2020年春季概率论与数理统计期末考试

- 一共有9大题。共100分。考试时间为3个小时。
- 请将1-3题，4-6题，7-9题的答卷分别上传到网络学堂。
- 在每份答卷上都必须在开始写上“我（姓名）保证本答卷为个人独立完成”。 如未书写，自动作废。
- 如无额外说明， $\bar{x}$, $\bar{y}$,...， 均表示样本均值，s_x^2, s_y^2,...,表示样本的无偏方差。$\Phi(\cdot)$, $K_n(\cdot)$ 分别表示标准正态分布与自由度为n的卡方分布的分布函数；u_α、$t_\alpha(\cdot)$、 $\chi^2_\alpha(\cdot)$、$F_\alpha(\cdot,\cdot)$分别表示标准正态分布、t分布、卡方分布与F分布的α分位数。这些符号均可以用来表述最终答案。
- 答题需给出必要的过程。
- 可能用到的数据： $\chi^2_{0.95}(2) = 5.9915$, $\chi^2_{0.95}(3) = 7.8147$, $t_{0.995}(16) = 2.9208$, $t_{0.995}(18) = 2.8784$, $t_{0.95}(34) = 1.6909$, $t_{0.95}(35) = 1.6896$, $F_{0.99}(1,16) = 8.53$, $F_{0.99}(1,18) = 8.29$.
- 开卷考试。
- 可使用计算器或相应的手机程序。
- 预祝暑期快乐。

1. （10分）某快递员要配送100份快递，但他早餐时错把白酒当白开水喝了，大醉。在酒精的作用下，他只能随机地配送这100份快递给相应的客户。（本题无需对最终答案进行化简。）

   (a) （4分）没有客户正确地收到自己的快递的概率是?

   (b) （4分）恰好有k个客户收到了自己的快递的概率是?

(c) (2分）记上一小问求得的概率为P_{100}^k (这里假设它为已知数), 如果错派一个快递，快递员要被扣2块钱，求这次犯错让他损失的金额的数学期望。

2. （8分）设随机向量(X,Y)的联合密度函数$f(x,y)$是球对称的，即存在一个函数$g(z)$使得 $f(x,y)=g(x^2+y^2)$。 g和f 均为光滑函数。设R与Θ是(X,Y)的极坐标，即$X=R\cos\Theta$, $Y=R\sin\Theta$.

   (a) （6分）求R与Θ的密度函数。

   (b) （2分）请问R与Θ是否相互独立？请说明理由。

3. （8分）某项研究想确定男性与女性饮酒者的啤酒偏好是否有差异，得到下面的数据

啤酒偏好 \ 性别	男性	女性	合计
淡啤酒	51	39	90
普通啤酒	56	21	77
黑啤酒	25	8	33
合计	132	68	200

设计一个近似显著水平为0.05的检验，并根据数据做出判断。

4. （10分） 盒子里有N个球，其中有a个黑球，b个白球，c个红球，$a+b+c=N$. 现在从盒子里有放回地取n个球。记X为取到黑球的数目，Y为取到红球的数目。

   (a) （4分）求随机变量X的分布列及数学期望。

   (b) （2分）求随机向量(X,Y)的联合分布列。

   (c) (4分）假如取出的球的个数n是也是随机变量，它服从参数为$\lambda>0$的柏松分布，记在这种情况下取出白球的个数为Z, 求Z的数学期望。

5. （12分）设总体为区间$[3\theta,4\theta]$上的均匀分布, $\theta>0$. $X_1,\ldots,X_n$是该总体的简单样本。

   (a) （4分）寻找θ的矩估计和最大似然估计。

   (b) （4分）它们是否无偏估计？请说明理由。

   (c) （4分）它们是否相合估计？请说明理由。

6. （12分）设$X_1,\ldots,X_n$是来自总体$X\sim N(0,\sigma^2)$的简单样本。

   (a) （4分） 试构造σ的一个置信水平为$1-\alpha$的置信区间，其中 $\alpha\in(0,1)$已知。

   (b) （8分） 考虑假设检验问题 $H_0:\sigma=\sigma_0$ vs $H_1:\sigma=\sigma_1$, 其中$\sigma_0<\sigma_1$已知. 令拒绝域为$\{\sum_{i=1}^n X_i^2>c\}$, 其中c 已知。 请描述该检验可能发生的错误，并计算相应错误发生的概率。

7. （12分） 设总体U服从开区间$(0,1)$ 上的均匀分布。$U_1,\ldots,U_n$是U的简单样本。

   (a) （6分）令$Y_n=\ln(\prod_{i=1}^n U_i)$. 当$n\to\infty$时， $\frac{1}{\sqrt{n}}(Y_n+n)$是否依分布收敛？请说明理由。

   (b) （6分）令$Z_n=(\sum_{i=1}^n U_i^2)(U_1+\ldots U_{\lfloor\frac{n}{\alpha}\rfloor})^{-1}$, 其中$\alpha>0$, $\lfloor x\rfloor$表示不超过x的最大整数。当$n\to\infty$时，Z_n 是否依概率收敛？请说明理由。

8. （12分）$X_1,\ldots,X_n$是来自总体$N(\mu_X,\sigma_X^2)$的简单样本， $Y_1,\ldots,Y_n$是来自总体$N(\mu_Y,\sigma_X^2)$的简单样本，且$X_1,\ldots,X_n$, $Y_1,\ldots,Y_n$相互独立。μ_X, μ_Y, σ_X, σ_Y 均未知。

   (a) （2分）考虑$W_i=X_i-Y_i$，求W_i的分布类型。

   (b) （4分） 考虑统计量$T=\frac{\bar{W}}{\sqrt{\frac{1}{n}S_W^2}}$, 其中$\bar{W}=\frac{1}{n}\sum_{i=1}^n W_i$, $S_W^2=\frac{1}{n-1}\sum_{i=1}^n(W_i-\bar{W})^2$. 假如$\mu_X=\mu_Y$, 求$T$的分布类型。

   (c) （6分）考虑假设检验$H_0:\mu_X-\mu_Y=5$ vs $H_1:\mu_X-\mu_Y\neq 5$. 参考上一问的统计量，设计一个显著水平为$\alpha\in(0,1)$的检验过程。

9. （16分）有人想研究某一时间段内降雨量x与风湿病发作的人数y之间的关系。 通过翻查18年的记录，得到了以下数据：

$$\sum_{i=1}^{18}x_i=195,\ \sum_{i=1}^{18}y_i=318,\ \sum_{i=1}^{18}x_i^2=2282,\ \sum_{i=1}^{18}x_iy_i=4362,\ \sum_{i=1}^{18}y_i^2=12548.$$

   假设y 与 x 满足线性回归模型：

$$y=b_0+b_1x+\epsilon,\quad \epsilon\sim N(0,\sigma^2),$$

   其中b_0，b_1， σ均为未知参数。（本题结果可以包含相应概率分布的分位数）

(a) （6分） 用最小二乘估计计算b_0， b_1的估计值，并建立线性回归方程。

(b) （3分）给出σ的置信水平为0.99的置信区间。

(c) （3分）在显著水平为0.01下，检验回归方程的显著性。

(d) （4分）假设回归方程显著，给出当$x = 10$时，y 的数学期望的置信水平为0.95的置信区间，和y 的概率为0.98的预测区间。