1. 设 X_1, X_2 是数学期望为 θ 的指数分布总体 X 的容量为 2 的样本，设 $Y = \sqrt{X_1 X_2}$，试证明:

$$E(\frac{4Y}{\pi}) = \theta$$

2. 设总体 $X \sim N(\mu, \sigma^2), X_1, X_2, \cdots, X_n$ 是一个样本，$\overline{X}, S^2$ 分别为样本均值和样本方差，试证

$$E[(\overline{X}S^2)^2] = \left(\frac{\sigma^2}{n} + \mu^2\right)\left(\frac{2\sigma^4}{n-1} + \sigma^4\right)$$

3. 设总体 X 具有概率密度

$$f(x) = \begin{cases} \dfrac{1}{\theta^2} x e^{-x/\theta}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

其中 $\theta > 0$ 为未知参数，$X_1, X_2, \cdots, X_n$ 是来自 X 的样本，$x_1, x_2, \cdots, x_n$ 是相应的样本观察值.

(1) 求 θ 的最大似然估计量.

(2) 求 θ 的矩估计量.

(3) 问求得的估计量是否是无偏估计量.

4. 设 $X_1, X_2, \cdots, X_n$ 以及 $Y_1, Y_2, \cdots, Y_n$ 为分别来自总体 $N(\mu_1, \sigma^2)$ 与 $N(\mu_2, \sigma^2)$ 的样本，且它们相互独立，μ_1, μ_2, σ^2 均未知，试求 μ_1, μ_2, σ^2 的最大似然估计量.

5. 为了研究一批贮存着的产品的可靠性，在产品投入贮存时，即在时刻 $t_0 = 0$ 时，随机地选定 n 件产品，然后在预先规定的时刻 $t_1, t_2, \cdots, t_k$ 取出来进行检测（检测时确定已失效的去掉，将未失效的继续投入贮存），今得到以下的寿命试验数据：

检测时刻（月）	t_1	t_2	$\cdots$	t_k		
区间 $(t_{i-1}, t_i]$	$(0, t_1]$	$(t_1, t_2]$	$\cdots$	$(t_{k-1}, t_k]$	(t_k, ∞)	
在 $(t_{i-1}, t_i]$ 的失效数	d_1	d_2	$\cdots$	d_k	s	$\sum_{i=1}^{k} d_i + s = n$

这种数据称为区间数据．设产品寿命 T 服从指数分布，其概率密度为

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & t > 0, \\ 0, & \text{其他}, \end{cases} \quad \lambda > 0 \text{未知}$$

(1) 试基于上述数据写出 λ 的对数似然方程．（提示：考虑事件“n 只产品分别在区间 $(0, t_1], (t_1, t_2], \cdots, (t_{k-1}, t_k]$ 失效 $d_1, d_2, \cdots, d_k$ 只，而直至 t_k 还有 s 只未失效”的概率.）

(2) 设 $d_1 < n, s < n$，我们可以用数值解法求得 λ 的最大似然估计值，在计算机上计算是容易的．特别，取检测时间是等间隔的，即取 $t_i = it_1, i = 1, 2, \cdots, k$．验证，此时可得 λ 的最大似然估计为

$$\hat{\lambda} = \frac{1}{t_1} \ln \left(1 + \frac{n-s}{\sum_{i=2}^{k} (i-1) d_i + sk} \right)$$

6. 设某种电子器件的寿命（以小时计）T 服从指数分布，概率密度为

$$f(t)=\begin{cases}\lambda e^{-\lambda t}, & t>0,\\ 0, & \text{其他},\end{cases}$$

其中 $\lambda>0$ 未知. 从这批器件中任取 n 只在时刻 $t=0$ 时投入独立寿命试验. 试验进行到预定时间 T_0 结束. 此时，有 $k(0<k<n)$ 只器件失效，试求 λ 的最大似然估计.（提示：考虑“试验直至时刻 T_0 为止，有 k 只器件失效，而有 $n-k$ 只未失效”这一事件的概率，从而写出 λ 的似然方程.）

7. 设系统由两个独立工作的成败型元件串联而成（成败型元件只有两种状态：正常工作或失效）. 元件 1、元件 2 的可靠性分别为 p_1,p_2，它们均未知. 随机地取 N 个系统投入试验，当系统中至少有一个元件失效时系统失效，现得到以下的试验数据：n_1——仅元件 1 失效的系统数；n_2——仅元件 2 失效的系统数；n_{12}——元件 1，元件 2 至少有一个失效的系统数；s——未失效的系统数. $n_1+n_2+n_{12}+s=N$. 这里 n_{12} 为隐蔽数据，也就是只知系统失效，但不能知道是由元件 1 还是元件 2 单独失效引起的，还是由元件 1，2 均失效引起的. 设隐蔽与系统失效的真正原因独立.

(1) 试写出 p_1,p_2 的似然函数.

(2) 设有系统寿命试验数据 $N=20,n_1=5,n_2=3,n_{12}=1,s=11$. 试求 p_1,p_2 的最大似然估计.（提示：p_1 应满足方程 $(p_1-1)(12p_1^2+11p_1-14)=0$.）

8. (1) 设总体 X 具有分布律

X	1	2	3
p_k	θ	θ	$1-2\theta$

$\theta>0$ 未知，今有样本

1 1 1 3 2 1 3 2 2 1 2 2 3 1 1 2

试求 θ 的最大似然估计值和矩估计值.

(2) 设总体 X 服从 Γ 分布. 其概率密度为

$$f(x)=\begin{cases}\dfrac{1}{\beta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}, & x>0\\ 0, & \text{其他}\end{cases}$$

其形状参数 $\alpha>0$ 为已知，尺度参数 $\beta>0$ 未知. 今有样本值 $x_1,x_2,\cdots,x_n$，求 β 的最大似然估计值.

9. (1) 设 $Z=\ln X\sim N(\mu,\sigma^2)$，即 X 服从对数正态分布，验证 $E(X)=\exp(\mu+\dfrac{1}{2}\sigma^2)$.

(2) 设自（1）中总体 X 中取一容量为 n 的样本 $x_1,x_2,\cdots,x_n$，求 $E(X)$ 的最大似然估计. 此处设 μ,σ^2 均为未知.

(3) 已知在文学家萧伯纳的《An Intelligent Woman’s Guide To Socialism》一书中，一个句子的单词数近似地服从对数正态分布，设 μ,σ^2 为未知. 今自该书中随机地取 20 个句子. 这些句子中的单词数分别为

52	24	15	67	15	22	63	26	16	32
7	33	28	14	7	29	10	6	59	30

问这本书中，一个句子单词数均值的最大似然估计值等于多少？

10. 考虑进行定数截尾寿命试验，假设将随机抽取的 n 件产品在时间 $t=0$ 时同时投入试验. 试验进行到 m 件 $(m<n)$ 产品失效时停止，m 件失效产品的失效时间分别为

$$0 \le t_1 \le t_2 \le \cdots \le t_m$$

t_m 是第 m 件产品的失效时间. 设产品的寿命分布为韦布尔分布，其概率密度为

$$f(x)=\begin{cases} \dfrac{\beta}{\eta^\beta}x^{\beta-1}e^{-(\frac{x}{\eta})^\beta}, & x>0 \\ 0, & \text{其他} \end{cases}$$

其中参数 β 已知. 求参数 η 的最大似然估计.

11. 设某大城市郊区的一条林荫道两旁开设了许多小商店，这些商店的开设延续时间（以月计）是一个随机变量，现随机地取 30 家商店，将它们的延续时间按自小到大排序，选其中前 8 家商店. 它们的延续时间分别是

$$3.2 \quad 3.9 \quad 5.9 \quad 6.5 \quad 16.5 \quad 20.3 \quad 40.4 \quad 50.9$$

假设商店开设延续时间的长度是韦布尔随机变量. 其概率密度为

$$f(x)=\begin{cases} \dfrac{\beta}{\eta^\beta}x^{\beta-1}e^{-(\frac{x}{\eta})^\beta}, & x>0 \\ 0, & \text{其他} \end{cases}$$

其中，$\beta=0.8$.

(1) 试用上题结果，写出 η 的最大似然估计.

(2) 按（1）的结果求商店开设延续时间至少为 2 年的概率的估计.

12. 设分别自总体 $N(\mu_1,\sigma^2)$ 和 $N(\mu_2,\sigma^2)$ 中抽取容量 n_1,n_2 的两独立样本，其样本方差分别为 S_1^2,S_2^2. 试证. 对于任意常数 $a,b(a+b=1),Z=aS_1^2+bS_2^2$ 都是 σ^2 的无偏估计，并确定常数 a,b，从使 $D(Z)$ 达到最小.

13. 设总体 $X\sim N(\mu,\sigma^2)$,$X_1,X_2,\cdots,X_n$ 是来自 X 的样本.已知样本方差 $S^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$ 是 σ^2 的无偏估计.验证样本标准差 S 不是标准差 σ 的无偏估计.(提示:记 $Y=\dfrac{(n-1)S^2}{\sigma^2}$，则 $Y\sim\chi^2(n-1)$，而 $S=\dfrac{\sigma}{\sqrt{n-1}}\sqrt{Y}$ 是 Y 的函数，利用 $\chi^2(n-1)$ 的概率密度可得 $E(S)=\dfrac{1}{(\sqrt{2})^n}\sqrt{\dfrac{2}{n-1}}\dfrac{\Gamma(n/2)\sigma}{\Gamma((n-1)/2)}\neq\sigma$.)

14. 设总体 X 服从指数分布，其概率密度为

$$f(x)=\begin{cases} \dfrac{1}{\theta}x^{-x/\theta}, & x>0 \\ 0, & \text{其他} \end{cases}$$

$\theta>0$ 未知. 从总体中抽取一容量为 n 的样本 $X_1,X_2,\cdots,X_n$.

(1) 证明 $\dfrac{2n\overline{X}}{\theta}\sim\chi^2(2n)$.

(2) 求 θ 的置信水平为 $1-\alpha$ 的单侧置信下限.

(3) 某种元件的寿命（以 h 计）服从上述指数分布，现从中抽得一容量 $n=16$ 的样本，测得样本均值为 5010 h, 试求元件的平均寿命的置信水平为 0.90 的单侧置信下限.

15. 设总体 $X \sim U(0,\theta)$,$X_1, X_2, \cdots, X_n$ 是来自 X 的样本.

(1) 验证 $Y=\max\{X_1, X_2, \cdots, X_n\}$ 的分布函数为

$$F_Y(y)=\begin{cases} 0, & y<0 \\ y^n/\theta^n, & 0 \leq y<\theta \\ 1, & y \geq \theta \end{cases}$$

(2) 验证 $U=Y/\theta$ 的概率密度为

$$f_U(u)=\begin{cases} nu^{n-1}, & 0 \leq u \leq 1 \\ 0, & \text{其他} \end{cases}$$

(3) 给定正数 $\alpha, 0<\alpha<1$，求 U 的分布的上 $\alpha/2$ 分位点 $h_{\alpha/2}$ 以及上 $1-\alpha/2$ 分位点 $h_{1-\alpha/2}$.

(4) 利用（2）,（3）求参数 θ 的置信水平为 $1-\alpha$ 的置信区间.

(5) 设某人上班的等车时间 $X \sim U(0,\theta)$，θ 未知. 现在有样本 $x_1=4.2, x_2=3.5, x_3=1.7, x_4=1.2, x_5=2.4$，求 θ 的置信水平为 0.95 的置信区间.

16. 设总体 X 服从指数分布，概率密度为

$$f(x)=\begin{cases} \dfrac{1}{\theta}e^{-x/\theta}, & x>0 \\ 0, & \text{其他} \end{cases} \qquad \theta>0$$

设 $X_1, X_2, \cdots, X_n$ 是来自 X 的一个样本. 试取第 14 题中当 $\theta=\theta_0$ 时的统计量 $\chi^2=\dfrac{2n\overline{X}}{\theta_0}$ 作为检验统计量，检验假设 $H_0: \theta=\theta_0, H_1: \theta \neq \theta_0$. 取显著性水平为 α（注意：$E(\overline{X})=\theta$）设某种电子元件的寿命（以小时计）服从均值为 θ 的指数分布，随机取 12 只元件测得它们的寿命分别为

340 430 560 920 1380 1520 1660 1770 2100 2320 2350 2650

试取显著性水平 $\alpha=0.05$，检验假设 $H_0: \theta=1450, H_1: \theta \neq 1450$.

17. 经过十一年的试验，达尔文于 1876 年得到 15 对玉米样品的数据如下表，每对作物除授粉方式不同外，其他条件都是相同的. 试用逐对比较法检验不同授粉方式对玉米高度是否有显著的影响 ($\alpha=0.05$). 问应增设什么条件才能用逐对比较法进行检验？

授粉方式	1	2	3	4	5	6	7	8
异株授粉的作物高度（x_i）	$23\frac{1}{8}$	12	$20\frac{3}{8}$	22	$19\frac{1}{8}$	$21\frac{4}{8}$	$22\frac{1}{8}$	$20\frac{3}{8}$
同株授粉的作物高度（y_i）	$27\frac{3}{8}$	21	20	20	$19\frac{3}{8}$	$18\frac{5}{8}$	$18\frac{5}{8}$	$15\frac{2}{8}$
授粉方式	9	10	11	12	13	14	15	
异株授粉的作物高度（x_i）	$18\frac{2}{8}$	$21\frac{5}{8}$	$23\frac{2}{8}$	21	$22\frac{1}{8}$	23	12	
同株授粉的作物高度（y_i）	$16\frac{4}{8}$	18	$16\frac{2}{8}$	18	$12\frac{6}{8}$	$15\frac{4}{8}$	18	

18. 一内科医生声称，如果病人每天傍晚聆听一种特殊的轻音乐会降低血压（舒张压，以 mmHg 计）．今选取了 10 个病人在试验之前和试验之后分别测量了血压，得到以下的数据：

病人	1	2	3	4	5	6	7	8	9	10
试验之前（x_i）	86	92	95	84	80	78	98	95	94	96
试验之后（y_i）	84	83	81	78	82	74	86	85	80	82

设 $D_i = X_i - Y_i (i = 1, 2, \cdots, 10)$ 为来自正态总体 $N(\mu_D, \sigma_D^2)$ 的样本，μ_D, σ_D^2 均未知．试检验是否可以认为医生的意见是对的（取 $\alpha = 0.05$）．

19. 以下是各种颜色的汽车的销售情况：

颜色	红	黄	蓝	绿	棕
车辆数	40	64	46	36	14

试检验顾客对这些颜色是否有偏爱，即检验销售情况是否是均匀的（取 $\alpha = 0.05$）．

20. 某种闪光灯，每盏灯含 4 个电池，随机地取 150 盏灯，经检测得到以下的数据：

一盏灯损坏的电池数 x	0	1	2	3	4
灯的盏数	26	51	47	16	10

试取 $\alpha = 0.05$ 检验一盏灯损坏的电池数 $X \sim b(4, \theta)$（θ 未知）．

21. 下面分别给出了某城市在春季（9 周）和秋季（10 周）发生的案件数．

春季	51	42	57	53	43	37	45	49	46	
秋季	40	35	30	44	33	50	41	39	36	38

试取 $\alpha = 0.03$ 用秩和检验法检验春季发生的案件数的均值是否较秋季的为多．

22. 临界闪烁频率 (cff) 是人眼对于闪烁光源能够分辨出它在闪烁的最高频率（以 Hz 计）．超过 cff 的频率，即使光源实际是在闪烁的，而人看起来是连续的（不闪烁的）．一项研究旨在判定 cff 的均值是否与人眼的虹膜颜色有关．所得数据如下：

虹膜颜色	棕色		绿色		蓝色	
	26.8	26.3	26.4	29.1	25.7	29.4
	27.9	24.8	24.2		27.2	28.3
	23.7	25.7	28.0		29.9	
	25.0	24.5	26.9		28.5	

试在显著性水平 0.05 下，检验各种虹膜颜色相应的 cff 的均值有无显著的差异．设各个总体服从正态分布，且方差相等，不同颜色下的样本之间相互独立．

23. 下面列出了挪威人自 1938 $\sim$ 1947 年间年人均脂肪消耗量与患动脉粥样硬化症而死亡的死亡率之间相关的一组数据．

年份	1938	1939	1940	1941	1942	1943	1944	1945	1946	1947
脂肪消耗量 x（千克/人年）	14.4	16.0	11.6	11.0	10.0	9.6	9.2	10.4	11.4	12.5
死亡率 y（1/（10^5 人年））	29.1	29.7	29.2	26.0	24.0	23.1	23.0	23.1	25.2	26.1

设对于给定的 x, Y 为正态变量，且方差与 x 无关．

(1) 求回归直线方程 $y = a + bx$.

(2) 在显著性水平 $\alpha = 0.05$ 下检验假设 $H_0 : b = 0, H_1 : b \neq 0$.

(3) 求 $\hat{y}|_{x=13}$.

(4) 求 $x = 13$ 处 $\mu(x)$ 置信水平为 0.95 的置信区间.

(5) 求 $x = 13$ 处 Y 的新观察值 Y_0 的置信水平为 0.95 的预测区间.

24. 下面给出 1924 $\sim$ 1992 年奥林匹克运动会女子 100 米仰泳的最佳成绩（以 s 计），（其中 1940 年及 1944 年未举行奥运会）

年份	1924	1928	1932	1936	1948	1952	1956	1960
成绩	83.2	82.2	79.4	78.9	74.4	74.3	72.9	69.3
年份	1964	1968	1972	1976	1980	1984	1988	1992
成绩	67.7	66.2	65.8	61.8	60.9	62.6	60.9	60.7

(1) 画出散点图.

(2) 求成绩关于年份的线性回归方程.

(3) 检验回归效果是否显著（取 $\alpha = 0.05$）.