2021 年秋统计学习题 01

1. 证明 $n \to \infty$ 时，t 分布收敛于标准正态分布.
2. 已知随机变量 $X \sim t(n)$，求证：$X^2 \sim F(1,n)$.
3. 求总体 $N(20,3)$ 的容量分别为 $10,15$ 的两个独立样本均值的差的绝对值大于 0.3 的概率.
4. 设 $X_1,\cdots,X_n,X_{n+1},\cdots,X_{n+m}$ 是分布 $N(0,\sigma^2)$ 的容量为 $n+m$ 的样本，试求下列统计量的概率分布

   (a).
$$Y_1 = \frac{\sqrt{m}\sum_{i=1}^{n} X_i}{\sqrt{n}\sqrt{\sum_{i=n+1}^{n+m} X_i^2}}.$$

   (b).
$$Y_2 = \frac{m\sum_{i=1}^{n} X_i^2}{n\sum_{i=n+1}^{n+m} X_i^2}$$

5. 设 $X_1,\cdots,X_n,X_{n+1}$ 是来自总体 $N(\mu,\sigma^2)$ 的样本，
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$
   试求统计量
$$T = \frac{X_{n+1} - \bar{X}}{S^*} \cdot \sqrt{\frac{n-1}{n+1}}$$
   的分布.
6. 设总体 $X \sim N(\mu,\sigma^2)$，$X_1,X_2,\cdots,X_n$ 为来自 X 的样本.

   (a). 写出 $X_1,X_2,\cdots,X_n$ 的联合概率密度.

   (b). 写出 $\bar{X}$ 的概率密度.

2021 年秋统计学习题 02

1. 设总体 X 服从参数为 N 和 p 的二项分布，$x_1, \cdots, x_n$ 为取自 X 的样本，试求参数 N, p 的矩估计.
2. 设总体概率密度为
$$f(x;a) = \begin{cases} (a+1)x^a, & 0 < x < 1, \\ 0, & \text{其他}, \end{cases}$$
其中 $a > -1$. 试用样本 $x_1, x_2, \cdots, x_n$ 求参数 a 的矩估计和极大似然估计.
3. 设总体的概率密度为
$$f(x;\theta) = \begin{cases} \exp(-(x-\theta)), & x \geq \theta, \\ 0, & \text{其他}, \end{cases}$$
试用样本 $x_1, x_2, \cdots, x_n$ 求参数 θ 的极大似然估计.
4. 设总体 X 服从几何分布
$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k = 1, 2, \cdots,$$
其中 $0 < p < 1$，试利用样本 $x_1, x_2, \cdots, x_n$ 求 p 的极大似然估计.

2021 年秋统计学习题 03

1. 设 $X_1, X_2, \cdots, X_n$ 是总体 $N(\mu, \sigma^2)$ 的一个样本.
   (a). 试求适当选择的常数 c 使得 $c\sum_{i=1}^{n-1}(X_{i+1} - X_i)^2$ 是 σ^2 的无偏估计;
   (b). 求 k 使得 $k\sum_{i=1}^{n}|X_i - \bar{X}|$ 为 σ 的无偏估计.
2. 设 $\hat{\theta}$ 是参数 θ 的无偏估计，且 $\mathrm{Var}\,\hat{\theta} > 0$，试证明 $\hat{\theta}^2 = (\hat{\theta})^2$ 不是 θ 的无偏估计.
3. 设总体 X 的数学期望 $\mu = \mathbb{E}X$ 已知，试证统计量 $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ 是总体方差 $\sigma^2 = \mathrm{Var}\,X$ 的无偏估计.
4. 设 X 为离散型随机变量，分布列为

$$\mathbb{P}(X = 1) = \theta, \quad \mathbb{P}(X = 2) = 1 - \theta.$$

   三个独立样本观测值为 $x_1 = 1, x_2 = 2, x_3 = 1$. 如果 Θ 的先验分布为 $[0, 1]$ 上的均匀分布，求其后验密度.

2021 年秋统计学习题 03

设 T 为参数 θ 的估计，称

$$\mathrm{MSE}(T) := \mathbb{E}_\theta(T-\theta)^2$$

为 T 的均分误差（MES 为 mean squarred error 的首字母缩写）.

1. 设 $X_1, X_2, \cdots, X_n$ 为独立同分布随机变量序列，期望为 μ，方差为 σ^2. 记

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2, \quad S^{*2} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2.$$

(a). 设 a 为常数，证明对任意形如 aS^2 的 σ^2 的估计，有

$$\mathrm{MSE}(aS^2) = \mathbb{E}(aS^2 - \sigma^2)^2 = a^2 \operatorname{Var} S^2 + (a-1)^2\sigma^4.$$

(b). 证明

$$\operatorname{Var}(S^2) = \frac{1}{n}\left(\kappa - \frac{n-3}{n-1}\right)\sigma^4,$$

其中 $\kappa = \frac{\mathbb{E}(X-\mu)^4}{\sigma^4}$ 为 X 的峰度 (kurtosis).

(c). 设 X_i 服从正态分布. 证明

(i) $\mathrm{MSE}(S^{*2}) < \mathrm{MSE}(S^2)$.

(ii) $\kappa = 3$.

(iii) 形如 aS^2 的估计中，MSE 最小的是 $\frac{n-1}{n+1}S^2$.

(d). 不做正态性假设，证明当

$$a = \frac{n-1}{(n+1) + \frac{(\kappa-3)(n-1)}{n}}$$

时 $\mathrm{MSE}(S^2)$ 取得最小值.

2021 年秋统计学习题 05

1. 设总体的概率密度函数如下

$$f(x|\theta) = \frac{1}{2}(1 + \theta x), \quad -1 < x < 1, -1 < \theta < 1,$$

$X_1, X_2, \cdots, X_n$ 为样本，求 θ 的矩估计，并证明它是弱相容的.

2. 设样本 $X_1, X_2, \cdots, X_n$ 来自正态总体 $N(\theta, \theta)$，其中 $\theta > 0$ 为参数.
   (a). 求 θ 的极大似然估计 $\hat{\theta}$.
   (b). 求 $\hat{\theta}$ 的近似方差.
3. 设 $X_1, X_2, \cdots, X_n$ 为来自几何分布的样本，用因子分解定理证明 $T = \sum_{i=1}^n X_i$ 是充分统计量.
4. 设 X 为来自正态总体 $N(0, \sigma^2)$ 的样本. $|X|$ 是否为充分统计量?
5. 设 $X_1, X_2, \cdots, X_n$ 为来自总体概率密度函数如下

$$f(x|\mu, \sigma) = \frac{1}{\sigma} \exp\left(-\frac{x - \mu}{\sigma}\right), \quad \mu < x < \infty, 0 < \sigma < \infty$$

的样本. 求 (μ, σ) 的二维充分统计量.

6. 设 X, Y 为期望有限的随机变量，证明

$$\min_{g(x)} \mathbb{E}(Y - g(X))^2 = \mathbb{E}(Y - \mathbb{E}(Y|X))^2,$$

其中 $g(x)$ 取遍所有的可测函数. $E(Y|X)$ 有时被称为 Y 在 X 上的回归，为给定条件 X 下，Y 的最好的预测.

7. 证明二项分布属于指数分布族.

2021 年秋统计学习题 06

1. 下面的 16 个数来自计算机的正态随机数生成器[1]

5.3885	3.2387	0.2195	0.3231	−3.6310	−0.3892	1.7771	1.4031
−1.5357	−1.6683	4.9765	2.0393	−0.5556	−0.5957	−0.9547	1.0123

   (a). 你认为这个正态分布的均值 μ 和方差 σ^2 是多少?
   (b). 给出 μ 的 $90\%, 95\%, 99\%$ 的置信区间.
   (c). 给出 σ^2 的 $90\%, 95\%, 99\%$ 的置信区间.
   (d). 给出 σ 的 $90\%, 95\%, 99\%$ 的置信区间.
   (e). 如果将 μ 的置信区间长度减半，样本需要增加多少?

2. 设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本，μ, σ 未知. 常数 c 应取多少时才能保证区间 $(-\infty, \bar{X} + c]$ 是 μ 的 95% 的置信区间，即求 c 使得
$$\mathbb{P}(-\infty < \mu \le \bar{X} + c) = 0.95.$$

3. 设 $L(x), U(x)$ 满足
$$\mathbb{P}(L(\mathbf{X}) \le \theta) = 1 - \alpha_1, \quad \mathbb{P}(U(\mathbf{X}) \ge \theta) = 1 - \alpha_2,$$
   且对任意 $\mathbf{x}$,
$$L(\mathbf{x}) \le U(\mathbf{x}).$$
   证明
$$\mathbb{P}(L(\mathbf{x}) \le \theta \le U(\mathbf{x})) = 1 - \alpha_1 - \alpha_2.$$

[1]本例使用 WPS 的电子表格，命令为 = norminv(rand(), μ, σ)).

4. 设 $X|P \sim B(n,P)$,$P \sim \mathrm{Beta}(\alpha,\beta)$，即成功的概率 P 是一个服从 $\mathrm{Beta}(\alpha,\beta)$ 的随机变量，在条件 $P=p$ 时，随机变量 X 服从二项分布 $B(n,p)$.

(a). 求 $\mathbb{E}X$.

(b). 求 $\operatorname{Var}X$.

(c). 证明 X 的方差可以写成

$$\operatorname{Var}X = n\mathbb{E}P(1-\mathbb{E}P) + n(n-1)\operatorname{Var}P.$$

2021 年秋统计学习题 07

1. 独立投掷硬币 10 次，记硬币正面朝上的概率为 p，假设检验问题为 $H_0: p = \frac{1}{2} \leftrightarrow p \neq \frac{1}{2}$. 如果正面出现 0 次或 10 次时拒绝检验原假设.
   (a). 检验的显著性水平是多少?
   (b). 如果 $p = 0.1$，检验的势是多少?
2. 设 $X_1, \cdots, X_n$ 为来自参数为 λ 的 Poisson 分布的样本.
   (a). 考虑检验问题 $H_0: \lambda = \lambda_0 \leftrightarrow H_1: \lambda = \lambda_1$ （$\lambda_0 < \lambda_1$）. 求检验的似然比.
   (b). 利用 Poisson 分布的可加性解释如何确定上述假设检验问题的显著性水平为 α 的拒绝域.
   (c). 证明对于假设 $H_0: \lambda = \lambda_0 \leftrightarrow H_1: \lambda > \lambda_0$，上述检验是一致最优势的.
3. 设样本 $X_1, X_2, \cdots, X_n$ 来自总体 $X \sim B(1, \theta)$. 现在有假设检验

$$H_0: \theta = 0.48 \leftrightarrow H_1: \theta = 0.52.$$

   检验方法如下：若 $\sum_{i=1}^n x_i$ 较大则拒绝 H_0. 利用中心极限定理进行近似计算，求样本量应该至少为多大才能使两类错误发生的概率都约为 0.01.
4. 设 $X_1, X_2, \cdots, X_n$ 为来自 $\mathrm{Beta}(\mu, 1)$ 的样本, $Y_1, Y_2, \cdots, Y_m$ 为来自 $\mathrm{Beta}(\theta, 1)$ 的样本，且假设这两个样本之间也相互独立.
   (a). 求 $H_0: \theta = \mu \leftrightarrow H_1: \theta \neq \mu$ 的广义似然比检验.
   (b). 证明上述检验可以基于如下统计量

$$T = \frac{\sum_{i=1}^n \log X_i}{\sum_{i=1}^n \log X_i + \sum_{i=1}^m \log Y_i}.$$

5. 假设基因频率是均衡的，基因型 AA,Aa,aa 出现的频率分比为 $(1-\theta)^2$，$2\theta(1-\theta)$，θ^2. Plato 等发表了如下 190 个人类样本中触珠蛋白型的数据：

Hp1-1	Hp1-2	Hp2-2
10	68	112

表 1 触珠蛋白型

请验证遗传模型的拟合优度.

6. Bhattacharjee 等为研究在盈月期间动物是否更易咬人而收集了某医疗结构处理被动物咬伤的病例. 月亮周期被划分为 10 个阶段，每个阶段被咬伤的人数如下表所示. 29 日是盈月. 咬伤事件是否有时间趋势?

日期	2,3,4	5,6,7	8,9,10	11,12,13	14,15
人数	155	142	146	148	110

日期	16,17,18	19,20,21	22,23,24	25,26,27	28,29,1
人数	137	150	163	201	269

2021 年秋统计学习题 08

1. 文件 Distribution-Test-Beta-Data.xlsx 中有 100 个数据，请检验这些数据是否为来自 Beta 分布 Beta$(2,1)$ 的观察值.
2. 生成容量为 100 的指数分布随机数（参数可以自定或取为 1），作出概率图.
3. 设 X 为连续型随机变量，$X_1, X_2, \cdots, X_n$ 为来自总体 X 的样本，求 $X_{(n)}$ 大于 X 的中位数 m 的概率.
4. 设 X_1, X_2 独立同分布，为来自均匀分布 $(\theta, 1+\theta)$ 的样本. 为检验 $H_0{:}\,\theta = 0 \leftrightarrow H_1{:}\,\theta > 0$，考虑如下两种检验方法：

$$\phi_1\text{：}X_1 > 0.95\text{时，拒绝}H_0,$$
$$\phi_2\text{：}X_1 + X_2 > C\text{时，拒绝}H_0.$$

   (a). 求 C 的值使得检验 ϕ_1, ϕ_2 的显著性水平相同.
   (b). 求两个检验各自的势函数.
   (c). 证明否否定：ϕ_2 的势比 ϕ_1 的势更大.

2021 年秋统计学习题 09

1. 表1中的数据为两种类型机器的表现（参文件 Exe9-Type.xlsx）．（本题为 Rice 的书第 11 章习题 21）
   (a). 请基于正态性假设检验两种类型机器是否有差别
   (b). 用非参数方法进行同样的假设检验
   (c). 哪种方法好?

类型 I	类型 II
3.03	3.19
5.53	4.26
5.6	4.47
9.3	4.53
9.92	4.67
12.51	4.69
12.95	12.78
15.21	6.79
16.04	9.37
16.84	12.75

表 1 两种类型

2. 设 $\{X_i\}_{i=1}^n$ 为来自总体 X 的样本，期望 $\mathbb{E}X = \mu$，$\operatorname{Var} X = \sigma^2$ 都有限，记 S_n^2 为样本方差，$\bar{X}_n$ 为样本均值，证明
   (a). S_n^2 依分布收敛到 σ^2；
   (b).
$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n}$$
   依分布收敛于标准正态分布.

3. 对 $n = 4$ 的情形写成 W_+ 的精确分布.
4. 表2（文件 Exe9-Method.xlsx）为使用两种方法得到的数据. 请用如下方法考虑两种方法是否有系统差异. （本题为 Rice 的书第 11 章习题 36）
   (a). 配对 t 检验;
   (b). Wilcoxon 符号秩检验;
   (c). 用正态分布近似进行符号秩检验.

方法 A	方法 B
97.2	97.2
105.8	97.8
99.5	96.2
100.0	101.8
93.8	88.0
79.2	74.0
72.0	75.0
72.0	67.5
69.5	65.8
20.5	21.2
95.2	94.8
90.8	95.8
96.2	98.0
96.2	99.0
91.0	100.2

表 2　两种方法

2021 年秋统计学习题 10

1. 下表（参文件 tablets2.xls）给出了另一个制造商所生产的马来酸氯苯那敏片的测量数据. 实验室之间具有系统性差异吗？如果有，哪一对具有显著差异性？

Lab1	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
4.15	3.93	4.10	4.16	4.14	4.09	4.10
4.08	3.92	4.10	4.14	4.14	4.08	4.05
4.09	4.08	4.05	4.16	4.14	4.08	4.05
4.08	4.09	4.07	4.02	4.17	4.12	4.06
4.01	4.06	4.06	4.15	4.13	4.17	4.06
4.01	4.06	4.03	4.15	4.24	4.15	4.12
4.00	4.02	4.04	4.18	4.18	4.12	4.07
4.09	4.00	4.03	4.12	4.14	4.10	4.18
4.08	4.01	4.03	4.09	4.25	4.12	4.15
4.00	4.01	4.06	4.04	4.17	4.12	4.18

表 1 实验室

2. 利用异氟醚、氟烷和环丙烷麻醉 10 条狗，测量血浆肾上腺素的浓度（用纳克每毫升表示）. 测量结果由下表（参文件 Exe10-Dog）给出. 试验效果有差异吗？利用参数和非参数方法进行分析.

	狗 1	狗 2	狗 3	狗 4	狗 5	狗 6	狗 7	狗 8	狗 9	狗 10
异氟醚	0.28	0.51	1.00	0.39	0.29	0.36	0.32	0.69	0.17	0.33
氟烷	0.30	0.39	0.63	0.68	0.38	0.21	0.88	0.39	0.51	0.32
环丙烷	1.07	1.35	0.69	0.28	1.24	1.53	0.49	0.56	1.02	0.30

表 2 麻醉方法

3. 下表（参文件 Exe10-Poison.xlsx）给出了试验中动物的存活时间 (以小时为单位)，该试验设计包含三种毒药，4 种试验方法，每个单元有 4 个观测值.

毒药	试验方法							
	A		B		C		D	
I	3.1	4.5	8.2	11.0	4.3	4.5	4.5	7.1
	4.6	4.3	8.8	7.2	6.3	7.6	6.6	6.2
II	3.6	2.9	9.2	6.1	4.4	3.5	5.6	10.0
	4	2.3	4.9	12.4	3.1	4.0	7.1	3.8
III	2.2	2.1	3.0	3.7	2.3	2.5	3.0	3.6
	1.8	2.3	3.8	2.9	2.4	2.5	3.1	3.3

(a). 利用双因素方差分析检验两个主因素和它们交互的效应.

(b). Box 和 Cox(1964) 分析了数据的倒数，指出存活时间的倒数可以解释为死亡率. 进行双因素方差分析，并与 (a) 中的结果进行比较. 评述标准双因素方差分析模型的拟合优度，以及两个分析中的交互效因.