2019年春季概率论与数理统计期末练习题

1. (a) 一列随机变量X_n, $P(X_n=\frac{1}{n})=1-\frac{1}{n^2}$. X_n是是否依概率收敛。
   (b) X_n表示在连续抛n次硬币中出现连续出现三个正面紧接着三个反面的次数。请问$\frac{1}{n}X_n$是否依概率收敛？指出极限且说明理由。
   (c) X服从柯西分布， 令$Y_n=(-1)^{\varphi(n)}X$, $\varphi(n)$是n的不同素因子的个数。请问Y_n是否依概率收敛？是否以分布收敛？
   (d) X_n服从伽马分布$Ga(n,\lambda)$, 令$Y_n=\frac{\lambda X_n-n}{\sqrt{n}}$. 请问$Y_n$是否依分布收敛。
   (e) X_n是一列方差一致有界的随机变量序列，且当$i\neq j$时，$|Cov(X_i,X_j)|\leqslant\frac{1}{|i-j|^{0.0001}}$. 请问$\frac{1}{n}\sum_{i=1}^{n}X_i$是否依概率收敛。

2. 抛一块硬币，正面出现的概率是$p\in(0,1)$。连续投掷，直到两面都出现才停止。
   (a) 求抛掷次数的数学期望。
   (b) $X_1,\ldots,X_n$表示n次重复试验中停止时抛掷的次数。用矩方法寻找p的统计估计量。

3. 甲、乙二人进行象棋比赛，每局甲胜的概率为p，乙胜的概率为$q=1-p$。比赛进行到有一个人连胜两局为止，求平均的比赛局数。

4. 已知(X,Y)的联合分布为$f(0,10)=f(0,20)=\frac{2}{18}$, $f(1,10)=f(1,30)=\frac{3}{18}$, $f(1,20)=\frac{4}{18}$, $f(2,30)=\frac{4}{18}$,
   (a) 求X,Y的边际分布。
   (b) 求条件期望$E(X|Y)$的分布。

5. X 和 Y是相互独立的标准正态分布。令$U=\frac{X}{Y}$, $V=|Y|$, 若$Y\neq0$; $U=0, V=0$ 若$Y=0$.

(a) 求 (U,V)的联合密度。

(b) 求U的密度函数。

6. 黑盒子里面有1000个硬币，有500个是公平硬币，即抛掷时正反面出现的可能性一样，300个是不公平硬币，抛掷时正面出现的可能性比反面大一倍，200个是作弊硬币，抛掷时只会出现正面。你从黑盒子里面抽出一个硬币，连续抛掷10次都是正面，请问，你拿到的是作弊硬币的概率是多少？公平硬币？不公平硬币？

7. 一次抽奖节目，有三个黑盒子，其中只有一个放着奖品。你选择了其中一个。这时主持人从剩下的两个盒子挑一个出来，打开，里面没有奖品。主持人给你一次机会重选，请问你是否应该重选？给出理由？

8. 设$x_1,\ldots,x_n$是来自总体X的样本。X 有如下分布列：

$$P(X=k)=\frac{-1}{\ln(1-p)}\times\frac{p^k}{k},\quad k=1,2,\ldots,$$

其中$0<p<1$为未知参数。试构造参数p的矩估计。(提示：计算一阶和二阶矩。)

9. X_n是独立同分布随机变量序列, $h(x)$是有界函数。当n很大时，求$\frac{1}{n}\sum_{i=1}^{n}h(X_i)$的近似分布。

10. $X_1,\ldots,X_n$是来自均匀分布$U(a,b)$的一个样本。用矩方法构造a 与 b的统计估计量。他们是否各自的无偏估计？

11. X_i是独立同分布随机变量序列，$E(X_i)=\mu$, $Var(X_i)<\infty$。考虑随机变量 $Y_n=\frac{2}{n(n+1)}\sum_{k=1}^{n}kx_k$. 当$n\to\infty$, Y_n是否依概率收敛。

12. 设$x_1,\ldots,x_n$是来自总体$U(0,\theta)$的样本。$x_{(n)}$为样本的最大次序统计量。

(a) ）请问$x_{(n)}$是否为θ的充分统计量？说明你的理由。

(b) 请问$x_{(n)}$是否为θ的无偏估计？说明你的理由。

(c) 请问$x_{(n)}$是否为θ的相合估计？说明你的理由。

13. x_1, x_2为来自密度函数为$p(x;\theta)$的总体的样本，

$$p(x;\theta) = \frac{3x^2}{\theta^3}, \quad 0 < x < \theta,\ \theta > 0.$$

(a) 考虑$T_1 = \frac{2}{3}(x_1 + x_2)$, $T_2 = \frac{7}{6}\max\{x_1, x_2\}$. 请问它们是否参数$\theta$的无偏估计。

(b) 比较它们的有效性。

14. $X_1, \ldots, X_n$是来自柏松分布$P(\beta\tau)$的样本， $Y_1, \ldots, Y_m$是来自柏松分布$P(\beta\tau)$的样本，且X_i, Y_j相互独立。求β 与τ 的最大似然估计。

15. 设$x_1, \ldots, x_n$是来自正态总体$N(\mu, \sigma^2)$的一个样本，假设均值μ已知。

(a) 试求σ^2的最大似然估计。

(b) 试计算σ^2的费希尔信息量。

(c) 请问σ^2的最大似然估计是否为有效估计？请说明你的理由。（提示：自由度为n的卡方分布的方差为$2n$.）

16. 设$x_1, \ldots, x_n$是来自正态总体$N(\mu, 1)$的样本，考虑如下假设检验问题

$$H_0 : \mu = 2 \quad \text{vs} \quad H_1 : \mu = 3.$$

检验的拒绝域为$W = \{\bar{x} \geqslant 2.6\}$, 其中$\bar{x}$为样本均值。

(a) 这个假设检验问题可能发生的两类错误为?

(b) 当样本容量$n = 25$时，该检验犯这两类错误的概率分别为（用$\Phi(\cdot)$表示）?

(c) 如果想让犯这两类错误的概率都小于0.003，样本容量至少为?（取$\Phi(-3) \approx 0.003$)

17. 世界卫生组织建议超过50岁的男性每天的锌摄入量为15mg/天。某团队调查了一组60-65岁的老年男性每天的锌摄入量，给出了以下的数据

$$n = 144, \quad \bar{x} = 10, \quad s^2 = 6^2,$$

其中n为调查对象的人数，$\bar{x}$为调查对象的日均锌摄入量，s^2为对应的样本无偏方差。 请问这一项调查研究是否能得出结论认为所有60-65岁的老年男性的平均每天锌摄入量低于世卫的指引标准呢？（可以认为样本容量已经足够大。）

(a) 对这个问题设计一个（近似）显著水平为0.01的假设检验问题。（$\Phi(-2.33)\approx 0.01$）

(b) 利用给出的数据做出判断。

(c) 对于你所设计的假设检验问题，样本数据的（近似）p值是多少（用函数$\Phi(\cdot)$表示）？并说明它的意义。

18. $x_1,\ldots,x_n$是来自$U(0,\theta)$的样本。考虑如下的检验问题

$$H_0:\theta\leqslant\frac{1}{2},\quad \text{vs}\quad H_1:\theta>\frac{1}{2}.$$

假设给定拒绝域$W=\{x_{(n)}\geqslant c\}$.

(a) 求该检验的势函数;

(b) 如果要求犯第一类错误的概率不超过0.05， c应该取多大? 这种情况下如果要求在$\theta=\frac{3}{4}$时，凡第二类错误的概率也不超过0.02，n应该取多大?

19. 现有两组组样本数据。一组来自正态总体$N(\mu_1,\sigma_1^2)$:

$$n=10,\quad \bar{x}=12,\quad s_x=30.$$

另一组来自正态总体$N(\mu_2,\sigma_2^2)$:

$$m=20,\quad \bar{y}=13,\quad s_y=75.$$

设$\alpha\in(0,1)$是给定常数。（本题计算式子待入数据即可，无须具体计算）

(a) 给出μ_1置信水平为$1-\alpha$的置信区间。

(b) 如果已知$\sigma_1/\sigma_2=0.5$, 给出$\mu_1-\mu_2$的置信水平为$1-\alpha$的置信区间。

(c) 给出σ_1/σ_2的置信水平为$1-\alpha$的置信区间。

20. 设$X\sim N(\mu_1,\sigma_1^2)$, $Y\sim N(\mu_2,\sigma_2^2)$. 来自$X$的样本容量为7的数据: $\bar{x}=95.7$, $s_x^2=2208.57$; 来自Y的样本容量为5的数据: $\bar{y}=97.4$，$s_y^2=78.801$. 在显著水平0.05下，

(a) 检验 $H_0: \sigma_1^2 = 10\sigma_2^2$, vs $H_1: \sigma_1^2 \neq 10\sigma_2^2$.

(b) 利用前一个检验，检验$H_0: \mu_1 - \mu_2 = 10$ vs $H_1: \mu_1 - \mu_2 \neq 10$.

21. 某合金钢的抗拉强度y与碳含量x有关。对92个该合金钢的样品进行研究，得到以下数据

$$\bar{x} = 0.12,\ \bar{y} = 45,\ l_{xx} = 0.3,\ l_{yy} = 2900,\ l_{xy} = 27.$$

假设y与x有如下关系，

$$y = a + bx + \epsilon, \quad \epsilon \sim N(0, \sigma^2).$$

(a) 由数据计算a和b的最小二乘估计并给出一元线性回归方程。

(b) 对线性回归方程进行显著水平为$\alpha \in (0,1)$的显著性假设检验。（计算式子待入数据即可，无须具体计算）

(c) 假设线性回归方程显著，在$x = 0.1$时，求对应的y的概率为$1 - \alpha$的预测区间。（计算式子待入数据即可，无须具体计算）