厦门大学《概率统计(A)》期末试卷

信息科学与技术学院________系 2020 年级 计算机 专业

学年学期:20212 主考教师:概率统计教研组 A 卷()B 卷(√)

一、选择题（在各小题四个备选答案中选出一个正确答案，填在题末的括号中，本大题共 6 个小题，每小题 3 分，总计 18 分）

1. $X_1, X_2, \cdots, X_n$ 是独立同分布的随机序列，当 $n \to \infty$ 时，以下正确的是（）

A、若 $X_i \sim P(\lambda)$，则 $\bar{X} \xrightarrow{P} \frac{\lambda}{2}$　　　　B、若 $X_i \sim P(\lambda)$，则 $\bar{X} \xrightarrow{P} \frac{\lambda}{4}$

C、若 $X_i \sim \mathrm{U}(0,\theta), \theta > 0$，则 $\bar{X} \xrightarrow{P} \frac{\theta}{2}$　　　　D、若 $X_i \sim \mathrm{U}(0,\theta), \theta > 0$，则 $\bar{X} \xrightarrow{P} \frac{\theta}{4}$

分析：本题考察独立同分布大数定律，经计算 C 选项符合题意

若 $X_i \sim P(\lambda)$，则 $\bar{X} \xrightarrow{P} \lambda$

若 $X_i \sim \mathrm{U}(0,\theta), \theta > 0$，则 $\bar{X} \xrightarrow{P} \frac{\theta}{2}$

2. 设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 X 的样本，总体方差为 σ^2，$\bar{X}$ 是样本平均值，S^2 是样本方差，则下列各式中（　）为统计量。

A、$\sum_{i=1}^{n}(X_i - EX)^2$　　　　B、$(n-1)S^2/\sigma^2$

C、$\bar{X} - EX_i$　　　　D、$nX^2 + 1$

解析： A 中含总体期望 EX 是未知参数，C 中 EX_i 也是未知参数，都不是统计量；而 D 不是样本的函数，也不是统计量。应选 B。

3. 设总体 $X \sim N(m, \sigma^2)$（m、σ^2 均未知），$(X_1, X_2, \cdots, X_n)$ 为其样本，则检验问题

$$H_0: \sigma^2 \geqslant \sigma_0^2, \quad H_1: \sigma^2 < \sigma_0^2$$

的检验统计量为 $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$，若取显著性水平为 0.1 则检验的拒绝域为（　(b)　）

(a) $\{\chi^2 \le \chi^2_{0.9}(n)\}$；　　　　(b) $\{\chi^2 \le \chi^2_{0.9}(n-1)\}$；

(c) $\{\chi^2 \le \chi^2_{0.1}(n)\}$；　　　　(d) $\{\chi^2 \le \chi^2_{0.1}(n-1)\}$

4. 在其他条件不变的情况下，要使正态总体均值的置信区间长度缩小一半，样本量应增加（）。
A. 一半　　B. 一倍　　C. 三倍　　D. 四倍

C

以总体方差σ^2已知的总体均值区间估计为例，其置信区间为

$$\left(\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

置信区间宽度为$2z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$

故在其他条件不变的情况下，样本量应为原来的 4 倍，即增加 3 倍。

5. 设 X_1，X_2，…，X_n是来自正态总体 $N(\mu,\sigma^2)$的简单随机样本，其中 μ，σ均未知，记$\bar{X}$和S^2分别为样本均值和样本方差，当H_0：$\mu=\mu_0$成立时则有 B

(A) $\frac{\bar{X}-\mu_0}{S}\sqrt{n}\sim N(0,1)$

(B) $\frac{\bar{X}-\mu_0}{S}\sqrt{n}\sim t(n-1)$

(C) $\frac{\bar{X}-\mu_0}{S}\sqrt{n}\sim t(n)$

(D) $\frac{1}{\sigma^2}\sum_i^n(X_i-\mu_0)^2\sim\chi^2(n-1)$

解析：$X\sim N(\mu_1,\sigma_1^2)$，$\mu$，$\sigma$均未知. 当$\mu=\mu_0$成立时$\frac{\bar{X}-\mu_0}{S}\sqrt{n}\sim t(n-1)$，选 B

6. 设总体 X 服从参数为λ的泊松分布，$\lambda>0$未知，$X_1,X_2\Lambda\ X_n$是总体的一个样本，$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$，

$S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ 分别是样本均值和样本方差，则下列哪个不是λ的无偏估计（）。

A. $\bar{X}$　　B. S^2　　C. $\frac{1}{3}\bar{X}+\frac{2}{3}S^2$　　D. $\frac{1}{4}\bar{X}+\frac{1}{3}S^2$

D

$$E(X)=D(X)=\lambda$$

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

$E\left(\overline{X}\right)=\lambda$，$E\left(S^2\right)=\lambda$

D 项 $E\left(\frac{1}{4}\overline{X}+\frac{1}{3}S^2\right)=\frac{1}{4}\lambda+\frac{1}{3}\lambda\neq\lambda$

二、填空题（本大题共 6 小题，每小题 3 分，总计 18 分）

1. 一本书共有 100 万个印刷符号，在排版时每个符号被排错的概率为 10^{-4}，在校对时每个排版错误被改正的概率为 0.9，则校对后错误不超过 15 个的概率为______。

分析：本题考察棣莫弗中心极限定理的基本掌握

依题设，校对后出现错误的概率应为 $P=10^{-4}\times(1-0.9)=10^{-5}$，记 100 万个印刷符号经校对后出错数为 X，则 $X\sim B(10^6,P)$，因而有 $E(X)=10$，$D(X)=9.9999$

根据定理，X 近似服从于正态分布，即 $X\sim N(10,9.9999)$，因此

$$P\{X\leq 15\}=P\left\{\frac{X-nP}{\sqrt{nP(1-P)}}\leq\frac{15-nP}{\sqrt{nP(1-P)}}\right\}$$

$$=\Phi\left(\frac{5}{10^3\sqrt{10^{-5}-10^{-10}}}\right)\approx\Phi(1.58)=0.9429$$

（查表，补充表值！）

2. 设 $X_1,X_2,\cdots,X_{10}$ 是来自正态总体 $X\sim N(0,0.09)$ 的样本，求 $P\left\{\sum_{i=1}^{10}X_i^2>1.44\right\}$= <u>0.1</u>.

解析：$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$，故 $P\left\{\sum_{i=1}^{10}X_i^2>1.44\right\}=P\{\chi^2(10)>16\}=0.1$

（查表，补充表值！）

3. 设总体 $X\sim N(\mu_1,\sigma^2)$，总体 $Y\sim N(\mu_2,\sigma^2)$，$X_1,X_2,\cdots,X_{n_1}$ 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 $E\left[\frac{\sum_{i=1}^{n_1}\left(X_i-\bar{X}\right)^2+\sum_{i=1}^{n_2}\left(Y_i-\bar{Y}\right)^2}{n_1+n_2-2}\right]$= <u>$\sigma^2$</u>.

解析：$E\left[\frac{\sum_{i=1}^{n_1}\left(X_i-\bar{X}\right)^2}{n_1-1}\right]=\sigma^2$，$E\left[\sum_{i=1}^{n_1}\left(X_i-\bar{X}\right)^2\right]=(n_1-1)\sigma^2$，

类似的， $E\left[\sum_{i=1}^{n_2}\left(Y_i-\bar{Y}\right)^2\right]=(n_2-1)\sigma^2$，

因此， $$E\left[\frac{\sum_{i=1}^{n_1}\left(X_i-\bar{X}\right)^2+\sum_{i=1}^{n_2}\left(Y_i-\bar{Y}\right)^2}{n_1+n_2-2}\right]=\frac{1}{n_1+n_2-2}\left\{E\left[\sum_{i=1}^{n_2}\left(X_i-\bar{X}\right)^2\right]+E\left[\sum_{i=1}^{n_2}\left(Y_i-\bar{Y}\right)^2\right]\right\}=\sigma^2$$

4. 设总体 $X\sim N(\mu,\sigma^2)$，σ^2 未知，若样本容量和置信度 $1-\alpha$ 均不变，则对于不同样本的观察值，总体均值 μ 的置信区间的长度________。

（可能）不同/不确定/不固定

5. 设总体 X~$N(0，4)$，$X_1,X_2,\cdots,X_{15}$ 是 X 的样本，当 a=_______时，$F=a\dfrac{X_1^2+\cdots+X_6^2}{X_7^2+\cdots X_{15}^2}$ 服从__________分布。

解：

$$\frac{X_i}{2}\sim N(0,1),\ 故\ \frac{X_i^2}{4}\sim\chi^2(1)$$

$X_1,X_2,\cdots,X_{15}$ 相互独立，故 $F=a\dfrac{X_1^2+\cdots+X_6^2}{X_7^2+\cdots X_{15}^2}\sim F(6,9)$时， $a=\dfrac{3}{2}$

6. 某青工以往的记录是： 平均每加工 100 个零件，由 60 个是一等品，今年考核他，在他加工零件中随机抽取 100 件，发现有 70 个是一等品，这个成绩是否说明该青工的技术水平有了显著性的提高（取 α = 0.05）？对此问题，假设检验问题应设为________________。

H_0： $p\le 0.6$； H_1： $p>0.6$

解析：一般的，选取问题的对立事件为原假设，在本题中，需考查青工的技术水平是否有了显著性的提高，故选取原假设为H_0：p≤ 0.6，相应的对立假设为H_1：$p>0.6$ 选 B

三、（**10** 分） $X_1,X_2,\text{L}\ ,X_{100}$ 是独立同分布的随机序列，其概率密度函数为

$$f(x)=\begin{cases}x & 0\le x<1\\ 2-x & 1\le x<2\\ 0 & x\ge 2\ or\ x<0\end{cases}$$

记 $\bar{X}=\dfrac{1}{100}\sum_{i=1}^{100}X_i$， 求 $P\left\{\left|\bar{X}\right|\le 1.1\right\}$

（查表，补充表值!）

分析：本题考察独立同分布中心极限定理的灵活掌握。

由 $X_1, X_2, \cdots, X_{100}$ 独立同分布可知，

$$EX_i = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 x^2 dx + \int_1^2 x(2-x)dx = \left.\frac{x^3}{3}\right|_0^1 + \left.x^2\right|_1^2 - \left.\frac{x^3}{3}\right|_1^2 = 1$$

$$E(X_i^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_0^1 x^3 dx + \int_1^2 x^2(2-x)dx = \left.\frac{x^4}{4}\right|_0^1 + \left.\left(\frac{2x^3}{3} - \frac{x^4}{4}\right)\right|_1^2 = \frac{7}{6}$$

因此，$DX_i = \frac{7}{6} - 1 = \frac{1}{6}$，方差存在且有界，近刺服从列维-林德伯格中心极限定理，于是有

$$E\bar{X} = E\left(\frac{1}{100}\sum_{i=1}^{100} X_i\right) = \frac{1}{100}\sum_{i=1}^{100} EX_i = 1$$

$$D\bar{X} = D\left(\frac{1}{100}\sum_{i=1}^{100} X_i\right) = \frac{1}{10000}\sum_{i=1}^{100} DX_i = \frac{1}{600}$$

$\bar{X}$ 近似服从于正态分布 $N(1, \frac{1}{600})$，因此

$$P\left\{\left|\bar{X}\right| \le 1.1\right\} = P\left\{\frac{-1.1-1}{\frac{1}{10\sqrt{6}}} \le \frac{\left|\bar{X}\right|-1}{\frac{1}{10\sqrt{6}}} \le \frac{1.1-1}{\frac{1}{10\sqrt{6}}}\right\}$$
$$\approx \Phi(\sqrt{6}) - 1 + \Phi(21\sqrt{6}) \approx \Phi(2.45) = 0.9929$$

四、（12 分）设总体 X 的一个样本为 $(X_1, X_2, \cdots X_n)$，X 的分布密度为 $f(x) = \begin{cases} \frac{2x}{\theta^2}, & 0 \le x \le \theta \\ 0, & 其它 \end{cases}$ 参数 $\theta > 0$。

（1）求 θ 的矩估计量；

（2）求 θ 的最大似然估计量。

解：（1）$\mu_1 = E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{\theta} x\frac{2x}{\theta^2}dx = \frac{2}{3}\theta, A_1 = \bar{X}$，即 $\frac{2\theta}{3} = \bar{X}$

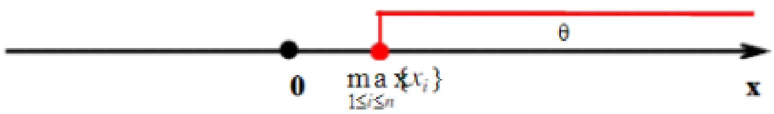
所以矩估计量 $\hat{\theta} = \frac{3}{2}\bar{X}$

（2）似然函数 $L(\theta) = \prod_{i=1}^{n}\frac{2x_i}{\theta^2}, 0 \le x_x \le \theta, i = 1,2,\cdots,n$

$$LnL(\theta) = \sum_{i=1}^{n}\left[\ln 2 + \ln x_i - 2\ln\theta\right], \frac{d}{d\theta}\ln L(\theta) = \sum_{i=1}^{n}\left[0 - \frac{2}{\theta}\right] = -\frac{2n}{\theta} < 0$$

$LnL(\theta)$ 单调递减，故 θ 取最小可能值时，$LnL(\theta)$ 有最大值，从而 $L(\theta)$ 有最大值。

注意到 $0 \le x_i \le \theta, i = 1,2,\Lambda\ n$ 等价于 $\theta \ge \max\limits_{1\le i\le n}\{x_i\} \ge 0$，如图



所以 θ 的最大似然估计值为 $\hat{\theta} = \max\limits_{1\le i\le n}\{x_i\}$，最大似然估计量为 $\hat{\theta} = \max\limits_{1\le i\le n}\{X_i\}$

五、(12 分）某公司雇用 3000 名推销员，为了发放外出补贴，需要估计推销员每年的平均乘车里程。从过去的经验可知，通常每位推销员乘车里程标准差为 4000 公里，随机选取 16 名推销员，得到他们的年平均里程为 12000 公里。

（1）确定总体均值 μ 的 95%的置信区间（已知 $z_{0.025} = 1.96$）

（2）公司经理们认为均值应介于 11000 到 13000 公里之间，如果希望该估计有 95%的置信水平，这时所要求的样本容量是多少？

（1）16 个样本，标准差为 4000，代入

$$\left[12000 \pm z_{\frac{1-\alpha}{2}} \frac{4000}{\sqrt{16}}\right] = [10040, 13960]$$

（2）要使均值处于 11000 到 13000 之间的概率为 0.95，则

$$\alpha = P\left(\frac{11000-12000}{4000\big/\sqrt{n}} < \frac{\hat{\mu}-12000}{4000/\sqrt{n}} < \frac{13000-12000}{4000/\sqrt{n}}\right) = 0.95$$

即 $\dfrac{13000-12000}{4000/\sqrt{n}} = 1.96$

解得 $n = 61.47$，

所以样本容量应为 62

六、(10 分）甲乙两台机床加工相同的产品，从这两台机床加工的产品中随机地抽取若干件，测得产品直径为（单位：mm）为

机床甲：20.5　19.8　19.7　20.4　20.1　20.0　19.0　19.9

机床乙：19.7　20.8　20.5　19.8　19.4　20.6　19.2

假定两台机床加工的产品直径都服从正态分布，且总体方差相等，问能否认为甲乙两

台机床加工的产品直径有显著差异（取 α = 0.05）？。

解：依题意，两总体 X 和 Y 分别服从正态分布$X\sim N$（μ_1， σ^2）和 $Y\sim N$（μ_2， σ^2），μ_1，μ_2,σ^2均未知，需要检验假设

H_0：$\mu_1 = \mu_2$；H_1：$\mu_1 \neq \mu_2$

拒绝域为 $|t| = \left|\frac{\bar{X}-\bar{y}}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right| \geq t_{\frac{\alpha}{2}}(n_1+n_2-2)$

$n_1=8$， $\bar{X} = 19.925$ ， $S_1{}^2 = 0.216$,

$n_2=7$， $\bar{y} = 19.925$ ， $S_2{}^2 = 0.397$,

且

$S_w{}^2=\frac{(8-1)S_1{}^2+(7-1)S_2{}^2}{8+7-2} = 0.547$

查表可知$t_{0.025}(13)=2.160$，

$\left|\frac{\bar{X}-\bar{y}}{S_w\sqrt{\frac{1}{8}+\frac{1}{7}}}\right| = 0.265 < 2.160$.所以接受H_0，即证明两台机床加工地产品直径无显著差异。

七、(**10** 分）甲乙两地相邻地段各取了 8 块和 9 块岩心进行磁化率测定，算出两样本标准差分别是 $s_1^2 = 0.0139$， $s_2^2 =0.0053$，问甲乙两段地标准差是否有显著性差异（$\alpha=0.05$）？

解析：做假设 H_0：$\sigma = \sigma_0$ ， 由题设有

$\frac{1}{8-1}\sum_{i=1}^{n}(X_j - \bar{X})^2 = \frac{8*S_1{}^2}{7} = 0.0159$,

$\frac{1}{9-1}\sum_{i=1}^{n}(Y_j - \bar{Y})^2 = \frac{9*S_2{}^2}{8} = 0.0060$;

从而统计量 $F=\frac{\frac{n_1S_1{}^2}{n_1-1}}{\frac{n_2S_2{}^2}{n_2-1}} = \frac{0.0159}{0.0060} = 2.65$,当 α = 0.05 ，查 F 分布表可得

$F_{\frac{\alpha}{2}}(7,8) = F_{0.025}(7,8) = 4.53$

$F_{1-\frac{\alpha}{2}}$ （7，8） = $F_{0.975}$ (7,8) = 0.5698

因为 F= 2.65 ＜ $F_{0.025}$ (7,8) = 4.53，故接受原假设H_0，即认为甲乙两段地标准差没有显著性差异。

八、（10 分）设 $X_1, X_2, \cdots X_9$ 是来自正态总体 X 的简单随机样本，

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

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

证明：统计量 Z 服从自由度为 2 的 t 分布

【分析】 要证 Z 服从自由度为 2 的 t 分布，根据定义应将 Z 表示为 $Z=\frac{U}{\sqrt{V/2}}$，其中 $U\sim N(0,1)$，$V\sim\chi^2(2)$. Z 的分子部分是正态随机变量的线性组合，仍服从正态分布，标准化后可得标准正态分布；而 Z 的分母部分的平方 S^2，是样本方差，乘以适当的常数后服从 χ^2 分布. 经过整理后，即可得 t 分布的定义形式.

【详解】 设 $X\sim N(\mu,\sigma^2)$，则有

$$E(Y_1)=E(Y_2)=\mu,\quad D(Y_1)=\frac{\sigma^2}{6},\quad D(Y_2)=\frac{\sigma^2}{3}.$$

由于 Y_1 和 Y_2 独立，因此有

$$E(Y_1-Y_2)=0,\quad D(Y_1-Y_2)=\frac{\sigma^2}{6}+\frac{\sigma^2}{3}=\frac{\sigma^2}{2},\quad Y_1-Y_2\sim N(0,\frac{\sigma^2}{2}).$$

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 $V=\frac{2S^2}{\sigma^2}\sim\chi^2(2)$.

又 Y_1-Y_2 与 S^2 独立，因此 $\frac{U}{\sqrt{V/2}}=\frac{\sqrt{2}(Y_1-Y_2)}{S}=Z\sim t(2)$.