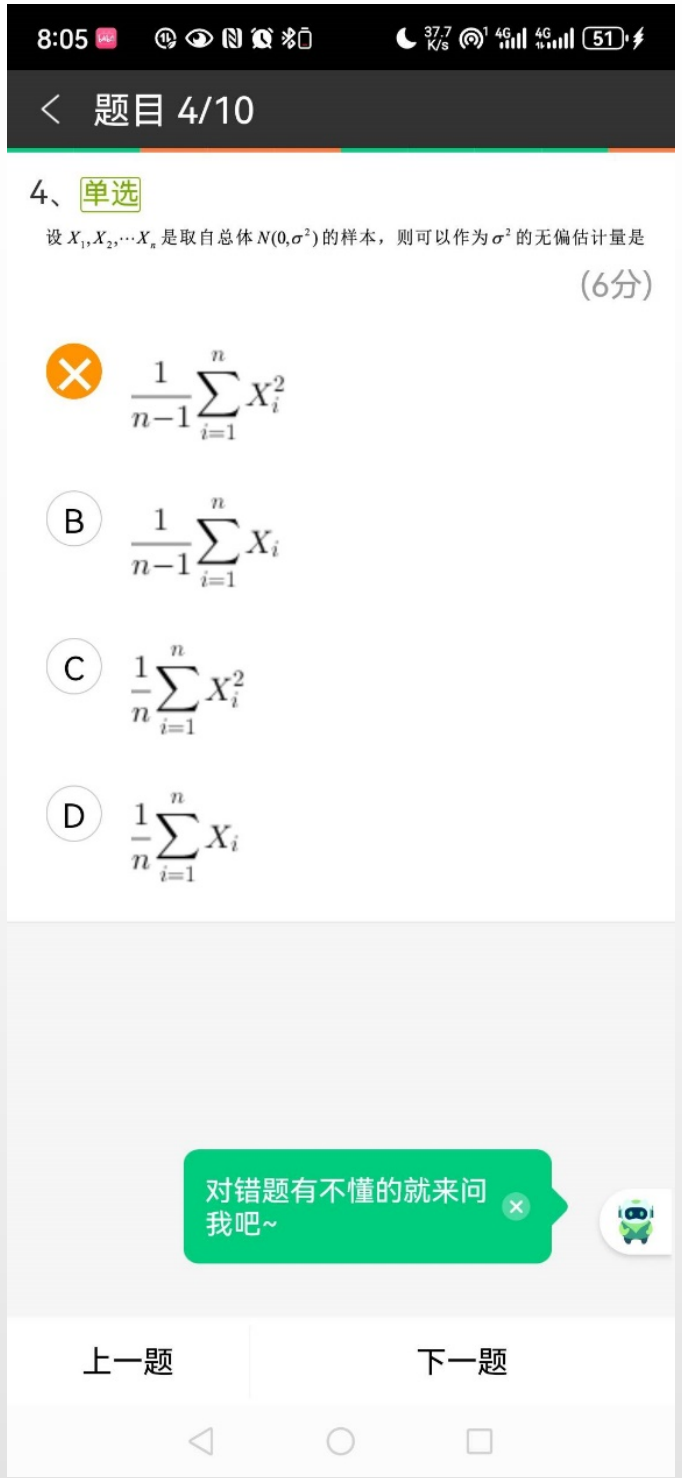
注2：单侧检验问题遵从备择有利原则(看8.1课件)。

注3：当参数已知时，区间估计可能不需要用到抽样分布定理。(能不用抽样分布定理就不用)。

注4：判断线性关系是否显著，用相关系数法。

总而言之：标准题型严格使用课本的标准步骤！

注5：关心数学建模的同学，看一下第九章课件。



9、设随机变量 X 和 Y 都服从标准正态分布，则必有（　　）.

（A）$X+Y$ 服从正态分布 $N(0,2)$；　（B）$\dfrac{X}{|Y|}$ 服从 $t(1)$ 分布；

（C）X^2 和 Y^2 都服从 $\chi^2(1)$ 分布；　（D）$\dfrac{X^2}{Y^2}$ 服从 $F(1,1)$ 分布.

感悟：备考和考试，重在细节

备考注意事项

1.抄过的部分重要习题，不看答案重新做一遍。
2.常见题型**严格**仿照课本上的过程解答。
3.重视**课本、课件、课后题**。
4.多动笔，提高计算能力。

固定答疑：
每周一晚7:30-9:30，清水河校区品A109，直到12.30

祝大家考试顺利！谢谢大家！