说明: 本习题集包含概率论与数理统计(A), 概率论与数理统计(B), 概率论3门课程的习题, 分别用PSA, PS, PB作为代号, 可以通过下方目录或书签链接跳转, 命名规则为 "代号(PS/PSA/PB).年份.学期.期中(M)/期末(F)", 部分参考答案有误, 请留意.

目录

中国科学技术大学

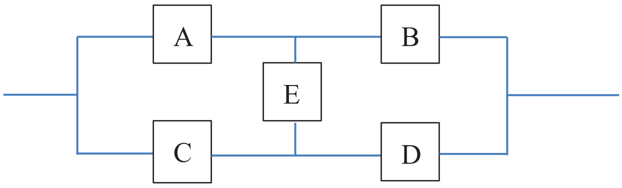
2002—2003 学年第二学期考试试卷

考试科目：概率论与数理统计　　　　　　得　分：____________

学生所在系：____________姓　名______________学　号：____________

（考期：2003 年 6 月 30 日，闭卷，可用计算器）

一、考虑如图所示的电路图：



其中开关 A、B、C、D、E 是独立工作的，每个开关以概率 p 开着，以概率 $q=1-p$ 关着，求一个输入的信号在输出处被接收到的概率；如果一个信号被接收到，那么开关 E 是开着的条件概率是多少？

二、设(X,Y)的联合密度函数为：

$$f(x,y)=\begin{cases}c, |y|<x, 0<x<1;\\ 0, others\end{cases}$$

求 (1) 常数 C 的值；(2) 条件密度函数$f_{X|Y}(x|y)$及$f_{Y|X}(y|x)$；(3) 讨论 X 与 Y 的独立性和相关性。

三、在一家保险公司里有 10000 个人参加保险，每人每年付 12 元保险费，在一年内一个人死亡的概率为 0.006，死亡时其家属可向保险公司领取 1000 元的保险金，问：
(1) 保险公司亏本的概率多大？
(2) 保险公司一年的利润不少于 40000 元、60000 元的概率各多大？

四、设$(X_1,X_2,\cdots,X_n)$是从总体 X 中抽取的一个简单随机样本，已知 X 的概率密度函数为：

$$f(x)=\begin{cases}e^{-(x-\theta)}, x>\theta;\\ 0, others\end{cases}$$

其中θ是未知参数，$-\infty<\theta<\infty$。
(1) 试求θ的极大似然估计$\hat{\theta}$和矩估计$\tilde{\theta}$；
(2) 求常数c_1和c_2，使得$c_1\hat{\theta}-c_2$为θ的无偏估计；
(3) 求常数c_3和c_4，使得$c_3\tilde{\theta}-c_4$为θ的无偏估计；
(4) 在均方误差意义下比较这两个无偏估计哪个更优。(注：上述常数可与 n 有关）

五、据信有一种疾病会导致病人的白细胞数目较常人少，假设正常人白细胞数服从均值为 7250（单位：个/立方毫米，下同）的正态分布，现有 16 个病人，其白细胞的样本均值为 4767，样本标准差为 3204，根据这批数据能否认为这种疾病使白细胞数

目减少？（显著性水平为$\alpha = 0.05$）

自由度为 n 的 t 分布的 p 分位数表

n \ p	0.90	0.95	0.975	0.99
15	1.341	1.753	2.131	2.602
16	1.337	1.746	2.120	2.583

六、在[0,1]区间上随机独立地投掷两点，设 X 与 Y 分别表示这两点的坐标，试求这两点间距离的概率密度函数、数学期望和方差。

中国科学技术大学

2003—2004 学年第一学期考试试卷

考试科目：概率论与数理统计　　　　　　得 分：__________

学生所在系：__________姓 名____________学 号：__________

（考期：2004 年 1 月 8 日，闭卷，可用计算器）

一、甲、乙、丙三人独立地向靶子各射击一次，其命中率分别为 0.6、0.5 和 0.4.现已知恰有两人命中靶子，问：

（1）此两人中包括丙的可能性大，还是不包括丙的可能性大？

（2）此两人中包括乙的可能性大，还是包括丙的可能性大？（要求写出计算过程）

二、某种商品一周的需求量是个随机变量，其概率密度为：

$$f(t)=\begin{cases} te^{-t}, t>0 \\ 0, \quad t\le 0 \end{cases}$$

各周的需求量相互独立，试求：

（1） 两周需求量的概率密度；

（2） 三周需求量的概率密度。

三、利用中心极限定理求解：

（1）设计算机在进行加法运算时，每次取整的误差相互独立，且服从[-0.5,0.5]上的均匀分布，若要保证误差总和的绝对值不超过 20 的概率大于或者等于 0.95，问至多只能进行多少次加法运算？

（2） $lim_{n\to\infty}\sum_{k=0}^{n}\frac{n^k}{k!}e^{-n}=?$

四、设样本$(X_1,X_2,\cdots,X_n)$抽自总体$X\sim f(x;\theta)$，其中：

$$f(x;\theta)=\frac{1}{2}e^{-\frac{x-\theta}{2}},(x>\theta;\theta\in R)$$

（1） 试求θ的矩估计$\hat{\theta}$和极大似然估计θ^*；

（2） 验证$\hat{\theta}$和θ^*是否为θ的无偏估计，若不是无偏估计，试将其分别修正为无偏估计$\hat{\theta}_1$和$\hat{\theta}_2$；

（3） 比较$\hat{\theta}_1$和$\hat{\theta}_2$何者为优？

五、为考察钢铁工人和电厂工人平均工资的差别，从两厂各抽取若干工人调查，结果如下：

钢厂：74，65，72，69（元）

电厂：75，78，74，76，72（元）

若钢厂工人与电厂工人工资分别服从正态分布$N(\mu_1,\sigma_1^2)$与$N(\mu_2,\sigma_2^2)$，总体独立且均值方差未知，试据上述数据判断：

（1） 是否可以认为$\sigma_1^2=\sigma_2^2$？（$\alpha=0.05$）

（2） 钢铁工人平均工资是否低于电厂工人平均工资？（$\alpha=0.05$）

中国科学技术大学

2003—2004 学年第二学期考试试卷

考试科目：概率论与数理统计　　　　　　得　分：＿＿＿＿＿＿

学生所在系：＿＿＿＿＿＿姓　名＿＿＿＿＿＿＿学　号：＿＿＿＿＿＿

（考期：2004 年 6 月 25 日，闭卷，可用计算器）

一、判断和填空：

（1）　设 P(A)=0，则 A 为不可能事件。

（2）　设(X,Y)服从二元正态，Cov(X,Y)=0,则 X、Y 相互独立。

（3）　设 X、Y 相互独立，则 X、Y 的联合分布可以由 X 和 Y 的边缘分布唯一确定。

（4）　设$X_1,\cdots,X_n$为从同一个总体中抽取的一个样本，则$max(X_1,\cdots,X_n)-min(X_1,\cdots,X_n)$+3 是统计量。

（5）　设$\theta>0$，X 的概率分布函数为：

$$F(x)=\begin{cases}1-exp\left\{-\dfrac{x-\mu}{\theta}\right\}, & x\geq\mu\\ 0 & ,x<\mu\end{cases}$$

则随机变量 X 的密度函数为（）。

（6）　设 X、Y 服从单位圆$x^2+y^2\leq 1$上的均匀分布，则在给定 Y=0.5 条件下的 X 的条件密度函数为（）。

（7）　设 X 和 Y 相互独立，它们的均值全为 0，方差全为 1，记 V=X-Y，则 X 与 V 的相关系数为（）。

二、求：（1）P(Y=2|X=1)；(2)X^2+Y^2的分布，其中 X、Y 的联合分布如下：

X \ Y	-1	0	1	2
-1	0.12	0.08	0.30	0.15
1	0.08	0.22	0	0.05

三、设 X 服从期望为 2 的指数分布，Y 服从(0,1)上的均匀分布，且 X 与 Y 相互独立，求：（1）X-Y 的概率密度函数；（2）P(X<Y)。

四、桌上有三个盒子，在甲盒中装有 2 支红芯圆珠笔，4 支蓝芯圆珠笔，乙盒中装有 4 支红芯圆珠笔，2 支蓝芯圆珠笔，丙盒中装有 3 支红芯圆珠笔，3 支蓝芯圆珠笔，今从三个盒子中任取一支笔，设甲乙丙三盒取笔的概率相等。试求：

（1）取得红笔的概率；（2）在已知取得红笔的条件下，问笔从哪个盒子中取出的概率最大？

五、某工厂生产线甲根据专利生产灯泡，生产线乙根据本厂原有技术生产。现分别在生产线甲和乙两条生产线各抽取 8 个灯泡，测得其寿命分别为（千小时）：

对生产线甲：10，9，3，11，5，7，9，11；

对生产线乙：4，9，6，5，3，5，7，7；

设灯泡寿命服从正态分布，且方差相等。试分别在显著性水平$\alpha=0.05$和$\alpha=0.01$下检验生产线甲的灯泡是否比生产线乙生产的寿命要长。

六、设总体 X 服从$(1,\theta+1)$上的均匀分布，$X_1,\cdots,X_n$为总体 X 中抽取的一个样本。试求：

（1）求θ的矩估计$\hat{\theta}_1$和极大似然估计$\hat{\theta}_2$；

（2）$\hat{\theta}_1$和$\hat{\theta}_2$是否为θ的无偏估计，若不是，请加以修正；

（3）$\hat{\theta}_3=2\hat{\theta}_4-2$是$\theta$的无偏估计，其中$\hat{\theta}_4=\frac{2X_1+X_{2+}\cdots+X_{n-1}+2X_n}{n+2}$，问$\hat{\theta}_1$的修正（如果需要修正的话）和$\hat{\theta}_3$哪个更有效？

中国科学技术大学

2004—2005 学年第一学期考试试卷

考试科目：概率论与数理统计　　　　　　　　得　分：___________

学生所在系：___________姓　名_____________学　号：___________

（考期：2005 年 1 月 20 日，闭卷，可用计算器）

一、甲、乙、丙三门火炮同时独立地向目标射击，其命中率分别为 0.2，0.3 和 0.5。目标被命中一发而被摧毁的概率为 0.2，被命中两发而被摧毁的概率为 0.6，被命中三发而被摧毁的概率 0.9，试求：

（1）三门火炮在一次射击中摧毁目标的概率；

（2）在目标被摧毁的条件下，其只由甲火炮击中的概率。

二、设 X 与 Y 独立同分布，都服从参数为λ的指数分布，试求 Z 的分布密度，其中：

（1）Z=min{X,Y}；　　　（2）Z=X+Y。

三、将一枚骰子独立地投掷 n 次，令 X 与 Y 分别表示其 1 点出现的次数和 6 点出现的次数，并记 Z=n-X。试求：

（1）X 与 Y 的协方差及相关系数；

（2）X 与 Z 的相关系数。

四、设样本$X_1,\cdots,X_n$抽自总体 X，总体的密度为：

$$X\sim f(x;\theta_1)=\begin{cases}\frac{1}{\theta_2}e^{-\frac{x-\theta_1}{\theta_2}}, & x\geq\theta_1\\ 0, & x<\theta_1\end{cases}$$，其中$\theta_1\in R$为未知参数，$\theta_2>0$为已知数。

（1）　求θ_1的矩估计$\hat{\theta}_1$和极大似然估计θ_1^*；

（2）　$\hat{\theta}_1$和θ_1^*是否为θ_1的无偏估计？是加以证明，不是请加以修正为无偏估计量。

五、某校组织学生参加英文词汇训练，并在年初与年底（即训练前与训后）各举行一次阅读考试，以考察训练的效果。现随机抽取 10 名同学，将其年初与年底的考试成绩记录如下：

学生	1	2	3	4	5	6	7	8	9	10
年初成绩	64	43	84	72	52	93	77	58	69	91
年底成绩	72	50	86	80	50	90	78	57	72	95

假定两次考分之差服从正态分布，试由此判断词汇训练是否有显著效果？（分别在$\alpha=0.05$与$\alpha=0.01$的水平下检验）

六、为了研究色盲是否与性别有关，随机抽取 1000 人进行调查，结果如下：

	男	女	和
正常	442	514	956

色盲	38	6	44
和	480	520	1000

（1） 试据此判断，色盲是否与性别有关？$(\alpha = 0.01)$

（2） 你认为是男性还是女性更容易患色盲？请说明理由。

中国科学技术大学

2005—2006 学年第一学期考试试卷

考试科目：概率论与数理统计　　　　　　　　得　分：___________

学生所在系：___________姓　名_____________学　号：___________

（考期：2006 年 1 月 22 日，闭卷，可用计算器）

一、设昆虫产卵个数服从参数为λ的 Possion 分布，而每个卵孵化成幼虫的概率为 p，且各卵是否成虫彼此之间没有关系。试求：

（1）一个昆虫产生 k 个后代的概率；

（2）若某个昆虫产生 k 个后代，求它产生 m 个卵的概率。

二、设二维随机变量(X,Y)的联合密度为：

$$f(x,y)=\begin{cases}0.25(1+xy), |x|<1, |y|<1\\ 0 \quad , others\end{cases}$$

（1）　求给定 X=1/2 时 Y 的条件概率密度；

（2）　求 Cov(X,Y)和 Var(Y|X=1/2)；

（3）　证明X^2与Y^2独立。

三、设某学校有 5000 名学生，在某一时间区间内每个学生去某个阅览室的概率为 0.05，且设每个学生是否去该阅览室是相互独立的。试问该阅览室至少需要设多少座位才能以 95%的概率保证每个到该阅览室来的同学均有座位？

四、设从总体

X	0	1	2	3
P	θ/2	θ	3θ/2	1-3θ

抽取的一个简单随机样本$X_1,\cdots,X_{10}$的观测值为(0,3,1,1,0,2,0,0,3,0)。

（1）求 θ的矩估计量$\hat{\theta}_M$和极大似然估计量$\hat{\theta}_L$；

（2）证明上述估计量都是无偏估计量；

（3）比较这两个估计量，指出哪个更有效。

五、假设某台精盐包装机生产的袋装盐的净重服从正态分布，按照要求每袋盐的标准重量为 500g，标准差不得超过 10g。某天开工后，从装好的盐中随机抽取 10 袋，测得其净重（单位：g）为：510，495，478，487，501，493，528，504，503，504。试据此判断这时机器的工作是否正常。（$\alpha=0.05$）

六、在著名的豌豆实验中，孟德尔（1822-1884）同时考虑豌豆的颜色和形状，共有四种组合：（黄、圆），（黄、皱），（绿、圆），（绿、皱）。按孟德尔的理论，这四类应该有 9：3：3：1 的比例。在一次实验中，发现这四类的观察数分别为 315，101，108 和 32.试据此判断孟德尔的理论是否正确？（$\alpha=0.05$）

中国科学技术大学

2005—2006 学年第二学期考试试卷

考试科目：概率论与数理统计　　　　　　得　分：____________

学生所在系：____________姓　名______________学　号：____________

（考期：2006 年 7 月 3 日，闭卷，可用计算器）

一、在空战中甲机先向乙机开火，击落乙机的概率为 0.2；若乙机未被击落，就进行还击，击落甲机的概率为 0.3；若甲机未被击落，则再进攻乙机，击落乙机的概率为 0.4.试求在这三回合中：

（1）乙机被击落的概率是多少？

（2）若乙机被击落，则它在第一回合中被击落的概率是多少？

二、设$X_1 \sim \begin{pmatrix} -1, & 0, & 1 \\ 0.25, & 0.5, & 0.25 \end{pmatrix}$，$X_2 \sim \begin{pmatrix} 0, & 1 \\ 0.5, & 0.5 \end{pmatrix}$，且$P\{X_1X_2 = 0\} = 1$。试求：

（1）试求(X_1, X_2)的分布；

（2）X_1与X_2是否独立？为什么？

（3）X_1与X_2是否不相关？为什么？

三、设 X 与 Y 相互独立，都服从指数分布，参数分别为λ与$\mu(\lambda \neq \mu)$，试求 Z 的概率密度$f_Z(z)$，其中：（1）Z=X+Y；（2）Z=X-Y。

四、设样本$X_1, \cdots, X_n$抽自总体 X，X 服从$(\theta, \theta + 1)$上的均匀分布：

（1）试求θ的矩估计$\hat{\theta}$和极大似然估计θ^*；

（2）证明$\hat{\theta}_1 = \bar{X} - \frac{1}{2}$与$\hat{\theta}_2 = X_{(n)} - \frac{n}{n+1}$均为$\theta$的无偏估计；

（3）$\hat{\theta}_1$和$\hat{\theta}_2$哪个更有效？

五、设样本$X_1, \cdots, X_n$抽自正态总体$N(\mu, \sigma^2)$，问在下列三个统计量中：

$$S_1^2 = \frac{1}{n-1}\sum_1^n (X_i - \bar{X})^2 \text{ ，} S_2^2 = \frac{1}{n}\sum_1^n (X_i - \bar{X})^2 \text{，} S_3^2 = \frac{1}{n+1}\sum_1^n (X_i - \bar{X})^2$$

谁是σ^2的无偏估计？谁对σ^2的均方误差$E\left(S_i^2 - \sigma^2\right)^2$最小？请证明你的结论。

六、某校组织学生参加英文词汇训练，并在年初与年底（即训练前与训后）各举行一次阅读考试，以考察训练的效果。现随机抽取 10 名同学，将其年初与年底的考试成绩记录如下：

（1）假定两次考分之差服从正态分布，试由此判断词汇训练是否有显著效果？（在$\alpha = 0.05$的水平下检验）

（2）若上述两组数据并非抽自相同的 10 名同学，而是分别从两次考分中各随机抽取 10

人，并假定两次考分分别服从正态分布（二总体独立），方差未知但相等，试据以判断词汇训练是否有显著效果？（在$\alpha = 0.05$的水平下检验）

学生	1	2	3	4	5	6	7	8	9	10
年初成绩	64	43	84	72	52	93	77	58	69	91
年底成绩	72	50	86	80	50	90	78	57	72	95

参考答案

一、（1）0.2+0.8*0.7*0.4=0.424

（2）0.2/0.424=0.472

二、（1）略；（2）不独立；（3）不相关

三、（1）$f_Z(z) = \frac{\lambda\mu}{\lambda-\mu}\left(e^{-\mu z} - e^{-\lambda z}\right), z \geq 0$；

(2) $f_Z(z) = \begin{cases} \frac{\lambda\mu}{\lambda+\mu} e^{-\lambda z}, z \geq 0 \\ \frac{\lambda\mu}{\lambda+\mu} e^{\mu z}, z < 0 \end{cases}$

四、（1）$\hat{\theta} = \bar{X} - \frac{1}{2}$　　$\theta^* \in \left[X_{(n)} - 1, X_{(1)}\right]$

（2）略

（3）$\mathrm{Var}(\hat{\theta}_1)=\frac{1}{12n}$　　$\mathrm{Var}\left(\hat{\theta}_2\right) = \frac{n}{(n+1)^2(n+2)}$　　$n \leq 7, \hat{\theta}_1$有效；$n \geq 8, \hat{\theta}_2$有效

五、（1）$S_1^2 = \frac{1}{n-1}\sum_1^n (X_i - \bar{X})^2$为无偏估计量；（2）均方误差排序$S_3^2 < S_2^2 < S_1^2$

六、（1）成对数据检验，拒绝原假设；（2）两样本 t 检验，无法拒绝原假设。

中国科学技术大学

2006—2007 学年第一学期考试试卷

考试科目：概率论与数理统计　　　　　　得 分：＿＿＿＿＿＿

学生所在系：＿＿＿＿＿＿姓 名＿＿＿＿＿＿＿学 号：＿＿＿＿＿＿

（考期：2007 年 1 月 31 日，闭卷，可用计算器）

一、有 12 个新的兵乓球，每次比赛时取出 3 个，用完之后再放回去。

（1）设第二次比赛时取到 X 个新球，试求 X 的分布律；

（2）若第三次比赛时取到 3 个新球，问第二次比赛时取出的 3 个球都是新球的概率是多少？

二、设 X 与 Y 独立，都服从指数分布，参数分别为λ 与 μ$(\lambda \neq \mu)$，试求 Z=X+Y 的分布密度$f_Z(z)$。

三、设 Y 服从参数为μ 与σ^2的对数正态分布（即 Y 满足：lnY~N(μ, σ^2)），试求 Y 的分布密度$f_Y(y)$及 E(Y)与 Var(Y)。

四、某蛋糕店出售三种生日蛋糕，单价分别为 12 元、20 元和 40 元，售出这三种蛋糕的概率分别为 0.3，0.2 和 0.5。某日该店售出 300 个蛋糕，问：

（1）该日总收入超过 8000 元的概率约为多少？

（2）该日售出单价为 20 元的蛋糕超过 60 的概率约为多少？

五、设样本$X_1, \cdots, X_n$抽自总体 X，其中：

$$X \sim f(x;\theta) = \begin{cases} \frac{1}{2} e^{-\frac{x-\theta}{2}}, x \geq \theta \\ 0 \qquad , \ x < \theta \end{cases}$$

(1) 试求θ的矩估计$\hat{\theta}$和极大似然估计θ^*；

(2) 验证$\hat{\theta}$和θ^*是否为θ的无偏估计；若否，试将其修正为无偏估计。

六、假设某台精盐包装机生产的袋装盐的净重服从正态分布，按照要求每袋盐的标准重量为 500g，标准差不得超过 10g。某天开工后，从装好的盐中随机抽取 10 袋，测得其净重（单位：g）为：510，495，478，487，501，493，528，504，503，504。试据此判断这时机器的工作是否正常。（$\alpha = 0.05$）

七、某一作业中可能发生两类事故：A（起火）和 B（爆炸），而该作业有三种不同的原料可供选择：L、M 和 N。下面给出的是事故记录：

	L	M	N	和
A	42	17	29	88
B	20	4	29	53
和	62	21	58	141

试据此判断事故类型是否与原料的种类有关？（$\alpha = 0.05$）

参考答案

一、（1） $P(X=k)=\frac{\binom{9}{k}\binom{3}{3-k}}{\binom{12}{3}}$,k=0,1,2,3

(2) Bayes formula =0.23

二、$f_Z(z)=\frac{\lambda\mu}{\lambda-\mu}\left(e^{-\mu z}-e^{-\lambda z}\right), z>0$

三、$f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma y}e^{-\frac{(lny-\mu)^2}{2\sigma^2}}, y>0$ $\quad$ $E(Y)=e^{\mu+\frac{\sigma^2}{2}}$ $\quad$ $Var(Y)=\left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2}$

四、（1）$E(X)=27.6$ $\quad$ $Var(X)=161.44$ $\quad$ $P(\sum_1^{300}X_i>8000)\approx\Phi(1.27)$

（2）$Y\sim B(300,0.2)$ $\quad$ $P(Y>60)\approx 0.5$

五、（1）$\hat{\theta}=\bar{X}-2$ $\quad$ $\theta^*=X_{(1)}$

(2) $E\hat{\theta}=\theta$ $\quad$ $E\theta^*=\theta+\frac{2}{n}$有偏，修正为$\tilde{\theta}=X_{(1)}-\frac{2}{n}$

中国科学技术大学

2006—2007 学年第二学期考试试卷

考试科目：概率论与数理统计　　　　　　　　　　得　分：___________

学生所在系：___________姓　名_____________学　号：___________

（考期：2007 年 7 月 13 日，闭卷，可用计算器）

一、（18 分）

（1）　举例说明：一般而言，$P(B\mid A)+P(B\mid\overline{A})=1$ 和 $P(B\mid A)+P(\overline{B}\mid\overline{A})=1$ 不成立；

（2）　举例说明：随机变量 X 与 Y 不独立，但 X^2 和 Y^2 独立；

（3）　设 A_1, A_2, A_3, A_4 相互独立，且 $P(A_i)=1/3$，$(i=1,2,3,4)$ 则

$P(\bigcup_{i=1}^{4}A_i)=$（　　　　　）；

（4）设随机变量 X 与 Y 独立，且 $E(X)=E(Y)=0, Var(X)=Var(Y)=1$。若命 $W=X-Y$，则 Y 与 W 的相关系数是（　　　　　）；

（5）判断正误：设 X 与 Y 都是正态随机变量，则 X 与 Y 的联合分布由 X 与 Y 的边缘分布唯一确定（　　　）；

（6）判断正误：在假设检验中，我们要检验两个正态总体均值差 $\mu_1-\mu_2=\delta$ 是否为零，则 $\overline{X}-\overline{Y}-\delta$ 是统计量（　　　）。

二、（10 分）有 100 个零件，其中 90 个为一等品，10 个为二等品。从中随机取出 2 个，安装在一台设备上。若 2 个零件中恰有 k 个二等品 $(k=0,1,2)$，则该设备的使用寿命服从参数为 $\lambda=k+1$ 的指数分布。若已知该设备寿命超过 1，试求安装的 2 个零件均为一等品的概率。

三、（20 分）设 $r.v.X\sim f(x)=6x(1-x),\quad(0\le x\le 1)$

（1）验证 $f(x)$ 是概率密度函数并画出其图形；

（2）求出 X 的概率分布函数；

（3）确定满足 $P(X<b)=P(X>3b/2)$ 的数 b，（$0<b<1$）；

（4）计算 $P\{X\le 1/2\mid 1/3<X<2/3\}$。

四、（7 分）设(X,Y)服从$D=\{(x,y)\mid -1\le x\le 1,0\le y\le 1\}$上的均匀分布，试求

$Z=\dfrac{Y}{3X}$的概率密度函数$f_Z(z)$。

五、（30 分）设样本$\left(X_1,X_2,\cdots,X_n\right)$抽自总体$X$，$X$服从三点分布：

$$P(X=-1)=p, P(X=0)=1-3p, P(X=1)=2p$$

（1） 试分别用样本一阶和二阶原点矩来估计未知参数p；

（2） 证明这两个估计都是无偏估计；

（3） 问这两个无偏估计，哪个更有效（即哪个方差更小）？

六、（15 分）为了解甲、乙二企业职工工资水平，分别从二企业各随机抽取若干名职工调查，得如下数据（单位：元）：

甲企业： 750，1060，750，1820，1140，1050，1000

乙企业：1000，1900，900，1800，1200，1700，1950，1200

设二企业职工工资分别服从正态分布$N(\mu_1,\sigma^2)$和$N(\mu_2,\sigma^2)$，二总体独立且均值、方差皆未知。试根据以上数据判断：甲企业职工平均工资是否低于乙企业职工平均工资？（分别在$\alpha=0.05$和$\alpha=0.01$两种水平下检验）

（完）

（ 参考数据：t分布上侧分位点$t_\alpha(n)$

α \ n	13	14	15
0.005	3.0123	2.9769	2.9467
0.01	2.6503	2.6245	2.6025
0.025	2.1604	2.1448	2.1315
0.05	1.7709	1.7613	1.7531

）

概率统计期末考题解答与评分标准

（2007 年 7 月 13 日考试）

一、（18 分）

（1）例如取：$\Omega=\{1,2,3,4,5,6\}, A=\{1,3,5\}, B=\{3,5\}$；

（2）如：$P\{X=-1\}=1-p, P\{X=1\}=p, 0<p<1, Y$ 为任意随机变量；

（3）$P(\bigcup_{i=1}^{4}A_i)=1-P(\bigcap_{i=1}^{4}\overline{A_i})=1-(2/3)^4=65/81$；

（4）$-1/2$；（5）误；（6）误。

二、（10 分）$89e^2/(89e^2+20e+1)$。

三、（20 分）：

（1）$\int_0^1 6x(1-x)dx=1$；（2）$F(x)=\begin{cases}1, & x>1\\ 3x^2-2x^3, & 0\le x\le 1\\ 0, & x<0\end{cases}$；（3）$b=2/5$；（4）$1/2$。

四、（7 分）：

$$f_Z(z)=\begin{cases}1/(12z^2), & |z|>1/3\\ 3/4, & |z|\le 1/3\end{cases}。$$

五、（30 分）：

（1）$\hat{p}_1=\overline{X}, \hat{p}_2=(1/3)\overline{X^2}$；（2）$E(\hat{p}_1)=E(\hat{p}_2)=p$；

（3）$Var(\hat{p}_1)=\dfrac{p(3-p)}{n}, Var(\hat{p}_2)=\dfrac{p(\frac{1}{3}-p)}{n}, (0<p<1/3)$，故 $\hat{p}_2$ 更有效。

六、（15 分）：

$H_0:\mu_1\ge\mu_2 \Leftrightarrow H_1:\mu_1<\mu_2$，算得：$\overline{x}\approx 1081.43, \overline{y}=1456.25, S_T\approx 396.5111$，代入计算统计量值得：

$$\frac{\overline{x}-\overline{y}}{S_T\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\approx -1.8265<-1.7709=-t_{0.05}(13), 拒绝H_0；$$

$$\frac{\overline{x}-\overline{y}}{S_T\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\approx -1.8265>-2.6503=-t_{0.01}(13), 无法拒绝H_0。$$

中国科学技术大学

2007—2008 学年第一学期考试试卷

考试科目：概率论与数理统计　　　　　　　得 分：___________

学生所在系：___________姓 名_____________学 号：___________

（考期：2008 年 1 月 22 日，闭卷，可用计算器）

一、（15 分）一串 0,1 数字（独立同分布）组成的序列中 1 的概率 p 代表了某种有用的信息，由于某种原因需要对其保密。现对该串数字进行随机加密，对序列中的每一个数字抛一枚硬币（每次正面出现的概率为 π），若抛出的为正面，则原序列的数字不变，若抛出的为反面，则原序列中相应的数字由 x 变成 $1-x$（即 0 变成 1，1 变成 0）。加密后的序列可以公布，其中 1 的概率 p^* 可以估计出来。若知道 π 的值，就可以从加密后的序列中的 1 的频率为 p^* 计算出原序列的 p，所以 π 称为“密钥”。

（1） 现已知 $p^*=0.7$，如果“密钥” $\pi=0.4$，试求 p；

（2） 试说明为什么均匀硬币（$\pi=0.5$）不适合用来加密。

二、（15 分）设随机变量 X 满足：$|X|\le 1$，$P(X=-1)=1/8$，$P(X=1)=1/4$，而且，X 在 $(-1,\ 1)$ 内任一子区间上取值的概率与该子区间的长度成正比。试求：

（1） X 的概率分布函数 $F(x)=P(X\le x)$；

（2） X 取负值的概率；　　（3） X 的数学期望 $E(X)$。

三、（20 分）二维随机变量 (X,Y) 的密度函数为：

$$f(x,y)=\begin{cases} Ae^{-(3x+4y)}, & (x>0,y>0) \\ 0, & \text{其他} \end{cases}$$

（1）试求系数 $A=?$；（2） X 与 Y 是否独立？

（3）试求 $Z=X+Y$ 的密度函数 $f_Z(z)$；

（4）试求 $Var\left(X\mid X+Y=1\right)$。

四、（20 分）设样本$(X_1, X_2, \cdots, X_n)$抽自正态总体$X \sim N(\mu\,,\,1)$，　μ为未知参数

（1）　试求$\theta = P(X \geq 2)$的极大似然估计θ^*（结果可用$\Phi(.)$的形式表示）；

（2）　写出μ的$(1-\alpha)$置信区间，并求θ的$(1-\alpha)$置信区间。

五、（15 分）为考查A, B两种制鞋材料的耐磨性，用它们制作了 10 双鞋，其中每双鞋的两只鞋分别用A和B两种材料制作(左、右脚两只鞋随机地采用A或B)。10 个男孩试穿这 10 双鞋之后的磨损情况如下表所示（数字代表磨损程度），假定A, B两组数据的差服从正态分布，问是否可以认为这两种材料的耐磨性无显著差异？（$\alpha = 0.05$）

男孩	1	2	3	4	5	6	7	8	9	10
A	13.2	8.2	10.9	14.3	10.7	6.6	9.5	10.8	8.8	13.3
B	14.0	8.8	11.2	14.2	11.8	6.4	9.8	11.3	9.3	13.6
差	-0.8	-0.6	-0.3	0.1	-1.1	0.2	-0.3	-0.5	-0.5	-0.3

六、（15 分）投资者感兴趣的一个问题，是上市公司股票价格的变化与其公司总部所在地是否有关。下表给出的是美国两个不同地区（公司总部所在地）的上市公司在 1998 年第三季度内股价变化情况。表格内的数字是相应的上市公司的个数。问股票价格的变化是否存在地区间的差异？（$\alpha = 0.05$）

股价变化 / 总部所在地	上升	不变	下降
新英格兰地区	100	7	561
西北地区	88	10	370

（完）

（参考数值：　$\chi^2_{0.025}(2) = 7.3778$；$\chi^2_{0.05}(2) = 5.9915$；

$\chi^2_{0.025}(6) = 14.4494$；$\chi^2_{0.05}(6) = 12.5916$；　$t_{0.025}(9) = 2.2622$；

$t_{0.05}(9) = 1.8331$；　$t_{0.025}(10) = 2.2281$；　$t_{0.05}(10) = 1.8125$。）

概率统计期末考试（2008 年 1 月 22 日）

（参考答案与评分标准）

一、（15 分）

（1） $p^* = p\pi + (1-p)(1-\pi)$， $p = (p^* - 1 + \pi)\big/(2\pi - 1)$，当 $p^* = 0.55$, $\pi = 0.4$ 时，

$p = 0.25$；

（2）当 $\pi = 0.5$ 时， $p^* \equiv 0.5$，由此无法解出 p 。

二、（15 分）

（1） $F(x) = \begin{cases} 1, & x \geq 1 \\ \frac{5x+7}{16}, & -1 < x < 1 \\ \frac{1}{8}, & x = -1 \\ 0, & x < -1 \end{cases}$；（2） $= F(0) = \dfrac{7}{16}$；（3） $E(X) = \dfrac{1}{8}$ 。

三、（20 分）

（1） $A = 12$；（2）独立；（3） $f_Z(z) = 12(e^{-3z} - e^{-4z})$, $(z > 0)$；

（4） $f_{X|Z}(x \mid Z = 1) = \dfrac{e^x}{e-1}$ ， $(0 < x < 1)$； $E(X \mid Z = 1) = \dfrac{1}{e-1}$ 。

四、（20 分）

（1） $\theta^* = 1 - \Phi(2 - \overline{X})$；（2） $\mu \in \overline{X} \pm u_{\frac{\alpha}{2}} \dfrac{1}{\sqrt{n}}$； $\theta \in \Phi(\overline{X} - 2 \pm u_{\frac{\alpha}{2}} \dfrac{1}{\sqrt{n}})$ 。

五、（15 分）

$$H_0 : \mu_Z = 0 \leftrightarrow H_1 : \mu_Z \neq 0$$

$\dfrac{|\overline{Z}|}{S_Z\big/\sqrt{n}} = \dfrac{|-0.41|}{0.3872\big/\sqrt{10}} \approx 3.3485 > 2.2622 = t_{0.025}(9)$，拒绝 H_0，有显著差异。

六、（15 分）

$Z \approx 5.4437 < 5.9915 = \chi^2_{0.05}(2)$，无法拒绝 H_0，未见有显著差异。

（完）

中国科学技术大学

2008—2009 学年第一学期考试试卷

考试科目：<u>概率论与数理统计</u>　　　　　　得 分：____________

学生所在系：____________姓 名______________学 号：____________

（考期：2009 年 1 月 7 日，闭卷，可用计算器）

一、填空与单项选择

（1）连续掷一枚不均匀硬币（掷出正面的概率为 p），直至正反面都掷出为止，设 X 为所掷的次数，则 X 的分布律为（）

（2）设 X 与 Y 独立，都服从 N(0,1)，则$(X+Y)^2/(X-Y)^2$的分布为（）

（3）设样本$X_1,\cdots,X_n$抽自正态总体$N(\mu,\sigma^2)$，σ^2未知，则μ的$(1-\alpha)$置信区间为（）

（4）设 A、B、C 两两独立，则 A、B、C 相互独立的充要条件为：

(a)A 与 BC 独立 (b)AB 与(A+B)独立 (c)AB 与 AC 独立 (d)(A+B)与(A+C)独立

（5）若 E(XY)=E(X)E(Y)，则必有：

(a)Var(XY)=Var(X)Var(Y)　　(b) Var(X+Y)=Var(X)+Var(Y)

(c)X 与 Y 独立　　(d) X 与 Y 相关

（6）将一枚硬币连掷 n 次，以 X 与 Y 表示出现正面和反面的次数，则$\rho_{X,Y}$=(　)。

（7）$\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$为$\theta$的无偏估计，且$\lim_{n\to\infty}Var(\hat{\theta})=0$，则$\frac{n+1}{n}\hat{\theta}$为$\theta$的：

(a)无偏估计 (b)最小方差无偏估计 (c)相合估计 (d)以上皆错

二、现有 4 白 6 黑共 10 个球，从中随机取 2 球，已知其中有一个白球，则另一个球也是白球的概率为多少？

三、设随机向量(X,Y)具有概率密度函数：$f(x,y)=\begin{cases}\frac{6}{7}\left(x^2+\frac{xy}{2}\right), 0<x<1, 0<y<2\\ 0\quad, \quad others\end{cases}$

试分别求 X、Y 的期望、方差及 X 与 Y 的协方差和相关系数。

四、(利用中心极限定理求解）某灯泡厂生产的灯泡的平均寿命原为 2000 小时，标准差为 250 小时，经过工艺改革，使平均寿命提高到 2250 小时，标准差不变。为了确认这一改革成果，主管部门派人来检查，办法是：任意挑选若干只灯泡来检测，若其平均寿命值超过 2200 小时，则认可这一成果。问：

（1）若挑选 160 只灯泡来检查，则其平均寿命值超过 2200 小时的概率约为多少？

（2）为了使检查通过的概率超过 0.997，问至少应检查多少只灯泡？

五、设样本$X_1,\cdots,X_n$抽自均匀分布$R(\theta,0),(\theta<0)$：

（1）试求θ的矩估计$\hat{\theta}$和极大似然估计θ^*；

（2）$\hat{\theta}$和θ^*是否为θ的无偏估计？若是请加以证明，若不是请加以修正。

（3）问(2)中所得的无偏估计，哪个更有效？

六、甲、乙、丙三个工厂生产同一种产品，产品质量分为 1、2、3 三个等级（分别代表高、中、低)。今从三个厂共抽得 300 件产品，逐一检测，的结果如下图所示：

（1）试问这三个厂产品质量是否一致？$(\alpha=0.01)$

（2）若不一致，试问哪个厂产品质量较优？哪个厂产品质量较劣？并请说明理由。

参考答案

一、（1）$\mathrm{P}(X=k)=p^{k-1}q+q^{k-1}p,(q=1-p,k=2,3,4\cdots)$　　（2）$F_{1,1}$

（3）$\bar{X}\pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}$　　（4）—（7）abac

二、1/5

三、E(X)=5/7　E(Y)=8/7　Var(X)=23/490　Var(Y)=46/147

Cov(X,Y)=-1/147　　$\rho_{X,Y}=-\frac{\sqrt{15}}{69}$

四、$\mathrm{P}(\bar{X}>2200)\approx\Phi(2.53)\approx 0.9943$　　$n\geq 189$

五、(1) $\hat{\theta}=2\bar{X}$　$\theta^*=X_{(1)}$　(2) $\hat{\theta}=2\bar{X}=\tilde{\theta}_1$无偏；$\theta^*$ =有偏 $E\theta^*=\frac{n}{n+1}\theta$，修正为

$\tilde{\theta}_2=\frac{n+1}{n}\theta^*=\frac{n+1}{n}X_{(1)}$　　(3) $Var(\tilde{\theta}_1)=\frac{1}{n}\theta^2$　$Var(\tilde{\theta}_2)=\frac{1}{n(n+1)}\theta^2$

六、（1）拒绝原假设，认为三个厂产品质量不一致；（2）甲厂最优，丙厂最劣，乙厂居间，可分别计算三个厂产品质量的算术平均数，愈小者愈优。

中国科学技术大学

2009—2010 学年第二学期考试试卷

考试科目：概率论与数理统计　　　　　　　得 分：__________

学生所在系：__________姓 名____________学 号：__________

（考期：2010 年 7 月 14 日，闭卷，可用计算器）

一、填空判断选择。

（1）掷 3 个骰子，已知三个点数各不相同，则其中至少有一个为 6 点的概率为（）。

（2）设$X_1,\cdots,X_4$为相互独立的 N(0,1)变量，$T=a(X_1-2X_2)^2+b(3X_3+4X_4)^2$服从卡方分布，则 a=___，b=___，此时 T 的自由度为___。

（3）设随机变量 X 与 Y 相互独立分别服从参数为μ和λ的 Possion 分布，则 P(X=k|X+Y=n)=___，即在给定 X+Y=n 的条件下，X 的条件分布为____。

（4）设 Var(X)=Var(Z)，Var(Y)=4Var(X)，相关系数$\rho_{X,Y}=-1$，$\rho_{X,Z}=1/2$，则$\rho_{X,Y+Z}=$_

（5）在假设检验中，第 I 类错误是指____；第 II 类错误是指______。

（6）设$X_1,\cdots,X_n$为正态总体$\mathrm{N}(\mu,\sigma^2)$中抽取的样本，σ^2未知，记样本均值和样本方差分别为$\bar{X}$和S^2，则假设检验$H_0:\mu\le\mu_0\leftrightarrow H_1:\mu>\mu_0$（$\mu_0$为已知数）的检验统计量为____。

（7）设$X_1,\cdots,X_n$为正态总体$\mathrm{N}(\mu,\sigma^2)$中抽取的样本，μ,σ^2为未知参数，则下面的量为（　　）统计量。

(A) $\bar{X}=\frac{1}{n}\sum_1^n X_i$ (B) $\bar{X}-\mu$ (C) $S^2=\frac{1}{n-1}\sum_1^n(X_i-\bar{X})^2$ (D) $\sqrt{n}(\bar{X}-\mu)/S$

（8）若总体密度函数为$\mathrm{f}(x)=\frac{1}{2}(x-\mu)^{-3/2}I(x>\mu+1)$，则可以使用矩估计法估计参数$\mu$，这种说法___。（A）对　　（B）错

（9）设$\mathrm{X}\sim\mathrm{B}(n,p)$,则当 n 趋于无穷时，$2\sqrt{n}\left(arcsin\sqrt{X/n}-arcsin\sqrt{p}\right)$的分布函数收敛到标准正态分布，据此作出在 X=90,n=150 情况下，参数 p 的置信系数为 0.95 的大样本区间估计________。

（10）设$X_1,\cdots,X_n$为从均匀总体$\mathrm{U}(0,\theta),\theta>0$中抽取的样本，则$\theta$的估计量$X_{(n)}$为$\theta$的：
（A）无偏估计 （B）相合估计 （C）似然估计 （D）矩估计

二、某工厂的第一、第二、第三号车间生产同一种产品，产量分别占总产量的 1/2，1/3 和 1/6，次品率分别为 1%、1%和 2%。现从该厂某批产品中随机抽取一件，则：

（1）求取的产品为次品的概率；

（2）若取出的产品为次品，求其是第二个车间生产的概率。

三、设二维随机变量(X,Y)的联合密度可以表示成$g(x^2+y^2)$，g 为连续函数。令极坐标变换 X=Rcosθ，Y=RSinθ，问 R 与θ是否相互独立，并求出各自的密度。

四、设$X_1,\cdots,X_n$为从均匀总体$\mathrm{U}(\theta,2\theta)$中抽取的简单随机样本，试求：

（1）求θ的矩估计$\tilde{\theta}$和极大似然估计$\hat{\theta}$；

（2）矩估计$\tilde{\theta}$和极大似然估计$\hat{\theta}$是否为无偏估计？若不是，请加以修正，并说明修正

中 国 科 学 技 术 大 学

2010—2011学年第一学期考试试卷

考试科目 <u>概率论与数理统计</u>　　得分 ________

所在系 ______　　姓名 ________　　学号 ________

考试时间: 2010年12月26日下午2:30—4:30;　使用简单计算器

装……………………订……………………线

一. 填空判断选择题(每题3分,答题请写在试卷上)：

1 掷3个骰子，恰好有两枚点数相同的概率为________________.

2 设 $X_1, \cdots, X_n$ 为相互独立的 $N(0, \sigma^2)$ 变量, 其中σ^2未知. 令$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i, S^2 = \frac{1}{n}\sum_{i=1}^n X_i^2$, 则$\bar{X}$的分布为________________, nS^2/σ^2的分布为______________.

3 设随机变量 X 和 Y 相互独立, 同分布于期望为 $\frac{1}{\lambda}$ 的指数分布, 则$\min\{X, Y\}$服从参数为________的____________分布.

4 设 $Var(X) = 4Var(Y) = 1$, $Cov(X, Y) = 0.25$, 则$X - Y$与$X + Y$的相关系数 $\rho_{X-Y,X+Y} =$ ____________________.

5 设 A, B 为互斥事件, 则 A, B 相互独立的充分必要条件为____________________.

6 参数估计量优良性的准则有________________(写出至少两个).

7 假设X, Y分别服从标准正态分布, 则$X+Y$的分布仍为正态分布. 该说法________.

(A) 正确　　　(B) 错误

8 总体参数的置信水平为95%的置信区间是指________

(A)总体参数落在一个特定的样本所构造的区间内的概率为95%

(B)总体参数落在一个特定的样本所构造的区间内的概率为5%

(C)在用同样方法构造的总体参数的多个区间中，包含总体参数的区间比例为95%

(D)在用同样方法构造的总体参数的多个区间中，包含总体参数的区间比例为5%

9 设 $X_1, \cdots, X_n$ 为来自于正态总体$N(\mu, 1)$的简单随机样本, 若要求参数 μ 的置信系数为0.95的置信区间长度不超过1, 则至少需要抽取的样本量n 为________.

(A) 14　(B) 16　(C) 18　(D) 20

10 进行1000次独立重复实验, 每次实验中事件A要么发生, 要么不发生, 且发生的概率为0.25, 则可以近似于95%的概率认为事件A发生的频率与概率相差不超过________.

(A) 2.12%　(B) 2.68%　(C) 1.08%　(D) 3.24%

二. (15分) 假定某种病菌在群体中的带菌率为1%. 在检测时, 带菌者和不带菌者被检测出阳性的概率分别为0.98和0.02.

(1) 现有某人被测出呈阳性反应, 则他是带菌者的概率是多少?

(2) 为了进一步确认, 这个人决定再独立的做一次测试,检测结果依然是阳性, 问在两次检测结果都呈阳性反应的情况下, 他确实为带菌者的概率是多少?

三. (15分) 设随机变量 (X, Y) 服从$A = \{(x,y) : |x+y| \le 1, |x-y| \le 1\}$内的均匀分布, 则

(1) 试求出X和Y的边际分布;

(2) X和Y是否相互独立? 不相关?

(3) 求在$X = x$ $(0 < x < 1)$ 时Y的条件密度.

四. (15分) 设总体X的分布律为

X	1	2	3
P	p	$2p$	$1-3p$

现从此总体中抽出一样本量为n的样本, 发现其中1出现了n_1次, 2出现了n_2次, 3出现了n_3次. 试

(1) 求p的极大似然估计量$\hat{p}$和矩估计量$\tilde{p}$.

(2) 证明所得的估计量均为无偏估计, 并说明两个估计量何者最优.

五. (15分) 某针灸减肥机构宣称疗程结束后可以使参加者平均减少体重5kg以上, 为检验该广告是否可信, 调查人员随机调查跟踪了10名参加者, 测得他们参加前和参加后的体重(kg)为

参加前	65.39	62.89	63.50	60.83	63.07	62.88	57.80	63.07	66.05	70.78
参加后	61.72	59.43	59.64	57.30	58.50	60.84	51.89	60.02	63.67	65.67

假设参加前和参加后的体重服从正态分布, 试

(1) 在显著性水平0.05下检验该机构的宣传是否可信.

(2) 给出平均减少体重的95%置信区间.

六. (10分) 为研究女性和男性在美国选举中的偏好差异，1991年美国普通社会调查随机调查了577名女性和403名男性，询问每人是倾向于“支持民主党”，“支持共和党”以及“中立”，得到的调查数据如下：

	所支持政党(Party)			
性别(Gender)	民主党(0)	中立(1)	共和党(2)	总计
女性(1)	279	73	225	577
男性(0)	165	47	191	403
总数	444	120	416	980

(1) 为了检验选民政治倾向是否与性别有关, 试写出此问题的原假设.

(2) 在显著性水平0.05下, 可否认为选民的政治倾向与性别无关?

附录 分位数: $u_{0.025} = 1.960$, $u_{0.05} = 1.645$, $t_{0.025}(10) = 2.228$, $t_{0.025}(9) = 2.262$, $t_{0.05}(10) = 1.812$, $t_{0.05}(9) = 1.833$, $\chi^2_{0.05}(1) = 3.841$, $\chi^2_{0.05}(2) = 5.991$.

2010-2011第一学期概率论与数理统计期末考试试卷答案

一. (30分, 每题3分) **1**. 5/12 **2**. $N(0,\sigma^2/n)$, χ_n^2 **3**. $2\lambda e^{-2\lambda z}I(z>0)$ **4**. $3/\sqrt{21}$ **5**. $P(A),P(B)$至少一个为0. **6**. 无偏性, 相合性, 均方误差准则, 渐近正态性 **7**. B **8**. C **9**. B **10**. B

二. (15分) (1) $\frac{0.98\times1\%}{0.98\times1\%+0.02\times99\%}=\frac{49}{148}=0.3311$

(2) $\frac{0.98^2\times1\%}{0.98^2\times1\%+0.02^2\times99\%}=0.9604$

三. (15分) (1)由对称性, X和Y有相同的边际密度, $f(x)=\begin{cases} x+1, & -1\le x<0 \\ 1-x, & 0\le x\le 1\end{cases}$

(2) 显然X 和Y不独立, 不相关.

(3) 易得$f(y|x)=\frac{1}{2(1-x)}I(x-1\le Y\le 1-x)$, 其中$0<x<1$

四. (15分) (1) $\hat{p}=\frac{n_1+n_2}{3n}$; $\tilde{p}=\frac{3-\bar{X}}{4}$, 其中$\bar{X}=\frac{n_1+2*n_2+3*n_3}{n}$;

(2) 由于$En_1=np, En_2=2np, En_3=n(1-3p)$, 故知$\hat{p}$和$\tilde{p}$均为无偏估计, 容易得到$var(\hat{p})=\frac{p(1-3p)}{3n}$, 而$var(\tilde{p})=\frac{3p-8p^2}{8n}$, 于是由$var(\hat{p})<var(\tilde{p})$知似然估计$\hat{p}$更有效.

五. (15分) 此为成对检验问题. 记X表示参加前后的体重差, 由题设知$X\sim N(\mu,\sigma^2)$. 从而从保护消费者角度来看, 考虑假设$H_0:\mu\le 5\leftrightarrow H_1:\mu>5$, 易知此假设的水平$\alpha$检验法则为

$$\text{当 }\sqrt{n}\frac{\bar{X}-5}{S}>t_\alpha(n-1)\text{时拒绝原假设, 否则不足以拒绝原假设}$$

(1) 当$\alpha=0.05$时, 计算得$\bar{X}=3.758$, $S=1.184575$, $n=10$,故

$\sqrt{n}\frac{\bar{X}-5}{S}=-3.315577<t_{0.05}(9)=1.833$, 从而在0.05水平下不足以拒绝原假设, 即该减肥机构的宣传不足以可信.

(2) 其95%置信区间为$[\bar{X}-\frac{S}{\sqrt{n}}t_{0.025}(9),\bar{X}+\frac{S}{\sqrt{n}}t_{0.025}(9)]$, 带入数据得到[2.91,4.61].

六. (10分) (1) 原假设可以表述为“H_0: 选民政治倾向与性别无关”.

(2) 在显著性水平0.05下, 对假设H_0, 根据拟合优度检验方法知

$$\begin{aligned}\chi^2 &= (279-444*577/980)^2/(444*577/980)+(73-120*577/980)^2/(120*577/980)\\ &\quad+(225-577*416/980)^2/(577*416/980)+(165-403*444/980)^2/(403*444/980)\\ &\quad+(47-120*403/980)^2/(120*403/980)+(191-403*416/980)^2/(403*416/980)\\ &= 7.009544\end{aligned}$$

自由度为2, 从而有$\chi^2=7.009544>5.99$, 因而拒绝原假设, 即拒绝“选民的政治倾向与性别无关”这一假设.

中国科学技术大学

2010—2011 学年第二学期考试试卷

考试科目：概率论与数理统计　　　　　　　得 分：____________

学生所在系：____________姓 名______________学 号：____________

（考期：2011 年 6 月 4 日，闭卷，可用计算器）

一、填空判断选择题。

（1）设 A,B,C 是三个相互独立的随机事件，且$0 < \mathrm{P}(C) < 1$，则在下列给定的四对事件中，不相互独立的是____

(A)$\overline{A+B}$和 C　(B)$\overline{AC}$和 C　(C)$\overline{A-B}$和$\bar{C}$　(D)$\overline{AB}$和$\bar{C}$

（2）设 A,B,C 为三个事件，则下面的等式中正确的是____

(A) $\mathrm{A} \cup \mathrm{B} - \mathrm{B} = \mathrm{A} - \mathrm{B}$　　(B) $(A-B) \cup B = A$

(C) $(A \cup B) - C = A \cup (B - C)$　　(D) $\mathrm{A} \cup \mathrm{B} = (A\bar{B}) \cup (\bar{A}B)$

（3）设 f(x)和 g(x)为两个概率密度函数，则 af(x)+bg(x)也是概率密度函数的充要条件为______。

（4）随机变量 X 与 Y 不相关，则必有_____

(A) Var(XY)=Var(X)Var(Y)　　(B) F(x,y)=$F_X(x)F_Y(y)$

(C) X 与 Y 相互独立　　(D) EXY=EX*EY

（5）设$\hat{\theta}_n$为未知参数θ的一个估计量，如果设$\lim_{n\to\infty} E\left|\hat{\theta}_n - \theta\right| = 0$，则$\hat{\theta}_n$为θ的__

(A) 无偏估计　（B）有效估计　（C）相合估计　（D）渐进正态估计

（6）在实验次数无穷大时，某个事件发生的频率就等于其发生的概率，该说法____

(A)正确　（B）错误

（7）连续型随机变量就是取值为连续区间的随机变量，该说法____

(A)正确　（B）错误

（8）设$X_1, \cdots, X_n$ iid~$\mathrm{N}(\mu, 1)$，考虑假设检验问题$H_0: \mu = 0 \leftrightarrow H_1: \mu = 1$，由$\mu$的极大似然估计可以得到一个水平α检验法则为____；该检验法则犯第 II 类错误的概率为_____

（9）设基于某组样本得到的总体均值μ的 95%置信区间为[0.234,1.03]，则我们可以在显著性水平____下_____（接受或拒绝）零假设$H_0: \mu = 0$。

（10）设某种产品的质量等级可以划分为“优”、“合格”和“不合格”，则使用拟合优度检验方法在检验生产此产品的三家工厂的产品没有差异这一假设时，检验统计量服从渐进卡方分布的自由度为______

二、假设有 4 个罐子，其中第 k 个罐子里有 k-1 个红球和 4-k 个蓝球，k=1,2,3,4。现随机取出一个罐子，然后不放回地从中取两球，求：

（1）取出的两个球颜色不同的概率；

（2）若已知其中一个球为红球，则另外一个球也为红球的概率为多少？

三、设二维随机变量 X,Y 的联合概率密度函数为：

$$f(x,y) = \begin{cases} 1, 0 < x < 1, 0 < y < 2x \\ 0, \quad others \end{cases}$$

(1)试求出 X,Y 的边际概率密度函数$f_X(x)$和$f_Y(y)$；

（2）试求出 Z=2X-Y 的概率密度函数$f_z(z)$；

（3）试求$P\left(Y \le \frac{1}{2} | X = \frac{1}{2}\right)$。

四、某种疾病的发病率为 0.005，现随机调查 1000 人，考虑事件 A=“在调查的人中发病人数在 3 至 7 个人”，试：

（1）使用 Possion 逼近方法求 P(A)；

（2）使用中心极限定理求 P(A)。

五、设样本$Y_1, \cdots, Y_n$相互独立，$Y_i \sim N(a_i\mu, \sigma^2), i = 1, \cdots, n$，其中$a_1, \cdots, a_n$为已知不全为零的常数。

（1）求μ和σ^2的极大似然估计$\hat{\mu}$和$\widehat{\sigma^2}$；

（2）$\hat{\mu}$是否为μ的无偏估计？

（3）$\widehat{\sigma^2}$是否为σ^2的无偏估计？若是请加以证明，若不是请加以修正。

六、为了了解甲乙两企业的职工工资水平，分别从两个企业各随机抽取若干名职工调查，的如下数据（单位：元）：

甲企业	750	1060	750	1820	1140	1050	1000	
乙企业	1000	1900	900	1800	1200	1700	1950	1200

假设两个企业的工资分别服从正态分布，且总体独立而均值方差未知。试根据以上数据判断：

（1）两企业职工工资的方差是否相等（$\alpha = 0.05$）

（2）甲企业职工平均工资是否低于乙企业职工平均工资（$\alpha = 0.05$）

2010-2011第二学期概率论与数理统计期末考试试卷答案

一. (30分, 每题3分)

1. B **2**. A **3**. $a+b=1, af(x)+bg(x)\ge 0, \forall x$ **4**. D **5**. C **6**. B **7**. B

8. 当$\bar{X} > u_\alpha/\sqrt{n}$时拒绝$H_0$,否则不足以拒绝. $\Phi(u_\alpha - \sqrt{n})$

9. 0.05, 拒绝 **10**. 4

二. (15分) (1) $\frac{1}{4}[0+\frac{2}{3}+\frac{2}{3}+0]=\frac{1}{3}$

(2) $\frac{1}{2}$

三. (15分) (1)X和Y的边际密度分别为, $f_X(x)=2xI(0<x<1)$, $f_Y(y)=[1-\frac{y}{2}]I(0<y<2)$.

(2) $f_Z(z)=[1-\frac{z}{2}]I(0<z<2)$.

(3) 1/2

四. (10分) (1) $\sum_{k=3}^{7}\frac{5^k}{k!}e^{-5}=0.742$

(2) $2\Phi(\frac{2}{\sqrt{5\times 0.995}})-1=2\Phi(0.897)-1=0.63$.

五. (15分) (1) $\hat{\mu}=\frac{\sum_i a_i y_i}{\sum_i a_i^2}$, $\widehat{\sigma^2}=\frac{1}{n}\sum_i(y_i-a_i\hat{\mu})^2$

(2) 是无偏估计.

(3) 不是无偏估计, 由于$E\widehat{\sigma^2}=\frac{1}{n}\sum_i E(y_i-a_i\hat{\mu})^2=\frac{1}{n}\sum_i E(y_i-a_i\mu+a_i(\mu-\hat{\mu}))^2=\frac{1}{n}[\sum_i E(y_i-a_i\mu)^2-\sum_i a_i^2 E(\mu-\hat{\mu})^2]=\frac{1}{n}[\sum_i Var(y_i)-\sum_i a_i^2 Var(\hat{\mu})]=\frac{n-1}{n}\sigma^2$. 从而可以修正为$\overline{\sigma^2}=\frac{n}{n-1}\widehat{\sigma^2}$

六. (15分)

记X_i, Y_j分别为甲企业和乙企业的样本, 则由$\bar{X}=1081.429, S_X^2=129447.6, \bar{Y}=1456.25, S_Y^2=181026.8$, 有

(1) 假设可以表述为$H_0:\sigma_1^2=\sigma_2^2 \leftrightarrow H_1:\sigma_1^2\ne\sigma_2^2$.

检验统计量$0.175=F_{0.975}(6,7)<T=\frac{S_X^2}{S_Y^2}=0.7151<F_{0.025}(6,7)=5.119$, 故没有足够的理由认为两家企业工人工资方差不同.

(2) 由(1)的结果,可以认为两组样本方差是相同的, 故对假设$H_0:\mu_1\ge\mu_2 \leftrightarrow H_1:\mu_1<\mu_2$

检验统计量$T=\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{15}{728}(6S_X^2+7S_Y^2)}}=-1.8265<t_{0.05}(13)=1.771$ 所以拒绝零假设, 故有充足的理由认为甲企业的平均工资低于乙企业的平均工资.

中国科学技术大学

2011—2012 学年第一学期考试试卷

考试科目：<u>概率论与数理统计</u>　　　　得 分：___________

学生所在系：___________ 姓 名_____________ 学 号：___________

（考期：2012 年 1 月 6 日，闭卷，可用计算器）

一、简单题（写出简要步骤）。

（1）设一批电子元件由甲工厂和乙工厂共同生产，其中甲、乙两厂的生产份额分别为 60%和 40%。根据经验可知甲、乙两工厂生产的电子元件的次品率分别为 1%和 2%，现从这批产品中随机抽取一件，发现是次品，则该次品是甲厂生产的概率是多少？

（2）从 1，2，3，4 四个数中任取一个数，记为 X，再从 1 到 X 中任取一个数，记为 Y，则{Y=2}这个事件发生的概率是多少？

（3）设 X 的概率密度函数为$f(x)=(1+\theta)x^{\theta}, 0<x<1$。现考虑假设检验问题$H_0:\theta=5 \leftrightarrow H_1:\theta=3$。该检验的否定域为$X>1/2$，则犯第一类错误和第二类错误的概率分别为多少？

（4）已知一批零件的长度 X（单位：cm）服从正态分布$N(\mu,1)$，从中随机抽取 16 个零件，得到长度的平均值为 40cm，试求μ的置信水平为 0.95 的置信区间？

（5）设$X_1,\cdots,X_n$是一组独立同分布样本，且$X_1 \sim N(\mu,\sigma^2)$，试问 c 取多少才使得$c\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2$为$\sigma^2$的无偏估计？

（6）进行 1000 次独立重复试验，每次实验中事件 A 发生的概率为 0.25，试问能以 95%的把握保证 1000 次实验中事件 A 发生的频率与概率相差不超过多少？

二、设随机向量(X,Y)服从区域 D 上的均匀分布，其中 D 是由直线 y=x,x=0,y=1 所围成的区域，试求：

（1）(X,Y)的联合密度$f(x,y)$　　（2）(X,Y)的边缘密度函数$f_1(x)$和$f_2(y)$

（3）条件密度$f(x|Y=y)$　　（4）E(X|Y=y)

三、设总体 X 的概率分布如下表，其中$0<\theta<1$为未知参数。现从此总体中随机抽取 100 个样本，发现有 17 个样本取值为 0，33 个样本取值为 1，50 个样本取值为 2。

X	0	1	2
P	$\theta/4$	$1-\theta$	$3\theta/4$

（1）求θ的矩估计$\hat{\theta}_1$和极大似然估计$\hat{\theta}_2$；并分别计算相应的计算值；

（2）$\hat{\theta}_1$和$\hat{\theta}_2$是否是无偏的？若否，请加以修正；

（3）请问修正后的估计哪个更有效？

四、为了解男性和女性对三种类型的啤酒：淡啤酒、普通啤酒和黑啤酒的偏好有没有差异，分别调查了 180 位男士和 120 位女士，得如下数据：

	淡啤酒	普通啤酒	黑啤酒
男性	49	31	100
女性	51	20	49

请问男性和女性对这三种类型的啤酒的偏好有显著性差异吗？（$\alpha=0.05$）

五、为了比较新旧两种肥料对小麦产量的影响，研究者选择了面积相等、土壤等条件相同的 12 块土地，分别在 6 块地上施用新旧两种肥料。对于旧肥料，得到的产量数据是 17，14，18，13，19 和 15；而新肥料的产量数据为：16，19，20，22，18 和 19。假设两种肥料的产量分别服从正态分布，且总体独立，均值和方差未知。试根据以上数据判断：

（1）两种肥料产量的方差是否相等？（$\alpha = 0.05$）

（2）新肥料获得的平均产量是否显著地高于旧肥料？（$\alpha = 0.05$）

中国科学技术大学

2011—2012 学年第二学期考试试卷

考试科目：概率论与数理统计　　　　　　　　得　分：__________

学生所在系：__________姓　名____________学　号：__________

（考期：2012 年 6 月 9 日，闭卷，可用计算器）

一、判断题

（1）若

二、(15 分)设随机变量 X 满足：$|X| \le 1$，$P(X=-1)=1/8$，$P(X=1)=1/4$，而且，X 在 $(-1,\ 1)$ 内任一子区间上取值的概率与该子区间的长度成正比。试求：

（1） X 的概率分布函数 $F(x)=P(X \le x)$；

（2） X 取负值的概率；　　（3） X 的数学期望 $E(X)$。

三、（20 分）二维随机变量 (X,Y) 的密度函数为：

$$f(x,y)=\begin{cases} Ae^{-(3x+4y)}, & (x>0, y>0) \\ 0, & \text{其他} \end{cases}$$

（1）试求系数 $A=$？；（2） X 与 Y 是否独立？

（3）试求 $Z=X+Y$ 的密度函数 $f_Z(z)$；

（4）试求 $Var\left(X \mid X+Y=1\right)$。

四、（20 分）设样本 $(X_1, X_2, \cdots, X_n)$ 抽自正态总体 $X \sim N(\mu\ ,\ 1)$，　μ 为未知参数

（1）　试求 $\theta=P(X \ge 2)$ 的极大似然估计 θ^*（结果可用 $\Phi(.)$ 的形式表示）；

（2）　写出 μ 的 $(1-\alpha)$ 置信区间，并求 θ 的 $(1-\alpha)$ 置信区间。

五、（15 分）为考查 A, B 两种制鞋材料的耐磨性，用它们制作了 10 双鞋，其中每双鞋的两只鞋分别用 A 和 B 两种材料制作(左、右脚两只鞋随机地采用 A 或 B)。10 个男孩试穿这 10 双鞋之后的磨损情况如下表所示（数字代表磨损程度），假定 A, B 两组数据的差服从正态

分布，问是否可以认为这两种材料的耐磨性无显著差异？（$\alpha = 0.05$）

男孩	1	2	3	4	5	6	7	8	9	10
A	13.2	8.2	10.9	14.3	10.7	6.6	9.5	10.8	8.8	13.3
B	14.0	8.8	11.2	14.2	11.8	6.4	9.8	11.3	9.3	13.6
差	-0.8	-0.6	-0.3	0.1	-1.1	0.2	-0.3	-0.5	-0.5	-0.3

六、（15 分）投资者感兴趣的一个问题，是上市公司股票价格的变化与其公司总部所在地是否有关。下表给出的是美国两个不同地区（公司总部所在地）的上市公司在 1998 年第三季度内股价变化情况。表格内的数字是相应的上市公司的个数。问股票价格的变化是否存在地区间的差异？（$\alpha = 0.05$）

股价变化 / 总部所在地	上升	不变	下降
新英格兰地区	100	7	561
西北地区	88	10	370

（完）

（参考数值： $\chi^2_{0.025}(2) = 7.3778$； $\chi^2_{0.05}(2) = 5.9915$；

$\chi^2_{0.025}(6) = 14.4494$； $\chi^2_{0.05}(6) = 12.5916$； $t_{0.025}(9) = 2.2622$；

$t_{0.05}(9) = 1.8331$； $t_{0.025}(10) = 2.2281$； $t_{0.05}(10) = 1.8125$。 ）

概率统计期末考试（2008 年 1 月 22 日）

（参考答案与评分标准）

一、（15 分）

（1） $p^* = p\pi + (1-p)(1-\pi)$， $p = (p^* - 1 + \pi)\big/(2\pi - 1)$，当 $p^* = 0.55, \pi = 0.4$ 时，$p = 0.25$；

（2）当 $\pi = 0.5$ 时， $p^* \equiv 0.5$，由此无法解出 p。

二、（15 分）

（1）$F(x)=\begin{cases}1, & x\ge 1\\ \frac{5x+7}{16}, & -1<x<1\\ \frac{1}{8}, & x=-1\\ 0, & x<-1\end{cases}$；（2）$=F(0)=\dfrac{7}{16}$；（3）$E(X)=\dfrac{1}{8}$。

三、（20 分）

（1）$A=12$；（2）独立；（3）$f_Z(z)=12(e^{-3z}-e^{-4z}),(z>0)$；

（4）$f_{X|Z}(x\,|\,Z=1)=\dfrac{e^x}{e-1}$，$(0<x<1)$；　$E(X\,|\,Z=1)=\dfrac{1}{e-1}$。

四、（20 分）

（1）$\theta^*=1-\Phi(2-\overline{X})$；（2）$\mu\in\overline{X}\pm u_{\frac{\alpha}{2}}\dfrac{1}{\sqrt{n}}$；$\theta\in\Phi(\overline{X}-2\pm u_{\frac{\alpha}{2}}\dfrac{1}{\sqrt{n}})$。

五、（15 分）

$$H_0:\mu_Z=0\leftrightarrow H_1:\mu_Z\ne 0$$

$\dfrac{|\overline{Z}|}{S_Z/\sqrt{n}}=\dfrac{|-0.41|}{0.3872/\sqrt{10}}\approx 3.3485>2.2622=t_{0.025}(9)$，拒绝 H_0，有显著差异。

六、（15 分）

$Z\approx 5.4437<5.9915=\chi^2_{0.05}(2)$，无法拒绝 H_0，未见有显著差异。

（完）

2012–2013 第一学期概率论期末考试试卷

一. 判断选择题 (每题 3 分,答题请写在试卷上):

1. 设 A,B,C 是三个随机事件,则在下列不正确的是________.

(A)$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$

(B)$(A\cup B)\cap C=A\cup(B\cap C)$

(C)$A\cap(B\cap C)=(A\cap B)\cap C$

(D)$A\cap(\overline{B\cap C})=(A\cap\bar{B})\cup(A\cap\bar{C})$

2. 设事件 A 与自身独立,则 A 的概率为________.

(A)0　(B)1　(C)0 或 1　(D)1/2

3. 设 $f(x)$ 和 $g(x)$ 为两个概率密度函数,则下述还是密度函数的是________.

(A)$f(x)/g(x)$　(B)$f(x)-g(x)$

(C)$(f(x)+g(x))/2$　(D)$(1+f(x))(1-g(x))$

4. 随机变量 X 和 Y 独立,Y 和 Z 独立,且都有期望方差,则必有________.

(A)X 和 Z 独立　(B)X 和 Z 不相关

(C)X 和 Z 相关　(D)$\mathrm{Cov}(X,Y)=0$

5. 设 $0<P(B)<1$,则 $P(A|B)=P(A|\bar{B})$ 成立的充分必要条件是________.

(A)$P(AB)=P(A)P(B)$　(B)$P(A+B)=P(A)+P(B)$

(C)$P(A)=P(B)$　(D)$P(A)=P(\bar{B})$

6. 设 $X_1,\ldots,X_n$ 为来自均匀分布 $U(-\theta,\theta)$ 的一组样本,θ 为未知参数,则下述量为统计量的是________.

(A)$\bar{X}-\theta$　(B)$\max\limits_{1\le i\le n}(X_i-\theta)-\min\limits_{1\le i\le n}(X_i-\theta)$

(C)$\max\limits_{1\le i\le n}(X_i-\theta)$　(D)$\min\limits_{1\le i\le n}(X_i-\theta)$

7. 假设从两个独立的正态总体中各得到样本量为 10 的两组样本,若两个总体的方差相同,则使用两样本 t 检验时 t 分布的自由度为________.

(A)9 (B)10 (C)18 (D)20

8. 当原假设 H_0 为真时,检验 ϕ 有可能________.

(A) 犯一类错误 (B) 犯第二类错误

(C) 犯第一类或者第二类错误 (D) 同时犯第一类和第二类错误

9. 假设总体密度为 $f_\theta(x)$,其中 θ 为参数. 若 X 为来自该总体的样本,则下述不正确的是________.

(A) 固定 x 时 $f_\theta(x)$ 为似然函数 (B) 固定 θ 时 $f_\theta(x)$ 为似然函数

(C) 固定 θ 时 $f_\theta(x)$ 为密度函数 (D)$f_\theta(x)$ 衡量了不同 θ 下观测到 x 的可能性大小

10. 假设总体 X 为取值 0,1,2 的离散型随机变量,且取各值的概率分别为 $P(X=0)=0.5$,$P(X=1)=p$,$P(X=2)=0.5-p$,其中 $0<p<0.5$ 为参数. 则当使用拟合优度检验时,检验统计量的渐近卡方分布的自由度为________.

(A)3 (B)2 (C)1 (D)0

二.(15 分) 设昆虫产卵数目服从参数为 1 的 Poisson 分布,而每个卵孵化为幼虫的概率为 p,各卵是否孵化相互独立,试求

(1) 一个昆虫产生 m 个幼虫的概率。

(2) 若已知某个昆虫产生了 m 个幼虫,求该昆虫产了 $n(n\geq m)$ 个卵的概率。

三.(15 分) 设随机变量 X,Y 相互独立,且 X 服从均匀分布 $U(-1,1)$,Y 服从均值为 1/2 的指数分布,则

(1) 求随机变量 $Z=(X+1)Y$ 和 X 的相关系数.

(2) 求条件概率 $P(Z>1|X=0)$.

四.(15 分) 当 PM2.5 值全天监测平均在 35 微克/立方米以内时,空气质量属于一级. 现观测到合肥市琥珀山庄过去 10 天的日平均 PM2.5 值分别为 28.24,31.48,33.85,39.34,37.78,30.21,29.92,31.21,30.17,37.84. 若假设琥珀山庄区域日均 PM2.5 值 X 服从正态分布,各天日均 PM2.5 值相互独立. 则

(1) 试给出日均 PM2.5 值的 95% 置信上限.

(2) 若感兴趣空气质量为一级的概率 $p=P(X\leq 35)$,试基于观测的日均数据给出 p 的极大似然估计.

五.(15 分) 设甲乙两家食用盐工厂生产的食盐每袋重量均服从正态分布 (忽略重量不可取负值). 现从这两家工厂产品中各随机抽出 10 件标称为 500 克的袋装食盐,分别测得抽出各袋食盐的重量 (单位为克) 为

甲厂: 495,494,500,502,501,492,495,495,499,503;

乙厂: 494,506,496,505,500,508,502,504,502,499.

试问甲乙两家工厂生产这种标称为 500 克的袋装盐重量上有无差异 ($\alpha = 0.05$).

六.(10 分) 为研究人们每天阅读电子书的时间 (T) 长短与购买实体书 (Y) 两者之间的关系,随机调查了 210 个人,结果如下

	$t<1$	$1<t<3$	$t>3$
购买	12	70	20
不购买	40	28	40

试在水平 $\alpha = 0.05$ 下判断每天阅读电子书的时间长短和购买实体书两者之间是否有关? 阅读电子书的时间长短和购买实体书之间呈现何种特点?

一. 判断选择题 (每题 3 分):

1. B　2. C　3. C　4. D　5. A　6. B　7. C　8. A　9. B　10. C

二.(15 分) (1)

$$P(X=m)=\sum_{n=m}^{\infty}P(X=m|Y=n)P(Y=n)=\sum_{n=m}^{\infty}\binom{n}{m}p^m q^{n-m}\frac{1}{n!}e^{-1}$$
$$=\frac{p^m}{m!}e^{-p},\quad m=0,1,2,\ldots$$

(2)

$$P(Y=n|X=m)=\frac{P(X=m|Y=n)P(Y=n)}{P(X=m)}$$
$$=\frac{q^{n-m}}{(n-m)!}e^{-q},\quad n=m,m+1,\ldots$$

三.(15 分) (1) 由于 $EZX=E[X(X+1)]EY=1/6,Var(Z)=5/12$, 因此

$$\rho_{Z,X}=[EZX-EZ\cdot EX]/\sqrt{Var(Z)Var(X)}=1/\sqrt{5}$$

求随机变量 $Z=(X+1)Y$ 的期望和方差.

(2) $P(Z>1|X=0)=P(Y>1)=e^{-2}$.

四.(15 分)(1) 由于 $\bar{x}=33.004,s=3.95$, 在题设下易知日均 PM2.5 的 95% 置信上限为 $\bar{x}+\frac{S}{\sqrt{n}}t_{0.05}(n-1)$, 带入样本值得到 35.5.

(2) $p=P(X\le 35)=P(\frac{x-\mu}{\sigma}\le\frac{35-\mu}{\sigma})=\Phi(\frac{35-\mu}{\sigma})$, 而 $\bar{x}$ 和 $\sqrt{(n-1)/ns^2}=\sqrt{9/10}s=3.75$ 为 μ 和 σ 的似然估计值,因此 p 的极大似然估计值为 $\hat{p}=\Phi(\frac{\bar{x}-\hat{\mu}}{\hat{\sigma}})=\Phi(0.53)=0.70$.

五.(15 分) 记 $N(\mu_1,\sigma_1^2)$ 和 $N(\mu_1,\sigma_1^2)$ 分别表示两家工厂袋装盐的重量分布

(1) 考虑方差是否一致:对假设 $H_0:\sigma_1^2=\sigma_2^2\leftrightarrow H_1:\sigma_1^2\neq\sigma_2^2$, 由检验统计量 $F=S_x^2/S_y^2=14.71/19.6=0.75>F_{0.975}(9,9)=1/F_{0.025}(9,9)=1/4.03=0.25$, 因此在 0.05 水平下不能拒绝零假设。

(2) 考虑均值是否一致:考虑假设 $H_0:\mu_1=\mu_2\leftrightarrow H_1:\mu_1\neq\mu_2$, 由 (1) 结果知可以使用两样本 t 检验,由检验统计量 $T=|\bar{x}-\bar{y}|/\sqrt{(s_x^2+s_y^2)/10}=2.16>t_{0.025}(18)=2.10$, 因此拒绝零假设。即在 0.05 水平下拒绝“两家工厂的袋装食盐平均重量一致”这一假设。

六.(10 分) 假设每天阅读电子书时间长短与购买实体书之间无关，则由 Pearson 卡方检验有 $T=\sum\frac{(O-E)^2}{E}=39.60>\chi_{0.05}(2)=5.99$，因此在 0.05 水平下拒绝“每天阅读电子书时间长短与购买实体书之间无关”这一假设。注意到在三类阅读时间下，购买实体书人的比例分别为 0.23，0.71 和 0.33, 因此每天阅读电子书时间在 1 小时和 3 小时之间的人群购买实体书的比例最高，而当每天阅读电子书时间长于 3 小时后，购买实体书的人比例反而下降为 0.33.

2012–2013 第二学期概率论期末考试试卷

一. 判断选择题 (每题 3 分，共 30 分，答题请写在试卷上)：

1. 设 A, B, C 为三个事件，则事件 $\overline{ABC}$ 表示的是________.

(A) A, B, C 不同时发生　　(B) A, B, C 中至少发生一个

(C) A, B, C 中至多发生一个　　(D) A, B, C 至少发生两个

2. 随机变量 $X \sim N(\mu, \sigma^2)$，且关于 y 的一元二次方程 $y^2 + 4y + X = 0$ 无实根的概率为 0.5，则 $\mu =$________.

(A) 2　　(B) 3　　(C) 4　　(D) 5

3. 在数字 1，2，3，4，5 中不放回地随机连取两个数，每次一个数. 则在第一次取出偶数的条件下，第二次取出奇数的概率为________.

(A) 1/2　　(B) 3/4　　(C) 1/4　　(D) 1/3

4. 设随机变量 X_1, X_2, X_3, X_4 独立同分布于标准正态分布，令 $T = a(2X_1 - X_2)^2 + b(X_3 + X_4)^2$. 则 $(a, b) =$________ 时候 T 服从自由度为 2 的卡方分布.

(A) $(0, \frac{1}{2})$　　(B) $(\frac{1}{5}, \frac{1}{2})$　　(C) $(\frac{1}{5}, 0)$　　(D) $(1, \frac{1}{2})$

5. 设随机变量 X, Y 的方差均为 σ^2，且两者的相关系数为 -0.5，则使得 $Z = \pi X + (1 - \pi)Y$ 的方差最小的 π 是________.

(A) $\frac{1}{2}$　　(B) $\frac{1}{3}$　　(C) $\frac{1}{4}$　　(D) $\frac{1}{5}$

6. 设 $X_1, \ldots, X_n$ 为来自某个存在期望 μ 和方差 σ^2 的总体的一组样本，$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 分别为样本均值和样本方差，则下述错误的是________.

(A) $\bar{X}$ 具有渐近正态性　　(B) S_n^2 为 σ^2 的无偏估计

(C) $\bar{X}$ 为 μ 的无偏估计　　(D) $\sqrt{n}(\bar{X} - \mu)/S_n$ 服从 t_{n-1} 分布

7. 设参数 θ 的 95% 置信区间在某组样本值下为 $[1.2, 2.2]$，则下述正确的是________.

(A) 区间 [1.2,2.2] 包含 θ 的概率为 95%

(B) 区间 [1.2,2.2] 包含 θ 的概率为 5%

(C) 区间 [1.2,2.2] 要么包含 θ 要么不包含 θ

(D) 以上都不对

8. 关于假设检验中检验方法的一类和二类两种错误,下述错误的是________.

(A) 两类错误不可避免

(B) 固定样本量时两类错误不可能同时很小

(C) 有可能同时犯一类和二类错误

(D) 限制第一类错误概率的原则是假设检验理论中的通用做法

9. 假设总体 $X \sim N(\mu,1)$,其均值 μ 的 95% 置信区间为 $[0.22,1.10]$,则概率 $P(X \leq 0)$ 的 95% 置信区间为________.

(A) $[\Phi(0.22),\Phi(1.10)]$ (B) $[1-\Phi(1.10),1-\Phi(0.22)]$

(C) $[0.22,1.10]$ (D) $[0,\Phi(1.10)]$

10. $X_1,\ldots,X_n$ 为来自正态总体 $N(\mu,1)$ 的样本,假设检验问题 $H_0:\mu=0 \leftrightarrow H_1:\mu=1$ 的 0.05 水平检验为 $\sqrt{n}\bar{X}>1.645$,若要求该检验犯二型错误的概率也不超过 0.05,则样本量 n 至少为________.

(A) 9 (B) 10 (C) 11 (D) 12

二.(15 分) 有甲乙两只口袋,甲袋中有 5 只白球和 2 只黑球,乙袋中有 4 只白球 5 只黑球. 先从甲袋中任取两球放入乙袋,然后再从乙袋中任取一球. 试

(1) 求从乙袋中取出的球为白球的概率.

(2) 若已知从乙袋中取出的球为白球,求从甲袋中取的两只球中有白色球的概率。

三.(15 分) 设随机变量 Y 的密度函数为 $f_Y(y)=4y^3I(0<y<1)$,随机变量 X 在给定 $Y=y\ (0<y<1)$ 时服从均匀分布 $U(0,y)$. 试

(1) 求随机变量 X 的边际密度.

(2) 求 X 和 Y 的相关系数.

四.(20 分) 假设总体 X 的概率分布为 $X_1,\ldots,X_n$ 为从该总体中抽取的一组简单样本,则

X	0	1	2
P	p	$1-2p$	p

(1) 据此给出参数 p 的矩估计量 $\hat{p}_1$ 和极大似然估计量 $\hat{p}_2$.

(2) $\hat{p}_1$ 和 $\hat{p}_2$ 是否为无偏估计?何者更有效?

(3) 若 $n = 100$,且一组样本值中统计发现其中等于 0 的有 23 个,等于 1 的有 53 个,等于 2 的有 24 个. 在显著性水平 $\alpha = 0.05$ 下,利用 $\hat{p}_2$ 和拟合优度检验方法,我们能否认为"该组样本来自于总体 X"?

五.(20 分) 假设某工厂产品的某个指标服从正态分布 $N(\mu, \sigma^2)$,μ, σ^2 均未知. 现从该厂某批产品中随机抽取了 50 件产品测得该指标值的平均值为 89.70 和样本标准差为 1.09. 据此

(1) 能否认为该批产品该指标的平均值为 $90(\alpha = 0.05)$.

(2) 能否认为该批产品该指标的标准差不超过 $1(\alpha = 0.05)$.

(3) 给出该批产品此指标均值的 95% 置信区间,并与 (1) 中假设检验结果比较,能得出什么结论?

一.(30 分, 每题 3 分)

1. A　2. C　3. B　4. B　5. A　6. D　7. C　8. C　9. B　10. C

二.(15 分)(1) $A=$ 从乙袋中取出白球, B_i 分别表示从甲袋中取出两只球中有 i 个白球 $(i=0,1,2)$, 则由全概率公式有

$$\begin{aligned}P(A) &= P(A|B_0)P(B_0)+P(A|B_1)P(B_1)+P(A|B_2)P(B_2)\\ &= \frac{4}{11}\cdot\frac{2}{42}+\frac{5}{11}\cdot\frac{20}{42}+\frac{6}{11}\cdot\frac{20}{42}\\ &= \frac{38}{77}\approx 0.494.\end{aligned}$$

(2) 由 Bayes 公式有

$$P(\bar{B}_0|A)=1-\frac{P(A|B_0)P(B_0)}{P(A)}=1-\frac{4/11\cdot 2/42}{38/77}=55/57\approx 0.965.$$

三.(15 分) (1) 由联合概率密度函数 $f(x,y)=4y^2I(0<x<y<1)$ 易得

$$f_X(x)=\frac{4}{3}(1-x^3)I(0<x<1)$$

(2) 易得 $EX=2/5, Var(X)=14/225$ 以及 $EY=4/5, Var(Y)=2/75$, $EXY=1/3$. 从而 $\rho_{XY}=\frac{1/3-8/25}{\sqrt{14/225\cdot 2/75}}=\sqrt{\frac{3}{28}}\approx 0.327$.

四.(20 分) (1) $\hat{p}_1=\frac{a_2-1}{2}$ 和极大似然估计量 $\hat{p}_2=\frac{n-n_1}{2n}$. 其中 $a_2=\frac{1}{n}\sum_{i=1}^n X_i^2$, n_1 为样本中等于 1 的个数.

(2) $E\hat{p}_1=p$ 和 $E\hat{p}_2=p$ 均为无偏估计. $Var(\hat{p}_1)=\frac{5p-2p^2}{2n}>Var(\hat{p}_2)=\frac{p(1-2p)}{2n}$, 故似然估计更有效.

(3) 卡方检验值为 $0.0213<\chi_1^2(0.05)=3.841$, 从而不能拒绝该组样本来自总体 X 的原假设.

五. (20 分) 该批产品该指标的平均值为 89.70 和样本标准差为 1.09. 据此

(1) 检验统计量 $|\sqrt{n}(\bar{X}-90)/S|=1.946<t_{49}(0.025)=2.010$, 因此在 0.05 水平下不能拒绝该批产品该指标的平均值为 90 的原假设.

(2) 检验统计量 $(n-1)S^2=58.22<\chi_{49}^2(0.05)=66.339$, 因此在 0.05 水平下不能能否认该批产品该指标的标准差不超过 1 的原假设.

(3) 置信区间为 $[\bar{X}-\frac{S}{\sqrt{n}}t_{n-1}(0.025), \bar{X}+\frac{S}{\sqrt{n}}t_{n-1}(0.025)]=[89.39, 90.01]$ 其包含了 90 这一点, 因此在 0.05 水平下不能拒绝该批产品该指标的平均值为 90 的原假设. 该置信区间为检验检验问题 (1) 的接受域.

2013-2014第一学期期末考试试卷及答案

一. 判断选择题 (每题3分,共30分,答题请写在试卷上):

1. 设事件A, B相互独立且$P(B) = 0.5$, $P(A-B) = 0.3$, 则$P(B-A) =$ ________.

(A) 0.1　(B) 0.2　(C) 0.3　(D) 0.4

2. 设A, B, C为三个事件,若$P(A) = p, P(B) = 2p, P(C) = 3p$ 且 $P(AB) = P(BC)$, 则 p 的最大值为 ______________.

(A) 1/3 (B) 1/4 (C) 1/5 (D) 1/6

3. 设$X_n \sim B(n, p)$, $0 < p < 1$, 则当n很大时, 下列叙述不正确的是_________.

(A) $\frac{X_n}{n}$依概率收敛到p　(B) 若$np \approx 5$, 则X_n近似服从参数为5的泊松分布

(C) $\frac{X_n}{n}$近似服从$N(p, \frac{p(1-p)}{n})$　(D) $\frac{X_n - np}{\sqrt{p(1-p)}}$近似服从$N(0,1)$

4. 袋中有10个球, 里面有0个, 1个, 2个,\ldots, 10个白球是等可能的. 今向袋中放入一个白球, 然后随机从袋中取出一个球, 则这个球为白球的概率是______________.

(A) 5/10 (B) 6/11 (C) 5/11 (D) 4/11

5. 设随机变量(X, Y)服从二元正态分布, 且有$Var(X) = 1, Var(Y) = 4$. 令$W = X - aY, V = X + aY$, 则当$a =$ ______________时W和V相互独立.

(A) 1 (B) $1/\sqrt{2}$ (C) 1/2 (D) 1/4

6. 假设θ的一个无偏估计量为θ, 其在一组样本下的值为1.2, 则下述描述正确的是_______.

(A) $E\hat{\theta} = 1.2$ (B) $\theta = 1.2$ (C) 估计量$\hat{\theta}$不存在系统性误差 (D) $\hat{\theta}$是θ的相合估计

7. 考虑假设检验问题$H_0 : \theta = 0 \leftrightarrow H_1 : \theta = -1$, 若$T = T(X)$为$\theta$的估计量, 则该假设的拒绝域有形式____ (其中c为合适的常数).

(A) $T > c$ (B) $T < c$ (C) $|T| > c$ (D) $|T| < c$

8. 某电子计算机有100个终端, 每个终端有15\%的可能处于闲置状态, 若各终端被使用与否是相互独立的, 则至少有15个终端空闲的概率约为
_________.

(A) 0.3 (B) 0.4 (C) 0.5 (D) 0.6

9. 设X_1,X_2,X_3为来自正态分布$N(0,\sigma^2)$的样本, 则$S=\frac{X_1-X_2}{\sqrt{2}|X_3|}$服从分布__________.

(A) $F(1,1)$ (B) $F(2,1)$ (C) $t(1)$ (D) $t(2)$

10. $X_1,\ldots,X_n$为来自正态总体$N(\mu,1)$ 的样本, 若要求μ的95%置信区间长度不超过0.2, 则样本量n至少为__________.

(A) 382 (B) 383 (C) 384 (D) 385

二.(10分) 假设某种品牌的饮料做促销活动, 消费者每买一瓶该饮料可获得奖品A和B之一, 且获得奖品A和B的概率分别为0.2和0.8. 若某人既想获得A又想获得B, 问他平均要买几瓶该品牌的饮料?

三.(15分) 设随机变量X的概率分布为$P(X=1)=P(X=2)=0.5$, 随机变量Y在给定$X=k$时服从均匀分布$U(0,k)$, $(k=1,2)$. 试

(1) 求随机变量Y的分布函数$F_Y(y)$.

(2) 求X和Y的相关系数.

四.(20分)
$X_1,\ldots,X_n$和$Y_1,\ldots,Y_m$为分别抽自正态总体$N(\theta,1)$和$N(\theta,4)$中抽取的独立样本, 记$\bar{X}=\sum_{i=1}^n X_i/n$, $\bar{Y}=\sum_{j=1}^m Y_j/m$.
试

(1) 证明θ的最大似然估计为$\hat{\theta}=\frac{4n}{4n+m}\bar{X}+\frac{m}{4n+m}\bar{Y}$.

(2) 证明$\hat{\theta}$在一切形如$c\bar{X}+d\bar{Y}$的无偏估计里方差最小.

(3) 基于$\hat{\theta}$, 作出θ的置信系数为$1-\alpha$的置信区间.

五. (15分) 装配一个部件可以采用不同的方法. 现在关心的是哪一种方法的效率更高. 从使用两种装配方法装配的部件中各独立随机的抽取12件, 记录它们的装配时间(单位:分钟), 得到

甲方法: 30 34 34 35 34 28 34 26 31 31 38 26
乙方法: 26 32 22 26 31 28 30 22 31 26 32 29

若假设两种装配方法的装配时间均服从正态分布, 则

(1) 两种装配方法装配时间的方差有无显著差异? ($\alpha=0.05$).

(2) 两种装配方法的平均装配时间有无显著差异? ($\alpha=0.05$).

六. (10分) 袋中有8个球, 其中红球数未知. 在其中任取3个, 记录红球的个数X, 然后放回再任取3个, 记录红球个数, 然后放回. 如此反复进行了112次, 得到结果如下:

X	0	1	2	3
次数	1	31	55	25

试在$\alpha=0.05$水平下检验假设H_0 : 红球的个数为5.

一. 1. (B) 2. (B) 3. (D) 4. (B) 5. (C) 6. (C) 7. (B) 8. (C) 9. (C) 10. (D)

二. 设需要买X瓶才能即得到A又得到B, 则对$n \geq 2$有

$$P(X=n)=0.2^{n-1}\times 0.8+0.8^{n-1}\times 0.2.$$

于是
$EX=\sum_{n=2}^{\infty} n\times 0.2^{n-1}\times 0.8+n\times 0.8^{n-1}\times 0.2=5.25.$

或者
$EX=0.2E(X|A)+0.8E(X|B)=0.2(1+1/0.8)+0.8(1+1/0.2)=5.25.$

三. (1). 当$y \geq 0$时有

$$F_Y(y)=\begin{cases}0, & y<0;\\ 0.75y, & 0\leq y\leq 1;\\ 0.5+0.25y, & 1<y<2;\\ 1, & y\geq 2.\end{cases}$$

(2). 容易得出
$EX=1.5, EX^2=0.5(1+4)=2.5, EY=E(Y|X=1)P(X=1)+E(Y|X=2)P(X=2)=0.5(0.5+1)=0.75,\ EY^2=\frac{5}{6}, E(XY)==1.25.$

进而
$\operatorname{Var}(X)=EX^2-(EX)^2=\frac{1}{4}$, $\operatorname{Var}(Y)=\frac{13}{48}$, $\operatorname{Cov}(X,Y)=EXY-EXEY=\frac{1}{8}$.
最后

$$\rho(X,Y)=\frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}=\sqrt{\frac{3}{13}}.$$

四. (1) 由对数似然函数

$$l(\theta)\propto -\frac{1}{2}\sum(X_i-\theta)^2-\frac{1}{8}\sum(Y_j-\theta)^2$$

看见其最大值为$\hat{\theta}=\frac{4n}{4n+m}\bar{X}+\frac{m}{4n+m}\bar{Y}$.

(2) 由无偏性知$c+d=1$, 再由方差

$$Var(c\bar{X}+(1-c)\bar{Y})=c^2\frac{1}{n}+(1-c)^2\frac{4}{m}$$

最小化该方差得到$c=\frac{4n}{4n+m}$, 对比$\hat{\theta}$从而得证.

(3) 由于$\hat{\theta} \sim N(\theta, \frac{4}{4n+m})$, 从而$\theta$的$1-\alpha$置信区间为
$\hat{\theta} \mp u_{\alpha/2}\sqrt{\frac{4}{4n+m}}$

五.(1) 考虑假设: $H_0 : \sigma_1^2 = \sigma_2^2 \leftrightarrow H_1 : \sigma_1^2 \neq \sigma_2^2$, 则使用$F$检验, 易知

$$F_{0.975}(11,11) = \frac{1}{F_{0.025}(11,11)} < F = S_1^2/S_2^2 = 14.02/12.63 = 1.11 < F_{0.025}(11,11)$$

因此不能拒绝H_0, 可以认为两种方法装配时间的方差没有差异.

(2) 考虑假设$H_0 : \mu_1 = \mu_2 \leftrightarrow H_1 : \mu_1 \neq \mu_2$, 由题设和(1)中的结论知应使用两样本$t$检验. 易知检验统计量值为$2.5722 > t_{0.025}(22)$, 因此拒绝零假设, 认为两种方法的平均装配时间有差异.

六. 在假设H_0下, 理论分布为

$$P(X=i) = \frac{C_5^i C_3^{3-i}}{C_8^3}, i = 0,1,2,3$$

即$P(X=0) = 1/56, P(X=1) = 15/56, P(X=2) = 30/56, P(X=3) = 10/56$. 从而由拟合优度检验方法

$$T = \sum \frac{(O-E)^2}{E} = 2.2 < \chi_{0.05}^2(3)$$

因此不能拒绝零假设.

2015-2016学年第一学期考试试卷和答案

一、单项选择填空题 (每题3分,共30分,答题请写在试卷上):

1. 考虑 $\{1,2,3,4,5,6\}$ 的一个随机排列，与原来位置序号相同的数字个数记为N（比如排列513246与原来位置序号只有数字3和6是相同的，所以 $N=2$），则 $EN=$____________

2. 设$X\sim N(1,9),Y\sim N(1,16)$, 且$X$与$Y$的相关系数为$\rho_{XY}=-1/2$，设$Z=\frac{X}{2}+\frac{Y}{3}$，则$X$和$Z$的相关系数$\rho_{XZ}=$____________

3. 下列哪些函数不是概率密度或者分布律？______________

(A) $f(x)=\frac{1}{2}I\{a-3/2\le x\le a\}+\frac{1}{4}I\{a-1\le x\le a\},a\in\mathbb{R}$

(B) $f(x)=\frac{5}{6}I\{a-3/2\le x\le a\}-\frac{1}{4}I\{a-1\le x\le a\},a\in\mathbb{R}$

(C) $P(X=k)=\frac{1}{2^k},k=0,1,2,\dots$

(D) $f(x)=\lambda\alpha x^{\alpha-1}e^{-\lambda x^\alpha}I\{x>0\},\lambda>0,\alpha>0$

4. 若两随机事件A,B相互独立，下列哪些说法不正确？______________

(A) $P(A|B)=P(A|B^c)$ (B) $P(A|B)=P(A^c|B)$

(C) $P(B|A)=P(B|A^c)$ (D) $P(B^c|A)=P(B^c|A^c)$

5. 假设盒子中有一个不知颜色是黑色还是白色的球，现在给盒子里再放一个白球，然后从中随机拿出一个球，发现是白色的。则此时盒中剩余的球是白色的概率为______________

(A) $\frac{1}{2}$ (B) $\frac{1}{3}$ (C) $\frac{2}{3}$ (D) $\frac{3}{4}$

6. 下述对正态总体均值的置信区间的表述错误的是______________

(A) 置信度愈高，则可靠性愈高

(B) 置信度愈高，则置信区间愈宽

(C) 置信区间的大小与测量次数的平方根成正比

(D) 置信区间的位置取决于测量的平均值

7. 设随机变量X和Y都服从标准正态分布，则下述正确的是______________

(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) $\frac{X^2}{Y^2}$服从F分布

8. 下述对一个检验方法的第二类错误描述错误的是________________

(A) 在给定样本量下，第二类错误的概率不可能任意小

(B) 在对立假设空间的子集下控制第二类错误的概率，可以用来确定样本量大小

(C) 在有限样本量下，第二类错误是不可以避免的

(D) 在一个检验结果是拒绝零假设时候，我们会有很大的风险犯第二类错误

9. 下述检验正态性假设的方法中错误的是________________

(A) 直方图方法 \qquad (B) 拟合优度检验方法

(C) 使用偏度系数和峰度系数 \qquad (D) t检验

10. 若一个总体的期望和方差分别为μ和σ^2，设$X_1,...,X_n$为来自该总体的一组简单样本，则总体变异系数$\frac{\sigma}{\mu}$的相合估计为_

二、(15分) 假设$Y \sim U(0,\theta), \theta > 1$，若随机变量

$$X = \begin{cases} Y, & Y \geq 1 \\ 0, & Y < 1 \end{cases}$$ 。试求

(1)X的分布函数。(2) 期望EX。

三、(15分) 设某生产线上组装每件产品的时间服从指数分布, 平均需要10分钟,且各产品的组装时间是相互独立的。

(1) 试求组装100件产品需要15小时至20小时的概率。

(2) 保证有95\%的可能性, 问16小时内最多可以组装多少件产品。

四、(15分) 称随机变量$X \sim Exp(a,b)$，如果X的概率密度函数为$f(x) = \frac{1}{b}e^{-(x-a)/b}I(x > a)$，其中$a \in \mathbb{R}, b > 0$为参数。现从总体$Exp(a,1)$中抽取简单样本$X_1,...,X_n$，从总体$Exp(a,2)$中抽取简单样本$Y_1,...,Y_m$，且两组样本相互独立。试

(1) 求a的矩估计和最大似然估计。

(2) 是否都为无偏估计？若不是，请修正为无偏估计，并比较修正后的估计何者最优。

五、(15分) StreetInsider.com报道了2002年一些著名公司的每股收益的数据.在2002年之前，财务分析家就预测了这些公司2002年的每股收益.利用下面数据评论实际的和预测的每股收益的差异.

公司	实际每股收益	预测每股收益	公司	实际每股收益	预测每股收益股收益
$AT\&T$	1.29	0.38	埃克森 — 美孚	2.72	2.19
美国运通	2.01	2.31	通用电气	1.51	1.71
花旗银行	2.59	3.43	强生	2.28	2.18
可口可乐	1.60	1.78	麦当劳	0.77	1.55
杜邦	1.84	2.18	沃尔玛	1.81	1.74

试

(1) 在显著性水平0.05下，检验实际的和预测的每股平均收益之间是否存在差异，你的结论是什么？

(2) 两均值之差的点估计是多少？分析家是低估还是高估了每股的收益？

(3) 给出两均值之差的95%置信区间，并据此对(1)的检验问题作出结论并解释。

六、(10分) 用甲、乙、丙、丁四种棉纱织成坯布，其中用甲纱织成的18匹坯布中17匹为上等品，乙纱织成的15匹坯布中11匹为上等品，丙纱织成的15匹坯布中8匹为上等品，丁纱织成的13匹坯布中11匹为上等品，问这四种棉纱的质量有无显著差异？（显著性水平为0.05)

2015-2016 第一学期答案：

一、 1. 1; 2. $\frac{5}{\sqrt{73}}$; 3. C; 4. B; 5. C; 6. C; 7. C; 8. D; 9. D; 10. $\frac{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}}{\sqrt{n}\overline{X}}$ (答案不唯一).

二、

1.

$$
\begin{aligned}
F_X(x) &= P(Y \le x | Y \ge 1)P(Y \ge 1) + P(0 \le x | Y < 1)P(Y < 1) \qquad (1)\\
&= \begin{cases} 0, & x < 0 \\ \frac{1}{\theta}, & 0 \le x \le 1; \\ \frac{x}{\theta}, & x > 1. \end{cases} \qquad (2)
\end{aligned}
$$

2.

$$
\begin{aligned}
E[X] &= E[X|Y \ge 1]P(Y \ge 1) + E[X|Y < 1]P(Y < 1) = E[Y|Y \ge 1]P(Y \ge 1) + E[0|Y < 1]P(Y < 1)\\
&= \frac{\theta - 1}{\theta}\frac{1+\theta}{2} = \frac{\theta^2 - 1}{2\theta}.
\end{aligned}
$$

三、

1. $X_1, \ldots, X_{100}$ i.i.d $\sim Exp(6)$.

$$
\begin{aligned}
P(15 \le \sum_{i=1}^{100} X_i \le 20) = \; & P(\frac{15 - 50/3}{10/6} \le \frac{\sum_{i=1}^{100} X_i - 100/6}{\sqrt{100 * 1/6}} \le \frac{20 - 50/3}{10/6})\\
& \approx \Phi(2) - \Phi(-1) = \Phi(2) - 1 + \Phi(1).
\end{aligned}
$$

2.

$$
P(\sum_{i=1}^{n} X_i \le 16) \;\; = P(\frac{\sum_{i=1}^{n} X_i - n/6}{\sqrt{n}/6} \le \frac{16 - n/6}{\sqrt{n}/6}) \approx \Phi(\frac{96-n}{\sqrt{n}}) = 0.95.
$$

$$
\frac{96-n}{\sqrt{n}} = 1.645
$$

$$
n = 81.19.
$$

最多可以组装~81个。

四、

1.

$$E[X] \;= a+1, E[Y]=a+2.$$

所以a 的矩估计为

$$\hat{a}_1 = \;2\overline{X}-\overline{Y}$$

其中 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$, $\overline{Y}=\frac{1}{n}\sum_{i=1}^{m}Y_i$. ($a$的矩估计不唯一，比如另一估计为$\hat{a}_1=(\overline{X}+\overline{Y}-3)/2$)

由于似然函数为

$$L=exp\{-\sum_{i=1}^{n}(X_i-a)-\frac{1}{2}\sum_{i=1}^{m}(Y_i-a)\}1_{\{X_{(1)}>a,Y_{(1)}>a\}}.$$

易见似然函数在 $a<\min\{X_{(1)},Y_{(1)}\}$ 时 是关于 a 单调递增的。从而得到 a的极大似然估计为

$$\hat{a}_2=\min\{X_{(1)},Y_{(1)}\}.$$

2.

$$E[\hat{a}_1] \;=\; E[\overline{X}-\overline{Y}]=a.$$

对 $x>a$,

$$P(\min\{X_{(1)},Y_{(1)}\}\le x)=1-P(\min\{X_{(1)},Y_{(1)}\}>x) \;=1-\prod_{i=1}^{n}e^{-(x-a)}\prod_{i=1}^{m}e^{-\frac{x-a}{2}}=1-\; e^{-\frac{(2n+m)(x-a)}{2}}$$

则 $\min\{X_{(1)},Y_{(1)}\}$ 的密度函数为

$$f(x) \;=\frac{2n+m}{2}e^{-\frac{(2n+m)(x-a)}{2}}1_{\{x>a\}}.$$

即$\hat{a}_2\sim Exp(a,\frac{2}{2n+m})$。从而

$$E[X] \;=\int_a^{\infty}x\;\frac{2n+m}{2}e^{-\frac{(2n+m)(x-a)}{2}}dx=a+\frac{2}{2n+m}$$

所以基于最大似然估计$\hat{a}_2$的修正后的无偏估计为

$$\hat{a}_2^*=\hat{a}_2-\frac{2}{2n+m}$$

容易得到

$$Var(\hat{a}_1)=4\frac{n+m}{nm}$$

以及

$$Var(\hat{a}_2^*) = Var(\hat{a}_2) = \frac{4}{(2n+m)^2}$$

因此，显然$\hat{a}_2^*$更有效。

五、

1. 此数据为成对数据， 利用一样本t检验：

$Z_i = X_i - Y_i, i = 1, \ldots, 10$， iid $\sim N(\mu, \sigma^2)$.

$$H_0 : \mu = 0 \longleftrightarrow H_1 : \mu \neq 0.$$

$$T = \frac{\sqrt{10}\sum_{i=1}^{10}(X_i - Y_i)}{S_{X-Y}}|_{H_0} \sim t_9.$$

代入样本数据有

$$t = -0.6048 > -t_9(0.025) = -2.262.$$

因此，我们不拒绝原假设，即在显著性水平 $\alpha = 0.05$下，实际与预测的平均收益不存在差异。

2. 实际的每股平均收益-预测的每股平均收益

$$\hat{\mu} = \frac{\sum_{i=1}^{10}(X_i - Y_i)}{10} = -0.103 < 0$$

，说明从点估计角度来说，分析家是高估了每股的收益。

3. 两个均值之差的 95% 的置信区间为

$$\left[-0.103 - \frac{S}{\sqrt{n}}t_9(0.025), -0.103 + \frac{S}{\sqrt{n}}t_9(0.025)\right] = [-0.488, 0.282]$$

此置信区间为检验问题（1）的接受域。 包含了 0 点， 因此检验问题的结论是接受原假设， 即可以认为实际的和预测的每股平均收益没有差异。

六、 这是齐一性检验，计算

$$\begin{aligned}\chi^2 = &\frac{(17 - 18*47/61)^2}{18*47/61} + \frac{(11 - 15*47/61)^2}{15*47/61} + \frac{(8 - 15*47/61)^2}{15*47/61} \\ &+ \frac{(11 - 13*47/61)^2}{13*47/61} + \frac{(1 - 18*14/61)^2}{18*14/61} + \frac{(4 - 15*14/61)^2}{15*14/61} + \frac{(7 - 15*14/61)^2}{15*14/61} \\ &+ \frac{(2 - 13*14/61)^2}{13*14/61} = 8.389 > 7.815 = \chi_3(0.05).\end{aligned}$$

在显著性水平 $\alpha = 0.05$下，四种棉纱的质量有显著性差异。

一、单项选择填空题 (每题 3 分, 共 30 分, 答题请写在试卷上):

1. 考虑 $\{1,2,3,4,5,6\}$ 的一个随机排列，与原来位置序号相同的数字个数记为 N (比如排列 513246 与原来位置序号只有数字 3 和 6 是相同的，所以 $N=2$)，则 $EN=$________________

2. 设 $X\sim N(1,9), Y\sim N(1,16)$, 且 X 与 Y 的相关系数为 $\rho_{XY}=-1/2$，设 $Z=\dfrac{X}{2}+\dfrac{Y}{3}$，则 X 和 Z 的相关系数 $\rho_{XZ}=$________________

3. 下列哪些函数不是概率密度或者分布律？________________

(A) $f(x)=\dfrac{1}{2}I\{a-3/2\le x\le a\}+\dfrac{1}{4}I\{a-1\le x\le a\},\ a\in\mathbb{R}$

(B) $f(x)=\dfrac{5}{6}I\{a-3/2\le x\le a\}-\dfrac{1}{4}I\{a-1\le x\le a\},\ a\in\mathbb{R}$

(C) $P(X=k)=\dfrac{1}{2^k}, k=0,1,2,\ldots$

(D) $f(x)=\lambda\alpha x^{\alpha-1}e^{-\lambda x^{\alpha}}I\{x>0\},\ \lambda>0,\ \alpha>0$

4. 若两随机事件 A,B 相互独立，下列哪些说法不正确？________________

(A) $P(A|B)=P(A|B^c)$ (B) $P(A|B)=P(A^c|B)$

(C) $P(B|A)=P(B|A^c)$ (D) $P(B^c|A)=P(B^c|A^c)$

5. 假设盒子中有一个不知颜色是黑色还是白色的球，现在给盒子里再放一个白球，然后从中随机拿出一个球，发现是白色的。则此时盒中剩余的球是白色的概率为 ________________

(A) $\dfrac{1}{2}$ (B) $\dfrac{1}{3}$ (C) $\dfrac{2}{3}$ (D) $\dfrac{3}{4}$

6. 下述对正态总体均值的置信区间的表述错误的是 ________________

(A) 置信度愈高，则可靠性愈高

(B) 置信度愈高，则置信区间愈宽

(C) 置信区间的大小与测量次数的平方根成正比

(D) 置信区间的位置取决于测量的平均值

7. 设随机变量 X 和 Y 都服从标准正态分布，则下述正确的是 ________________

(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) $\dfrac{X^2}{Y^2}$ 服从 F 分布

8. 下述对一个检验方法的第二类错误描述错误的是 ________________

(A) 在给定样本量下，第二类错误的概率不可能任意小

(B) 在对立假设空间的子集下控制第二类错误的概率，可以用来确定样本量大小

(C) 在有限样本量下，第二类错误是不可以避免的

(D) 在一个检验结果是拒绝零假设时候，我们会有很大的风险犯第二类错误

9. 下述检验正态性假设的方法中错误的是 ________________

(A) 直方图方法　　(B) 拟合优度检验方法

(C) 使用偏度系数和峰度系数　　(D) t 检验

10. 若一个总体的期望和方差分别为 μ 和 σ^2，设 $X_1,\ldots,X_n$ 为来自该总体的一组简单样本，则总体变异系数 $\dfrac{\sigma}{\mu}$ 的相合估计为 ________________

二、(15 分) 假设 $Y\sim U(0,\theta),\theta>1$，若随机变量 $X=\begin{cases} Y, & Y\ge 1\\ 0, & Y<1\end{cases}$ 。试求

(1)X 的分布函数。(2) 期望 EX。

三、(15 分) 设某生产线上组装每件产品的时间服从指数分布, 平均需要 10 分钟, 且各产品的组装时间是相互独立的。

(1) 试求组装 100 件产品需要 15 小时至 20 小时的概率。

(2) 保证有 95% 的可能性, 问 16 小时内最多可以组装多少件产品。

四、(15 分) 称随机变量 $X \sim Exp(a,b)$，如果 X 的概率密度函数为 $f(x)=\dfrac{1}{b}e^{-(x-a)/b}I(x>a)$，其中 $a\in\mathbb{R}, b>0$ 为参数。现从总体 $Exp(a,1)$ 中抽取简单样本 $X_1,\ldots,X_n$, 从总体 $Exp(a,2)$ 中抽取简单样本 $Y_1,\ldots,Y_m$，且两组样本相互独立。试

(1) 求 a 的矩估计和最大似然估计。

(2) 是否都为无偏估计？若不是，请修正为无偏估计，并比较修正后的估计何者最优。

五、(15 分) StreetInsider.com 报道了 2002 年一些著名公司的每股收益的数据. 在 2002 年之前，财务分析家就预测了这些公司 2002 年的每股收益. 利用下面数据评论实际的和预测的每股收益的差异.

公司	实际每股收益	预测每股收益	公司	实际每股收益	预测每股收益股收益
AT&T	1.29	0.38	埃克森 -美孚	2.72	2.19
美国运通	2.01	2.31	通用电气	1.51	1.71
花旗银行	2.59	3.43	强生	2.28	2.18
可口可乐	1.60	1.78	麦当劳	0.77	1.55
杜邦	1.84	2.18	沃尔玛	1.81	1.74

试

(1) 在显著性水平 0.05 下，检验实际的和预测的每股平均收益之间是否存在差异，你的结论是什么？

(2) 两均值之差的点估计是多少？分析家是低估还是高估了每股的收益？

(3) 给出两均值之差的 95% 置信区间，并据此对 (1) 的检验问题作出结论并解释。

六、(10 分) 用甲、乙、丙、丁四种棉纱织成坯布，其中用甲纱织成的 18 匹坯布中 17 匹为上等品，乙纱织成的 15 匹坯布中 11 匹为上等品，丙纱织成的 15 匹坯布中 8 匹为上等品，丁纱织成的 13 匹坯布中 11 匹为上等品，问这四种棉纱的质量有无显著差异？(显著性水平为 0.05)

一、1. 1; 2. $\dfrac{5}{\sqrt{73}}$; 3. C; 4. B; 5. C; 6. C; 7. C; 8. D; 9. D; 10. $\dfrac{\sqrt{\sum_{i=1}^n (X_i-\overline{X})^2}}{\sqrt{n}\overline{X}}$ (答案不唯一).

二、1.

$$\begin{aligned} F_X(x) &= P(Y\le x|Y\ge 1)P(Y\ge 1)+P(0\le x|Y<1)P(Y<1) \\ &= \begin{cases} 0, & x<0 \\ \dfrac{1}{\theta}, & 0\le x\le 1; \\ \dfrac{x}{\theta}, & x>1. \end{cases} \end{aligned}$$

2.

$$\begin{aligned} E[X] &= E[X|Y\ge 1]P(Y\ge 1)+E[X|Y<1]P(Y<1) = E[Y|Y\ge 1]P(Y\ge 1)+E[0|Y<1]P(Y<1) \\ &= \frac{\theta-1}{\theta}\frac{1+\theta}{2} = \frac{\theta^2-1}{2\theta}. \end{aligned}$$

三、

1. $X_1,\dots,X_{100}$ i.i.d $\sim Exp(6)$.

$$\begin{aligned} P(15\le \sum_{i=1}^{100} X_i \le 20) &= P(\frac{15-50/3}{10/6}\le \frac{\sum_{i=1}^{100} X_i - 100/6}{\sqrt{100*1/6}} \le \frac{20-50/3}{10/6}) \\ &\approx \Phi(2)-\Phi(-1) = \Phi(2)-1+\Phi(1). \end{aligned}$$

2.

$$P(\sum_{i=1}^n X_i \le 16) = P(\frac{\sum_{i=1}^n X_i - n/6}{\sqrt{n}/6} \le \frac{16-n/6}{\sqrt{n}/6}) \approx \Phi(\frac{96-n}{\sqrt{n}}) = 0.95.$$

$$\frac{96-n}{\sqrt{n}} = 1.645$$

$$n = 81.19.$$

最多可以组装 81 个。

四、1.

$$E[X]=a+1, E[Y]=a+2.$$

所以 a 的矩估计为

$$\hat{a}_1 = 2\overline{X}-\overline{Y}$$

其中 $\overline{X}=\dfrac{1}{n}\sum_{i=1}^n X_i$, $\overline{Y}=\dfrac{1}{n}\sum_{i=1}^m Y_i$. ($a$ 的矩估计不唯一，比如另一估计为 $\hat{a}_1=(\overline{X}+\overline{Y}-3)/2$)

由于似然函数为

$$L = exp\{-\sum_{i=1}^n (X_i-a) - \frac{1}{2}\sum_{i=1}^m (Y_i-a)\} 1_{\{X_{(1)}>a, Y_{(1)}>a\}}.$$

易见似然函数在 $a<\min\{X_{(1)},Y_{(1)}\}$ 时是关于 a 单调递增的。从而得到 a 的极大似然估计为

$$\hat{a}_2 = \min\{X_{(1)},Y_{(1)}\}.$$

2.

$$E[\hat{a}_1] = E[\overline{X}-\overline{Y}] = a.$$

对 $x > a$,

$$P(\min\{X_{(1)}, Y_{(1)}\} \le x) = 1 - P(\min\{X_{(1)}, Y_{(1)}\} > x) = 1 - \prod_{i=1}^{n} e^{-(x-a)} \prod_{i=1}^{m} e^{-\frac{x-a}{2}} = 1 - e^{-\frac{(2n+m)(x-a)}{2}}.$$

则 $\min\{X_{(1)}, Y_{(1)}\}$ 的密度函数为

$$f(x) = \frac{2n+m}{2} e^{-\frac{(2n+m)(x-a)}{2}} 1_{\{x>a\}}.$$

即 $\hat{a}_2 \sim Exp(a, \frac{2}{2n+m})$。从而

$$E[X] = \int_a^\infty x \frac{2n+m}{2} e^{-\frac{(2n+m)(x-a)}{2}} dx = a + \frac{2}{2n+m}$$

所以基于最大似然估计 $\hat{a}_2$ 的修正后的无偏估计为

$$\hat{a}_2^* = \hat{a}_2 - \frac{2}{2n+m}$$

容易得到

$$Var(\hat{a}_1) = 4\frac{n+m}{nm}$$

以及

$$Var(\hat{a}_2^*) = Var(\hat{a}_2) = \frac{4}{(2n+m)^2}$$

因此，显然 $\hat{a}_2^*$ 更有效。

五、1. 此数据为成对数据，利用一样本 t 检验：$Z_i = X_i - Y_i$, $i = 1, \ldots, 10$，iid $\sim N(\mu, \sigma^2)$.

$$H_0 : \mu = 0 \longleftrightarrow H_1 : \mu \neq 0.$$

$$T = \frac{\sqrt{10}\sum_{i=1}^{10}(X_i - Y_i)}{S_{X-Y}}|_{H_0} \sim t_9.$$

代入样本数据有

$$t = -0.6048 > -t_9(0.025) = -2.262.$$

因此，我们不拒绝原假设，即在显著性水平 $\alpha = 0.05$ 下，实际与预测的平均收益不存在差异。

2. 实际的每股平均收益 -预测的每股平均收益

$$\hat{\mu} = \frac{\sum_{i=1}^{10}(X_i - Y_i)}{10} = -0.103 < 0$$

, 说明从点估计角度来说，分析家是高估了每股的收益。

3. 两个均值之差的 95% 的置信区间为

$$\left[-0.103 - \frac{S}{\sqrt{n}} t_9(0.025), -0.103 + \frac{S}{\sqrt{n}} t_9(0.025)\right] = [-0.488, 0.282]$$

此置信区间为检验问题（1）的接受域。包含了 0 点，因此检验问题的结论是接受原假设，即可以认为实际的和预测的每股平均收益没有差异。

六、这是齐一性检验，计算

$$\begin{aligned}\chi^2 = &\frac{(17 - 18*47/61)^2}{18*47/61} + \frac{(11 - 15*47/61)^2}{15*47/61} + \frac{(8 - 15*47/61)^2}{15*47/61} \\ &+ \frac{(11 - 13*47/61)^2}{13*47/61} + \frac{(1 - 18*14/61)^2}{18*14/61} + \frac{(4 - 15*14/61)^2}{15*14/61} + \frac{(7 - 15*14/61)^2}{15*14/61} \\ &+ \frac{(2 - 13*14/61)^2}{13*14/61} = 8.389 > 7.815 = \chi_3(0.05).\end{aligned}$$

在显著性水平 $\alpha = 0.05$ 下，四种棉纱的质量有显著性差异。

中 国 科 学 技 术 大 学

2015—2016学年第二学期考试试卷

考试科目 概率论与数理统计 得分 ____________

所在系 _______ 姓名 ____________ 学号 ____________

考试时间: 2016年6月24日上午8:30–10:30; 使用简单计算器

装 …… 订 …… 线

一. (10分) 判断题, 答案可以直接写在试卷上.

1. 若事件A和B满足$\mathrm{P}(A|\overline{B}) = \mathrm{P}(A)$, 则$A$和$B$相互独立. ( )

2. 一维连续型随机变量的分布函数是$(-\infty, \infty)$上的连续函数. ( )

3. 若一个参数θ的某个估计量$\hat{\theta}$的值刚好等于θ, 则$\hat{\theta}$为θ的一个无偏估计. ( )

4. 在假设检验中所遵循的原则是哪个假设成立的概率大就接受哪个假设. ( )

5. 在假设检验中接受原假设则一定不会犯第一类错误. ( )

二. (20分) 填空题或单选题, 答案可以直接写在试卷上.

1. (4分) 主人要外出, 委托邻居给花园的花浇水. 若不浇水, 则花死去的概率为0.8; 若浇水, 则花死去的概率为0.2. 假设有0.9的把握确定邻居会记得浇水. 那么, 主人回来时花还活着的概率为_____; 若主人回来花已死, 则邻居忘记浇水的概率为_____.

2. (3分) 设X服从参数为λ的Poisson分布. 若已知$\mathrm{E}[(X-1)(X-2)] = 1$, 则$\lambda =$_____.

3. (3分) 设随机向量(X, Y)的联合分布函数为

$$F(x,y) = \begin{cases} 1 - \mathrm{e}^{-\frac{x}{2}} - \mathrm{e}^{-\frac{y}{2}} + \mathrm{e}^{-\frac{x+y}{2}}, & x, y > 0; \\ 0, & \text{其它}. \end{cases}$$

那么X与Y是否相互独立? _____(填“是”或“否”).

4. (3分) 设随机变量X和Y的方差分别是4和9, 且它们的相关系数为-0.5, 则$X - Y$的方差为_____.

5. (3分) 设随机变量X和Y均服从标准正态分布, 则( ).

(A) $X + Y$服从正态分布 (B) $X^2 + Y^2$服从χ^2分布

(C) X^2/Y^2服从F分布 (D) X^2和Y^2服从χ^2分布

6. (4分) 设总体X的分布函数为

$$F(x) = \begin{cases} 1 - x^{-\beta}, & x > 1; \\ 0, & x \leq 1, \end{cases}$$

其中$\beta > 1$为未知参数, 且$X_1, X_2, \cdots, X_n$是来自总体X的一组简单随机样本, 样本均值为$\overline{X}$. 那么, 参数β的矩估计量和极大似然估计量分别是_______和________.

三. (18分) 设X和Y分别表示两只不同型号的灯泡的寿命, X服从参数为λ的指数分布, Y服从参数为μ的指数分布, 且它们相互独立.

1. 求先失效的那只灯泡的寿命(即$Z = \min\{X, Y\}$)的分布.

2. 求$U = X/Y$的概率密度函数.

四. (16分) 某人寿保险公司推出一项团购活动, 每个参保人在年初付150元保险费, 而在该年死亡时家属可从公司获得5万元赔偿金. 假设有1万名几乎同龄和同阶层的人参加该保险, 且已知该类人群在一年内死亡的概率是0.002. 利用中心极限定理求在此项业务中,

1. 保险公司亏本的概率是多少?

2. 保险公司获得的毛利润不少于50万的概率是多少?

五. (16分) 设总体X的期望为μ, 方差为σ^2. 分别抽取容量为n_1和n_2的两个独立样本, 样本均值依次为$\overline{X}_1$和$\overline{X}_2$.

1. 证明: 对满足$a + b = 1$的任意常数a和b, 统计量$Y = a\overline{X}_1 + b\overline{X}_2$均为$\mu$的无偏估计.

2. 试确定a和b的值使得上述估计量Y最有效.

六. (20分) 热处理车间工人为提高振动板的硬度, 对淬火温度进行试验, 在两种淬火温度A与B中分别测得硬度如下:

温度A	85.6	85.9	85.9	85.7	85.8	85.7	86.0	85.5	85.4	85.5
温度B	86.2	85.7	86.5	86.0	85.7	85.8	86.3	86.0	86.0	85.8

设振动板的硬度服从正态分布, 利用假设检验的知识判断淬火温度的改变是否对振动板的硬度具有显著性影响(显著性水平取$\alpha = 0.05$).

附录

$u_{0.025} = 1.960$, $u_{0.05} = 1.645$, $\Phi(2.24) = 0.9874$

$F_{10,10}(0.025) = 3.72$, $F_{9,9}(0.025) = 4.03$, $F_{10,10}(0.05) = 2.98$, $F_{9,9}(0.05) = 3.18$,

$t_{20}(0.025) = 2.086$, $t_{19}(0.025) = 2.093$, $t_{18}(0.025) = 2.101$,

中 国 科 学 技 术 大 学

2016—2017学年第一学期考试试卷

考试科目 概率论与数理统计(B) 得分 ____________

所在系 _______ 姓名 ____________ 学号 ____________

考试时间: 2017年1月3日上午8:30–10:30; 使用简单计算器

装 …… 订 …… 线

一. (30分, 每小题均3分) 填空题或单选题, 答案可以直接写在试卷上.

1. 设A和B为随机事件, $\mathrm{P}(A)=0.4, \mathrm{P}(B)=0.3, \mathrm{P}(A|B)=0.4$, 则( ).

(A) A与B相互独立 (B) A与B互斥

(C) $A \supset B$ (D) $\mathrm{P}(A\cup B)=\mathrm{P}(A)+\mathrm{P}(B)$

2. 甲乙二人进行网球比赛, 每回合胜者得 1 分, 且每回合甲胜的概率为 $p(0<p<1)$, 乙胜的概率为 $1-p$, 比赛进行到有一人比另外一个人多 2 分就终止, 多 2 分者最终获胜, 则甲最终获胜的概率为___________.

3. 设随机变量X服从参数为λ的Poisson分布, 且已知$\mathrm{P}(X=1)=\mathrm{P}(X=2)$, 则$X$取值为 3 的概率为___________.

4. 设随机变量X和Y均服从正态分布$N(0,\sigma^2)$, 且已知$\mathrm{P}(X\le 2,\ Y\le -2)=0.25$, 则$\mathrm{P}(X>2, Y>-2)=$ ___________.

5. 设X和Y相互独立且分别服从均值为 1 和 1/4 的指数分布, 则$\mathrm{P}(X<Y)=$ _______.

6. 设随机变量X_1和X_2相互独立, 方差均存在, 且概率密度函数分别为$f_1(x)$和$f_2(x)$. 若Y_1的概率密度函数为$[f_1(y)+f_2(y)]/2$, 而$Y_2=(X_1+X_2)/2$, 则( ).

(A) $\mathrm{E}[Y_1]>\mathrm{E}[Y_2], \mathrm{Var}[Y_1]>\mathrm{Var}[Y_2]$ (B) $\mathrm{E}[Y_1]=\mathrm{E}[Y_2], \mathrm{Var}[Y_1]=\mathrm{Var}[Y_2]$

(C) $\mathrm{E}[Y_1]=\mathrm{E}[Y_2], \mathrm{Var}[Y_1]<\mathrm{Var}[Y_2]$ (D) $\mathrm{E}[Y_1]=\mathrm{E}[Y_2], \mathrm{Var}[Y_1]>\mathrm{Var}[Y_2]$

7. 在假设检验中, 下列关于拒绝域和接受域说法错误的是( ).

(A) 与显著性水平 α 有关 (B) 与所构造的统计量的分布有关

(C) 随样本观测值的不同而改变 (D) 互不相交

8. 设X_1,X_2,X_3,X_4为来自总体$N(1,\sigma^2)$的一组简单随机样本, 则$T=\frac{(X_1-X_2)^2}{(X_3+X_4-2)^2}$的分布为( ).

(A) t_1 (B) $F_{1,1}$ (C) $F_{2,2}$ (D) 以上皆不正确.

9. 设$X_1,X_2,\cdots,X_n$为来自二项总体$B(n,p)$的一组简单随机样本, 且记 $\overline{X}$ 和 S^2 为样本均值和样本方差. 若 $\overline{X}+kS^2$ 为 np^2 的一个无偏估计, 则常数 $k=$ ___________.

10. 设正态总体$X\sim N(\mu,\sigma^2)$, 其中参数 μ 和 σ^2 均未知. 若样本容量 n 和置信系数 $1-\alpha$ 均保持不变, 对于不同的样本观测值, 则总体均值 μ 的置信区间长度( ).

(A) 始终保持不变 (B) 与 μ 的真值有关

(C) 与样本均值有关 (D) 不固定.

二. (11分) 设二维随机向量 (X, Y) 的概率密度函数为 $f(x,y) = \begin{cases} Ax^2, & 0 < |x| < y < 1; \\ 0, & \text{其它}. \end{cases}$

试求常数 A 的值及条件概率 $\mathrm{P}(X \le 0.25 | Y = 0.5)$.

三. (16分) 设随机变量 X 和 Y 相互独立, 且 X 的概率分布为 $\mathrm{P}(X=0) = \mathrm{P}(X=2) = 0.5$, Y的概率密度函数为$f(y) = \begin{cases} 2y, & 0 < y < 1; \\ 0, & \text{其它}. \end{cases}$

1. 求 $\mathrm{P}(Y \le \mathrm{E}Y)$;

2. 求 $Z = X + Y$ 的概率密度函数.

四. (15分) 某工程师为了解一台天平的精度, 用该天平对一物体进行 n 次测量, 该物体的质量 μ 是已知的. 设 n 次测量结果 $X_1, X_2, \cdots, X_n$ 相互独立且服从正态分布$N(\mu, \sigma^2)$, 该工程师记录的是 n 次测量的绝对误差 $Z_i = |X_i - \mu|$, $i = 1, 2, \cdots, n$. 现利用这些绝对误差来估计标准差 σ.

1. 求 Z_i 的概率密度函数;

2. 利用一阶矩求 σ 的矩估计量;

3. 求 σ 的极大似然估计量.

五. (18分) 在质量管理中, 产品的优良性和稳定性可以分别用均值和方差来体现. 要比较甲乙两厂所生产的电视机的质量是否有差异, 对两厂所生产的各 10 台产品的使用情况进行了追踪调查, 得到这 20 台电视机的寿命(单位: 年)数据如下:

甲	8	7	9	5	12	10	9	10	8	7
乙	10	8	5	7	8	7	11	4	5	6

设电视机的寿命服从正态分布, 利用假设检验的知识判断两厂所生产电视机的寿命是否有显著差异(显著性水平取 $\alpha = 0.05$).

六. (10分) 对截止目前为止一共 44 位美国总统(含候任总统Trump)的星座进行分析, 发现十二星座中天蝎座和水瓶座各有 5 人, 双子座和射手座各有 3 人, 处女座和白羊座各有两人, 而其余星座均有 4 人. 于是有人宣称有些星座擅长当美国总统, 而有些星座则不擅长. 结合你所学的知识, 说明该说法是否有统计学上的依据?(显著性水平取 $\alpha = 0.05$)

附录: 上分位数表

$F_{10,10}(0.025) = 3.72$, $F_{9,9}(0.025) = 4.03$, $F_{10,10}(0.05) = 2.98$, $F_{9,9}(0.05) = 3.18$,

$t_{20}(0.025) = 2.086$, $t_{19}(0.025) = 2.093$, $t_{18}(0.025) = 2.101$, $\chi^2_{11}(0.05) = 24.725$.

参考答案

一. 每小题3分

1. A **2.** $p^2/[p^2+(1-p)^2]$ 或 $p^2/(1-2p+2p^2)$ **3.** $\frac{4}{3}e^{-2}$ 或 0.18 **4.** 0.25 或 1/4
5. 0.2 或 1/5 **6.** D **7.** C **8.** B **9.** -1 **10.** D

二. 常数$A=6$. (5分)

由$f_Y(y)=4y^3, 0<y<1$及$f_{X|Y}(x|y)=\frac{3x^2}{2y^3}$, $-y<x<y$, 可知$f_{X|Y}(x|\frac{1}{2})=12x^2, -\frac{1}{2}<x<\frac{1}{2}$. 故

$$\mathrm{P}(X\le 1/4|Y=1/2)=\int_{-1/2}^{1/4}12x^2\mathrm{d}x=\frac{9}{16}.\quad (6分)$$

三. **1.** 由$\mathrm{E}Y=\frac{2}{3}$, 知 $\mathrm{P}(Y\le \mathrm{E}Y)=\int_0^{2/3}2y\mathrm{d}y=4/9$. (6分)

2. 对任意$0<z<3$, 由全概率公式可知

$$\begin{aligned}\mathrm{P}(Z\le z) &= \frac{1}{2}\mathrm{P}(Y\le z)+\frac{1}{2}\mathrm{P}(Y\le z-2)\\ &= \begin{cases} z^2/2, & 0\le z<1;\\ 1/2, & 1\le z<2;\\ ((z-2)^2+1)/2, & 2\le z<3.\end{cases}\qquad (5分)\end{aligned}$$

从而Z的概率密度函数为

$$l(z)=\begin{cases} z, & 0\le z<1;\\ z-2, & 2\le z<3;\\ 0, & 其它.\end{cases}\qquad (5分)$$

四. **1.** 概率密度函数为$f(z)=\frac{2}{\sqrt{2\pi}\sigma}\exp\{-\frac{z^2}{2\sigma^2}\}, \quad z>0.$ (5分)

2. 由于$\mathrm{E}[Z_i]=\frac{2\sigma}{\sqrt{2\pi}}$, 故$\sigma$的矩估计量为$\hat{\sigma}=\sqrt{\frac{\pi}{2}}\,\overline{Z}$, 其中$\overline{Z}=\frac{1}{n}\sum_{i=1}^n Z_i$. (5分)

3. 由对数似然函数为$l(\sigma)=c-n\ln\sigma-\frac{1}{2\sigma^2}\sum_{i=1}^n z_i^2$, 其中$c$为一与$\sigma$无关的常数, 令$\frac{\mathrm{d}l(\sigma)}{\mathrm{d}\sigma}=0$, 可知$\sigma$的极大似然估计量为$\hat{\sigma}=\sqrt{\frac{1}{n}\sum_{i=1}^n Z_i^2}$.

五. 由样本观测值得$\bar{x}=8.5, s_1^2=3.83, n_1=10; \bar{y}=7.1, s_2^2=4.99, n_2=10.$ (4分)

先检验两总体的方差是否相等. 此时$F=\frac{3.83}{4.99}=0.768$, 而由附表可得$F_{9,9}(0.025)=4.03, F_{9,9}(0.975)=1/4.03=0.248$. 因为$0.248<F<4.03$, 我们可以认为两总体方差相等. (7分)

再检验两总体均值是否相等. 此时$s_w^2=4.411$, 而统计量$t=\frac{\bar{x}-\bar{y}}{s_w\sqrt{1/n_1+1/n_2}}=-1.235$. 故由$|t|<t_{18}(0.025)=2.101$, 我们也可认为两总体均值相等. 综上可知, 我们可以认为两厂所生产电视机的寿命没有显著性差异. (7分)

六. 首先建立假设H_0：每个星座当上美国总统的可能性是一样的. 计算χ^2统计量的值为$2.91<\chi^2_{11}(0.05)=24.725$. 接受$H_0$, 故我们可以认为该说法没有统计学上的依据.

中 国 科 学 技 术 大 学

2017—2018学年第一学期考试试卷

考试科目 概率论与数理统计(B)　　　得分 ____________

所在系 ________　　姓名 ____________　　学号 ____________

考试时间: 2018年1月10日上午8:30–10:30;　使用简单计算器

装 …………………… 订 …………………… 线

一. (30分, 每小题3分) 填空题或单选题, 答案可以直接写在试卷上.

(1) 设随机事件A和B相互独立, A和C相互独立, 且B和C互斥. 若$\mathrm{P}(A)=\mathrm{P}(B)=1/2$, $\mathrm{P}(AC|AB\cup C)=1/4$, 则$\mathrm{P}(C)=$ ______.

(2) 一只蚂蚁从等边三角形$\triangle ABC$的顶点A出发开始沿着边爬行, 设它每次爬行到一个顶点后, 会休憩片刻再随机选择一条边继续爬行, 则第n次爬行是往A爬的概率为________.

(3) 设连续型随机变量X的密度函数$f(x)$满足$f(1+x)=f(1-x)$, 且$\int_0^2 f(x)\mathrm{d}x=0.4$, 则$\mathrm{P}(X<0)=$ (　)

(A) 0.2　　(B) 0.3　　(C) 0.4　　(D) 0.5

(4) 设随机变量X的分布函数为$F(x)=0.5\Phi(x)+0.5\Phi(\frac{x-4}{2})$, 其中$\Phi(x)$为标准正态分布函数, 则$X$的数学期望$\mathrm{E}X=$ ______.

(5) 设随机变量X与Y相互独立, X的概率分布为$\mathrm{P}(X=1)=\mathrm{P}(X=-1)=1/2$, Y服从参数为$\lambda>0$的Poisson分布. 若记$Z=XY$, 则$\mathrm{Cov}(X,Z)=$ ______.

(6) 设将1米长的木棒随机截成两段, 其中一段的长度记为X, 另一段长度的1/3记为Y, 则X与Y的相关系数为(　)

(A) 1　　(B) -1　　(C) $-1/3$　　(D) $1/3$

(7) 设$X_1,X_2,\cdots,X_9$是来自总体$X\sim N(0,1)$的一组简单随机样本, 则下列统计量中服从F分布的是(　)

(A) $\frac{X_1^2+X_2^2+X_3^2}{X_4^2+X_5^2+\cdots+X_9^2}$　(B) $\frac{X_1^2+X_2^2+X_3^2+X_4^2+X_5^2}{X_5^2+X_6^2+X_7^2+X_8^2+X_9^2}$　(C) $\frac{X_1^2+X_2^2+X_3^2}{2(X_4^2+X_5^2+\cdots+X_9^2)}$　(D) $\frac{2(X_1^2+X_2^2+X_3^2)}{X_4^2+X_5^2+\cdots+X_9^2}$

(8) 设$X_1,X_2,\cdots,X_n$是来自总体X的一组简单随机样本, 以$\overline{X}$和S^2分别表示样本均值和样本方差. 若记$\mathrm{Var}(X)=\sigma^2$, 则(　)

(A) S是σ的无偏估计量　　(B) S是σ的极大似然估计量

(C) S与$\overline{X}$相互独立　　(D) 以上均不对

(9) 设来自总体$X\sim N(\mu,0.9^2)$一组容量为9的简单随机样本, 其样本均值$\overline{X}=5$, 则未知参数μ的置信度为0.95的置信区间为________________(保留到小数点后三位).

(10) 设$X_1,X_2,\cdots,X_n$是来自正态总体$N(\mu,\sigma^2)$的一组简单随机样本, 据此样本做假设检验$H_0:\mu=\mu_0\leftrightarrow H_1:\mu\neq\mu_0$, 其中$\mu_0$是给定的已知常数, 则(　)

(A) 如果在检验水平$\alpha=0.05$下接受H_0, 那么在检验水平$\alpha=0.01$下必接受H_0

(B) 如果在检验水平$\alpha=0.05$下拒绝H_0, 那么在检验水平$\alpha=0.01$下必接受H_0

(C) 如果在检验水平$\alpha=0.05$下接受H_0, 那么在检验水平$\alpha=0.01$下必拒绝H_0

(D) 如果在检验水平$\alpha=0.05$下拒绝H_0, 那么在检验水平$\alpha=0.01$下必拒绝H_0

二. (16分)设二维随机向量(X,Y)的联合密度函数为

$$f(x,y)=C\mathrm{e}^{-2x^2+2xy-y^2},\quad -\infty<x,y<\infty.$$

(1) 求常数C的值;

(2) 在$X=x$的条件下, 求Y的条件密度$f_{Y|X}(y|x)$.

三. (16分)设二维随机向量(X,Y)服从二元正态分布$N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$, 其中$\mu_1=\mu_2=1$, $\sigma_1^2=\sigma_2^2=0.5$, $\rho=0.5$. 记

$$Z=|X-Y|,\quad U=\max(X,Y),\quad V=\min(X,Y).$$

(1) 求Z的密度函数$f_Z(z)$;

(2) 求数学期望$\mathrm{E}(U+V)$;

(3) 分别求数学期望$\mathrm{E}U$和$\mathrm{E}V$.

四. (18分)设总体X的密度函数为

$$f(x)=\frac{2x}{a^2},\quad 0\le x\le a,$$

其中$a>0$为未知参数, 而$X_1,X_2,\cdots,X_n$是来自该总体的一组简单随机样本.

(1) 求a的矩估计量$\hat{a}_1$和极大似然估计量$\hat{a}_2$;

(2) 求$p=\mathrm{P}(0<X<\sqrt{a})$的极大似然估计量$\hat{p}$;

(3) 问$\hat{a}_1$和$\hat{a}_2$是否为无偏估计? 若是, 请证明你的结论; 若不是, 请修正之.

五. (10分) 为了检验某种体育锻炼对减肥的效果, 随机抽取了10名减肥者进行测试. 在进行体育锻炼前后这些减肥者的体重(单位:千克)数据列表如下, 问该体育锻炼方法对降低体重是否具有显著性(设人的体重服从正态分布, 取显著性水平α=0.05)?

锻炼前体重	70	65	67	58	69	72	74	61	63	67
锻炼后体重	68	60	68	58	67	70	70	60	60	65

六. (10分)上海证券综合指数简称"上证指数", 反映了上海证券交易所上市股票价格的变动情况. 自上证指数诞生的二十七年(1991年1月至2017年12月)以来, 所有月份上涨或下跌的情况如下:

月份	一	二	三	四	五	六	七	八	九	十	十一	十二
上涨月数	14	21	16	15	14	14	13	15	11	13	18	13
下跌月数	13	6	11	12	13	13	14	12	16	14	9	14

结合你所学的知识, 我们能否认为上证指数的涨跌与月份有关?

附录: 上分位数表

$u_{0.025}=1.96,\ u_{0.05}=1.645;$

$t_8(0.025)=2.306,\ t_8(0.05)=1.86,\ t_9(0.025)=2.262,\ t_9(0.05)=1.833;$

$\chi^2_{11}(0.05)=19.675.$

参考答案

一. (每小题3分)

$\frac{1}{4}$; $\frac{1}{3}[1-(-\frac{1}{2})^{n-1}]$; B; 2; λ; B; D; C; $[4.412, 5.588]$; A.

二. (1) (8分) 由

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-2x^2+2xy-y^2}\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{\infty} \mathrm{e}^{-x^2}\mathrm{d}x \int_{-\infty}^{\infty} \mathrm{e}^{-(x-y)^2}\mathrm{d}y = \pi$$

可知$C=\frac{1}{\pi}$;

(2) (8分) 由于X的边缘密度函数为

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

从而,

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{1}{\sqrt{\pi}}\mathrm{e}^{-(x-y)^2}, \quad -\infty < y < \infty.$$

三. (1) (6分) 由$\mathrm{E}(X-Y)=0$,

$$\begin{aligned}\mathrm{Var}(X-Y) &= \mathrm{Var}(X)+\mathrm{Var}(Y)-2\mathrm{Cov}(X,Y)\\ &= 0.5+0.5-2\times 0.25 = 0.5,\end{aligned}$$

及二元正态分布的性质可知$X-Y\sim N(0,0.5)$, 从而$Z=|X-Y|$的密度函数为

$$f_Z(z) = \frac{2}{\sqrt{\pi}}\mathrm{e}^{-z^2}, \quad z>0.$$

(2) (4分) 易知, $\mathrm{E}(U+V)=\mathrm{E}(X+Y)=2$.

(3) (6分) 由$\mathrm{E}U-\mathrm{E}V=\mathrm{E}Z=\frac{1}{\sqrt{\pi}}$, 可知$\mathrm{E}U=1+\frac{1}{2\sqrt{\pi}}$, $\mathrm{E}V=1-\frac{1}{2\sqrt{\pi}}$.

四. (1) (6分) 矩估计量$\hat{a}_1=\frac{3}{2}\overline{X}$, 极大似然估计量$\hat{a}_2=X_{(n)}$;

(2) (4分) 由$p=\frac{1}{a}$知其极大似然估计量为$\hat{p}=1/X_{(n)}$;

(3) (8分) 矩估计$\hat{a}_1$是无偏的, 因$\mathrm{E}(\hat{a}_1)=\frac{3}{2}\mathrm{E}(\overline{X})=\frac{3}{2}\mathrm{E}(X)=a$; 而由$X_{(n)}$的密度函数为

$$h(x)=n[F(x)]^{n-1}f(x)=\frac{2n}{a^{2n}}x^{2n-1}, \quad 0<x<a,$$

知$\mathrm{E}(\hat{a}_2)=\frac{2n}{2n+1}a$. 故$\hat{a}_2$不是无偏估计, 可修正为$\hat{a}_2^*=\frac{2n+1}{2n}X_{(n)}$.

五. (10分) 成对数据. 首先可算得相减之后, 有$\overline{X}=2, S^2=28/9$. 故由

$$t=\frac{\sqrt{n}\overline{X}}{S}=3.59>t_9(0.05)=1.833,$$

可拒绝原假设(H_0: 锻炼前后体重无显著变化), 即认为该体育锻炼方法对降低体重具有显著性.

六. (10分) 列联表齐一性检验. 两行的和分别为177和147, 每列之和均为27. 由此可算得χ^2统计量的值为$11.394<\chi^2_{11}(0.05)=19.675$, 故可认为"无充分证据表明上证指数的涨跌与月份有关"或"上证指数的涨跌与月份无关".

中国科学技术大学

2018—2019学年第一学期考试试卷

考试科目 概率论与数理统计(B)　　得分 ______

所在系 ______　姓名 ______　学号 ______

考试时间: 2019 年1月9日上午8:30–10:30;　使用简单计算器

装 …… 订 …… 线

一、 (30分, 每小题3分) **填空题或单选题, 答案可以直接写在试卷上.**

(1) 已知 10 台洗衣机中有 7 台一等品和 3 台二等品, 现已售出一台, 在余下的 9 台中随机抽取 2 台后发现均为一等品, 则原先售出的一台为二等品的概率为______.

(2) 设随机变量 X 的概率密度 $f(x) = Ae^{-x^2+x}$, $-\infty < x < \infty$, 则常数 $A =$ ______.

(3) 设随机变量 X 与 Y 相互独立, 且它们的取值范围分别是 $\{1,2\}$ 和 $\{1,2,3\}$. 已知

$$\mathrm{P}(Y=1) = \frac{1}{6}, \quad \mathrm{P}(X=1, Y=2) = \mathrm{P}(Y=1, X=2) = \frac{1}{8},$$

则 $\mathrm{P}(Y=3) =$ ______.

(4) 设随机变向量 (X,Y) 的分布函数为 $\Phi(2x)\Phi(y-1)$, 其中 $\Phi(x)$ 为标准正态分布函数, 则 (X,Y) 服从二元正态分布(　)

(A) $N(0,1;\frac{1}{4},1;0)$　(B) $N(0,-1;\frac{1}{4},1;0)$　(C) $N(0,1;4,1;0)$　(D) $N(0,-1;4,1;0)$

(5) 设随机变量 X 与 Y 相互独立, 且 X 服从参数为 2 的泊松分布, Y 服从区间 $[-3,3]$ 上的均匀分布, 则它们的乘积的方差 $\mathrm{Var}(XY) =$ ______.

(6) 设 X_1, X_2, X_3, X_4 是来自标准正态总体的简单随机样本, $a > 0$为某个常数. 若已知

$$Y = a\Big(\frac{1}{2}X_1^2 + \frac{1}{2}X_2^2 + \frac{1}{2}X_3^2 + \frac{1}{2}X_4^2 + X_1X_2 + X_3X_4\Big)$$

服从 χ_n^2 分布, 则 $n + a =$ _____.

(7) 已知随机变量 X 服从 $F_{3,4}$ 分布. 设对给定的 $\alpha(0 < \alpha < 1)$, 实数 $F_{3,4}(\alpha)$ 满足 $\mathrm{P}(X > F_{3,4}(\alpha)) = \alpha$. 若有 $\mathrm{P}(X \le x) = 1 - \alpha$, 则 x 等于(　)

(A) $\frac{1}{F_{4,3}(1-\alpha)}$　(B) $\frac{1}{F_{3,4}(1-\alpha)}$　(C) $F_{4,3}(\alpha)$　(D) $F_{4,3}(1-\alpha)$

(8) 设 $X_1, X_2, \cdots, X_n$ 是来自均匀总体 $U[-\theta,\theta]$ 的简单随机样本, 则参数 θ 的极大似然估计量 $\hat{\theta}$ 为(　)

(A) $\max\limits_{1\le i\le n} X_i$　(B) $\max\limits_{1\le i\le n} |X_i|$　(C) $-\min\limits_{1\le i\le n} X_i$　(D) $-\min\limits_{1\le i\le n} |X_i|$

(9) 设 $X_1, X_2, \cdots, X_n$ $(n > 2)$ 是来自正态总体 $N(\mu,\sigma^2)$ 的简单随机样本, 且 $\overline{X}$ 为样本均值. 若统计量 $T = c(X_1 + X_n - 2\overline{X})^2$ 为 σ^2 的无偏估计, 则常数 $c =$ _____.

(10) 已知两个正态总体 X_1 和 X_2 分别为 $N(\mu_1,\sigma_1^2)$ 和 $N(\mu_2,\sigma_2^2)$, 为了检验总体 X_1 的均值大于 X_2 的均值, 则应作检验的假设为(　)

(A) $H_0: \mu_1 > \mu_2 \leftrightarrow H_1: \mu_1 \le \mu_2$　(B) $H_0: \mu_1 \ge \mu_2 \leftrightarrow H_1: \mu_1 < \mu_2$

(C) $H_0: \mu_1 < \mu_2 \leftrightarrow H_1: \mu_1 \ge \mu_2$　(D) $H_0: \mu_1 \le \mu_2 \leftrightarrow H_1: \mu_1 > \mu_2$

二、 (24分) 设二维随机向量 (X,Y) 的联合密度函数为

$$f(x,y)=\frac{1}{5}(2x+y),\quad 0\le x\le 2, 0\le y\le 1.$$

(1) 分别求 X 和 Y 的边缘密度函数 $f_X(x)$ 和 $f_Y(y)$;

(2) 在给定 $X=1$ 的条件下, 求 Y 在点 $y=0.5$ 处的概率密度 $f_{Y|X}(0.5\,|\,1)$;

(3) 求 X 和 Y 的协方差 $\mathrm{Cov}(X,Y)$;

(4) 求随机变量 $Z=\max\{X,Y\}$ 的密度函数 $f_Z(z)$.

三、 (12分) 某药厂试制了一种新药, 声称对贫血患者的治疗有效率达到 80%. 医药监管部门随机抽取 200 个贫血患者进行此药的临床试验, 若至少有 152 人用药有效, 就批准此药的生产. 试利用中心极限定理, 求解如下问题:

(1) 若该药的有效率确实达到 80%, 此药被批准生产的概率大约是多少?

(2) 若监管部门的方案是 200 个人中要有 160 人用药有效才批准, 这对药厂是否公平? 需说明理由.

四、 (18分) 已知总体 X 的密度函数为

$$f(x;\theta)=\begin{cases}-\theta^x\ln\theta, & x>0;\\ 0, & x\le 0,\end{cases}$$

其中 $0<\theta<1$ 为未知参数. 设 $X_1,X_2,\cdots,X_n$ 为来自该总体的一组简单随机样本.

(1) 求 θ 的矩估计量 $\widehat{\theta}$;

(2) 求 $h(\theta)=(\ln\theta)^{-1}$ 的极大似然估计量 $\widehat{h}_\theta$;

(3) 试求实数 a, 使得 $\widehat{h}_\theta$依概率收敛到 a, 即对任何$\varepsilon>0$, 都有$\lim\limits_{n\to\infty}\mathrm{P}\left(\left|\widehat{h}_\theta-a\right|\ge\varepsilon\right)=0$.

五、 (8分) 为比较A和B两种型号步枪子弹的枪口速度, 随机地抽取A型子弹 10 发, 得到枪口速度的平均值为 $\overline{x}=500(\mathrm{m/s})$, 样本标准差 $s_1=1.10(\mathrm{m/s})$; 随机地抽取B型子弹 20 发, 得到枪口速度的平均值为 $\overline{y}=496(\mathrm{m/s})$, 样本标准差 $s_2=1.20(\mathrm{m/s})$. 假设A和B型号子弹的枪口速度分别近似服从方差相等的正态分布 $N(\mu_1,\sigma^2)$ 和 $N(\mu_2,\sigma^2)$. 试在显著性水平 $\alpha=0.05$下检验假设 $H_0:\mu_1-\mu_2=5\leftrightarrow H_1:\mu_1-\mu_2\ne 5$.

六、 (8分) 某机构为了研究鼻咽癌是否与血型有关, 随机调查了一些患者和健康人, 得到的数据如下:

	A	B	O	AB
患者	64	86	130	20
健康人	125	138	210	26

请你根据所学的统计知识给出适当的结论(显著性水平设为 $\alpha=0.05$).

附录: 上分位数表

$u_{0.025}=1.96,\ u_{0.05}=1.645,\ \Phi(1.414)=0.9214;$

$t_{28}(0.025)=2.0484,\ t_{28}(0.05)=1.7011,\ t_{29}(0.025)=2.0452,\ t_{29}(0.05)=1.6991;$

$\chi^2_3(0.95)=0.3518,\ \chi^2_3(0.05)=7.8147.$

参考答案

一. (每小题3分)

3/8; $\frac{1}{\sqrt{\pi}}\mathrm{e}^{-\frac{1}{4}}$; 1/3; A; 18; 3; A; B; $\frac{n}{2(n-2)}$; D.

二. (1) (6分) 由边缘密度和联合密度的关系可知,

$$\begin{aligned} f_X(x) &= \int_0^1 f(x,y)dy = \frac{2}{5}x + \frac{1}{10}, \quad 0 \le x \le 2; \\ f_Y(y) &= \int_0^2 f(x,y)dx = \frac{2}{5}y + \frac{4}{5}, \quad 0 \le y \le 1. \end{aligned}$$

(2) (6分) 由边缘密度、条件密度和联合密度的关系可知,

$$f_{Y|X}(0.5\,|\,1) = \frac{f(1,0.5)}{f_X(1)} = 1.$$

(3) (6分) 由

$$\mathrm{E}[XY] = \int_0^1\int_0^2 xyf(x,y)dxdy = \frac{2}{3},$$

及 $\mathrm{E}X = \frac{19}{15}$ 和 $\mathrm{E}Y = \frac{8}{15}$, 可知 $\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}X\mathrm{E}Y = -\frac{2}{225}$.

(4) (6分) 由

$$F_Z(z) = \mathrm{P}(Z \le z) = \iint\limits_{x,y\le z} f(x,y)dxdy$$

可知,

$$F_Z(z) = \begin{cases} 0, & z < 0; \\ \frac{3}{10}z^3, & 0 \le z < 1; \\ \frac{1}{5}z^2 + \frac{1}{10}z, & 1 \le z < 2; \\ 1, & z \ge 2. \end{cases}$$

从而, 所求密度函数为

$$f_Z(z) = \begin{cases} \frac{9}{10}z^2, & 0 \le z < 1; \\ \frac{2}{5}z + \frac{1}{10}, & 1 \le z < 2. \end{cases}$$

三. (1) (6分) 所求概率为 $\Phi(\sqrt{2}) = 0.9214$.

(2) (6分) 不公平. 对药厂而言, 在治疗有效率达到80%的情况下被批准的概率大约为 $\Phi(0) = 0.5$, 这相当于用掷硬币的方式来决定是否得到批准.

四. (1) (6分) 矩估计量 $\widehat{\theta} = \exp\{-\frac{1}{\overline{X}}\}$, 其中 $\overline{X}$ 为样本均值;

(2) (6分) 参数 θ 的极大似然估计量同样为 $\exp\{-\frac{1}{\overline{X}}\}$, 从而 $h(\theta)$ 的极大似然估计量为 $\widehat{h}_\theta = -\overline{X}$;

(3) (6分) 由弱大数律可知, 所求的实数 $a = -\mathrm{E}X = \frac{1}{\ln\theta}$.

五. (8分) 两样本 t 检验, 其检验统计量为

$$T = \frac{\overline{X} - \overline{Y} - \mu_0}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

代入数据计算可知, $s_w = 1.169, t = -2.209$. 由于 $|t| > t_{28}(0.025) = 2.0484$, 在显著性水平 $\alpha = 0.05$下我们应该拒绝原假设 H_0.

六. (8分) 拟合优度联列表检验. 原假设为鼻咽癌与血型无关, 而其检验统计量为

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{4} \frac{(nn_{ij} - n_{i\cdot}n_{\cdot j})^2}{nn_{i\cdot}n_{\cdot j}}.$$

代入数据计算可知, $\chi^2 = 1.921 < \chi_3^2(0.05) = 7.8147$. 故在显著性水平 $\alpha = 0.05$下我们不能拒绝原假设, 即可以认为鼻咽癌与血型无关.

装 …… 订 …… 线

中 国 科 学 技 术 大 学

2018—2019学年第二学期考试试卷

考试科目 概率论与数理统计 得分 ____________

所在系 _______ 姓名 ____________ 学号 ____________

考试时间: 2019年6月19日下午14:30—16:30; 使用简单计算器

一. 简答(写出简要步骤, 40分)

1. (6分) 设一批电子产品由甲和乙两工厂共同生产，其中甲，乙两工厂生产的份额分别为60%和40%。根据经验可知甲，乙两工厂生产该产品的次品率分别为1%和2%，现从这批产品中随机抽取一件，发现是次品，则该次品是甲厂生产的概率是多少？

2. (6分) 设某保险公司每个月受理的索赔事件个数服从参数为λ的Poisson分布，而每个索赔成功的概率为p,且各个索赔彼此之间没有关系。二月份该公司有k个索赔成功的概率是多少? 若该月份有k个索赔成功，则它受理了m $(m \geq k)$个索赔事件的概率是多少?

3. (6分) 设X的概率密度函数为$f(x) = (1+\theta)x^{\theta}$, $0 < x < 1$. 现考虑假设检验问题$H_0 : \theta = 5 \leftrightarrow H_1 : \theta = 3$. 该检验的否定域为$X > 1/2$, 则犯第一类错误的概率和第二类错误的概率分别为多少?

4. (6分) 已知一批零件的长度X(单位：cm)服从正态分布$N(\mu, 1)$, 从中随机抽取16个零件, 得到长度的平均值为40cm, 试求μ的置信水平为0.95的置信区间?

5. (8分) 设$X_1, ..., X_n$是来自总体X的独立同分布样本, 且都只取正值, 试求数学期望:

$$E\left(\frac{X_1}{\sum_{i=1}^n X_i}\right).$$

6. (8分) 某工厂为了保证设备正常工作, 需配备适量的维修工人(工人配备多了就浪费, 配备少了又要影响生产),现有同类型设备300台, 各台工作是相互独立的,发生故障的概率都是0.01. 在通常情况下一台设备的故障可由一个人来处理(我们也只考虑这种情况), 问至少需配备多少工人, 才能保证设备发生故障但不能及时维修的概率小于0.01?

二. (20分) 设随机向量(ξ, η)服从区域D上的均匀分布, 其中D是由直线$y = x$, $x = 0$, $y = 1$所围成的区域. 试求:

(1) (ξ, η)的联合密度$p(x, y)$; (2) (ξ, η)的边缘密度$p_1(x)$和$p_2(y)$;

(3) 条件密度$p(x|\eta = y)$; (4) $E(\xi|\eta = y)$.

三. (15分) 设总体X的概率分布如下表, 其中$0 < \theta < 1$为未知参数. 现从此总体中随机抽取100个样本，发现有17个样本取值为0, 33个样本取值为1, 50个样本取值为2.

X	0	1	2
P	$\frac{\theta}{4}$	$1-\theta$	$\frac{3\theta}{4}$

(1) 求θ的矩估计$\hat{\theta}_1$和极大似然估计$\hat{\theta}_2$; 并分别计算相应的估计值。

(2) $\hat{\theta}_1$ 和$\hat{\theta}_2$是否是无偏的？若否，请修正。

(3) 请问修正后的估计那个更有效？

四. (10分)简· 奥斯汀(1775 - 1817), 英国女作家，作品有：《理智与情感》，《傲慢与偏见》，《爱玛》等，在其身后，她的哥哥亨利主持了遗作《劝导》和《诺桑觉寺》两部作品出版。下面表格收集了代表作《理智与情感》，《爱玛》以及《劝导》前两章中常用代表词的出现频数，

单词	理智与情感	爱玛	劝导
a	147	186	184
an	25	26	40
this	32	39	30
that	94	105	59

请问作品《理智与情感》,《爱玛》以及《劝导》之间在选择常用词比例是否存在差异？($\alpha = 0.05$)

五. (15分) 为比较甲和乙两处矿石的含灰率(%), 分别从甲、乙两处随机抽取矿石6块，甲处矿石含灰率数据是: 17, 14, 18, 13, 19 和15; 而乙处矿石含灰率数据为: 16, 19, 20, 22, 18 和19. 假设两处矿石含灰率分别服从正态分布, 且总体独立, 均值和方差未知. 试根据以上数据判断：

(1)(6 分) 在显著水平0.05下, 甲、乙两处矿石含灰率的方差是否相等?

(2)(9 分) 在显著水平0.05下, 乙处矿石含灰率的平均量是否显著地高于甲处矿石含灰率的平均量?

附录 分位数: $u_{0.025} = 1.960$, $u_{0.05} = 1.645$, $t_{0.025}(10) = 2.228$, $t_{0.05}(10) = 1.812$, $t_{0.025}(11) = 2.201$, $t_{0.05}(11) = 1.796$, $t_{0.025}(12) = 2.178$, $t_{0.05}(12) = 1.782$, $\chi^2_{0.05}(1) = 3.841$, $\chi^2_{0.05}(2) = 5.991$, $\chi^2_{0.05}(6) = 12.591$, $F_{0.05}(5,5) = 5.050$, $F_{0.025}(5,5) = 7.146$, $F_{0.05}(6,6) = 4.284$ $F_{0.025}(6,6) = 5.820$.

中 国 科 学 技 术 大 学

2018—2019学年第二学期考试试卷答案

一. 简答(写出简要步骤)

1. (6分) 解：设X表示随机抽取一件产品为次品的事件，A和B分别表示产品是由甲，乙工厂生产的事件，则$P(A)=0.6$, $P(B)=0.4$, $P(X|A)=0.01$, $P(X|B)=0.02$. 于是有

$$P(X)=P(X|A)P(A)+P(X|B)P(B)=0.014.$$

那么随机抽取一件，发现是次品，而该次品是A厂生产的概率是：

$$P(A|X)=\frac{P(X|A)P(A)}{P(X)}=\frac{3}{7}.$$

2. (6分) 解：二月份该公司有k个索赔成功的概率是:

$$\sum_{n=k}^{\infty}\frac{\lambda^n e^{-\lambda}}{n!}C_n^k p^k(1-p)^{n-k}=\frac{(\lambda p)^k e^{-\lambda p}}{k!}.$$

若该月份有k个索赔成功，则它受理了$m\ (m\geq k)$个索赔事件的概率是:

$$\frac{C_m^k p^k(1-p)^{m-k}\frac{\lambda^m e^{-\lambda}}{m!}}{\sum_{n=k}^{\infty}\frac{\lambda^n e^{-\lambda}}{n!}C_n^k p^k(1-p)^{n-k}}=\frac{(\lambda(1-p))^{m-k}e^{-(\lambda(1-p))}}{(m-k)!}$$

3. (6分) 解：由犯第一类错误和犯第二类错误的定义可知，犯第一类错误的概率为

$$P(X>1/2|\theta=5)=\int_{1/2}^{1}(1+5)x^5dx=1-\frac{1}{2^6}=0.984375;$$

犯第二类错误的概率为

$$P(X<1/2|\theta=3)=\int_{0}^{1/2}(1+3)x^3dx=\frac{1}{2^4}=0.0625.$$

4. (6分) 解：μ的置信水平为0.95的置信区间为$[\bar{x}-u_{0.025}/\sqrt{n},\ \bar{x}+u_{0.025}/\sqrt{n}]$，即

$$[40-\frac{1.96}{4},\ 40+\frac{1.96}{4}]=[39.51,\ 40.49].$$

5. (8分) 解:

$$E\left(\frac{X_1}{\sum_{i=1}^{n}X_i}\right)=\frac{1}{n}.$$

装 订 线

6. (8分) 解: 设发生故障的设备数为X和维修工人人数为k, 则$X\sim B(300,0.01)$，由题意可知

$$P(X\le k)=\sum_{i=0}^{k}C_{300}^{i}0.01^{i}0.99^{300-i}\ge 0.99$$

由泊松分布近似可知

$$P(X\le k)\approx\sum_{i=0}^{k}\frac{3^i}{i!}e^{-3}\ge 0.99$$

计算可知，$k=8$，至少需配备8工人, 才能保证设备发生故障但不能及时维修的概率小于0.01.

注：此题也可以用正态分布近似计算$k=8$。

二. (20分) 解：

(1) (ξ,η)的联合密度为$p(x,y)=2,\ 0<x\le y<1;\ p(x,y)=0$, 其它.

(2) ξ的边缘密度$p_1(x)=\int_x^1 2dy=2(1-x),\ 0<x<1;$

η的边缘密度$p_2(y)=\int_0^y 2dx=2y,\ 0<y<1.$

(3)当$0<y<1$时, 条件密度$p(x|\eta=y)=\frac{p(x,y)}{p_2(y)}=\frac{1}{y},\ 0<x\le y.$

(4) $E(\xi|\eta=y)=\int_0^y xp(x|\eta=y)dx=\int_0^y\frac{x}{y}dx=\frac{y}{2}.$

三. (15分)解：

(1) 由$E(X)=0\times\frac{\theta}{4}+1\times(1-\theta)+2\times\frac{3\theta}{4}=1+\frac{\theta}{2}$可得,

$$\hat{\theta}_1=2(\bar{X}-1),$$

其中$\bar{X}$是样本均值. 代入观测数据可知, $\hat{\theta}_1=0.66.$

令n_i表示样本中取值为i的个数, $i=0,1,2,$则样本量$n=n_0+n_1+n_2$. 那么似然函数为

$$L(\theta)=\left(\frac{\theta}{4}\right)^{n_0}(1-\theta)^{n_1}\left(\frac{3\theta}{4}\right)^{n_2}.$$

于是, 解方程

$$\frac{\partial\log(L(\theta))}{\partial\theta}=\frac{n_0}{\theta}-\frac{n_1}{1-\theta}+\frac{n_2}{\theta}=0,$$

得到

$$\hat{\theta}_2=\frac{n-n_1}{n}.$$

代入观测数据可知, $\hat{\theta}_2=0.67.$

(2) $E(\hat{\theta}_1)=E(2(\bar{X}-1))=2(E(\bar{X})-1)=2(E(X)-1)=\theta$, 由此可知$\hat{\theta}_1$ 是θ的无偏估计. $E(\hat{\theta}_2)=E(\frac{n-n_1}{n})=\frac{n-E(n_1)}{n}=\frac{n-n(1-\theta)}{n}=\theta$, 所以$\hat{\theta}_2$ 也是θ的无偏估计.

(3) $E(X^2)=0\times\frac{\theta}{4}+1\times(1-\theta)+4\times\frac{3\theta}{4}=1+2\theta$, 则$Var(X)=E(X^2)-(E(X))^2=$

$1+2\theta-(1+\frac{\theta}{2})^2=\theta-\frac{\theta^2}{4}$. 因此,

$$Var(\hat{\theta}_1)=Var(2(\bar{X}-1))=4Var(\bar{X})=4Var(X)/n=\frac{4\theta-\theta^2}{n}.$$

很容易算出$Var(n_1)=n(1-\theta)\theta$, 则

$$Var(\hat{\theta}_2)=Var(\frac{n-n_1}{n})=\frac{Var(n_1)}{n^2}=\frac{\theta-\theta^2}{n}.$$

因为$Var(\hat{\theta}_2)<Var(\hat{\theta}_1)$,所以估计$\hat{\theta}_2$比$\hat{\theta}_1$更有效.

四. (10分) 解：设原假设为: 作品《理智与情感》,《爱玛》以及《劝导》之间在选择常用词比例没有差异. 在原假设下,《理智与情感》,《爱玛》以及《劝导》选择常用词比例的期望值为 那么, 检验统计量为,

单词	理智与情感	爱玛	劝导
a	159.32368	190.33299	167.34333
an	28.04343	33.50155	29.45502
this	31.12513	37.18304	32.69183
that	79.50776	94.98242	83.50982

$$T=\sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(n_{ij}-E(n_{ij}))^2}{E(n_{ij})}=19.722,$$

自由度为6. 由于$T>\chi^2_{0.05}(6)=12.591$,则拒绝原假设,即作品《理智与情感》,《爱玛》以及《劝导》之间在选择常用词比例有显著差异.

五. (15分) 解: 分别用X和Y表示甲和乙两处矿石的含灰率, 且由题意知$X\sim N(\mu_1,\sigma_1^2)$, $Y\sim N(\mu_2,\sigma_2^2)$. 利用数据可得, $\hat{\mu}_1=16$, $\hat{\mu}_2=19$, $\hat{\sigma}_1^2=5.6$ 和$\hat{\sigma}_2^2=4$.

(1)假设检验问题为: $H_0:\sigma_1^2=\sigma_2^2\longleftrightarrow H_1:\sigma_1^2\neq\sigma_2^2$. F检验统计量为

$$F=\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}=1.4,$$

自由度为$(5,5)$. 因为$F_{0.025}(5,5)=7.146$ 和$F_{0.975}(5,5)=1/F_{0.025}(5,5)=0.140$, 所以$F_{0.975}(5,5)<F<F_{0.025}(5,5)$, 则接受原假设, 即甲和乙两处矿石的含灰率的方差相等.

(2)假设检验问题为: $H_0:\mu_1=\mu_2\longleftrightarrow H_1:\mu_1<\mu_2$. 记$S_w^2=(5\hat{\sigma}_1^2+5\hat{\sigma}_2^2)/10=4.8$, 两样本$t$检验统计量为

$$T=\frac{\hat{\mu}_1-\hat{\mu}_2}{S_w\sqrt{1/6+1/6}}=-2.3717,$$

自由度为10. 因为$t_{0.05}(10)=1.812$, 所以$T<-t_{0.05}(10)$, 则拒绝原假设,即乙处矿石含灰率的平均量显著地高于甲处矿石含灰率的平均量.

中 国 科 学 技 术 大 学

2019—2020学年第一学期考试试卷

考试科目 概率论与数理统计(B) 得分 ____________

所在系 _______ 姓名 ____________ 学号 ____________

考试时间: 2020年1月13日上午8:30–10:30; 使用简单计算器

装 …… 订 …… 线

一、 (30分, 每小题3分) **填空题或单选题, 答案可以直接写在试卷上.**

(1) 设 $\mathrm{P}(A)=\mathrm{P}(B)=0.4$, 且 $\mathrm{P}(B|A)+\mathrm{P}(\overline{B}|\overline{A})=1$, 则 $\mathrm{P}(AB)=$ ________.

(2) 甲乙二人抛掷一枚均匀的硬币, 甲抛了101 次, 乙抛了100 次, 则甲抛出的正面次数比乙多的概率是________.

(3) 设随机变量 X 的密度函数为 $f(x)=\frac{2}{\pi}\sqrt{1-x^2}$, $x\in(-1,1)$. 对任意 $x\in(-1,1)$, 若在条件$X=x$下, 随机变量 Y 的条件分布律为

$$\mathrm{P}(Y=-\sqrt{1-x^2})=\mathrm{P}(Y=\sqrt{1-x^2})=1/2,$$

则 Y __ 连续型随机变量, (X,Y) __ 连续型随机向量. ( )
(A) 是, 是 (B) 是, 不是 (C) 不是, 是 (D) 不是, 不是

(4) 在单位圆盘 $\{(x,y): x^2+y^2\le 1\}$ 上随机取两个点, 以随机变量 X 表示它们之间的距离, 则 $\mathrm{E}(X^2)=$ ________.

(5) 设 $X_1,X_2,\cdots,X_n$ 为一列独立同分布的随机变量, 且均服从参数为 $\lambda>0$ 的指数分布. 记 $\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i$ 且 $\Phi(x)$ 为标准正态分布函数, 则对任意 $x\in\mathbb{R}$, 有( )
(A) $\lim\limits_{n\to\infty}\mathrm{P}\left(\frac{\sqrt{n}}{\lambda}(\overline{X}-\lambda)\le x\right)=\Phi(x)$ (B) $\lim\limits_{n\to\infty}\mathrm{P}\left(\sqrt{\frac{n}{\lambda}}(\overline{X}-\lambda)\le x\right)=\Phi(x)$
(C) $\lim\limits_{n\to\infty}\mathrm{P}\left(\sqrt{n}(\lambda\overline{X}-1)\le x\right)=\Phi(x)$ (D) $\lim\limits_{n\to\infty}\mathrm{P}\left(\sqrt{n\lambda}(\overline{X}-\frac{1}{\lambda})\le x\right)=\Phi(x)$

(6) 设 $X_1,X_2,\cdots,X_n$ 是来自标准正态总体的简单随机样本, 且 $1\le m<n$, 则当常数 $c=$ ________ 时, 统计量 $c\,(\sum_{i=1}^m X_i)^2/\sum_{i=m+1}^n X_i^2$ 服从 F 分布.

(7) 设 $X_1,X_2,\cdots,X_n$ 是来自正态总体 $N(\mu,\sigma^2)$ 的简单随机样本, 其中 μ 为已知常数, 记 $\overline{X}$ 和 S^2 分别为样本均值和样本方差, 则下列统计量中与 $\overline{X}$ 不独立的是( )
(A) 样本标准差 S (B) $\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})^2$ (C) $\frac{1}{n}\sum_{i=1}^n(X_i-\mu)^2$ (D) X_1-X_2

(8) 设 X_1,X_2,X_3 是来自正态总体 $N(\mu,\sigma^2)$ 的简单随机样本, 则下列统计量中, ( )为 μ 的无偏估计且方差最小.
(A) $\frac{1}{2}X_1+\frac{1}{3}X_2+\frac{1}{6}X_3$ (B) $\frac{1}{3}X_1+\frac{1}{3}X_2+\frac{1}{3}X_3$
(C) $\frac{1}{5}X_1+\frac{2}{5}X_2+\frac{2}{5}X_3$ (D) $\frac{1}{7}X_1+\frac{2}{7}X_2+\frac{3}{7}X_3$

(9) 对一正态总体 $N(\mu,100)$ 的均值 μ 求置信水平为 95% 的置信区间, 若要求其区间长度不大于 4, 则样本容量 n 至少应取_____.

(10) 假设检验中, 在显著性水平$\alpha=0.05$ 下若原假设 H_0 被接受, 这说明( )
(A) 有充分的理由表明 H_0 是正确的 (B) 没有充分的理由表明 H_0 是错误的
(C) 有充分的理由表明 H_1 是错误的 (D) 没有充分的理由表明 H_1 是正确的

二、 (20分)设 $X_1, X_2, \cdots, X_n$ 为一列独立的随机变量, 且均服从 $U(0,1)$ 分布. 记

$$Y = \min\{X_1, X_2, \cdots, X_n\}, \quad Z = \max\{X_1, X_2, \cdots, X_n\}.$$

(1) 试证明: 对任意常数 $0 < y, z < 1$, 有

$$\mathrm{P}(Y \le y, Z \le z) = \begin{cases} z^n - (z-y)^n, & y < z; \\ z^n, & y \ge z. \end{cases}$$

(2) 利用上述结果, 试求随机变量 Y 和 Z 的联合密度函数 $f(y,z)$.

(3) 在 $Y = y$ 条件下 $(0 < y < 1)$, 试求 Z 的条件密度函数 $f_{Z|Y}(z|y)$.

(4) 若 $n = 2$, 试求 Y 和 Z 的协方差 $\mathrm{Cov}(Y, Z)$.

三、 (15分)设随机变量 X, Y 和 Z 相互独立, 且均服从参数为 1 的指数分布. 记

$$U = \frac{X}{X+Y}, \quad V = \frac{X+Y}{X+Y+Z}, \quad W = X+Y+Z.$$

(1) 计算随机向量 (U, V, W) 的联合密度函数.

(2) 随机变量 U, V 和 W 是否相互独立? 请证明你的结论.

四、 (15分)设某种元件的使用寿命 T 的分布函数为 $F(t) = \begin{cases} 1 - \exp\{-(\frac{t}{\theta})^m\}, & t \ge 0; \\ 0, & t < 0, \end{cases}$ 其中 $m > 0$ 为已知参数, 而 $\theta > 0$ 为未知参数. 随机取 n 个这种元件, 测得它们的寿命分别为 $T_1, T_2, \cdots, T_n$. 记 $g(\theta) = \theta^m$.

(1) 试求 $g(\theta)$ 的极大似然估计 $\hat{g}(T_1, T_2, \cdots, T_n)$.

(2) 上述估计是否为无偏估计? 请证明你的结论.

五、 (12分) 经大量调查, 已知一般健康成年男子每分钟脉搏的次数服从正态分布 $N(72, 6^2)$. 现测得 16 例成年男子慢性铅中毒患者的脉搏平均 67 次/分钟, 标准差为 7 次/分钟. 问在显著性水平 0.05 下, 这群患者每分钟脉搏的次数(假设也服从正态分布)和正常人有无显著性差异? (要求对均值和方差都进行检验.)

六、 (8分) 中国科学技术大学 2019 级本科新生入学考试中, 某学院两个班级的英语科目各档成绩(从低到高)人数如下表所示:

档次	I	II	III	IV	V	VI	合计
一班	8	27	10	6	8	6	65
二班	15	25	8	7	6	4	65

我们能否认为这两个班级的英语水平大致相当? 显著性水平设为 $\alpha = 0.05$.

附录:

$\Phi(1.645) = 0.95,\ \Phi(1.96) = 0.975;$

$t_{15}(0.025) = 2.131,\ t_{15}(0.05) = 1.753,\ t_{16}(0.025) = 2.12,\ t_{16}(0.05) = 1.746;$

$\chi^2_5(0.95) = 1.145,\ \chi^2_5(0.05) = 11.071,\ \chi^2_{15}(0.975) = 6.262,\ \chi^2_{15}(0.025) = 27.488.$

参考答案

一. (1) 0.16　(2) 0.5　(3) B　(4) 1　(5) C
(6) $\frac{n-m}{m}$　(7) C　(8) B　(9) 97 (若答 96 也算对)　(10) B.

二. (1) 当 $0<y,z<1$ 时, 由 $X_1,\cdots,X_n$ 独立同分布可知

$$\begin{aligned}\mathrm{P}(Y\le y,Z\le z) &= \mathrm{P}(Z\le z)-\mathrm{P}(Y>y,Z\le z)\\ &= \mathrm{P}(X_1\le z,\cdots,X_n\le z)-\mathrm{P}(y<X_1\le z,\cdots,y<X_n\le z)\\ &= [\mathrm{P}(X_1\le z)]^n-[\mathrm{P}(y<X_1\le z)]^n.\end{aligned}$$

再由 $X_1\sim U(0,1)$ 即知

$$\mathrm{P}(Y\le y,Z\le z)=\begin{cases} z^n-(z-y)^n, & y<z;\\ z^n, & y\ge z.\end{cases}$$

(2) 注意到联合密度函数 $f(y,z)$ 的取值范围是 $0<y<z<1$, 将 (1) 中联合分布函数 $\mathrm{P}(Y\le y,Z\le z)$ 对变量 y 和 z 求一阶偏导数, 即得

$$f(y,z)=n(n-1)(z-y)^{n-2},\quad 0<y<z<1.$$

(如果上述表达式正确, 但没有说明取值范围为 $0<y<z<1$, 扣 1–2 分.)
(3) 由 (2) 可知, 随机变量 Y 的密度函数为

$$f_Y(y)=\int_y^1 f(y,z)\mathrm{d}z=n(1-y)^{n-1},\quad 0<y<1.$$

从而

$$f_{Z|Y}(z|y)=\frac{f(y,z)}{f_Y(y)}=\frac{(n-1)(z-y)^{n-2}}{(1-y)^{n-1}},\quad y<z<1.$$

(如果上述表达式正确, 但没有说明取值范围为 $y<z<1$, 扣 1–2 分.)
(4) 当 $n=2$ 时, 随机变量 Y 的密度函数为 $f_Y(y)=2(1-y)$, $0<y<1$, 故

$$\mathrm{E}Y=\int_0^1 2y(1-y)\mathrm{d}y=\frac{1}{3}.$$

类似地, 可求得随机变量 Z 的密度函数为 $f_Z(z)=2z$, $0<z<1$, 及 $\mathrm{E}Z=\frac{2}{3}$. 此外, 由于此时联合密度函数退化为 $f(y,z)=2$, $0<y<z<1$, 我们有

$$\mathrm{E}[YZ]=\int_0^1\int_y^1 2yz\mathrm{d}z\mathrm{d}y=\frac{1}{4}.$$

所以,

$$\mathrm{Cov}(Y,Z)=\mathrm{E}[YZ]-\mathrm{E}Y\cdot\mathrm{E}Z=\frac{1}{4}-\frac{1}{3}\cdot\frac{2}{3}=\frac{1}{36}.$$

三. (1) 由

$$u=\frac{x}{x+y},\quad v=\frac{x+y}{x+y+z},\quad w=x+y+z,$$

可得 $x=uvw,\ y=(1-u)vw,\ z=(1-v)w$, 从而Jacobi行列式

$$J=\begin{vmatrix} vw & uw & uv \\ -vw & (1-u)w & (1-u)w \\ 0 & -w & 1-v \end{vmatrix}=vw^2.$$

由 (X,Y,Z) 的联合密度函数 $f(x,y,z)=\mathrm{e}^{-(x+y+z)},\ x,y,z>0$, 及密度变换公式可得所求随机向量 (U,V,W) 的联合密度函数为

$$p(u,v,w)=vw^2\mathrm{e}^{-w},\quad 0<u,v<1,\ w>0.$$

(如果上述表达式正确, 但没有说明取值范围, 扣 1–2 分.)

(2) 随机变量 U,V 和 W 相互独立. 事实上, 由上述联合密度函数可以分解为

$$p(u,v,w)=p_U(u)p_V(v)p_W(w)$$

即知该结论成立, 其中

$$p_U(u)=1,\ 0<u<1;\quad p_V(v)=2v,\ 0<v<1;\quad p_W(w)=\frac{1}{2}w^2\mathrm{e}^{-w},\ w>0.$$

四. (1) 由总体 T 的概率密度函数为 $f(t)=\frac{mt^{m-1}}{\theta^m}\exp\{-(\frac{t}{\theta})^m\},\ t\geq 0$, 故似然函数为

$$L(\theta)=\frac{m^n}{\theta^{mn}}\prod_{i=1}^{n}t_i^{m-1}\exp\Big\{-\Big(\frac{t_i}{\theta}\Big)^m\Big\}.$$

对其取对数, 得对数似然函数为

$$l(\theta)=\ln L(\theta)=C-n\ln(\theta^m)-\frac{1}{\theta^m}\sum_{i=1}^{n}t_i^m.$$

将上式对 θ^m 求导数, 并令其等于0, 可得(也可对 θ 求导, 结果相同)

$$\frac{\mathrm{d}l(\theta)}{\mathrm{d}(\theta^m)}=-\frac{n}{\theta^m}+\frac{1}{\theta^{2m}}\sum_{i=1}^{n}t_i^m=0.$$

解之即可得所求极大似然估计量为

$$\hat{g}(T_1,T_2,\cdots,T_n)=\frac{1}{n}\sum_{i=1}^{n}T_i^m.$$

(2) 由

$$\begin{aligned}\mathrm{E}[\hat{g}(T_1,T_2,\cdots,T_n)]&=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[T_i^m]=\mathrm{E}[T^m]\\ &=\int_0^{\infty}\frac{mt^{2m-1}}{\theta^m}\exp\Big\{-\Big(\frac{t}{\theta}\Big)^m\Big\}\mathrm{d}t=\theta^m\int_0^{\infty}x\mathrm{e}^{-x}\mathrm{d}x=\theta^m,\end{aligned}$$

可知, $\hat{g}(T_1,T_2,\cdots,T_n)$ 是 $g(\theta)$ 的一个无偏估计.

五. (1) 均值的检验. $H_0 : \mu = 72, \quad \longleftrightarrow \quad H_1 : \mu \neq 72$. 由 t 检验统计量

$$t = \frac{\sqrt{n}(\overline{X} - \mu_0)}{S} = \frac{\sqrt{16}(67 - 72)}{7} = -2.857,$$

可知 $|t| > t_{15}(0.025) = 2.131$, 故应拒绝原假设, 即认为患者每分钟脉搏的平均次数与正常人有显著性差异.

(2) 方差的检验. $H_0 : \sigma^2 = 6^2, \quad \longleftrightarrow \quad H_1 : \sigma^2 \neq 6^2$. 由 χ^2 检验统计量

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{15 \cdot 7^2}{6^2} = 20.417,$$

可知 $\chi^2_{15}(0.975) = 6.262 < \chi^2 < \chi^2_{15}(0.025) = 27.488$, 故应接受原假设, 即认为患者每分钟脉搏次数的方差与正常人相同. 结合均值和方差两个方面, 我们最终可认为患者每分钟脉搏的次数与正常人有显著性差异.

(注: 如果先进行方差的检验, 认定方差可以等于 6^2, 然后利用一样本 u 检验也算正确答案, 均值的检验结果与上面相同.)

六. 拟合优度联列表齐一性检验. 原假设为两个班级的英语水平相当, 而其检验统计量为

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{6} \frac{(nn_{ij} - n_{i\cdot}n_{\cdot j})^2}{nn_{i\cdot}n_{\cdot j}}.$$

代入数据计算可知, $\chi^2 = 3.1922 < \chi^2_5(0.05) = 11.071$, 故在显著性水平 $\alpha = 0.05$ 下我们不能拒绝原假设, 即可认为两个班级的英语水平相当.

中 国 科 学 技 术 大 学

2020—2021学年第一学期考试试卷

考试科目 概率论与数理统计(B) 得分 __________

所在院系 _______ 姓名 __________ 学号 __________

考试时间: 2021 年 3 月 6 日上午 8:30–10:30; 可使用简单计算器

一、 (30分, 每小题3分) **填空题或单选题, 答案可以直接写在试卷上.**

(1) 设 A, B 为随机事件, 且 $0 < \mathrm{P}(B) < 1$, 下列为假命题的是( )

(A) 若 $\mathrm{P}(A|B) = \mathrm{P}(A)$, 则 $\mathrm{P}(A|\overline{B}) = \mathrm{P}(A)$

(B) 若 $\mathrm{P}(A|B) > \mathrm{P}(A)$, 则 $\mathrm{P}(\overline{A}|\overline{B}) > \mathrm{P}(\overline{A})$

(C) 若 $\mathrm{P}(A|B) > \mathrm{P}(A|\overline{B})$, 则 $\mathrm{P}(A|B) > \mathrm{P}(A)$

(D) 若 $\mathrm{P}(A|A \cup B) > \mathrm{P}(\overline{A}|A \cup B)$, 则 $\mathrm{P}(A) > \mathrm{P}(B)$

(2) 设平面上有 n 个点, 编号分别为 $1, 2, \cdots, n$. 现一质点在此点集上做随机游动, 每次它在一点上停留片刻后就会在其余各点中随机地选择一个并移动到该点上. 已知其初始位置为点 1, 则它在第一次返回点 1 之前访问过点 2 的概率是_______.

(3) 设随机变量 X 服从参数为 $0 < p < 1$ 的几何分布, 且在条件 $X = k$ 下, Y 服从参数为 k 的指数分布. 对任一实数 $y > 0$, 则 $\mathrm{P}(Y > y) =$( ).

(A) $\mathrm{e}^{-y/p}$ (B) $p\mathrm{e}^{-y}$ (C) $p/(y+p)$ (D) $p/(\mathrm{e}^y - 1 + p)$

(4) 若随机变量 X 和 Y 满足 $\mathrm{P}(X^2 + Y^2 = 2) = 1$, 则下列说法中一定不成立的是( )

(A) (X, Y) 为连续型随机向量 (B) $\mathrm{P}(X + Y = 0) = 1/2$

(C) $\mathrm{E}X = \mathrm{E}Y = 0$ (D) X 和 Y 相互独立

(5) 设 X 和 Y 为相互独立的标准正态随机变量, 则 $\mathrm{P}(\max\{X, Y\} \geq 0) =$_______.

(6) 设随机变量 X 和 Y 分别服从参数为 λ 和 μ 的 Poisson 分布, 且相互独立. 对任一非负整数 n, 则条件期望 $\mathrm{E}[X|X+Y=n] =$_______.

(7) 设随机向量 (X, Y) 服从二维正态分布 $N(a, a; \sigma^2, \sigma^2; 0)$, 则 $\mathrm{Cov}(X, XY^2) =$______.

(8) 设 $X_1, X_2, \cdots, X_n, X_{n+1}$ 是来自总体 $N(\mu, \sigma^2)$ 的简单随机样本, 且 $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$, $m_2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$. 已知 $T = c(X_{n+1} - \overline{X})/\sqrt{m_2}$ 服从 t 分布, 则 $c =$_______.

(9) 设正态总体 $X \sim N(\mu, \sigma^2)$, 其中 σ^2 未知, 若样本容量 n 和置信水平 $1 - \alpha$ 均保持不变, 对不同的样本观察值, 则总体均值 μ 的置信区间长度( )

(A) 与样本均值有关 (B) 与 μ 本身有关 (C) 保持不变 (D) 不确定

(10) 设 $X_1, X_2, \cdots, X_{16}$ 是来自总体 $N(\mu, 4)$ 的简单随机样本, $\overline{X}$ 为其样本均值, $\Phi(x)$ 为标准正态分布函数. 考虑假设检验问题: $H_0 : \mu \leq 10 \leftrightarrow H_1 : \mu > 10$. 若其拒绝域为 $W = \{\overline{X} \geq 11\}$, 则当 $\mu = 11.5$ 时, 该检验犯第二类错误的概率为( )

(A) $1 - \Phi(0.5)$ (B) $1 - \Phi(1)$ (C) $1 - \Phi(1.5)$ (D) $1 - \Phi(2)$

二、 (10分) 设有两个罐子, 一罐中有 m 个红球和 n 个黑球, 另一罐中有 n 个红球和 m 个黑球, 且 $m > n$. 某人随机选取一个罐子并从中随机抽取一球, 发现为红球. 现将该球放回原罐后并摇匀, 然后再次在此罐中随机抽取一球, 则它仍为红色的概率是否比 $1/2$ 大? 通过计算事件的概率来证明你的结论.

三、 (20分) 将区间 $(0,2)$ 随机截成两段, 记较短一段的长度为 X, 较长一段的长度为 Y.

(1) 求 X 和 Y 的相关系数 $\mathrm{Corr}(X,Y)$;

(2) 求 X 的概率密度函数 $f(x)$;

(3) 求 $Z = \frac{Y}{X}$ 的概率密度函数 $g(z)$;

(4) 设随机变量 X^* 与 X 独立同分布, 试求 $V = 2|X - X^*|$ 的概率密度函数 $h(v)$.

四、 (16分) 已知总体 X 的概率密度函数为 $f(x) = (\theta+1)x^\theta$, $0 < x < 1$, 其中 $\theta > -1$ 为一未知参数. 设 $X_1, X_2, \cdots, X_n$ 是来自该总体的一组简单随机样本.

(1) 求 θ 的矩估计量 $\hat{\theta}$;

(2) 求 $g(\theta) = \frac{1}{\theta+1}$ 的极大似然估计量 $\hat{g}$;

(3) 问 $\hat{g}$ 是否为 $g(\theta)$ 的一个无偏估计? 证明你的结论.

(4) 求常数 b, 使得对任意实数 x, 都有 $\lim\limits_{n\to\infty} \mathrm{P}\left(\sqrt{n}(\hat{g} - g(\theta))/b \le x\right) = \Phi(x)$ 成立, 其中 $\Phi(x)$ 为标准正态分布函数.

五、 (14分) 在 1970 年代后期, 人们发现酿造啤酒时麦芽干燥的过程中会形成致癌物质亚硝基二甲胺(NDMA). 在 1980 年代初期为此开发了一种新麦芽干燥工艺. 独立地随机抽查了新旧工艺下各一组样本, 得到NDMA含量(以10亿份中的份数计)的结果如下:

旧工艺	6	4	5	5	6	5	5	6	4	6	7	4
新工艺	2	1	2	2	1	0	3	2	1	0	1	3

设旧、新工艺下的两样本均来自正态总体. 在显著性水平 $\alpha = 0.05$ 下,

(1) 是否可以认为两个总体的方差相等?

(2) 是否可以认为旧工艺下NDMA平均含量比新工艺下显著地大 3?

六、 (10分) 某种鸟在起飞前, 双足齐跳的次数 X 服从参数为 p 的几何分布, 即其分布律为 $\mathrm{P}(X = k) = p(1-p)^{k-1}$, $k = 1, 2, \cdots$. 某人观测 130 次后, 获得一组样本如下:

k	1	2	3	4	5	6	7	8	9	10	11	12	≥ 13
频数	48	31	20	9	6	5	4	2	1	1	2	1	0

(1) 求 p 的最大似然估计值(精确到小数点后三位);

(2) 在拟合优度检验中频数一般不能小于 5, 故需将上述所有 $k \ge 7$ 情形下的频数进行合并, 此时请检验假设“X 服从几何分布”是否成立(显著性水平 $\alpha = 0.05$).

附录: $t_{22}(0.025) = 2.074$, $t_{22}(0.05) = 1.717$, $t_{23}(0.025) = 2.069$, $t_{23}(0.05) = 1.714$
$F_{11,11}(0.025) = 3.474$, $F_{11,11}(0.05) = 2.818$, $F_{12,12}(0.025) = 3.277$, $F_{12,12}(0.05) = 2.687$
$\chi^2_5(0.05) = 11.071$, $\chi^2_5(0.95) = 1.145$, $\chi^2_6(0.05) = 12.592$, $\chi^2_6(0.95) = 1.635$

参考答案

一. (1) D　(2) $\frac{n}{2(n-1)}$　(3) D　(4) A　(5) $\frac{3}{4}$　(6) $\frac{n\lambda}{\lambda+\mu}$　(7) $(a^2+\sigma^2)\sigma^2$　(8) $\sqrt{\frac{n-1}{n+1}}$
(9) D　(10) B

二. 以 A 表示选取的罐子为甲罐 (m 红 n 黑) 的事件, B 表示第一次取出的球为红球的事件, 则由 Bayes 公式可知

$$\mathrm{P}(A|B)=\frac{\mathrm{P}(B|A)\mathrm{P}(A)}{\mathrm{P}(B|A)\mathrm{P}(A)+\mathrm{P}(B|A^c)\mathrm{P}(A^c)}=\frac{m}{m+n}.$$

再由全概率公式可知, 第二次抽取的球仍为红色的概率为

$$\frac{m}{m+n}\mathrm{P}(A|B)+\frac{n}{m+n}\mathrm{P}(A^c|B)=\frac{m^2+n^2}{(m+n)^2}>\frac{1}{2}.$$

三. (1) 由 $\mathrm{P}(X+Y=2)=1$ 立知 $\mathrm{Corr}(X,Y)=-1$.

(2) 设随机变量 $U\sim U(0,2)$, 而 X 取值范围为 $(0,1)$, 故对任意 $0<x<1$,

$$\mathrm{P}(X\le x)=\mathrm{P}(U\le x)+\mathrm{P}(2-U\le x)=\frac{x}{2}+\frac{x}{2}=x,$$

即 $X\sim U(0,1)$. 故 X 的概率密度函数 $f(x)=1$, $0<x<1$.

(3) 易知 $Z=\frac{Y}{X}=\frac{2-X}{X}$ 及 $X=\frac{2}{Z+1}$, 由 (2) 和密度变换公式可知 $g(z)=\frac{2}{(z+1)^2}$, $z>1$.

(4) 利用密度变换公式或者几何概型, 可知 V 的概率密度函数为

$$h(v)=\begin{cases}1-\dfrac{v}{2}, & 0<v<2,\\ 0, & \text{else.}\end{cases}$$

注：若上述密度函数表达式中变量范围缺乏或不正确, 按每处扣分.

四. (1) 由 $\mathrm{E}X=\frac{\theta+1}{\theta+2}$, 解方程 $\frac{\theta+1}{\theta+2}=\overline{X}$ 可知 θ 的矩估计量 $\hat{\theta}=\frac{2\overline{X}-1}{1-\overline{X}}$.

(2) 由题意, 似然函数 $L(\theta)=(\theta+1)^n\prod_{i=1}^n x_i^\theta$, 从而对数似然函数为

$$l(\theta)=\ln L(\theta)=n\ln(\theta+1)+\theta\sum_{i=1}^n\ln x_i.$$

令 $\frac{\mathrm{d}l(\theta)}{\mathrm{d}\theta}=0$, 可得 $\frac{n}{\theta+1}+\sum_{i=1}^n\ln x_i=0$. 由此可知, $g(\theta)=\frac{1}{\theta+1}$ 的极大似然估计量

$$\hat{g}=-\frac{1}{n}\sum_{i=1}^n\ln X_i.$$

(3) 记 $Y_i=-\ln X_i$, $i=1,2,\cdots,n$, 则易知 $\{Y_i,1\le i\le n\}$ 独立同分布于参数为 $\theta+1$ 的指数分布, 由此即知 $\mathrm{E}\hat{g}=\frac{1}{\theta+1}$. 故 $\hat{g}$ 是 $g(\theta)$ 的一个无偏估计.

(4) 由上可知, $\hat{g}$ 可表示为一列独立同分布随机变量 $\{Y_i,1\le i\le n\}$ 的平均, 故由经典场合下的中心极限定理可知, 常数 $b=\sqrt{\mathrm{Var}(Y_1)}=\frac{1}{\theta+1}$.

五. 先计算一些统计量的值. 旧工艺: $n_1 = 12$, $\overline{x} = 5.25$, $(n_1-1)S_1^2 = 10.25$; 新工艺: $n_2 = 12$, $\overline{y} = 1.5$, $(n_2-1)S_2^2 = 11$; $S_w^2 = 0.983^2$.

(1) $H_0 : \sigma_1^2 = \sigma_2^2 \ \leftrightarrow\ H_1 : \sigma_1^2 \neq \sigma_2^2$.
由

$$\frac{1}{3.474} = \frac{1}{F_{11,11}(0.025)} < \frac{S_1^2}{S_2^2} = 0.932 < F_{11,11}(0.025) = 3.474,$$

接受 H_0, 即可以认为两个总体的方差相等.

(2) $H_0 : \overline{X} - \overline{Y} \leq 3 \ \leftrightarrow\ H_1 : \overline{X} - \overline{Y} > 3$.
由

$$t = \frac{\overline{X} - \overline{Y} - 3}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 1.867 > t_{22}(0.05) = 1.717,$$

拒绝 H_0, 即可以认为旧工艺下NDMA平均含量比新工艺下显著地大 3.

六. (1) 对几何分布总体, 易知其参数 p 的极大似然估计值为

$$\hat{p} = \frac{1}{\overline{X}} = \frac{130}{363} = 0.358.$$

(2) 合并后的数据为

k	1	2	3	4	5	6	≥ 7
频数	48	31	20	9	6	5	11

从而检验统计量

$$\chi^2 = \frac{(48 - 130 \times \hat{p})^2}{130 \times \hat{p}} + \frac{[31 - 130 \times \hat{p}(1-\hat{p})]^2}{130 \times \hat{p}(1-\hat{p})} + \cdots + \frac{[11 - 130 \times (1-\hat{p})^6]^2}{130 \times (1-\hat{p})^6}$$
$$= 1.868 < \chi_5^2(0.05) = 11.071.$$

故接受原假设.

中 国 科 学 技 术 大 学

2019—2020学年第一学期考试试卷

考试科目 概率论与数理统计 得分 ____________
所在系 ________ 姓名 ____________ 学号 ____________

考试时间: 2020年1月9日下午2:30—4:30; 使用简单计算器

装 …… 订 …… 线

一. 填空判断选择题 (每题3分,答题请写在试卷上):

1 设$Z(t) = X + Yt$, 其中X与Y相互独立，且都服从$N(0, \sigma^2)$, 则$Z(t)$与$Z(s)(s < t)$的相关系数= ______________.

2 三人独立地去破译一份密码，已知各人能译出的概率分别为1/5, 1/3, 1/4, 则破译出密码的概率是__________________.

3 设随机变量 X 和 Y 都服从正态分布, 且他们不相关, 则 _________

(A) X 与 Y一定独立 (B) (X, Y)服从二元正态

(C) X和Y未必独立 (D) $X + Y$服从一维正态分布

4 设$X_1, \ldots, X_{16}$服从正态分布$N(\mu, 0.16)$, 检验问题

$$H_0 : \mu = 0.5 \leftrightarrow H_1 : \mu > 0.5,$$

显著水平为0.05. 检验的拒绝域是______________, $\mu = 0.65$时犯第二类错误的概率为 ______________.

5 设$X \sim N(\mu, \sigma^2)$, 样本$X_1, \cdots, X_n$来自总体X的一组样本, $\overline{X}$是样本均值, 则$c =$ ______________时, $c(X_1 + X_2 + X_3 - 3\overline{X})^2$为$\sigma^2$的无偏估计.

6 总体参数θ的置信水平为95%的置信区间是指______________.

(A) θ落在该区间的概率为95% (B) θ落在该区间的概率为5%

(C) 在用同样方法构造的多个区间中包含θ 的区间比例为95%

(D) 在用同样方法构造的多个区间中包含θ 的区间比例为5%

7 用一组有30个观测值的样本去估计模型$y = a + bx + \epsilon$. 在0.05的显著性水平下对b的显著性作t检验, 则b显著不为0 的条件是其统计量T ______________.

(A) $T > t_{29}(0.05)$ (B) $|T| > t_{29}(0.025)$

(C) $T > t_{28}(0.05)$ (D) $|T| > t_{28}(0.025)$

8 设$X_1, \ldots, X_n$为来自概率密度函数为$f(x) = \theta x^{\theta-1} I(0 < x < 1)$的总体的样本, θ未知, 则$U = e^{-1/\theta}$的极大似然估计为______________.

9 设总体$X \sim N(\mu, \sigma^2)$, σ^2未知, $x_1, \cdots, x_n$为来自X的样本值, 先对μ进行假设检验。若在显著性水平$\alpha = 0.01$下拒绝了$H_0 : \mu = \mu_0$, 则当显著性水平改为$\alpha = 0.05$时, 下列结论正确的是

(A) 必拒绝H_0 (B)必接受H_0

(C)第一类错误的概率变小 (D) 可能接受也可能拒绝H_0

二. (12分) 某次考试共有100道选择题，每道题都是4选1.

(1)设某考生一点也不懂, 完全靠蒙，求答对20题以上40题以下的概率是多少？

(2)设另一考生平常模拟考试的正确率时85%, 求他答对90题以上的概率有多大？

三. (15分) 设$X_1, \dots, X_n$为来自均匀总体$U(-\theta, \theta)$的简单随机样本,

(1) 求θ的矩估计$\hat{\theta}_1$和极大似然估计$\hat{\theta}_2$;

(2) $\hat{\theta}_1$ 和 $\hat{\theta}_2$是否是无偏的？若否，请修正。

(3) 请问修正后的估计那个更有效？

四. (15分) 博尔特科维茨研究了从1875年到1894年20年间14个军团（共计280个军团)士兵被马踢死的人数记录如下：

军团死亡人数	0	1	2	3	4	≥ 5
观察数	144	91	32	11	2	0

博尔特科维茨推断这个每年每军团死亡数是服从Poisson分布.

(1) 求出Poisson分布的参数λ的极大似然估计.

(2) 试检验在水平0.05下博尔特科维茨的推断是否正确？

五. (13分) 做一下试验比较人对红光或绿光的反应时间(单位s). 随机抽取了8人做实验, 结果如下:

红光	0.30	0.23	0.41	0.53	0.24	0.36	0.38	0.51
绿光	0.43	0.32	0.58	0.46	0.27	0.41	0.38	0.61

试问是否可以认为人对红光的反应比绿光快？ ($\alpha = 0.05$)

六. (15分) 考虑日本汇率与汽车出口数量的数据: (X:汇率(日元/美元), Y: 数量(万辆))

年度	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
X	168	145	128	138	145	135	127	111	102	94
Y	661	631	610	588	583	575	567	502	446	379

已知$\sum X_i = 1293$, $\sum Y_i = 5542$, $\sum X_i^2 = 171617$, $\sum Y_i^2 = 3139490$, $\sum X_i Y_i = 732776$. 满足经典线性回归模型的假设 ($\alpha = 0.05$).

(1) 求X与Y的样本相关系数.

(2) 求出一元线性回归方程, 即求a和b的最小二乘估计. 试给出参数的经济意义.

(3) 检验回归模型的水平α的显著性检验.

(4) 若1996年日元对美元是104.8, 预测当年的出口汽车数量和其$1-\alpha$的置信区间.

附录 分布及分位数: $\Phi(3.46) = 0.9997$, $\Phi(-2/\sqrt{3}) = 0.124, \Phi(1.4) = 0.92$, $u_{0.025} = 1.960$, $u_{0.05} = 1.645$, $t_{0.025}(13) = 2.16$, $t_{14}(0.025) = 2.145$, $t_{13}(0.05) = 1.771$, $t_7(0.05) = 1.89$, $t_7(0.025) = 2.36$, $t_8(0.025) = 2.306$, $t_9(0.025) = 2.262$

答案： 1. $\frac{1+ts}{\sqrt{(1+t^2)(1+s^2)}}$, 2. 3/5, 3. C, 4. $\bar{X} > 0.6645, \Phi(0.145)$ 5. $\frac{n}{3(n-3)}$ 6. C 7. D 8. $\exp\{\frac{1}{n}\sum \ln X_j\}$ 9. A

二、 (1)令X_i=I(第i题答对),

$$P(20 \leq \sum_{i=1}^{100} X_i \leq 40) = P(\frac{20-25}{\sqrt{100*0.25*0.75}} \leq \frac{\sum_{i=1}^{100} X_i - 25}{\sqrt{100*0.25*0.75}} \leq \frac{40-25}{\sqrt{100*0.25*0.75}})$$
$$= \Phi(2\sqrt{3}) - \Phi(-2/\sqrt{3}) = 0.8785.$$

(2)令Y_i=I(第i题答对),

$$P(\sum_{i=1}^{100} Y_i \geq 90) = P(\frac{\sum_{i=1}^{100} Y_i - 85}{\sqrt{100*0.85*0.15}} \geq \frac{90-85}{\sqrt{100*0.85*0.15}})$$
$$= 1 - \Phi(1.4) = 0.08.$$

三

(1)令$Y_i = |X_i|$,则$Y_i \sim U(0,\theta)$,$EY_i = \theta/2$.

所以矩估计$\hat{\theta}_1 = 2\bar{Y}$.

似然函数$L(y,\theta) = (\frac{1}{\theta})^n \prod_{i=1}^n I(0 < Y_i < \theta) = (\frac{1}{\theta})^n I(Y_{(n)} < \theta)$.

所以极大似然估计$\hat{\theta}_2 = \max Y_i = Y_{(n)}$.

(2)$E\hat{\theta}_1 = 2EY_1 = \theta$, 所以$\hat{\theta}_1$是无偏的.

$\hat{\theta}_2 \sim f(y) = 1/\theta(y/\theta)^{n-1}$, $E\hat{\theta}_2 = \int_0^\theta n(y/\theta)^n dy = \frac{n}{n+1}\theta$.

$\hat{\theta}_2$不是无偏的,修正$\hat{\theta}_2^* = \frac{n+1}{n}\hat{\theta}_2$.

(3)$var(\hat{\theta}_1) = \frac{4}{n}var(Y_1) = \frac{\theta^2}{3n}$,

$var(\hat{\theta}_2^*) = \frac{(n+1)^2}{n^2}(E\hat{\theta}_2^2 - (E\hat{\theta}_2)^2) = \frac{\theta^2}{n(n+2)}$,

$var(\hat{\theta}_1) \geq var(\hat{\theta}_2)$, 当且仅当$n=1$时取等, 所以修正后极大似然估计更有效.

四.

(1)令n_i为死亡人数为i的观察数,似然函数$L = \prod_i (e^{-\lambda}\frac{\lambda^i}{i!})^{n_i} = C_1 e^{-\lambda \sum_i n_i} \lambda^{\sum_i in_i}$,

对数似然函数$l = C_2 - \lambda \sum_i n_i + \sum_i in_i log\lambda$,

$\hat{\lambda} = \frac{\sum_i in_i}{n} = 0.7$.

(2)重新分组,

H_0: 服从参数为0.7的Poisson分布.

军团死亡人数	0	1	2	3	≥ 4
观察数ν_i	144	91	32	11	2
理论频数np_i	139.04	97.33	34.07	7.95	1.61

$\chi^2 = \sum_{i=1}^{5} \frac{(np_i - \nu_i)^2}{np_i} = 1.98$, $\chi^2 \le \chi_3^2(0.05) = 7.81$,无法拒绝$H_0$.

五.

记红光反应时间为X_i, 绿光反应时间为Y_i, $Z_i = X_i - Y_i$,

Z	-0.13	-0.09	-0.17	0.07	-0.03	-0.05	0	-0.10

H_0:$\mu \ge 0$, $H_1 : \mu < 0$

$\bar{Z} = -0.0625$, $S^2 = 0.00585$, $T = \frac{\sqrt{n}(\bar{Z}-\mu_0)}{S} = \frac{\sqrt{8}(-0.0625)}{\sqrt{0.00585}} = -2.31 < -t_7(0.05) = -1.85$.

拒绝H_0, 红光反应速度快.

六

(1)

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} = 0.932.$$

(2)

$$\hat{b} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = 3.654,$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} = 81.738.$$

日本汇率每增长1, 汽车出口数量增加3.654万辆.

(3) 提出假设

$H_0 : b = 0, H_1 : b \ne 0$

$$F = \frac{SSR/1}{SSE/(n-2)} = \frac{59179.84}{8933.76/8} = 52.99 > F_{1,8}(0.05) = 5.317.$$

拒绝H_0, 模型显著.

(4) $\hat{y}_0 = 81.738 + 3.654 * 104.8 = 464.677$,

$Se = \sqrt{\frac{SSE}{n-2}} = 33.417$,

预测区间为

$$\hat{y}_0 \pm t_{0.025}(n-2)Se\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{X})^2}{\sum(X_i - \bar{X})^2}}$$

[379.03,550.33].

中 国 科 学 技 术 大 学

2020—2021学年第一学期期中试卷

考试科目 概率论与数理统计 得分 ____________

所在系 _______ 姓名 ____________ 学号 ____________

考试时间: 2020年11月26日上午9:45—11:45; 使用简单计算器

装……订……线

一. 填空判断选择题(每题3分,答题请写在试卷上):

1 设一次试验中A发生的概率为p, 现独立重复进行n次，则A至少发生一次的概率为______________, 至多发生一次的概率为______________.

2 已知随机事件A 的概率$P(A) = 0.5$, 随机事件B 的概率$P(B) = 0.6$ 及条件概率$P(B \mid A) = 0.8$, 则和事件$A \cup B$ 的概率$P(A \cup B) =$___________________.

3 设随机变量 X 和 Y 都服从正态分布, 且他们不相关, 则__________.

(A) X 与Y一定独立 (B) (X, Y)服从二元正态

(C) X和Y未必独立 (D) $X + Y$服从一维正态分布

4 设$F_1(x)$ 与$F_2(x)$ 分别为随机变量X_1 与X_2 的分布函数. 为使$F(x) = aF_1(x) - bF_2(x)$ 是某一随机变量的分布函数，在下列给定的各组数值中应取

(A) $a = \frac{3}{5}$, $b = -\frac{2}{5}$. (B) $a = \frac{2}{3}$, $b = \frac{2}{3}$.

(C) $a = -\frac{1}{2}$, $b = \frac{3}{2}$. (D) $a = \frac{1}{2}$, $b = -\frac{3}{2}$.

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

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

6 设随机变量X 服从正态分布$N(\mu, \sigma^2)\ (\sigma > 0)$, 且二次方程$y^2 + 4y + X = 0$ 无实根的概率为$\frac{1}{2}$, 则$\mu =$______________.

7 一实习生用同一台机器接连独立地制造3 个同种零件，第i 个零件是不合格品的概率$p_i = \frac{1}{i+1}$(i = 1, 2, 3), 以X 表示3 个零件中合格品的个数, 则$P(X = 2) =$______________.

8 设随机变量X 和Y 相互独立, 同分布于期望为$\frac{1}{\lambda}$ 的指数分布,则$\min\{X, Y\}$ 服从参数为______________ 的______________ 分布.

9 设 $\mathrm{Var}(X) = \mathrm{Var}(Z), \mathrm{Var}(Y) = 4\,\mathrm{Var}(X)$, 相关系数$\rho_{X,Y} = -1, \rho_{X,Z} = 1/2$, 则$\rho_{X,Y+Z} =$______________.

10 设$f(x)$ 和$g(x)$ 为两个概率密度函数，则$af(x) + bg(x)$也是概率密度函数的充要条件为______________.

二. 假定某种病菌在全人口中的带菌率为10%, 又在检测时. 带菌者呈阳、阴性反应的概率为0.95 和0.05, 而不带菌者呈阳、阴性反应的概率则为0.01 和0.99.

(1). 现某人被测出呈阳性反应, 则该人确为带菌者的概率多少？

(2). 今某人又独立地检测两次，发现1次呈阳性反应, 1次呈阴性反应. 问在三次检测中2次阳性1次阴性的情况下, “该人为带菌者”的概率是多少?

三. 二维随机变量(X, Y) 的密度函数为:

$$f(x, y)=\begin{cases} Ae^{-(3x+4y)}, & (x>0, y>0) \\ 0, & \text{其他} \end{cases}$$

(1) 试求系数$A=$?

(2)X 与Y 是否独立?

(3) 试求$Z=X+Y$ 的密度函数$f_Z(z)$;

(4) 试求$\operatorname{Var}(X \mid X+Y=1)$.

四. 在一家保险公司里有10000 个老人参加保险，每人每年付200元保险费，在一年内一个人死亡的概率为0.017, 死亡时其家属可向保险公司领取10000元的保险金，问:

(1) 保险公司亏本的概率多大?

(2) 保险公司一年的利润不少于10万元且不超过20万元的概率多大?

以下两题选做一题, 只算一题得分

五. 设某两个风险(X, Y) 服从二元正态分布$N\left(\mu, 2\mu, \sigma^2, 2\sigma^2, \sqrt{2}/4\right)$, 某投资者购买了一个基于这两只风险和的金融衍生品(欧式看涨期权), 即到期收益为

$$(X+Y-3\mu)_+=\max\{X+Y-3\mu, 0\}.$$

(1) 求到期收益的均值$\mathbb{E}\left[(X+Y-3\mu)_+\right]$.

(2) 求到期收益的方差$\operatorname{Var}\left[(X+Y-3\mu)_+\right]$.

六. 记 $U_1, \cdots, U_n$ 为在 $(0,1)$ 中均匀分布的独立随机变量. 对 $0<x<1$ 定义

$$I(X \leq x)=\begin{cases} 1, & X \leq x, \\ 0, & X>x, \end{cases}$$

并记 $F_n(x)=\dfrac{1}{n} \sum_{k=1}^n I(U_k \leq x)$, $0 \leq x \leq 1$, 这是 $U_1, \cdots, U_n$ 的经验分布函数。试求

(1) $F_n(x)$的均值和方差.

(2) $F_n(x)$与$F_n(y)$的协方差$(0<x, y<1)$.

附录 分布及分位数: $\Phi(0.77)=0.78$, $\Phi(1.55)=0.94$, $\Phi(2.32)=0.99$

期中考试

Jiaqi Xia

一.

1.

记 A 发生的次数为 X，则 $X \sim B(n,p)$

$$\begin{aligned}
P(\text{至少发生一次}) &= P(X \geq 1) \\
&= 1 - P(X = 0) \\
&= 1 - C_n^0 p^0 (1-p)^n \\
&= 1 - (1-p)^n
\end{aligned}$$

$$\begin{aligned}
P(\text{至多发生一次}) &= P(X = 0) + P(X = 1) \\
&= C_n^0 p^0 (1-p)^n + C_n^1 p (1-p)^{n-1} \\
&= (1-p)^n + np(1-p)^{n-1}
\end{aligned}$$

答案：$1-(1-p)^n$；$(1-p)^n + np(1-p)^{n-1}$

2.

$$\begin{aligned}
P(A \cup B) &= P(A) + P(B) - P(AB) \\
&= P(A) + P(B) - P(B|A)P(A) \\
&= 0.5 + 0.6 - 0.8 \times 0.5 \\
&= 0.7
\end{aligned}$$

2

答案：0.7

3.

由于 X 与 Y 的联合分布未知，故不相关未必独立

答案：C

4.

由分布函数的规范性，有

$$\lim_{x\to+\infty} F(x) = a - b = 1$$

由分布函数的非负性，有

$$aF_1(x) - bF_2(x) \geq 0,\ \forall x \in R$$

即对二元函数 $f(u,v) = au - bv$，有 $\min\limits_{(u,v)\in[0,1]^2} f(u,v) \geq 0$，从而

$$\begin{cases} a \geq 0 \\ b \leq 0 \\ a - b \geq 0 \end{cases}$$

综上，有

$$\begin{cases} a \geq 0 \\ b \leq 0 \\ a - b = 1 \end{cases}$$

符合条件的仅有 A

答案：A

5.

由正态分布的对称性，有

$$\alpha = P(|X| \leq x) = 1 - 2P(X > x)$$

即 $P(X > x) = \frac{1-\alpha}{2}$，则 $x = u_{\frac{1-\alpha}{2}}$

答案：B

6.

二次方程无实根的概率 p 为

$$\begin{aligned} p &= P(16 - 4X < 0) \\ &= P(X > 4) \\ &= P(X - \mu > 4 - \mu) = \frac{1}{2} \end{aligned}$$

则 $4 - \mu = 0$，即 $\mu = 4$

答案：4

7.

记 $A_i =$" 第 i 个零件合格 "，则

$$\begin{aligned} P(X = 2) &= P(\bar{A}_1 A_2 A_3) + P(A_1 \bar{A}_2 A_3) + P(A_1 A_2 \bar{A}_3) \\ &= (1 - p_1)p_2 p_3 + p_1(1 - p_2)p_3 + p_1 p_2 (1 - p_3) \\ &= \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{4} \\ &= \frac{11}{24} \end{aligned}$$

答案：$\frac{11}{24}$

8.

$F_X(t) = F_Y(t) = (1 - e^{-\lambda t})I(t > 0)$ 则 $Z := \min\{X, Y\}$ 的分布函数为

$$\begin{aligned} F_Z(z) &= P(\min\{X, Y\} \le z) \\ &= 1 - P(X > z, Y > z) \\ &= 1 - (1 - F_X(z))(1 - F_Y(z)) \\ &= (1 - e^{-2\lambda z})I(z > 0) \end{aligned}$$

即 $Z \sim Exp(2\lambda)$

答案：2λ；指数

9.

要求 $\rho_{X,Y+Z} = \frac{Cov(X,Y+Z)}{\sqrt{Var(X)Var(Y+Z)}}$

$$\begin{aligned} Cov(X, Y + Z) &= Cov(X, Y) + Cov(X, Z) \\ &= \rho_{X,Y}\sqrt{Var(X)Var(Y)} + \rho_{X,Z}\sqrt{Var(X)Var(Z)} \\ &= -2Var(X) + \frac{1}{2}Var(X) \\ &= -\frac{3}{2}Var(X) \end{aligned}$$

由于 $\rho_{X,Y} = -1$，则存在 $a < 0, b \in R$，使得 $Y = aX + b$，又 $Var(Y) = 4Var(X)$，则 $a = -2$，从而

$$\begin{aligned} Var(Y + Z) &= Var(Y) + Var(Z) + 2Cov(Y, Z) \\ &= 5Var(X) - 4Cov(X, Z) \\ &= 3Var(X) \end{aligned}$$

则 $\rho_{X,Y+Z} = \frac{-\frac{3}{2}Var(X)}{\sqrt{3}Var(X)} = -\frac{\sqrt{3}}{2}$

答案：$-\frac{\sqrt{3}}{2}$

10.

$$af(x)+bg(x)\text{为密度函数}\Longleftrightarrow\int_{-\infty}^{\infty}af(x)+bg(x)dx=a+b=1;$$
$$af(x)+bg(x)\geq 0,\forall x\in R$$
$$\Longleftrightarrow a+b=1,a\geq 0,b\geq 0$$

答案：$a+b=1,a\geq 0,b\geq 0$

二.

(1).

记 $A=$“检测呈阳性”，$B=$“该人带菌”，则所求概率即为

$$\begin{aligned}P(B|A)&=\frac{P(AB)}{P(A)}\\&=\frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|\overline{B})P(\overline{B})}\\&=\frac{0.95\times 0.1}{0.95\times 0.1+0.01\times 0.9}\\&=\frac{95}{104}\\&\approx 0.9135\end{aligned}$$

(2).

记 $C=$“三次检测中 2 次阳性 1 次阴性”，则所求概率即为

$$\begin{aligned}P(B|C)&=\frac{P(C|B)P(B)}{P(C|B)P(B)+P(C|\overline{B})P(\overline{B})}\\&=\frac{0.95\cdot C_2^1\cdot 0.05\cdot 0.95\times 0.1}{0.95\cdot C_2^1\cdot 0.05\cdot 0.95\times 0.1+0.01\cdot C_2^1\cdot 0.01\cdot 0.99\times 0.9}\\&=\frac{45125}{46016}\\&\approx 0.9806\end{aligned}$$

三.

(1).

由规范性，有

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = A\int_0^{\infty}\int_0^{\infty} e^{-3x-4y}dxdy$$
$$= \frac{A}{12}$$

则 $A = 12$

(2).

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\ &= 3e^{-3x}I(x>0)\int_0^{\infty} 4e^{-4y}dy \\ &= 3e^{-3x}I(x>0) \end{aligned}$$

同理有 $f_Y(y) = 4e^{-4y}I(y>0)$

则 $f(x,y) = f_X(x)f_Y(y)$，X 与 Y 独立

(3).

由卷积公式，有

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f(x, z-x)dx \\ &= \int_0^{z} 12e^{x-4z}I(z>0)dx \\ &= 12(e^{-3z} - e^{-4z})I(z>0) \end{aligned}$$

(4).

X 与 $Z=X+Y$ 的联合分布函数为

$$\begin{aligned}
F_{XZ}(x,z) &= P(X\leq x, X+Y\leq z)\\
&= \iint\limits_{\substack{u\leq x\\ u+v\leq z}} f(u,v)dudv\\
&= \begin{cases}\int_0^z\int_0^{z-u}12e^{-3u-4v}dvdu, & 0<z\leq x\\ \int_0^x\int_0^{z-u}12e^{-3u-4v}dvdu, & 0<x<z\\ 0, & else\end{cases}\\
&= \begin{cases}1-e^{-3z}+3e^{-4z}-3e^{-5z}, & 0<z\leq x\\ 1-e^{-3x}+3e^{-4z}-3e^{x-4z}, & 0<x<z\\ 0, & else\end{cases}
\end{aligned}$$

则给定 $Z=X+Y=1$ 的条件下，X 的条件密度为

$$\begin{aligned}
f_{X|Z=1}(x|z=1) &= \frac{f_{XZ}(x,1)}{f_Z(1)}\\
&= \frac{e^x}{e-1}I(0<x<1)
\end{aligned}$$

从而 $E(X|X+Y=1)=\frac{1}{e-1}$，$E(X^2|X+Y=1)=\frac{e-2}{e-1}$

故

$$\begin{aligned}
Var(X|X+Y=1) &= E(X^2|X+Y=1)-(E(X|X+Y=1))^2\\
&= \frac{e^2-3e+1}{(e-1)^2}
\end{aligned}$$

四.

(1).

对 $n = 10000, p = 0.017$，记

$$X_i = \begin{cases} 1, & i\text{老人在年内死亡} \\ 0, & \text{否则} \end{cases}$$

则 $X_1, \dots, X_n \overset{i.i.d.}{\sim} B(1, p)$

由中心极限定理，亏本概率 p 为

$$\begin{aligned} p &= P\left(10000 \sum_{i=1}^{n} X_i > 2000000\right) \\ &= P\left(\sum_{i=1}^{n} X_i > 200\right) \\ &= P\left(\frac{\sum\limits_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} > \frac{200 - np}{\sqrt{np(1-p)}}\right) \\ &\approx 1 - \Phi(\frac{200 - np}{\sqrt{np(1-p)}}) \\ &\approx 1 - \Phi(2.32) \\ &\approx 0.01 \end{aligned}$$

(2).

利用中心极限定理，所求概率 p' 为

$$
\begin{aligned}
p' &= P\left(100000 \le 2000000 - 10000\sum_{i=1}^{n} X_i \le 200000\right) \\
&= P\left(180 \le \sum_{i=1}^{n} X_i \le 190\right) \\
&= P\left(\frac{180-np}{\sqrt{np(1-p)}} \le \frac{\sum\limits_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \le \frac{190-np}{\sqrt{np(1-p)}}\right) \\
&\approx \Phi(1.55) - \Phi(0.77) \\
&\approx 0.16
\end{aligned}
$$

五.

(1)

由于 $Z := X + Y - 3\mu \sim N(0, 4\sigma^2)$

则

$$
\begin{aligned}
E[(X+Y-3\mu)_+] &= EZ_+ \\
&= \int_0^{\infty} z f_Z(z) dz \\
&= \int_0^{\infty} \frac{z}{\sqrt{8\pi\sigma^2}} e^{-\frac{z^2}{8\sigma^2}} dz \\
&= \sqrt{\frac{2}{\pi}}\sigma
\end{aligned}
$$

(2)

$$
\begin{aligned}
E[(X+Y-3\mu)_+^2] &= EZ_+^2 \\
&= \int_0^\infty z^2 f_Z(z)dz \\
&= \int_0^\infty \frac{z^2}{\sqrt{8\pi\sigma^2}} e^{-\frac{z^2}{8\sigma^2}} dz \\
&= 2\sigma^2
\end{aligned}
$$

则 $Var[(X+Y-3\mu)_+] = EZ_+^2 - (EZ_+)^2 = (2-\frac{2}{\pi})\sigma^2$

六.

(1)

由于 $0 \le x \le 1$，则有

$$
\begin{aligned}
EF_n(x) &= \frac{1}{n}\sum_{k=1}^{n} EI(U_k \le x) \\
&= P(U_1 \le x) \\
&= x
\end{aligned}
$$

$$
\begin{aligned}
Var(F_n(x)) &= \frac{1}{n} Var(I(U_1 \le x)) \\
&= \frac{1}{n}\left(EI^2(U_1 \le x) - [EI(U_1 \le x)]^2\right) \\
&= \frac{1}{n}\left(EI(U_1 \le x) - [EI(U_1 \le x)]^2\right) \\
&= \frac{x-x^2}{n}
\end{aligned}
$$

(2)

由于 $0 < x, y < 1$，则

$$
\begin{aligned}
Cov(F_n(x), F_n(y)) &= Cov\left(\frac{1}{n}\sum_{k=1}^{n} I(U_k \le x), \frac{1}{n}\sum_{l=1}^{n} I(U_l \le y)\right) \\
&= \frac{1}{n^2}\sum_{k=1}^{n} Cov\left(I(U_k \le x), \sum_{l=1}^{n} I(U_l \le y)\right) \\
&= \frac{1}{n^2}\sum_{k=1}^{n} Cov\left(I(U_k \le x), I(U_k \le y)\right) \\
&= \frac{1}{n} Cov(I(U_1 \le x), I(U_1 \le y)) \\
&= \frac{1}{n}\left(E[I(U_1 \le x)I(U_1 \le y)] - EI(U_1 \le x)EI(U_1 \le y)\right) \\
&= \frac{\min\{x, y\} - xy}{n}
\end{aligned}
$$

2020–2021学年《概率论》期中考试试卷

学生所在院系_________ 姓 名_________ 学 号_________ 得 分_________

考试时间: 2020 年 11 月 11 日下午 14:00–16:00

一、 (30分, 每小题3分) **填空题或单选题, 答案可以直接写在试卷上.**

(1) 有三个外观完全一致的袋子, 第一个袋子里有一块金币和一块银币, 第二个袋子里有两块金币, 而第三个袋子里有两块银币. 某人随机选一个袋子后, 从中摸出了一枚钱币, 发现是金币. 那么此袋中另一枚钱币也是金币的概率是________.

(2) 平面上有 n 个不同的点, 编号分别为 $1, 2, \cdots, n$. 现有一个质点在这些点上做随机游动, 假设每次它在某点上停留片刻之后就会在其余所有点中等概率选择一个并移动到该点上. 设其初始位置为点 1 上, 则它在第一次返回此点之前访问过点 2 的概率是________.

(3) 设随机变量 X 的分布函数为 $F(x)$, 则下列说法正确的是( )
(A) 若 X 为连续型, 则函数 $F(x)$ 连续 (B) 若 X 为连续型, 则函数 $F(x)$ 可导
(C) 若函数 $F(x)$ 连续, 则 X 为连续型 (D) 若函数 $F(x)$ 不连续, 则 X 为离散型

(4) 设随机变量 X 和 Y 分别服从参数为 λ 和 μ 的 Poisson 分布, 且相互独立. 对正整数 $m \le n$, $\mathbf{P}(X = m | X + Y = n) =$________.

(5) 设随机变量 X 的概率密度函数为 $p(x) = \begin{cases} a + x, & -1 < x \le 0; \\ a - x, & 0 < x < 1, \end{cases}$ 其中 a 为某个常数, 则概率 $\mathbf{P}(|X| < 0.5) =$( )
(A) 0.25 (B) 0.5 (C) 0.75 (D) 无法确定

(6) 设随机变量 $X \sim \mathrm{Exp}(\lambda), Y \sim \mathrm{Exp}(\mu)$且相互独立, 则 $\mathbf{P}(X \le Y) =$ ________.

(7) 已知随机变量 $X \sim \mathrm{Poi}(\lambda)$. 设对任一非负整数 k, 在 $X = k$ 条件下, 随机变量 $Y \sim \mathrm{Exp}(k+1)$, 则 Y 的概率密度函数 $p(y) =$_______________________, $y > 0$.

(8) 设随机变量 $X \sim \mathcal{N}(1,1), Y \sim \mathcal{N}(0,1)$ 且相互独立, 则下列与随机变量 $Z = X - Y$ 同分布的是( )
(A) $Y - X$ (B) $X + Y$ (C) $X + 2Y$ (D) $X - 2Y$

(9) 设随机变量 X_1, X_2, X_3, X_4 iid $\sim \mathcal{N}(0, 4)$, 若 $T = a(X_1 - 2X_2)^2 + b(3X_3 - 4X_4)^2$ 服从 χ^2 分布, 则常数 a 和 b 的乘积 $ab =$________.

(10) 设随机变量 $X_1, X_2, \cdots, X_n$ iid $\sim \mathrm{Exp}(\lambda)$, 则 $\min\{X_1, X_2, \cdots, X_n\}$ 的分布为( )
(A) $\mathrm{Exp}(\lambda)$ (B) $\mathrm{Exp}(n\lambda)$ (C) $\mathrm{Exp}(\lambda/n)$ (D) 非指数分布

二、 (10分) 设有甲和乙两个罐子, 甲罐中有 m 个红球和 n 个黑球, 乙罐中有 n 个红球和 m 个黑球, 且 $m > n$. 随机选取一个罐子再从中随机抽取一球, 发现为红球, 将其放回后并摇匀. 若再次在该罐中随机抽取一球, 则该球仍为红色的概率是否比 $1/2$ 大? 通过计算这一事件的概率来证明你的结论.

三、 (18分) 设二维随机向量(X, Y)的概率密度为$f(x, y) = Ax^2$, $0 < |x| < y < 1$. 试求

(1) 常数A.

(2) 条件概率密度$p_{X|Y}(x|y)$.

(3) 条件概率$\mathbf{P}(X \leq 1/4 | Y = 1/2)$.

四、 (18分) 设在单位圆盘 $\{(x, y) : x^2 + y^2 \leq 1\}$ 上独立地随机取两点, 极坐标分别记为 (R_1, Θ_1) 和 (R_2, Θ_2).

(1) 分别求 R_1 和 Θ_1 的概率密度函数.

(2) 问 R_1 与 Θ_1 是否独立? 请证明你的结论.

(3) 试求夹角 $\Theta = |\Theta_1 - \Theta_2|$ 的概率密度函数.

五、 (24分) 设随机变量 $X_1, X_2, \cdots, X_n$ iid $\sim \mathrm{Exp}(1)$, 而 $U_1, U_2, \cdots, U_n$ iid $\sim U[0, 1]$, 且这两组随机变量也相互独立. 记

$$S_n = X_1 + X_2 + \cdots + X_n, \quad n \geq 2.$$

(1) 证明：S_n 的概率密度函数为

$$p_n(x) = \mathrm{e}^{-x} \frac{x^{n-1}}{(n-1)!}, \quad x > 0.$$

(2) 对任一整数 $1 \leq m < n$, 试求 $Y_m = S_m / S_n$ 的概率密度函数.

(3) 设随机变量 N 服从参数为 $p\,(0 < p < 1)$ 的几何分布, 且与 $X_i\,(1 \leq i \leq n)$ 独立, 试问 S_N 是否服从指数分布？证明你的结论.

(4) (附加题, 10分) 试求 $Z_m = (U_1 U_2 \cdots U_n)^{-Y_m}$ 的概率密度函数.