2019年秋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。（提示：考虑定义）

   (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。（提示：考虑马尔可夫大数定律）

   (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？ （提示：考虑定义）

   (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服从卡方分布$\chi^2(n)$, 令$Y_n = \frac{X_n - n}{\sqrt{n}}$,请问$Y_n$是否依分布收敛。（提示：考虑特征函数）

   (f) X_n为一列服从参数为$\frac{1}{n}$的柏松分布的随机变量序列。令$Y_n = n^{100}X_n$, Y_n是否依概率收敛。（提示：考虑定义）

   (g) X服从$[0, 1]$上的均匀分布。令

   $$Y_n := \begin{cases} 1, & X \in [\frac{k}{2^i}, \frac{k+1}{2^i}), \\ 0, & \text{其它}, \end{cases}$$

   其中$n = 2^i + k$, $0 < k < 2^i$. 请问Y_n是否依概率收敛。（提示：考虑依概率收敛的定义）

   (h) 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。（提示：考虑马尔可夫大数定律）

   (i) $f(x)$ 是光滑函数，Y_i, $i = 1, \ldots, n$是独立同分布随机变量序列, $E(Y_1) = a$, $Var(Y_1) = b < \infty$，请问$\sqrt{n}(f(\frac{1}{n}\sum_{i=1}^n Y_n) - f(a))$是否依分布收敛？（提示：考虑中心极限定理和泰勒展开）

(j) $X_n = aX_{n-1} + K_n$, 其中$|a| < 1$, K_n是相互独立的标准正态分布，请问$\frac{1}{n}\sum_{i=1}^{n} X_n$是否依概率收敛？（提示：计算协方差，考虑马尔可夫大数定律）

(k) 设$\{X_n\}$为独立同分布的随机变量序列，$P(X_k = \pm 1) = \frac{1}{2}$. 考虑$U_n = \sum_{k=1}^{n} \frac{X_k}{2^k}$, 请问$U_n$是否依分布收敛到$[-1, 1]$上的均匀分布？请说明理由。（提示：计算特征函数）

(l) $X_1, Y_1, \ldots X_n, Y_n$是独立同分布的标准正态分布序列。考虑$T_n = \frac{\sum_{i=1}^{n} X_i}{\sqrt{\sum_{i=1}^{n} Y_i^2}}$, 请问$T_n$是否依分布收敛？（考虑$t$分布的定义和性质）

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$Y \neq 0$; $U = 0$, $V = 0$ 若$Y = 0$.

   (a) 求 (U, V)的联合密度。

   (b) 求U的密度函数。

6. $X_1, \ldots, X_{100}$为独立同分布随机变量序列，均服从伽马分布$Ga(1, \frac{1}{2})$. 求随机变量$\frac{1}{100}E(\sum_{i=1}^{100} X_i | \sum_{i=1}^{50} X_i)$的密度函数。（提示：考虑伽玛分布的可加性）

7. 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？（考虑：条件概率，贝叶斯公式）

8. 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？（提示：直接计算各自得奖的概率）

9. n个人在聚会上摘下他们的帽子。帽子混在一起后，每人随机地取一顶，如果一个人取回了自己的帽子，我们就说发生了一次匹配，那么，没有发生匹配的概率是多少？恰巧有k次匹配的概率是多少？（提示：可以参考教材第36页例题）

10. 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。）

11. 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。（考虑中心极限定理）

12. 设$\{X_n\}$为独立同分布的随机变量序列。已知$E(X^2) = 3$, $E(X^4) = 16$. 当n很大时，求事件$\{\sum_{i=1}^n X_i^2 \geqslant 3n + 10\sqrt{n}\}$的近似概率。（考虑中心极限定理）

13. $X_1, \ldots, X_n$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？（考虑定义）

14. 某总体服从指数分布$Exp(\frac{1}{\lambda})$, $\lambda > 0$, $X_1, \ldots, X_n$是其样本。请问统计量$n \times \min\{X_1, \ldots, X_n\}$是否参数$\lambda$的无偏估计。说明理由。（考虑定义）

15. 某总体服从二项分布$b(k, p)$, $X_1, \ldots, X_n$是其样本。请使用矩方法寻找参数k 与p的统计估计量。

16. $X_1,...,X_n$是来自总体X的样本， X的密度函数为

$$f(x;\theta)=\theta x^{-2}, 0<\theta\leqslant x<\infty.$$

(a) 请说明$f(x;\theta)$是密度函数。

(b) 寻找θ的最大似然估计。

(c) 寻找θ的置信水平为$1-\alpha$的置信区间。（通过最大似然估计构造枢轴量）

17. 假设$X_1,...,X_n$是总体X的样本，X的密度函数为

$$f(x;\mu)=e^{-(x-\mu)},\ x\geqslant\mu.$$

(a) 请说明$Y_n=\min\{X_1,...,X_n\}$是参数μ的充分统计量。

(b) Y_n是否μ的相合估计。

(c) 试通过Y_n来构造μ的置信水平为$1-\alpha$的置信区间。

18. 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。（提示：直接计算数学期望和方差）

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

(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。

20. 设 设$x_1,...,x_n$是来自总体$U(-\theta,\theta)$的样本, $\theta>0$。

(a) 寻找θ的矩估计和最大似然估计。他们是和否无偏估计？是否相合？

(b) 寻找θ的置信水平为$1-\alpha$的置信区间。（尝试通过最大似然估计寻找枢轴量）

21. 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。（计算各自方差）

22. X和Y是相互独立的指数分布，

$$f(x|\lambda) = \frac{1}{\lambda}e^{-\frac{x}{\lambda}},\ x > 0, \quad f(y|\mu) = \frac{1}{\mu}e^{-y/\mu}, y > 0.$$

定义

$$Z = \min(X, Y) \text{ and } W = \begin{cases} 1 & \text{if } Z = X, \\ 0, & \text{if } Z = Y. \end{cases}$$

(a) 求(Z, W)的联合分布。

(b) (Z_i, W_i), $i = 1, \ldots, n$是来自总体(Z, W)的样本， 求λ 与μ的最大似然估计。

23. $X_1, \ldots, X_n$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。（最大似然估计的标准求法）

24. 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$2n$.)

25. 某正态总体的方差σ^2已知。 均值μ 有两种可能性$\mu \leqslant \mu_0$或者$\mu = \mu_1 > \mu_0$. $\bar{x}$是容量为n的样本均值。考虑检验问题： $H_0: \mu \leqslant \mu_0 \quad vs \quad H_1: \mu = \mu_1$. 设计一个显著水平为0.01的假设检验过程。发生第二类型错误的概率是?

26. 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$\Phi(-3) \approx 0.003$）

27. 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。

28. $x_1, \ldots, x_n$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?

29. $X_1, \ldots, X_n$是来自正态分布$N(0, \sigma^2)$的样本，考虑检验 $H_0 : \sigma = \sigma_1$, $H_1 : \sigma = \sigma_2$, $\sigma_1 < \sigma_2$。考虑拒绝域$\sum_{i=1}^n X_i^2 > c$, 请求该检验的势函数和犯第一第二类错误的概率。

30. 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。

31. 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 $H_1:\mu_1-\mu_2\neq 10$.

32. 某人想研究臀部大小和智商的关系，得到以下数据:

智 / 臀	< 80	80 —100	> 100	合计
大	18	15	33	66
小	20	19	45	84
合计	38	34	78	150

设计一个（近似）显著水平为0.05的假设检验过程检验原假设H_0:臀部大小与智商不相关，并作出判断。

33. 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) 寻找b的置信水平为$1-\alpha$的置信区间。

(c) 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）

(d) 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）