第五章　统计量及其分布

习题 5.1

1．某地电视台想了解某电视栏目（如：每日九点至九点半的体育节目）在该地区的收视率情况，于是委托一家市场咨询公司进行一次电话访查．

（1）该项研究的总体是什么？

（2）该项研究的样本是什么？

解：（1）总体是该地区的全体用户；

（2）样本是被访查的电话用户．

2．某市要调查成年男子的吸烟率，特聘请 50 名统计专业本科生作街头随机调查，要求每位学生调查 100 名成年男子，问该项调查的总体和样本分别是什么，总体用什么分布描述为宜？

解：总体是任意 100 名成年男子中的吸烟人数；样本是这 50 名学生中每一个人调查所得到的吸烟人数；总体用二项分布描述比较合适．

3．设某厂大量生产某种产品，其不合格品率 p 未知，每 m 件产品包装为一盒．为了检查产品的质量，任意抽取 n 盒，查其中的不合格品数，试说明什么是总体，什么是样本，并指出样本的分布．

解：总体是全体盒装产品中每一盒的不合格品数；样本是被抽取的 n 盒产品中每一盒的不合格品数；

总体的分布为 $X \sim b(m, p)$，$P\{X = x\} = \binom{m}{x} p^x q^{m-x}$，$x = 0, 1, \cdots, n$，

样本的分布为 $P\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\} = \binom{m}{x_1} p^{x_1} q^{m-x_1} \cdot \binom{m}{x_2} p^{x_2} q^{m-x_2} \cdots \binom{m}{x_n} p^{x_n} q^{m-x_n}$

$$= \prod_{i=1}^{n} \binom{m}{x_i} \cdot p^{\sum\limits_{i=1}^{n} x_t} q^{mn - \sum\limits_{i=1}^{n} x_t} .$$

4．为估计鱼塘里有多少鱼，一位统计学家设计了一个方案如下：从鱼塘中打捞出一网鱼，计有 n 条，涂上不会被水冲刷掉的红漆后放回，一天后再从鱼塘里打捞一网，发现共有 m 条鱼，而涂有红漆的鱼则有 k 条，你能估计出鱼塘里大概有多少鱼吗？该问题的总体和样本又分别是什么呢？

解：设鱼塘里有 N 条鱼，有涂有红漆的鱼所占比例为 $\dfrac{n}{N}$，

而一天后打捞出的一网鱼中涂有红漆的鱼所占比例为 $\dfrac{k}{m}$，估计 $\dfrac{n}{N} \approx \dfrac{k}{m}$，

故估计出鱼塘里大概有 $N \approx \dfrac{mn}{k}$ 条鱼；

总体是鱼塘里的所有鱼；样本是一天后再从鱼塘里打捞出的一网鱼．

5．某厂生产的电容器的使用寿命服从指数分布，为了了解其平均寿命，从中抽出 n 件产品测其使用寿命，试说明什么是总体，什么是样本，并指出样本的分布．

解：总体是该厂生产的全体电容器的寿命；

样本是被抽取的 n 件电容器的寿命；

总体的分布为 $X \sim \mathrm{e}(\lambda)$，$p(x) = \lambda \mathrm{e}^{\lambda x}$，$x > 0$，

样本的分布为 $p(x_1, x_2, \cdots, x_n) = \lambda \mathrm{e}^{\lambda x_1} \cdot \lambda \mathrm{e}^{\lambda x_2} \cdots \lambda \mathrm{e}^{\lambda x_n} = \lambda^n \mathrm{e}^{\lambda \sum\limits_{i=1}^{n} x_i}$，$x_i > 0$.

6．美国某高校根据毕业生返校情况纪录，宣布该校毕业生的年平均工资为 5 万美元，你对此有何评论？

解：返校的毕业生只是毕业生中一部分特殊群体，样本的抽取不具有随机性，不能反应全体毕业生的情况．

习题 5.2

1．以下是某工厂通过抽样调查得到的 10 名工人一周内生产的产品数

149　156　160　138　149　153　153　169　156　156

试由这批数据构造经验分布函数并作图.

解：经验分布函数

$$F_n(x)=\begin{cases}0, & x<138,\\ 0.1, & 138\le x<149,\\ 0.3, & 149\le x<153,\\ 0.5, & 153\le x<156,\\ 0.8, & 156\le x<160,\\ 0.9, & 160\le x<169,\\ 1, & x\ge 169.\end{cases}$$

作图略.

2．下表是经过整理后得到的分组样本

组序	1	2	3	4	5
分组区间	(38,48]	(48,58]	(58,68]	(68,78]	(78,88]
频数	3	4	8	3	2

试写出此分布样本的经验分布函数.

解：经验分布函数

$$F_n(x)=\begin{cases}0, & x<37.5,\\ 0.15, & 37.5\le x<47.5,\\ 0.35, & 47.5\le x<57.5,\\ 0.75, & 57.5\le x<67.5,\\ 0.9, & 67.5\le x<77.5,\\ 1, & x\ge 77.5.\end{cases}$$

3．假若某地区 30 名 2000 年某专业毕业生实习期满后的月薪数据如下：

909	1086	1120	999	1320	1091
1071	1081	1130	1336	967	1572
825	914	992	1232	950	775
1203	1025	1096	808	1224	1044
871	1164	971	950	866	738

（1）构造该批数据的频率分布表（分 6 组）；

（2）画出直方图.

解：（1）最大观测值为 1572，最小观测值为 738，则组距为 $d=\dfrac{1572-738}{6}\approx 140$，

区间端点可取为 735，875，1015，1155，1295，1435，1575，

频率分布表为

组序	分组区间	组中值	频数	频率	累计频率
1	(735, 875]	805	6	0.2	0.2
2	(875, 1015]	945	8	0.2667	0.4667
3	(1015, 1155]	1085	9	0.3	0.7667
4	(1155, 1295]	1225	4	0.1333	0.9

5	(1295, 1435]	1365	2	0.06667	0.9667
6	(1435, 1575]	1505	1	0.03333	1
合计			30	1	

（2）作图略.

4． 某公司对其 250 名职工上班所需时间（单位：分钟）进行了调查，下面是其不完整的频率分布表：

所需时间	频率
0~10	0.10
10~20	0.24
20~30	
30~40	0.18
40~50	0.14

（1）试将频率分布表补充完整.

（2）该公司上班所需时间在半小时以内有多少人？

解：（1）频率分布表为

组序	分组区间	组中值	频数	频率	累计频率
1	(0, 10]	5	25	0.1	0.1
2	(10, 20]	15	60	0.24	0.34
3	(20, 30]	25	85	0.34	0.68
4	(30, 40]	35	45	0.18	0.86
5	(40, 50]	45	35	0.14	1
合计			250	1	

（2）上班所需时间在半小时以内有 $25 + 60 + 85 = 170$ 人.

5． 40 种刊物的月发行量（单位：百册）如下：

5954	5022	14667	6582	6870	1840	2662	4508
1208	3852	618	3008	1268	1978	7963	2048
3077	993	353	14263	1714	11127	6926	2047
714	5923	6006	14267	1697	13876	4001	2280
1223	12579	13588	7315	4538	13304	1615	8612

（1）建立该批数据的频数分布表，取组距为 1700（百册）;

（2）画出直方图.

解：（1）最大观测值为 353，最小观测值为 14667，则组距为 $d = 1700$，

区间端点可取为 0，1700，3400，5100，6800，8500，10200，11900，13600，15300，

频率分布表为

组序	分组区间	组中值	频数	频率	累计频率
1	(0, 1700]	850	9	0.225	0.225
2	(1700, 3400]	2550	9	0.225	0.45
3	(3400, 5100]	4250	5	0.125	0.575
4	(5100, 6800]	5950	4	0.1	0.675
5	(6800, 8500]	7650	4	0.1	0.775
6	(8500, 10200]	9350	1	0.025	0.8
7	(10200, 11900]	11050	1	0.025	0.825
8	(11900, 13600]	12750	3	0.075	0.9
9	(13600, 15300]	14450	4	0.1	1
合计			30	1	

（2）作图略.

6．对下列数据构造茎叶图

472	425	447	377	341	369	412	399
400	382	366	425	399	398	423	384
418	392	372	418	374	385	439	408
429	428	430	413	405	381	403	479
381	443	441	433	399	379	386	387

解：茎叶图为

34	1
35	
36	9, 6
37	7, 2, 4, 9
38	2, 4, 5, 1, 1, 6, 7
39	9, 8, 2
40	0, 5, 3
41	2, 9, 8, 8, 3, 9
42	5, 5, 3, 8, 9, 8
43	9, 0, 3
44	7, 3, 1
45	
46	
47	2, 9

7．根据调查，某集团公司的中层管理人员的年薪（单位：千元）数据如下：

40.6	39.6	37.8	36.2	38.8
38.6	39.6	40.0	34.7	41.7
38.9	37.9	37.0	35.1	36.7
37.1	37.7	39.2	36.9	38.3

试画出茎叶图.

解：茎叶图为

34.	7
35.	1
36.	2, 7, 9
37.	0, 1, 7
38.	6
39.	6, 6, 2
40.	6, 8, 0
41.	7
42.	
43.	8
44.	9, 5
45.	4

习题 5.3

1．在一本书上我们随机的检查了 10 页，发现每页上的错误数为：

4　5　6　0　3　1　4　2　1　4

试计算其样本均值、样本方差和样本标准差.

解：样本均值 $\bar{x}=\dfrac{1}{10}(4+5+6+\cdots+1+4)=3$；

样本方差 $s^2=\dfrac{1}{9}[(4-3)^2+(5-3)^2+(6-3)^2+\cdots+(1-3)^2+(4-3)^2]\approx 3.7778$；

样本标准差 $s=\sqrt{3.7778}\approx 1.9437$.

2．证明：对任意常数 c, d，有 $\sum\limits_{i=1}^{n}(x_i-c)(y_i-d)=\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})+n(\bar{x}-c)(\bar{y}-d)$.

证：$\sum\limits_{i=1}^{n}(x_i-c)(y_i-d)=\sum\limits_{i=1}^{n}[(x_i-\bar{x})+(\bar{x}-c)][(y_i-\bar{y})+(\bar{y}-d)]$

$$=\sum_{i=1}^{n}[(x_i-\bar{x})(y_i-\bar{y})+(\bar{x}-c)(y_i-\bar{y})+(x_i-\bar{x})(\bar{y}-d)+(\bar{x}-c)(\bar{y}-d)]$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})+(\bar{x}-c)\sum_{i=1}^{n}(y_i-\bar{y})+(\bar{y}-d)\sum_{i=1}^{n}(x_i-\bar{x})+n(\bar{x}-c)(\bar{y}-d)$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})+0+0+n(\bar{x}-c)(\bar{y}-d)=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})+n(\bar{x}-c)(\bar{y}-d)\ .$$

3．设 $x_1, \cdots, x_n$ 和 $y_1, \cdots, y_n$ 是两组样本观测值，且有如下关系：$y_i=3x_i-4$，$i=1, \cdots, n$，试求样本均值 $\bar{x}$ 和 $\bar{y}$ 间的关系以及样本方差 s_x^2 和 s_y^2 间的关系.

解：$\bar{y}=\dfrac{1}{n}\sum\limits_{i=1}^{n}y_i=\dfrac{1}{n}\sum\limits_{i=1}^{n}(3x_i-4)=\dfrac{1}{n}\left(3\sum\limits_{i=1}^{n}x_i-4n\right)=\dfrac{3}{n}\sum\limits_{i=1}^{n}x_i-4=3\bar{x}-4$；

$$s_y^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2=\frac{1}{n-1}\sum_{i=1}^{n}[(3x_i-4)-(3\bar{x}-4)]^2=\frac{9}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2=9s_x^2\ .$$

4．记 $\bar{x}_n=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i$ ， $s_n^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2$ ，$n=1, 2, \cdots$，证明

$$\bar{x}_{n+1}=\bar{x}_n+\frac{1}{n+1}(x_{n+1}-\bar{x}_n)\text{，}\quad s_{n+1}^2=\frac{n-1}{n}s_n^2+\frac{1}{n+1}(x_{n+1}-\bar{x}_n)^2\ .$$

证：$\bar{x}_{n+1}=\dfrac{1}{n+1}\sum\limits_{i=1}^{n+1}x_i=\dfrac{n}{n+1}\cdot\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i+\dfrac{1}{n+1}x_{n+1}=\dfrac{n}{n+1}\bar{x}_n+\dfrac{1}{n+1}x_{n+1}=\bar{x}_n+\dfrac{1}{n+1}(x_{n+1}-\bar{x}_n)$；

$$s_{n+1}^2=\frac{1}{n}\sum_{i=1}^{n+1}(x_i-\bar{x}_{n+1})^2=\frac{1}{n}\left[\sum_{i=1}^{n+1}(x_i-\bar{x}_n)^2-(n+1)(\bar{x}_n-\bar{x}_{n+1})^2\right]$$

$$=\frac{1}{n}\left[\sum_{i=1}^{n}(x_i-\bar{x}_n)^2+(x_{n+1}-\bar{x}_n)^2-(n+1)\cdot\frac{1}{(n+1)^2}(x_{n+1}-\bar{x}_n)^2\right]$$

$$=\frac{1}{n}\left[(n-1)\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2+\frac{n}{n+1}(x_{n+1}-\bar{x}_n)^2\right]=\frac{n-1}{n}s_n^2+\frac{1}{n+1}(x_{n+1}-\bar{x}_n)^2\ .$$

5．从同一总体中抽取两个容量分别为 n, m 的样本，样本均值分别为 $\bar{x}_1$，$\bar{x}_2$，样本方差分别为 s_1^2，s_2^2，将两组样本合并，其均值、方差分别为 $\bar{x}$, s^2，证明：

$$\bar{x}=\frac{n\bar{x}_1+m\bar{x}_2}{n+m}\,,\quad s^2=\frac{(n-1)s_1^2+(m-1)s_2^2}{n+m-1}+\frac{nm(\bar{x}_1-\bar{x}_2)^2}{(n+m)(n+m-1)}\,.$$

证：$\bar{x}=\dfrac{1}{n+m}\left(\sum\limits_{i=1}^{n}x_{1i}+\sum\limits_{j=1}^{m}x_{2j}\right)=\dfrac{1}{n+m}\left(\sum\limits_{i=1}^{n}x_{1i}+\sum\limits_{j=1}^{m}x_{2j}\right)=\dfrac{n\bar{x}_1+m\bar{x}_2}{n+m}$；

$$s^2=\frac{1}{n+m-1}\left[\sum_{i=1}^{n}(x_{1i}-\bar{x})^2+\sum_{j=1}^{m}(x_{2j}-\bar{x})^2\right]$$

$$=\frac{1}{n+m-1}\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2+n(\bar{x}_1-\bar{x})^2+\sum_{j=1}^{m}(x_{2j}-\bar{x}_2)^2+m(\bar{x}_2-\bar{x})^2\right]$$

$$=\frac{1}{n+m-1}\left[(n-1)s_1^2+n\left(\bar{x}_1-\frac{n\bar{x}_1+m\bar{x}_2}{n+m}\right)^2+(m-1)s_2^2+m\left(\bar{x}_2-\frac{n\bar{x}_1+m\bar{x}_2}{n+m}\right)^2\right]$$

$$=\frac{(n-1)s_1^2+(m-1)s_2^2}{n+m-1}+\frac{1}{n+m-1}\cdot\frac{nm^2(\bar{x}_1-\bar{x}_2)^2+mn^2(\bar{x}_2-\bar{x}_1)^2}{(n+m)^2}$$

$$=\frac{(n-1)s_1^2+(m-1)s_2^2}{n+m-1}+\frac{nm(\bar{x}_1-\bar{x}_2)^2}{(n+m)(n+m-1)}\,.$$

6．设有容量为 n 的样本 A，它的样本均值为 $\bar{x}_A$，样本标准差为 s_A，样本极差为 R_A，样本中位数为 m_A．现对样本中每一个观测值施行如下变换：$y=ax+b$，如此得到样本 B，试写出样本 B 的均值、标准差、极差和中位数．

解：$\bar{y}_B=\dfrac{1}{n}\sum\limits_{i=1}^{n}y_i=\dfrac{1}{n}\sum\limits_{i=1}^{n}(ax_i+b)=\dfrac{1}{n}(a\sum\limits_{i=1}^{n}x_i+nb)=a\cdot\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i+b=a\bar{x}_A+b$；

$$s_B=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y}_B)^2}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(ax_i+b-a\bar{x}_A-b)^2}=|a|\cdot\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}_A)^2}=|a|\,s_A;$$

$R_B=y_{(n)}-y_{(1)}=a\,x_{(n)}+b-a\,x_{(1)}-b=a\,[x_{(n)}-x_{(1)}]=a\,R_A$；

当 n 为奇数时，$m_{B0.5}=y_{\left(\frac{n+1}{2}\right)}=ax_{\left(\frac{n+1}{2}\right)}+b=am_{A0.5}+b$，

当 n 为偶数时，$m_{B0.5}=\dfrac{1}{2}[y_{\left(\frac{n}{2}\right)}+y_{\left(\frac{n}{2}+1\right)}]=\dfrac{1}{2}[ax_{\left(\frac{n}{2}\right)}+b+ax_{\left(\frac{n}{2}+1\right)}+b]=\dfrac{a}{2}[x_{\left(\frac{n}{2}\right)}+x_{\left(\frac{n}{2}+1\right)}]+b=am_{A0.5}+b$，

故 $m_{B0.5}=a\,m_{A0.5}+b$.

7．证明：容量为 2 的样本 x_1 , x_2 的方差为 $s^2=\dfrac{1}{2}(x_1-x_2)^2$．

证：$s^2=\dfrac{1}{2-1}[(x_1-\dfrac{x_1+x_2}{2})^2+(x_2-\dfrac{x_1+x_2}{2})^2]=\dfrac{(x_1-x_2)^2}{4}+\dfrac{(x_2-x_1)^2}{4}=\dfrac{1}{2}(x_1-x_2)^2$．

8．设 $x_1,\cdots,x_n$ 是来自 $U(-1,1)$ 的样本，试求 $E(\bar{X})$ 和 $\mathrm{Var}(\bar{X})$．

解：因 $X_i \sim U(-1, 1)$，有 $E(X_i)=\dfrac{-1+1}{2}=0$， $\mathrm{Var}(X_i)=\dfrac{(1+1)^2}{12}=\dfrac{1}{3}$，

故 $E(\overline{X})=E(\dfrac{1}{n}\sum_{i=1}^{n}X_i)=\dfrac{1}{n}\sum_{i=1}^{n}E(X_i)=0$， $\mathrm{Var}(\overline{X})=\mathrm{Var}\left(\dfrac{1}{n}\sum_{i=1}^{n}X_i\right)=\dfrac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i)=\dfrac{1}{n^2}\cdot n\cdot\dfrac{1}{3}=\dfrac{1}{3n}$.

9． 设总体二阶矩存在，$X_1, \cdots, X_n$ 是样本，证明 $X_i-\overline{X}$ 与 $X_j-\overline{X}$ $(i\neq j)$ 的相关系数为 $-(n-1)^{-1}$.

证：因 $X_1, X_2, \cdots, X_n$ 相互独立，有 $\mathrm{Cov}\,(X_l, X_k)=0$，$(l\neq k)$，

则 $\mathrm{Cov}(X_i-\overline{X}, X_j-\overline{X})=\mathrm{Cov}(X_i, X_j)-\mathrm{Cov}(X_i, \overline{X})-\mathrm{Cov}(\overline{X}, X_j)+\mathrm{Cov}(\overline{X}, \overline{X})$

$$=0-\mathrm{Cov}(X_i, \frac{1}{n}X_i)-\mathrm{Cov}(\frac{1}{n}X_j, X_j)+\mathrm{Var}(\overline{X})$$

$$=-\frac{1}{n}\mathrm{Var}(X_i)-\frac{1}{n}\mathrm{Var}(X_j)+\mathrm{Var}(\overline{X})=-\frac{1}{n}\sigma^2-\frac{1}{n}\sigma^2+\frac{1}{n}\sigma^2=-\frac{1}{n}\sigma^2,$$

且 $\mathrm{Var}(X_i-\overline{X})=\mathrm{Var}(X_i)+\mathrm{Var}(\overline{X})-2\mathrm{Cov}(X_i, \overline{X})=\sigma^2+\dfrac{1}{n}\sigma^2-2\mathrm{Cov}(X_i, \dfrac{1}{n}X_i)$

$$=\sigma^2+\frac{1}{n}\sigma^2-\frac{2}{n}\sigma^2=\frac{n-1}{n}\sigma^2=\mathrm{Var}(X_j-\overline{X}),$$

故 $\mathrm{Corr}(X_i-\overline{X}, X_j-\overline{X})=\dfrac{\mathrm{Cov}(X_i-\overline{X}, X_j-\overline{X})}{\sqrt{\mathrm{Var}(X_i-\overline{X})}\cdot\sqrt{\mathrm{Var}(X_j-\overline{X})}}=\dfrac{-\dfrac{1}{n}\sigma^2}{\sqrt{\dfrac{n-1}{n}\sigma^2}\cdot\sqrt{\dfrac{n-1}{n}\sigma^2}}=-\dfrac{1}{n-1}$.

10． 设 $x_1, x_2, \cdots, x_n$ 为一个样本， $s^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ 是样本方差，试证：

$$\frac{1}{n(n-1)}\sum_{i<j}(x_i-x_j)^2=s^2.$$

证：因 $s^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2=\dfrac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)$，

则 $\sum_{i<j}(x_i-x_j)^2=\dfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i-x_j)^2=\dfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i^2+x_j^2-2x_ix_j)=\dfrac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}x_i^2+\sum_{i=1}^{n}\sum_{j=1}^{n}x_j^2-2\sum_{i=1}^{n}\sum_{j=1}^{n}x_ix_j\right)$

$$=\frac{1}{2}\left(n\sum_{i=1}^{n}x_i^2+n\sum_{j=1}^{n}x_j^2-2\sum_{i=1}^{n}x_i\sum_{j=1}^{n}x_j\right)=\frac{1}{2}\left(2n\sum_{i=1}^{n}x_i^2-2n\bar{x}\cdot n\bar{x}\right)=n\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)=n(n-1)s^2,$$

故 $\dfrac{1}{n(n-1)}\sum_{i<j}(x_i-x_j)^2=s^2$.

11． 设总体 4 阶中心矩 $\nu_4=E[X-E(X)]^4$ 存在，试对样本方差 $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$，有

$$\mathrm{Var}(S^2)=\frac{n(\nu_4-\sigma^4)}{(n-1)^2}-\frac{2(\nu_4-2\sigma^4)}{(n-1)^2}+\frac{\nu_4-3\sigma^4}{n(n-1)^2},$$

其中σ^2为总体 X 的方差.

证：因 $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}[(X_i-\mu)-(\overline{X}-\mu)]^2=\dfrac{1}{n-1}\left[\sum_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2\right]$，其中$\mu=E(X)$，

则 $\operatorname{Var}(S^2)=\dfrac{1}{(n-1)^2}\operatorname{Var}\left[\sum_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2\right]$

$$=\frac{1}{(n-1)^2}\left\{\operatorname{Var}\left[\sum_{i=1}^{n}(X_i-\mu)^2\right]-2\operatorname{Cov}\left(\sum_{i=1}^{n}(X_i-\mu)^2,n(\overline{X}-\mu)^2\right)+\operatorname{Var}[n(\overline{X}-\mu)^2]\right\}$$

$$=\frac{1}{(n-1)^2}\left\{\sum_{i=1}^{n}\operatorname{Var}(X_i-\mu)^2-2n\sum_{i=1}^{n}\operatorname{Cov}((X_i-\mu)^2,(\overline{X}-\mu)^2)+n^2\operatorname{Var}(\overline{X}-\mu)^2\right\},$$

因 $E(X_i-\mu)^2=\sigma^2$，$E(X_i-\mu)^4=\nu_4$，

则 $\sum_{i=1}^{n}\operatorname{Var}(X_i-\mu)^2=\sum_{i=1}^{n}\{E(X_i-\mu)^4-[E(X_i-\mu)^2]^2\}=\sum_{i=1}^{n}\{\nu_4-(\sigma^2)^2\}=n(\nu_4-\sigma^4)$，

因 $E(X_i-\mu)=0$， $E(\overline{X}-\mu)^2=\operatorname{Var}(\overline{X})=\dfrac{1}{n}\sigma^2$，且当 $i\neq j$ 时，$X_i-\mu$ 与 $X_j-\mu$ 相互独立，

则 $\sum_{i=1}^{n}\operatorname{Cov}((X_i-\mu)^2,(\overline{X}-\mu)^2)=\sum_{i=1}^{n}\{E[(X_i-\mu)^2(\overline{X}-\mu)^2]-E(X_i-\mu)^2E(\overline{X}-\mu)^2\}$

$$=\sum_{i=1}^{n}\left\{E\left[(X_i-\mu)^2\cdot\left(\frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)\right)^2\right]-\sigma^2\cdot\frac{1}{n}\sigma^2\right\}$$

$$=\sum_{i=1}^{n}\left\{\frac{1}{n^2}\left[E(X_i-\mu)^4+E(X_i-\mu)^2\cdot\sum_{k\neq i}E(X_k-\mu)^2\right]-\frac{1}{n}\sigma^4\right\}$$

$$=\sum_{i=1}^{n}\left\{\frac{1}{n^2}[\nu_4+\sigma^2\cdot(n-1)\sigma^2]-\frac{1}{n}\sigma^4\right\}=\frac{1}{n}(\nu_4-\sigma^4),$$

且 $\operatorname{Var}(\overline{X}-\mu)^2=E(\overline{X}-\mu)^4-[E(\overline{X}-\mu)^2]^2=E\left[\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\right]^4-\left[\dfrac{1}{n}\sigma^2\right]^2$

$$=\frac{1}{n^4}E\left[\sum_{i=1}^{n}(X_i-\mu)^4+\binom{4}{2}\sum_{i<j}(X_i-\mu)^2(X_j-\mu)^2\right]-\frac{1}{n^2}\sigma^4$$

$$=\frac{1}{n^4}\left[\sum_{i=1}^{n}E(X_i-\mu)^4+6\sum_{i<j}E(X_i-\mu)^2E(X_j-\mu)^2\right]-\frac{1}{n^2}\sigma^4$$

$$=\frac{1}{n^4}\left[n\nu_4+6\cdot\binom{n}{2}\sigma^2\cdot\sigma^2\right]-\frac{1}{n^2}\sigma^4=\frac{1}{n^4}[n\nu_4+3n(n-1)\sigma^4]-\frac{1}{n^2}\sigma^4=\frac{1}{n^3}(\nu_4-3\sigma^4)+\frac{2}{n^2}\sigma^4,$$

故 $\operatorname{Var}(S^2)=\dfrac{1}{(n-1)^2}\left\{n(\nu_4-\sigma^4)-2n\cdot\dfrac{1}{n}(\nu_4-\sigma^4)+n^2\left[\dfrac{1}{n^3}(\nu_4-3\sigma^4)+\dfrac{2}{n^2}\sigma^4\right]\right\}$

$$=\frac{1}{(n-1)^2}\left\{n(\nu_4-\sigma^4)-2(\nu_4-\sigma^4)+\frac{1}{n}(\nu_4-3\sigma^4)+2\sigma^4\right\}$$

$$=\frac{1}{(n-1)^2}\left\{n(\nu_4-\sigma^4)-2(\nu_4-2\sigma^4)+\frac{1}{n}(\nu_4-3\sigma^4)\right\}=\frac{n(\nu_4-\sigma^4)}{(n-1)^2}-\frac{2(\nu_4-2\sigma^4)}{(n-1)^2}+\frac{\nu_4-3\sigma^4}{n(n-1)^2}.$$

12．设总体 X 的 3 阶矩存在，设 $X_1, X_2, \cdots, X_n$ 是取自该总体的简单随机样本，$\overline{X}$ 为样本均值，S^2 为样本方差，试证：$\mathrm{Cov}(\overline{X}, S^2)=\dfrac{\nu_3}{n}$，其中 $\nu_3=E[X-E(X)]^3$.

证：因 $S^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}[(X_i-\mu)-(\overline{X}-\mu)]^2=\dfrac{1}{n-1}\left[\sum\limits_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2\right]$，其中 $\mu=E(X)$，

则 $\mathrm{Cov}(\overline{X}, S^2)=\mathrm{Cov}(\overline{X}-\mu, S^2)=\mathrm{Cov}\left(\overline{X}-\mu, \dfrac{1}{n-1}\left[\sum\limits_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2\right]\right)$

$$=\frac{1}{n-1}\left[\sum_{i=1}^{n}\mathrm{Cov}(\overline{X}-\mu,(X_i-\mu)^2)-n\,\mathrm{Cov}(\overline{X}-\mu,(\overline{X}-\mu)^2)\right],$$

因 $E(\overline{X}-\mu)=E(X_i-\mu)=0$，$E(X_i-\mu)^2=\sigma^2$，$E(X_i-\mu)^3=\nu_3$，且当 $i\neq j$ 时，$X_i-\mu$ 与 $X_j-\mu$ 相互独立，

则 $\sum\limits_{i=1}^{n}\mathrm{Cov}(\overline{X}-\mu,(X_i-\mu)^2)=\sum\limits_{i=1}^{n}\mathrm{Cov}\left(\dfrac{1}{n}\sum\limits_{k=1}^{n}(X_k-\mu),(X_i-\mu)^2\right)=\dfrac{1}{n}\sum\limits_{i=1}^{n}\mathrm{Cov}(X_i-\mu,(X_i-\mu)^2)$

$$=\frac{1}{n}\sum_{i=1}^{n}[E(X_i-\mu)^3-E(X_i-\mu)E(X_i-\mu)^2]=\frac{1}{n}\cdot n\nu_3=\nu_3,$$

且 $\mathrm{Cov}(\overline{X}-\mu,(\overline{X}-\mu)^2)=E(\overline{X}-\mu)^3-E(\overline{X}-\mu)E(\overline{X}-\mu)^2=E\left[\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\mu)\right]^3$

$$=\frac{1}{n^3}E\left[\sum_{i=1}^{n}(X_i-\mu)^3\right]=\frac{1}{n^3}\sum_{i=1}^{n}E(X_i-\mu)^3=\frac{1}{n^3}\cdot n\nu_3=\frac{1}{n^2}\nu_3,$$

故 $\mathrm{Cov}(\overline{X}, S^2)=\dfrac{1}{n-1}\left(\nu_3-n\cdot\dfrac{1}{n^2}\nu_3\right)=\dfrac{1}{n-1}\cdot\dfrac{n-1}{n}\nu_3=\dfrac{\nu_3}{n}$.

13．设 $\overline{X}_1$ 与 $\overline{X}_2$ 是从同一正态总体 $N(\mu, \sigma^2)$独立抽取的容量相同的两个样本均值．试确定样本容量 n，使得两样本均值的距离超过 σ 的概率不超过 0.01．

解：因 $E(\overline{X}_1)=E(\overline{X}_2)=\mu$，$\mathrm{Var}(\overline{X}_1)=\mathrm{Var}(\overline{X}_2)=\dfrac{\sigma^2}{n}$，$\overline{X}_1$ 与 $\overline{X}_2$ 相互独立，且总体分布为 $N(\mu, \sigma^2)$，

则 $E(\overline{X}_1-\overline{X}_2)=\mu-\mu=0$，$\mathrm{Var}(\overline{X}_1-\overline{X}_2)=\dfrac{\sigma^2}{n}+\dfrac{\sigma^2}{n}=\dfrac{2\sigma^2}{n}$，即 $\overline{X}_1-\overline{X}_2\sim N\left(0, \dfrac{2\sigma^2}{n}\right)$，

因 $P\{|\overline{X}_1-\overline{X}_2|>\sigma\}=2\left[1-\Phi\left(\dfrac{\sigma}{\sigma\sqrt{2/n}}\right)\right]=2-2\Phi\left(\sqrt{\dfrac{n}{2}}\right)\le 0.01$，有 $\Phi\left(\sqrt{\dfrac{n}{2}}\right)\ge 0.995$，$\sqrt{\dfrac{n}{2}}\ge 2.5758$，

故 $n \ge 13.2698$，即 n 至少 14 个．

14．利用切比雪夫不等式求抛均匀硬币多少次才能使正面朝上的频率落在 (0.4, 0.6) 间的概率至少为 0.9．如何才能更精确的计算这个次数？是多少？

解：设 $X_i = \begin{cases} 1, & 第\,i\,次正面朝上, \\ 0, & 第\,i\,次反面朝上, \end{cases}$ 有 $X_i \sim B(1, 0.5)$，且正面朝上的频率为 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ，

则 $E(X_i) = 0.5$，$\text{Var}(X_i) = 0.25$，且 $E(\overline{X}) = 0.5$ ， $\text{Var}(\overline{X}) = \frac{0.25}{n}$ ，

由切比雪夫不等式得 $P\{0.4 < \overline{X} < 0.6\} = P\{|\overline{X} - 0.5| < 0.1\} \ge 1 - \frac{0.25}{0.1^2 n} = 1 - \frac{25}{n}$ ，

故当 $1 - \frac{25}{n} \ge 0.9$ 时，即 $n \ge 250$ 时， $P\{0.4 < \overline{X} < 0.6\} \ge 0.9$ ；

利用中心极限定理更精确地计算，当 n 很大时 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 的渐近分布为正态分布 $N(0.5, \frac{0.25}{n})$ ，

则 $P\{0.4 < \overline{X} < 0.6\} = F(0.6) - F(0.4) = \Phi(\frac{0.6-0.5}{\sqrt{\frac{0.25}{n}}}) - \Phi(\frac{0.4-0.5}{\sqrt{\frac{0.25}{n}}}) = \Phi(0.2\sqrt{n}) - \Phi(-0.2\sqrt{n})$

$= 2\Phi(0.2\sqrt{n}) - 1 \ge 0.9$ ，

即 $\Phi(0.2\sqrt{n}) \ge 0.95$ ， $0.2\sqrt{n} \ge 1.64$ ，

故当 $n \ge 67.24$ 时，即 $n \ge 68$ 时， $P\{0.4 < \overline{X} < 0.6\} \ge 0.9$ ．

15．从指数总体 $Exp(1/\theta)$ 抽取了 40 个样品，试求 $\overline{X}$ 的渐近分布．

解：因 $E(\overline{X}) = E(X) = \theta$ ， $\text{Var}(\overline{X}) = \frac{\text{Var}(X)}{n} = \frac{1}{40}\theta^2$ ，故 $\overline{X}$ 的渐近分布为 $N(\theta, \frac{1}{40}\theta^2)$ ．

16．设 $X_1, \cdots, X_{25}$ 是从均匀分布 $U(0, 5)$ 抽取的样本，试求样本均值 $\overline{X}$ 的渐近分布．

解：因 $E(\overline{X}) = E(X) = \frac{5}{2}$ ， $\text{Var}(\overline{X}) = \frac{\text{Var}(X)}{n} = \frac{(5-0)^2}{25 \times 12} = \frac{1}{12}$ ，故 $\overline{X}$ 的渐近分布为 $N(\frac{5}{2}, \frac{1}{12})$ ．

17．设 $X_1, \cdots, X_{20}$ 是从二点分布 $b(1, p)$ 抽取的样本，试求样本均值 $\overline{X}$ 的渐近分布．

解：因 $E(\overline{X}) = E(X) = p$ ， $\text{Var}(\overline{X}) = \frac{\text{Var}(X)}{n} = \frac{p(1-p)}{20}$ ，故 $\overline{X}$ 的渐近分布为 $N(p, \frac{p(1-p)}{20})$ ．

18．设 $X_1, \cdots, X_8$ 是从正态分布 $N(10, 9)$ 中抽取的样本，试求样本均值 $\overline{X}$ 的标准差．

解：因 $\text{Var}(\overline{X}) = \frac{\text{Var}(X)}{n} = \frac{9}{8}$ ，故 $\overline{X}$ 的标准差为 $\sqrt{\text{Var}(\overline{X})} = \frac{3\sqrt{2}}{4}$ ．

19．切尾均值也是一个常用的反映样本数据的特征量，其想法是将数据的两端的值舍去，而用剩下的当中的值为计算样本均值，其计算公式是

$$\overline{X}_\alpha = \frac{X_{([n\alpha]+1)} + X_{([n\alpha]+2)} + \cdots + X_{(n-[n\alpha])}}{n - 2[n\alpha]} ,$$

其中 $0 < \alpha < 1/2$ 是切尾系数，$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ 是有序样本．现我们在高校采访了 16 名大学生，了解他们平时的学习情况，以下数据是大学生每周用于看电视的时间：

15 14 12 9 20 4 17 26 15 18 6 10 16 15 5 8

取 $\alpha = 1/16$，试计算其切尾均值．

解：因 $n\alpha=1$，且有序样本为 4, 5, 6, 8, 9, 10, 12, 14, 15, 15, 15, 16, 17, 18, 20, 26，

故切尾均值 $\bar{x}_{1/16}=\dfrac{1}{16-2}(5+6+8+\cdots+20)=12.8571$.

20．有一个分组样本如下：

区间	组中值	频数
(145,155)	150	4
(155,165)	160	8
(165,175)	170	6
(175,185)	180	2

试求该分组样本的样本均值、样本标准差、样本偏度和样本峰度.

解：$\bar{x}=\dfrac{1}{20}(150\times4+160\times8+170\times6+180\times2)=163$；

$$s=\sqrt{\frac{1}{19}[(150-163)^2\times4+(160-163)^2\times8+(170-163)^2\times6+(180-163)^2\times2]}=9.2338;$$

因 $b_2=\dfrac{1}{20}[(150-163)^2\times4+(160-163)^2\times8+(170-163)^2\times6+(180-163)^2\times2]=81$,

$b_3=\dfrac{1}{20}[(150-163)^3\times4+(160-163)^3\times8+(170-163)^3\times6+(180-163)^3\times2]=144$,

$b_4=\dfrac{1}{20}[(150-163)^4\times4+(160-163)^4\times8+(170-163)^4\times6+(180-163)^4\times2]=14817$,

故样本偏度 $\gamma_1=\dfrac{b_3}{b_2^{3/2}}=0.1975$，样本峰度 $\gamma_2=\dfrac{b_4}{b_2^2}-3=-0.7417$.

21．检查四批产品，其批次与不合格品率如下：

批号	批量	不合格品率
1	100	0.05
2	300	0.06
3	250	0.04
4	150	0.03

试求这四批产品的总不合格品率.

解：$\bar{p}=\dfrac{1}{800}(100\times0.05+300\times0.06+250\times0.04+150\times0.03)=0.046875$.

22．设总体以等概率取 1, 2, 3, 4, 5，现从中抽取一个容量为 4 的样本，试分别求 $X_{(1)}$ 和 $X_{(4)}$ 的分布.

解：因总体分布函数为

$$F(x)=\begin{cases}0, & x<1,\\ \dfrac{1}{5}, & 1\le x<2,\\ \dfrac{2}{5}, & 2\le x<3,\\ \dfrac{3}{5}, & 3\le x<4,\\ \dfrac{4}{5}, & 4\le x<5,\\ 1, & x\ge5,\end{cases}$$

则 $F_{(1)}(x)=P\{X_{(1)}\le x\}=1-P\{X_{(1)}>x\}=1-P\{X_1>x,X_2>x,X_3>x,X_4>x\}=1-[1-F(x)]^4$

$$
=\begin{cases} 0, & x<1, \\ \dfrac{369}{625}, & 1\le x<2, \\ \dfrac{544}{625}, & 2\le x<3, \\ \dfrac{609}{625}, & 3\le x<4, \\ \dfrac{624}{625}, & 4\le x<5, \\ 1, & x\ge 5, \end{cases}
$$

且 $F_{(4)}(x)=P\{X_{(4)}\le x\}=P\{X_1\le x, X_2\le x, X_3\le x, X_4\le x\}=[F(x)]^4$

$$
=\begin{cases} 0, & x<1, \\ \dfrac{1}{625}, & 1\le x<2, \\ \dfrac{16}{625}, & 2\le x<3, \\ \dfrac{81}{625}, & 3\le x<4, \\ \dfrac{256}{625}, & 4\le x<5, \\ 1, & x\ge 5, \end{cases}
$$

故 $X_{(1)}$ 和 $X_{(4)}$ 的分布为

$X_{(1)}$	1	2	3	4	5
P	$\frac{369}{625}$	$\frac{175}{625}$	$\frac{65}{625}$	$\frac{15}{625}$	$\frac{1}{625}$

；

$X_{(4)}$	1	2	3	4	5
P	$\frac{1}{625}$	$\frac{15}{625}$	$\frac{65}{625}$	$\frac{175}{625}$	$\frac{369}{625}$

.

23．设总体 X 服从几何分布，即 $P\{X=k\}=pq^{k-1}$，$k=1, 2, \cdots$，其中 $0<p<1$，$q=1-p$，$X_1, X_2, \cdots, X_n$ 为该总体的样本．求 $X_{(n)}, X_{(1)}$的概率分布．

解：因 $P\{X\le k\}=\sum_{j=1}^{k} pq^{j-1}=\dfrac{p(1-q^k)}{1-q}=1-q^k$，$k=1, 2, \cdots$，

故 $P\{X_{(n)}=k\}=P\{X_{(n)}\le k\}-P\{X_{(n)}\le k-1\}=\prod_{i=1}^{n}P\{X_i\le k\}-\prod_{i=1}^{n}P\{X_i\le k-1\}=(1-q^k)^n-(1-q^{k-1})^n$；

且 $P\{X_{(1)}=k\}=P\{X_{(1)}>k-1\}-P\{X_{(1)}>k\}=\prod_{i=1}^{n}P\{X_i>k-1\}-\prod_{i=1}^{n}P\{X_i>k\}=q^{n(k-1)}-q^{nk}$．

24．设 $X_1, \cdots, X_{16}$ 是来自 $N(8, 4)$ 的样本，试求下列概率

（1）$P\{X_{(16)}>10\}$；

（2）$P\{X_{(1)}>5\}$．

解：（1） $P\{X_{(16)}>10\}=1-P\{X_{(16)}\le 10\}=1-\prod_{i=1}^{16}P\{X_i\le 10\}=1-[F(10)]^{16}=1-[\Phi(\dfrac{10-8}{2})]^{16}$

$=1-[\Phi(1)]^{16}=1-0.8413^{16}=0.9370$；

（2） $P\{X_{(1)}>5\}=\prod_{i=1}^{16}P\{X_i>5\}=[1-F(5)]^{16}=[1-\Phi(\dfrac{5-8}{2})]^{16}=[\Phi(1.5)]^{16}=0.9332^{16}=0.3308$．

25．设总体为韦布尔分布，其密度函数为

$$p(x;m,\eta)=\frac{mx^{m-1}}{\eta^m}\exp\left\{-\left(\frac{x}{\eta}\right)^m\right\}，x>0, m>0, \eta>0.$$

现从中得到样本 $X_1, \cdots, X_n$ ，证明 $X_{(1)}$ 仍服从韦布尔分布，并指出其参数.

解：总体分布函数 $F(x)=\int_0^x p(t)\mathrm{d}t=\int_0^x \frac{mt^{m-1}}{\eta^m}\mathrm{e}^{-\left(\frac{t}{\eta}\right)^m}\mathrm{d}t=\int_0^x \mathrm{e}^{-\left(\frac{t}{\eta}\right)^m}\mathrm{d}\left(\frac{t}{\eta}\right)^m=-\mathrm{e}^{-\left(\frac{t}{\eta}\right)^m}\Bigg|_0^x=1-\mathrm{e}^{-\left(\frac{x}{\eta}\right)^m}$，$x>0$，

则 $X_{(1)}$ 的密度函数为

$$p_1(x)=n[1-F(x)]^{n-1}p(x)=n\mathrm{e}^{-(n-1)\left(\frac{x}{\eta}\right)^m}\cdot\frac{mx^{m-1}}{\eta^m}\mathrm{e}^{-\left(\frac{x}{\eta}\right)^m}=\frac{mnx^{m-1}}{\eta^m}\mathrm{e}^{-n\left(\frac{x}{\eta}\right)^m}=\frac{mx^{m-1}}{(\eta/\sqrt[m]{n})^m}\mathrm{e}^{-\left(\frac{x}{\eta/\sqrt[m]{n}}\right)^m},$$

故 $X_{(1)}$ 服从参数为 $\left(m,\frac{\eta}{\sqrt[m]{n}}\right)$ 的韦布尔分布.

26．设总体密度函数为 $p(x)=6x(1-x), 0<x<1$，$X_1, \cdots, X_9$ 是来自该总体的样本，试求样本中位数的分布.

解：总体分布函数 $F(x)=\int_0^x p(t)\mathrm{d}t=\int_0^x 6t(1-t)\mathrm{d}t=(3t^2-2t^3)\Big|_0^x=3x^2-2x^3$，$0<x<1$，

因样本容量 $n=9$，有样本中位数 $m_{0.5}=x_{\left(\frac{n+1}{2}\right)}=x_{(5)}$，其密度函数为

$$p_5(x)=\frac{9!}{4!4!}[F(x)]^4[1-F(x)]^4p(x)=\frac{9!}{4!4!}(3x^2-2x^3)^4(1-3x^2+2x^3)^4\cdot 6x(1-x).$$

27．证明公式

$$\sum_{k=0}^r\binom{n}{k}p^k(1-p)^{n-k}=\frac{n!}{r!(n-r-1)!}\int_p^1 x^r(1-x)^{n-r-1}dx，\text{其中 } 0\le p\le 1.$$

证：设总体 X 服从区间(0, 1)上的均匀分布，$X_1, X_2, \cdots, X_n$ 为样本，$X_{(1)}, X_{(2)}, \cdots, X_{(n)}$是顺序统计量，
则样本观测值中不超过 p 的样品个数服从二项分布 $b(n, p)$，即最多有 r 个样品不超过 p 的概率为

$$P\{X_{(r+1)}>p\}=\sum_{k=0}^r\binom{n}{k}p^k(1-p)^{n-k},$$

因总体 X 的密度函数与分布函数分别为

$$p(x)=\begin{cases}1, & 0<x<1;\\ 0, & \text{其他.}\end{cases}\qquad F(x)=\begin{cases}0, & x<0;\\ x, & 0\le x<1;\\ 1, & x\ge 1.\end{cases}$$

则 $X_{(r+1)}$的密度函数为

$$p_{r+1}(x)=\frac{n!}{r!(n-r-1)!}[F(x)]^r[1-F(x)]^{n-r-1}p(x)=\begin{cases}\frac{n!}{r!(n-r-1)!}x^r(1-x)^{n-r-1}, & 0<x<1,\\ 0, & \text{其他.}\end{cases}$$

故 $\sum_{k=0}^r\binom{n}{k}p^k(1-p)^{n-k}=P\{X_{(r+1)}>p\}=\frac{n!}{r!(n-r-1)!}\int_p^1 x^r(1-x)^{n-r-1}dx$.

28．设总体 X 的分布函数 $F(x)$是连续的，$X_{(1)}, \cdots, X_{(n)}$为取自此总体的次序统计量，设$\eta_i=F(X_{(i)})$，试证：
（1）$\eta_1\le\eta_2\le\cdots\le\eta_n$，且$\eta_i$是来自均匀分布 $U(0, 1)$总体的次序统计量；

（2） $E(\eta_i)=\dfrac{i}{n+1}$， $\mathrm{Var}(\eta_i)=\dfrac{i(n+1-i)}{(n+1)^2(n+2)}$， $1\le i\le n$；

（3） η_i 和 η_j 的协方差矩阵为

$$\begin{pmatrix} \dfrac{a_1(1-a_1)}{n+2} & \dfrac{a_1(1-a_2)}{n+2} \\ \dfrac{a_1(1-a_2)}{n+2} & \dfrac{a_2(1-a_2)}{n+2} \end{pmatrix}$$

其中 $a_1=\dfrac{i}{n+1}$， $a_2=\dfrac{j}{n+1}$.

注：第（3）问应要求 $i<j$.

解：（1） 首先证明 $Y=F(X)$的分布是均匀分布 $U(0, 1)$，

因分布函数 $F(x)$连续，对于任意的 $y\in(0, 1)$，存在 x，使得 $F(x)=y$，

则 $F_Y(y)=P\{Y=F(X)\le y\}=P\{F(X)\le F(x)\}=P\{X\le x\}=F(x)=y$，

即 $Y=F(X)$的分布函数是

$$F_Y(y)=\begin{cases}0, & y<0;\\ y, & 0\le y<1;\\ 1, & y\ge 1.\end{cases}$$

可得 $Y=F(X)$的分布是均匀分布 $U(0, 1)$，即 $F(X_1), F(X_2), \cdots, F(X_n)$是均匀分布总体 $U(0, 1)$的样本，

因分布函数 $F(x)$单调不减， $\eta_i=F(X_{(i)})$， 且 $X_{(1)}\le X_{(2)}\le\cdots\le X_{(n)}$是总体 X 的次序统计量，

故 $\eta_1\le\eta_2\le\cdots\le\eta_n$， 且 η_i 是来自均匀分布 $U(0, 1)$总体的次序统计量；

（2） 因均匀分布 $U(0, 1)$ 的密度函数与分布函数分别为

$$p_Y(y)=\begin{cases}1, & 0<y<1;\\ 0, & \text{其他}.\end{cases}\quad F_Y(y)=\begin{cases}0, & y<0;\\ y, & 0\le y<1;\\ 1, & y\ge 1.\end{cases}$$

则 $\eta_i=F(X_{(i)})$的密度函数为

$$p_i(y)=\frac{n!}{(i-1)!(n-i)!}[F_Y(y)]^{i-1}[1-F_Y(y)]^{n-i}p_Y(y)=\begin{cases}\dfrac{n!}{(i-1)!(n-i)!}y^{i-1}(1-y)^{n-i}, & 0<y<1,\\ 0, & \text{其他}.\end{cases}$$

即 η_i 服从贝塔分布 $Be(i, n-i+1)$， 即 $Be(a, b)$， 其中 $a=i$， $b=n-i+1$，

故 $E(\eta_i)=\dfrac{a}{a+b}=\dfrac{i}{n+1}$， $\mathrm{Var}(\eta_i)=\dfrac{ab}{(a+b)^2(a+b+1)}=\dfrac{i(n+1-i)}{(n+1)^2(n+2)}$， $1\le i\le n$；

（3） 当 $i<j$ 时， (η_i, η_j)的联合密度函数为

$$p_{ij}(y,z)=\frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F_Y(y)]^{i-1}[F_Y(z)-F_Y(y)]^{j-i-1}[1-F_Y(z)]^{n-j}p_Y(y)p_Y(z)\mathrm{I}_{y<z}$$

$$=\frac{n!}{(i-1)!(j-i-1)!(n-j)!}y^{i-1}(z-y)^{j-i-1}(1-z)^{n-j}\mathrm{I}_{0<y<z<1},$$

则 $E(\eta_i\eta_j)=\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yz\cdot p_{ij}(y,z)dydz=\frac{n!}{(i-1)!(j-i-1)!(n-j)!}\int_0^1 dz\int_0^z y^i(z-y)^{j-i-1}\cdot z(1-z)^{n-j}dy$，

令 $y=zu$， 有 $dy=zdu$， 且当 $y=0$ 时， $u=0$； 当 $y=z$ 时， $u=1$，

则 $\displaystyle\int_0^z y^i(z-y)^{j-i-1}\cdot z(1-z)^{n-j}dy=z(1-z)^{n-j}\int_0^1(zu)^i(z-zu)^{j-i-1}\cdot zdu$

$$=z(1-z)^{n-j}\cdot z^j\int_0^1 u^i(1-u)^{j-i-1}du=z^{j+1}(1-z)^{n-j}\cdot B(i+1, j-i)=\frac{i!(j-i-1)!}{j!}z^{j+1}(1-z)^{n-j},$$

即 $E(\eta_i\eta_j)=\dfrac{n!}{(i-1)!(j-i-1)!(n-j)!}\int_0^1\dfrac{i!(j-i-1)!}{j!}z^{j+1}(1-z)^{n-j}dz$

$$=\frac{n!}{(i-1)!(j-i-1)!(n-j)!}\cdot\frac{i!(j-i-1)!}{j!}B(j+2,n-j+1)$$

$$=\frac{n!}{(i-1)!(j-i-1)!(n-j)!}\cdot\frac{i!(j-i-1)!}{j!}\cdot\frac{(j+1)!(n-j)!}{(n+2)!}=\frac{i(j+1)}{(n+1)(n+2)},$$

可得 $\mathrm{Cov}(\eta_i,\eta_j)=E(\eta_i\eta_j)-E(\eta_i)E(\eta_j)=\dfrac{i(j+1)}{(n+1)(n+2)}-\dfrac{i}{n+1}\cdot\dfrac{j}{n+1}=\dfrac{i(n+1-j)}{(n+1)^2(n+2)}$,

因 $a_1=\dfrac{i}{n+1}$，$a_2=\dfrac{j}{n+1}$,

则 $\mathrm{Cov}(\eta_i,\eta_j)=\dfrac{i(n+1-j)}{(n+1)^2(n+2)}=\dfrac{a_1(1-a_2)}{n+2}$,

且 $\mathrm{Var}(\eta_i)=\dfrac{i(n+1-i)}{(n+1)^2(n+2)}=\dfrac{a_1(1-a_1)}{n+2}$，$\mathrm{Var}(\eta_j)=\dfrac{j(n+1-j)}{(n+1)^2(n+2)}=\dfrac{a_2(1-a_2)}{n+2}$,

故 η_i 和 η_j 的协方差矩阵为

$$\begin{pmatrix}\mathrm{Var}(\eta_i) & \mathrm{Cov}(\eta_i,\eta_j)\\ \mathrm{Cov}(\eta_i,\eta_j) & \mathrm{Var}(\eta_j)\end{pmatrix}=\begin{pmatrix}\dfrac{a_1(1-a_1)}{n+2} & \dfrac{a_1(1-a_2)}{n+2}\\ \dfrac{a_1(1-a_2)}{n+2} & \dfrac{a_2(1-a_2)}{n+2}\end{pmatrix}.$$

29．设总体 X 服从 $N(0, 1)$，从此总体获得一组样本观测值

$$x_1=0, x_2=0.2, x_3=0.25, x_4=-0.3, x_5=-0.1, x_6=2, x_7=0.15, x_8=1, x_9=-0.7, x_{10}=-1.$$

（1）计算 $x=0.15$（即 $x_{(6)}$）处的 $E[F(X_{(6)})]$，$\mathrm{Var}[F(X_{(6)})]$；

（2）计算 $F(X_{(6)})$在 $x=0.15$ 的分布函数值.

解：（1）根据第 28 题的结论知 $E[F(X_{(i)})]=\dfrac{i}{n+1}$，$\mathrm{Var}[F(X_{(i)})]=\dfrac{i(n+1-i)}{(n+1)^2(n+2)}$，且 $n=10$，

故 $E[F(X_{(6)})]=\dfrac{6}{11}$，$\mathrm{Var}[F(X_{(6)})]=\dfrac{6\times5}{11^2\times12}=\dfrac{5}{242}$；

（2）因 $F(X_{(i)})$服从贝塔分布 $Be(i, n-i+1)$，即这里的 $F(X_{(6)})$服从贝塔分布 $Be(6, 5)$，

则 $F(X_{(6)})$在 $x=0.15$ 的分布函数值为 $F_6(0.15)=\dfrac{10!}{5!\cdot4!}\int_0^{0.15}x^5(1-x)^4dx$,

故根据第 27 题的结论知

$$F_6(0.15)=\frac{10!}{5!\cdot4!}\int_0^{0.15}x^5(1-x)^4dx=1-\sum_{k=0}^{5}\binom{10}{k}\times0.15^k\times0.85^{10-k}=0.0014.$$

30．在下列密度函数下分别寻求容量为 n 的样本中位数 $m_{0.5}$ 的渐近分布.

（1）$p(x)=6x(1-x)$，$0<x<1$；

（2）$p(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\dfrac{(x-\mu)^2}{2\sigma^2}\right\}$；

（3）$p(x)=\begin{cases}2x, & 0<x<1;\\ 0, & \text{其他}.\end{cases}$

（4）$p(x)=\dfrac{\lambda}{2}\mathrm{e}^{-\lambda|x|}$.

解：样本中位数 $m_{0.5}$ 的渐近分布为 $N\left(x_{0.5}, \dfrac{1}{4n\cdot p^2(x_{0.5})}\right)$，其中 $p(x)$是总体密度函数，$x_{0.5}$ 是总体中位数，

（1）因 $p(x)=6x(1-x)$，$0<x<1$，有 $0.5=F(x_{0.5})=\int_0^{x_{0.5}}6x(1-x)dx=(3x^2-2x^3)\Big|_0^{x_{0.5}}=3x_{0.5}^2-2x_{0.5}^3$，

则 $x_{0.5}=0.5$，有 $\dfrac{1}{4n\cdot p^2(0.5)}=\dfrac{1}{4n\times(6\times0.5\times0.5)^2}=\dfrac{1}{9n}$，

故样本中位数 $m_{0.5}$ 的渐近分布为 $N\left(0.5,\dfrac{1}{9n}\right)$；

（2）因 $p(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\dfrac{(x-\mu)^2}{2\sigma^2}\right\}$，有 $0.5=F(x_{0.5})=F(\mu)$，

则 $x_{0.5}=\mu$，有 $\dfrac{1}{4n\cdot p^2(\mu)}=\dfrac{1}{4n\times\left(\dfrac{1}{\sqrt{2\pi}\sigma}\right)^2}=\dfrac{\pi\sigma^2}{2n}$，

故样本中位数 $m_{0.5}$ 的渐近分布为 $N\left(\mu,\dfrac{\pi\sigma^2}{2n}\right)$；

（3）因 $p(x)=\begin{cases}2x, & 0<x<1;\\ 0, & 其他.\end{cases}$ 有 $0.5=F(x_{0.5})=\int_0^{x_{0.5}}2xdx=x^2\Big|_0^{x_{0.5}}=x_{0.5}^2$，

则 $x_{0.5}=\dfrac{1}{\sqrt2}$，有 $\dfrac{1}{4n\cdot p^2\left(\dfrac{1}{\sqrt2}\right)}=\dfrac{1}{4n\times\left(2\times\dfrac{1}{\sqrt2}\right)^2}=\dfrac{1}{8n}$，

故样本中位数 $m_{0.5}$ 的渐近分布为 $N\left(\dfrac{1}{\sqrt2},\dfrac{1}{8n}\right)$；

（4）因 $p(x)=\dfrac{\lambda}{2}e^{-\lambda|x|}$，有 $0.5=F(x_{0.5})=F(0)$，

则 $x_{0.5}=0$，有 $\dfrac{1}{4n\cdot p^2(0)}=\dfrac{1}{4n\times\left(\dfrac{\lambda}{2}\right)^2}=\dfrac{1}{n\lambda^2}$，

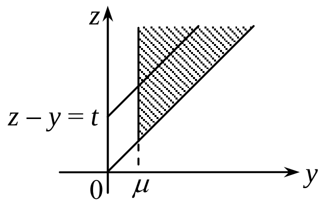
故样本中位数 $m_{0.5}$ 的渐近分布为 $N\left(0,\dfrac{1}{n\lambda^2}\right)$.

31．设总体 X 服从双参数指数分布，其分布函数为

$$F(x)=\begin{cases}1-\exp\left\{-\dfrac{x-\mu}{\sigma}\right\}, & x>\mu;\\ 0, & x\le\mu.\end{cases}$$

其中，$-\infty<\mu<+\infty$，$\sigma>0$，$X_{(1)}\le\cdots\le X_{(n)}$为样本的次序统计量．试证明 $(n-i-1)\dfrac{2}{\sigma}(X_{(i)}-X_{(i-1)})$ 服从

自由度为 2 的χ^2分布（$i=2, \cdots, n$）.



注：此题有误，讨论的随机变量应为 $(n-i+1)\dfrac{2}{\sigma}(X_{(i)}-X_{(i-1)})$.

证：因$(X_{(i-1)}, X_{(i)})$的联合密度函数为

$$p_{(i-1)i}(y,z)=\frac{n!}{(i-2)!(n-i)!}[F(y)]^{i-2}[1-F(z)]^{n-i}p(y)p(z)\mathrm{I}_{y<z}$$

$$=\frac{n!}{(i-2)!(n-i)!}\left[1-\exp\left\{-\frac{y-\mu}{\sigma}\right\}\right]^{i-2}\left[\exp\left\{-\frac{z-\mu}{\sigma}\right\}\right]^{n-i}\cdot\frac{1}{\sigma}\exp\left\{-\frac{y-\mu}{\sigma}\right\}\cdot\frac{1}{\sigma}\exp\left\{-\frac{z-\mu}{\sigma}\right\}\mathrm{I}_{\mu<y<z}$$

$$=\frac{n!}{(i-2)!(n-i)!\sigma^2}\exp\left\{-\frac{y-\mu}{\sigma}\right\}\left[1-\exp\left\{-\frac{y-\mu}{\sigma}\right\}\right]^{i-2}\left[\exp\left\{-\frac{z-\mu}{\sigma}\right\}\right]^{n-i+1}\mathrm{I}_{\mu<y<z},$$

则 $T=X_{(i)}-X_{(i-1)}$的密度函数为

$$p_T(t)=\int_{-\infty}^{+\infty}p_{(i-1)i}(y,y+t)\cdot 1\cdot dy$$

$$=\frac{n!}{(i-2)!(n-i)!\sigma^2}\int_{\mu}^{+\infty}\exp\left\{-\frac{y-\mu}{\sigma}\right\}\left[1-\exp\left\{-\frac{y-\mu}{\sigma}\right\}\right]^{i-2}\left[\exp\left\{-\frac{y+t-\mu}{\sigma}\right\}\right]^{n-i+1}dy$$

$$=\frac{n!}{(i-2)!(n-i)!\sigma^2}\left[\exp\left\{-\frac{t}{\sigma}\right\}\right]^{n-i+1}\int_{\mu}^{+\infty}\left[\exp\left\{-\frac{y-\mu}{\sigma}\right\}\right]^{n-i+1}\left[1-\exp\left\{-\frac{y-\mu}{\sigma}\right\}\right]^{i-2}(-\sigma)d\left[\exp\left\{-\frac{y-\mu}{\sigma}\right\}\right]$$

$$=\frac{n!}{(i-2)!(n-i)!\sigma^2}\left[\exp\left\{-\frac{t}{\sigma}\right\}\right]^{n-i+1}\int_{1}^{0}u^{n-i+1}(1-u)^{i-2}(-\sigma)du$$

$$=\frac{n!}{(i-2)!(n-i)!\sigma}\exp\left\{-\frac{(n-i+1)t}{\sigma}\right\}\int_{0}^{1}u^{n-i+1}(1-u)^{i-2}du$$

$$=\frac{n!}{(i-2)!(n-i)!\sigma}\exp\left\{-\frac{(n-i+1)t}{\sigma}\right\}B(n-i+2,i-1)$$

$$=\frac{n!}{(i-2)!(n-i)!\sigma}\exp\left\{-\frac{(n-i+1)t}{\sigma}\right\}\cdot\frac{(n-i+1)!(i-2)!}{n!}=\frac{n-i+1}{\sigma}\exp\left\{-\frac{(n-i+1)t}{\sigma}\right\},\quad t>0,$$

可得 $S=(n-i+1)\dfrac{2}{\sigma}(X_{(i)}-X_{(i-1)})=(n-i+1)\dfrac{2}{\sigma}T$ 的密度函数为

$$p_S(s)=p_T\left(\frac{\sigma}{2(n-i+1)}s\right)\cdot\frac{\sigma}{2(n-i+1)}=\frac{n-i+1}{\sigma}\exp\left\{-\frac{s}{2}\right\}\cdot\frac{\sigma}{2(n-i+1)}=\frac{1}{2}\exp\left\{-\frac{s}{2}\right\},\quad s>0,$$

故 $S=(n-i+1)\dfrac{2}{\sigma}(X_{(i)}-X_{(i-1)})$服从参数为 $\dfrac{1}{2}$ 的指数分布，也就是服从自由度为 2 的χ^2分布.

32．设总体 X 的密度函数为

$$p(x)=\begin{cases}3x^2, & 0<x<1;\\ 0, & \text{其他}.\end{cases}$$

$X_{(1)}\le X_{(2)}\le\cdots\le X_{(5)}$为容量为 5 的取自此总体的次序统计量，试证 $\dfrac{X_{(2)}}{X_{(4)}}$ 与 $X_{(4)}$相互独立.

证：因总体 X 的密度函数和分布函数分别为

$$p(x)=\begin{cases}3x^2, & 0<x<1;\\ 0, & 其他.\end{cases}\qquad F(x)=\begin{cases}0, & x<0;\\ x^3, & 0\le x<1;\\ 1, & x\ge 1.\end{cases}$$

则$(X_{(2)}, X_{(4)})$的联合密度函数为

$$p_{24}(x_{(2)},x_{(4)})=\frac{5!}{1!\cdot 1!\cdot 1!}[F(x_{(2)})]^1[F(x_{(4)})-F(x_{(2)})]^1[1-F(x_{(4)})]^1\,p(x_{(2)})p(x_{(4)})\,\mathrm{I}_{x_{(2)}<x_{(4)}}$$

$$=120x_{(2)}^3(x_{(4)}^3-x_{(2)}^3)(1-x_{(4)}^3)\cdot 3x_{(2)}^2\cdot 3x_{(4)}^2\,\mathrm{I}_{0<x_{(2)}<x_{(4)}<1}=1080x_{(2)}^5x_{(4)}^2(x_{(4)}^3-x_{(2)}^3)(1-x_{(4)}^3)\,\mathrm{I}_{0<x_{(2)}<x_{(4)}<1},$$

设 $Y_1=\dfrac{X_{(2)}}{X_{(4)}}$， $Y_2=X_{(4)}$， 有 $X_{(2)}=Y_1Y_2$， $X_{(4)}=Y_2$，

则$(X_{(2)}, X_{(4)})$关于(Y_1, Y_2)的雅可比行列式为

$$J=\frac{\partial(x_{(2)},x_{(4)})}{\partial(y_1,y_2)}=\begin{vmatrix}y_2 & y_1\\ 0 & 1\end{vmatrix}=y_2,$$

且 $0<X_{(2)}\le X_{(4)}<1$ 对应于 $0<Y_1<1, 0<Y_2<1$，
可得(Y_1, Y_2)的联合密度函数为

$$p(y_1,y_2)=p_{24}(y_1y_2,y_2)\cdot|J|=1080(y_1y_2)^5y_2^2[y_2^3-(y_1y_2)^3](1-y_2^3)\,\mathrm{I}_{0<y_1<1,\,0<y_2<1}\cdot y_2$$

$$=1080y_1^5(1-y_1^3)\,\mathrm{I}_{0<y_1<1}\cdot y_2^{11}(1-y_2^3)\,\mathrm{I}_{0<y_2<1},$$

由于$(Y_1, Y_2, \cdots, Y_n)$的联合密度函数 $p(y_1, y_2)$可分离变量，

故 $Y_1=\dfrac{X_{(2)}}{X_{(4)}}$ 与 $Y_2=X_{(4)}$相互独立.

33．（1）设 $X_{(1)}$和 $X_{(n)}$分别为容量 n 的最小和最大次序统计量，证明极差 $R_n=X_{(n)}-X_{(1)}$的分布函数

$$F_{R_n}(x)=n\int_{-\infty}^{+\infty}[F(y+x)-F(y)]^{n-1}p(y)dy$$

其中 $F(y)$与 $p(y)$分别为总体的分布函数与密度函数；

（2）利用（1）的结论，求总体为指数分布 $Exp(\lambda)$时，样本极差 R_n 的分布.

注：第（1）问应添上 $x>0$ 的要求.

解：（1）方法一：增补变量法

因$(X_{(1)}, X_{(n)})$的联合密度函数为

$$p_{1n}(y,z)=\frac{n!}{(n-2)!}[F(z)-F(y)]^{n-2}p(y)p(z)\,\mathrm{I}_{y<z}=n(n-1)[F(z)-F(y)]^{n-2}p(y)p(z)\,\mathrm{I}_{y<z},$$

对于其函数 $R_n=X_{(n)}-X_{(1)}$， 增补变量 $W=X_{(1)}$，

$$\begin{cases}w=y;\\ r=z-y.\end{cases}\quad 反函数为\begin{cases}y=w;\\ z=w+r.\end{cases}$$

其雅可比行列式为

$$J=\begin{vmatrix}1 & 0\\ 1 & 1\end{vmatrix}=1,$$

则 R_n 的密度函数为

$$p_{R_n}(r)=\int_{-\infty}^{+\infty}n(n-1)[F(w+r)-F(w)]^{n-2}\,p(w)p(w+r)\,\mathrm{I}_{r>0}\,dw\,,$$

故 $R_n = X_{(n)} - X_{(1)}$的分布函数为

$$F_{R_n}(x)=\int_{-\infty}^{x}p_{R_n}(r)dr=\int_{-\infty}^{x}dr\int_{-\infty}^{+\infty}n(n-1)[F(w+r)-F(w)]^{n-2}\,p(w)p(w+r)\,\mathrm{I}_{r>0}\,dw$$

$$=\int_{-\infty}^{+\infty}dw\int_{-\infty}^{x}n(n-1)[F(w+r)-F(w)]^{n-2}\,p(w)p(w+r)\,\mathrm{I}_{r>0}\,dr$$

$$=\int_{-\infty}^{+\infty}n(n-1)\,p(w)dw\int_{0}^{x}[F(w+r)-F(w)]^{n-2}\,p(w+r)dr$$

$$=\int_{-\infty}^{+\infty}n(n-1)\,p(w)dw\int_{0}^{x}[F(w+r)-F(w)]^{n-2}\,dF(w+r)$$

$$=\int_{-\infty}^{+\infty}n(n-1)\,p(w)dw\cdot\frac{1}{n-1}[F(w+r)-F(w)]^{n-1}\Big|_{0}^{x}$$

$$=n\int_{-\infty}^{+\infty}[F(w+x)-F(w)]^{n-1}\,p(w)dw$$

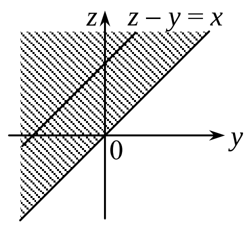
$$=n\int_{-\infty}^{+\infty}[F(y+x)-F(y)]^{n-1}\,p(y)dy\,,\ \ x>0;$$

方法二：分布函数法

因$(X_{(1)}, X_{(n)})$的联合密度函数为

$$p_{1n}(y,z)=\frac{n!}{(n-2)!}[F(z)-F(y)]^{n-2}\,p(y)p(z)\,\mathrm{I}_{y<z}=n(n-1)[F(z)-F(y)]^{n-2}\,p(y)p(z)\,\mathrm{I}_{y<z}\,,$$

故 $R_n = X_{(n)} - X_{(1)}$的分布函数为



$$F_{R_n}(x)=P\{R_n=X_{(n)}-X_{(1)}\le x\}=\int_{-\infty}^{+\infty}dy\int_{-\infty}^{y+x}p_{1n}(y,z)dz$$

$$=n(n-1)\int_{-\infty}^{+\infty}dy\int_{y}^{y+x}[F(z)-F(y)]^{n-2}\,p(y)p(z)dz$$

$$=n(n-1)\int_{-\infty}^{+\infty}dy\cdot p(y)\int_{y}^{y+x}[F(z)-F(y)]^{n-2}\,d[F(z)]$$

$$=n(n-1)\int_{-\infty}^{+\infty}dy\cdot p(y)\cdot\frac{1}{n-1}[F(z)-F(y)]^{n-1}\Big|_{y}^{y+x}=n\int_{-\infty}^{+\infty}[F(y+x)-F(y)]^{n-1}\,p(y)dy\,,\ \ x>0;$$

（2）因指数分布 $Exp(\lambda)$的密度函数与分布函数分别为

$$p(x)=\begin{cases}\lambda\,\mathrm{e}^{-\lambda x}, & x>0;\\ 0, & x\le 0.\end{cases}\qquad F(x)=\begin{cases}1-\mathrm{e}^{-\lambda x}, & x>0;\\ 0, & x\le 0.\end{cases}$$

故 $R_n = X_{(n)} - X_{(1)}$的分布函数为

$$F_{R_n}(x)=n\int_{-\infty}^{+\infty}[F(y+x)-F(y)]^{n-1}\,p(y)dy=n\int_{0}^{+\infty}[(1-\mathrm{e}^{-\lambda(y+x)})-(1-\mathrm{e}^{-\lambda y})]^{n-1}\cdot\lambda\,\mathrm{e}^{-\lambda y}\,dy$$

$$=n\int_{0}^{+\infty}(\mathrm{e}^{-\lambda y})^{n-1}(1-\mathrm{e}^{-\lambda x})^{n-1}\cdot(-1)d\,\mathrm{e}^{-\lambda y}=n(1-\mathrm{e}^{-\lambda x})^{n-1}\cdot\left(-\frac{1}{n}\right)(\mathrm{e}^{-\lambda y})^{n}\Big|_{0}^{+\infty}=(1-\mathrm{e}^{-\lambda x})^{n-1}\,,\ \ x>0.$$

34．设 $X_1, \cdots, X_n$ 是来自 $U(0, \theta)$ 的样本，$X_{(1)} \le \cdots \le X_{(n)}$ 为次序统计量，令

$$Y_i=\frac{X_{(i)}}{X_{(i+1)}}\,,\ \ i=1,\cdots,n-1\,,\ \ Y_n=X_{(n)}\,,$$

证明 $Y_1, \cdots, Y_n$ 相互独立.

解：总体密度函数 $p(x)=\frac{1}{\theta}I_{0<x<\theta}$，

且 $(X_{(1)},X_{(2)},\cdots,X_{(n)})$ 联合密度函数为 $p(x_{(1)},x_{(2)},\cdots,x_{(n)})=n!\cdot\frac{1}{\theta^n}I_{0<x_{(1)}\le x_{(2)}\le\cdots\le x_{(n)}<\theta}$，

由于 $Y_i=\frac{X_{(i)}}{X_{(i+1)}}$，$i=1,2,\ldots,n-1$，$Y_n=X_{(n)}$，

有 $X_{(1)}=Y_1Y_2\cdots Y_n$，$X_{(2)}=Y_2\cdots Y_n$，$\cdots$，$X_{(n-1)}=Y_{n-1}Y_n$，$X_{(n)}=Y_n$，

则 $(X_{(1)},X_{(2)},\cdots,X_{(n)})$ 关于 $(Y_1,Y_2,\cdots,Y_n)$ 的雅可比行列式为

$$\frac{\partial(x_{(1)},x_{(2)},\cdots,x_{(n)})}{\partial(y_1,y_2,\cdots,y_n)}=\begin{vmatrix} y_2\cdots y_n & y_1y_3\cdots y_n & \cdots & y_1y_2\cdots y_{n-1} \\ 0 & y_3\cdots y_n & \cdots & y_2y_3\cdots y_{n-1} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{vmatrix}=y_2y_3^2\cdots y_n^{n-1},$$

且 $0<X_{(1)}\le X_{(2)}\le\cdots\le X_{(n)}<\theta$ 对应于 $0<Y_1\le 1, 0<Y_2\le 1,\cdots,0<Y_{n-1}\le 1, 0<Y_n<\theta$，

可得 $(Y_1,Y_2,\cdots,Y_n)$ 的联合密度函数为

$$p(y_1,y_2,\cdots,y_n)=n!\cdot\frac{1}{\theta^n}y_2y_3^2\cdots y_n^{n-1}I_{0<y_1\le 1}I_{0<y_2\le 1}\cdots I_{0<y_{n-1}\le 1}I_{0<y_n<\theta},$$

由于 $(Y_1,Y_2,\cdots,Y_n)$ 的联合密度函数 $p(y_1,y_2,\cdots,y_n)$可分离变量，

故 $Y_1,Y_2,\cdots,Y_n$ 相互独立.

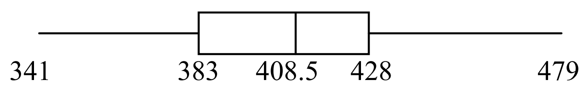
35．对下列数据构造箱线图

472	425	447	377	341	369	412	419
400	382	366	425	399	398	423	384
418	392	372	418	374	385	439	428
429	428	430	413	405	381	403	479
381	443	441	433	419	379	386	387

解：$x_{(1)}=341$，$m_{0.25}=\frac{1}{2}(x_{(10)}+x_{(11)})=383$，$m_{0.5}=\frac{1}{2}(x_{(20)}+x_{(21)})=408.5$，$m_{0.75}=\frac{1}{2}(x_{(30)}+x_{(31)})=428$，

$x_{(n)}=479$，

箱线图



36．根据调查，某集团公司的中层管理人员的年薪数据如下（单位：千元）

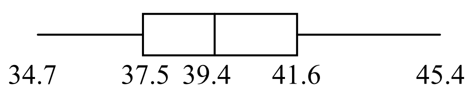
40.6	39.6	43.8	36.2	40.8	37.3	39.2	42.9
38.6	39.6	40.0	34.7	41.7	45.4	36.9	37.8
44.9	45.4	37.0	35.1	36.7	41.3	38.1	37.9
37.1	37.7	39.2	36.9	44.5	40.4	38.4	38.9
39.9	42.2	43.5	44.8	37.7	34.7	36.3	39.7
42.1	41.5	40.6	38.9	42.2	40.3	35.8	39.2

试画出箱线图.

解：$x_{(1)}=34.7$，$m_{0.25}=\frac{1}{2}(x_{(12)}+x_{(13)})=37.5$，$m_{0.5}=\frac{1}{2}(x_{(24)}+x_{(25)})=39.4$，$m_{0.75}=\frac{1}{2}(x_{(36)}+x_{(37)})=41.6$，

$x_{(n)}=45.4$，

箱线图



习题 5.4

1． 在总体 $N(7.6, 4)$ 中抽取容量为 n 的样本，如果要求样本均值落在 (5.6，9.6) 内的概率不小于 0.95，则

n 至少为多少？

解：因总体 $X \sim N(7.6, 4)$，有 $\overline{X} \sim N(7.6, \frac{4}{n})$，$\frac{\overline{X} - 7.6}{2/\sqrt{n}} \sim N(0,1)$，

则 $P\{5.6 < \overline{X} < 9.6\} = P\{-\sqrt{n} < \frac{\overline{X} - 7.6}{2/\sqrt{n}} < \sqrt{n}\} = \Phi(\sqrt{n}) - \Phi(-\sqrt{n}) = 2\Phi(\sqrt{n}) - 1 \ge 0.95$，

即 $\Phi(\sqrt{n}) \ge 0.975$，$\sqrt{n} \ge 1.96$，$n \ge 3.8416$，

故取 $n \ge 4$.

2． 设 $x_1, \cdots, x_n$ 是来自 $N(\mu, 16)$ 的样本，问 n 多大时才能使得 $P\{|\overline{X} - \mu| < 1\} \ge 0.95$ 成立？

解：因总体 $X \sim N(\mu, 16)$，有 $\overline{X} \sim N\left(\mu, \frac{16}{n}\right)$，$\frac{\overline{X} - \mu}{4/\sqrt{n}} \sim N(0,1)$，

则 $P\{|\overline{X} - \mu| < 1\} = P\{\left|\frac{\overline{X} - \mu}{4/\sqrt{n}}\right| < \frac{\sqrt{n}}{4}\} = \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 \ge 0.95$，

即 $\Phi\left(\frac{\sqrt{n}}{4}\right) \ge 0.975$，$\frac{\sqrt{n}}{4} \ge 1.96$，$n \ge 61.4656$，

故取 $n \ge 62$.

3． 由正态总体 $N(100, 4)$ 抽取二个独立样本，样本均值分别为 $\overline{x}$，$\overline{y}$，样本容量分别为 15, 20，试求 $P\{|\overline{x} - \overline{y}| > 0.2\}$.

解：因 $\overline{X} \sim N(100, \frac{4}{15})$，$\overline{Y} \sim N(100, \frac{4}{20})$，即 $\overline{X} - \overline{Y} \sim N(0, \frac{4}{15} + \frac{4}{20})$，$\frac{\overline{X} - \overline{Y}}{\sqrt{\frac{4}{15} + \frac{4}{20}}} \sim N(0,1)$，

故 $P\{|\overline{X} - \overline{Y}| > 0.2\} = P\{\frac{|\overline{X} - \overline{Y}|}{\sqrt{\frac{4}{15} + \frac{4}{20}}} > \frac{0.2}{\sqrt{\frac{4}{15} + \frac{4}{20}}} = 0.29\} = 2[1 - \Phi(0.29)] = 2 - 2 \times 0.6141 = 0.7718$.

4． 由正态总体 $N(\mu, \sigma^2)$ 抽取容量为 20 的样本，试求 $P\{10\sigma^2 < \sum_{i=1}^{20} (X_i - \mu)^2 < 30\sigma^2\}$.

解：因 $\frac{\sum_{i=1}^{20} (X_i - \mu)^2}{\sigma^2} \sim \chi^2(20)$，

故 $P\{10\sigma^2 < \sum_{i=1}^{20} (X_i - \mu)^2 < 30\sigma^2\} = P\{10 < \frac{\sum_{i=1}^{20} (X_i - \mu)^2}{\sigma^2} < 30\} = \int_{10}^{30} p_{\chi^2(20)}(x) \mathrm{d}x = 0.8983$.

注：最后一步的积分利用 MATLAB 计算，命令窗口输入：chi2cdf(30,20)- chi2cdf(10,20)
这里 chi2cdf(x, n) 表示自由度为 n 的 χ^2 分布在点 x 处的分布函数值.

5． 设 $x_1, \cdots, x_{16}$ 是来自 $N(\mu, \sigma^2)$ 的样本，经计算 $\overline{x} = 9$，$s^2 = 5.32$，试求 $P\{|\overline{X} - \mu| < 0.6\}$.

解：因 $\dfrac{\bar{X}-\mu}{s/\sqrt{n}}=\dfrac{\bar{X}-\mu}{\sqrt{5.32}/\sqrt{16}}\sim t(15)$，

故 $P\{|\bar{X}-\mu|<0.6\}=P\{\dfrac{|\bar{X}-\mu|}{\sqrt{5.32}/\sqrt{16}}<\dfrac{0.6}{\sqrt{5.32}/\sqrt{16}}=1.0405\}=\int_{-1.0405}^{1.0405}p_{t(15)}(x)\mathrm{d}x=0.6854$.

注：最后一步的积分利用 MATLAB 计算，命令窗口输入：2*tcdf(1.0405,15)-1
这里 tcdf(x, n) 表示自由度为 n 的 t 分布在点 x 处的分布函数值.

6. 设 $x_1, \cdots, x_n$ 是来自 $N(\mu, 1)$ 的样本，试确定最小的常数 c，使得对任意的$\mu\geq 0$，有 $P\{|\bar{X}|<c\}\leq\alpha$.

解：因 $\bar{X}\sim N(\mu,\dfrac{1}{n})$，$\dfrac{\bar{X}-\mu}{1/\sqrt{n}}\sim N(0,1)$，

则 $P\{|\bar{X}|<c\}=P\{\sqrt{n}(-c-\mu)<\dfrac{|\bar{X}-\mu|}{1/\sqrt{n}}<\sqrt{n}(c-\mu)\}=\Phi(\sqrt{n}(c-\mu))-\Phi(\sqrt{n}(-c-\mu))\leq\alpha$，

设 $f(\mu)=\Phi(\sqrt{n}(c-\mu))-\Phi(\sqrt{n}(-c-\mu))$，

令 $f'(\mu)=-\sqrt{n}\varphi(\sqrt{n}(c-\mu))+\sqrt{n}\varphi(\sqrt{n}(-c-\mu))=0$，其中$\varphi(x)$ 是标准正态分布的密度函数，

得 $\varphi(\sqrt{n}(c-\mu))=\varphi(\sqrt{n}(-c-\mu))$，由$\varphi(x)$ 的对称性得 $\sqrt{n}(c-\mu)=\sqrt{n}(c+\mu)$，即$\mu=0$，

因 $f''(\mu)=n\varphi'(\sqrt{n}(c-\mu))-n\varphi'(\sqrt{n}(-c-\mu))$，且当 $x<0$ 时，$\varphi'(x)>0$，当 $x>0$ 时，$\varphi'(x)<0$，

则 $f''(0)=n\varphi'(\sqrt{n}c)-n\varphi'(-\sqrt{n}c)<0$，即$\mu=0$ 时，$f(\mu)$ 达到最大值，

当$\mu=0$ 时，$f(0)=\Phi(\sqrt{n}c)-\Phi(-\sqrt{n}c)=2\Phi(\sqrt{n}c)-1\leq\alpha$，即 $\Phi(\sqrt{n}c)\leq\dfrac{1+\alpha}{2}$，$\sqrt{n}c\leq u_{\frac{1+\alpha}{2}}$，

故取 $c=\dfrac{u_{\frac{1+\alpha}{2}}}{\sqrt{n}}$.

7. 设随机变量 $X\sim F(n, n)$，证明 $P\{X<1\}=0.5$.

证：因 $X\sim F(n, n)$，有 $\dfrac{1}{X}\sim F(n,n)$，且 $X>0$，

则 $P\{X<1\}=P\{\dfrac{1}{X}<1\}=P\{X>1\}$，且显然 $P\{X<1\}+P\{X>1\}=1$，

故 $P\{X<1\}=0.5$.

8. 设 $X\sim F(n, m)$，证明 $Z=\dfrac{n}{m}X\Big/\left(1+\dfrac{n}{m}X\right)$服从贝塔分布，并指出其参数.

证：因 $X\sim F(n, m)$，密度函数 $p_F(x)=\dfrac{\Gamma\left(\dfrac{n+m}{2}\right)\left(\dfrac{n}{m}\right)^{\frac{n}{2}}}{\Gamma\left(\dfrac{n}{2}\right)\Gamma\left(\dfrac{m}{2}\right)}x^{\frac{n}{2}-1}\left(1+\dfrac{n}{m}x\right)^{-\frac{n+m}{2}}$，$x>0$，

而 $z=\dfrac{n}{m}x\Big/\left(1+\dfrac{n}{m}x\right)$ 在 $x>0$ 时严格单调增加，反函数为 $x=\dfrac{m}{n}\cdot\dfrac{z}{1-z}$，其导数 $\dfrac{\mathrm{d}x}{\mathrm{d}z}=\dfrac{m}{n}\cdot\dfrac{1}{(1-z)^2}$，

则 Z 的密度函数为

$$p_Z(z)=\frac{\Gamma\left(\frac{n+m}{2}\right)\left(\frac{n}{m}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}\left(\frac{m}{n}\cdot\frac{z}{1-z}\right)^{\frac{n}{2}-1}\left(1+\frac{z}{1-z}\right)^{-\frac{n+m}{2}}\cdot\frac{m}{n}\cdot\frac{1}{(1-z)^2}$$

$$=\frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}\left(\frac{z}{1-z}\right)^{\frac{n}{2}-1}\left(\frac{1}{1-z}\right)^{-\frac{n+m}{2}}\cdot\frac{1}{(1-z)^2}=\frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}z^{\frac{n}{2}-1}(1-z)^{\frac{m}{2}-1},$$

故 Z 服从参数为 $\left(\dfrac{n}{2},\dfrac{m}{2}\right)$ 的 β 分布.

注：分布 $\beta(p,q)$ 的密度函数为 $p_\beta(x)=\dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}x^{p-1}(1-x)^{q-1}$.

9．设是来自 $N(0,\sigma^2)$ 的样本，试求 $Y=\left(\dfrac{x_1+x_2}{x_1-x_2}\right)^2$ 的分布.

解：因 $X_1\sim N(0,\sigma^2)$，$X_2\sim N(0,\sigma^2)$，有 $X_1+X_2\sim N(0,2\sigma^2)$，$X_1-X_2\sim N(0,2\sigma^2)$，

则 $\dfrac{X_1+X_2}{\sqrt{2}\sigma}\sim N(0,1)$，$\dfrac{X_1-X_2}{\sqrt{2}\sigma}\sim N(0,1)$，即 $\dfrac{(X_1+X_2)^2}{2\sigma^2}\sim\chi^2(1)$，$\dfrac{(X_1-X_2)^2}{2\sigma^2}\sim\chi^2(1)$，

因 (X_1,X_2) 服从二维正态分布，知 (X_1+X_2,X_1-X_2) 也服从二维正态分布，

且 $\mathrm{Cov}(X_1+X_2,X_1-X_2)=\mathrm{Cov}(X_1,X_1)-\mathrm{Cov}(X_2,X_2)=\mathrm{Var}(X_1)-\mathrm{Var}(X_2)=\sigma^2-\sigma^2=0$，

则 X_1+X_2 与 X_1-X_2 相互独立，有 $\dfrac{(X_1+X_2)^2}{2\sigma^2}$ 与 $\dfrac{(X_1-X_2)^2}{2\sigma^2}$ 相互独立，

故由 F 分布定义知 $Y=\left(\dfrac{X_1+X_2}{X_1-X_2}\right)^2=\dfrac{(X_1+X_2)^2}{2\sigma^2}\Big/\dfrac{(X_1-X_2)^2}{2\sigma^2}\sim F(1,1)$.

注：F 分布结构为 $F=\dfrac{X/n}{Y/m}\sim F(n,m)$，其中 $X\sim\chi^2(n)$，$Y\sim\chi^2(m)$，且 X 与 Y 相互独立.

10．设总体为 $N(0,1)$，x_1,x_2 为样本，试求常数 k，使得

$$P\left\{\frac{(X_1+X_2)^2}{(X_1-X_2)^2+(X_1+X_2)^2}>k\right\}=0.05.$$

解：因 $X_1\sim N(0,1)$，$X_2\sim N(0,1)$，有 $\dfrac{(X_1+X_2)^2}{2}\Big/\dfrac{(X_1-X_2)^2}{2}=\dfrac{(X_1+X_2)^2}{(X_1-X_2)^2}\sim F(1,1)$，

则 $P\left\{\dfrac{(X_1+X_2)^2}{(X_1-X_2)^2+(X_1+X_2)^2}>k\right\}=P\left\{\dfrac{(X_1-X_2)^2}{(X_1+X_2)^2}+1<\dfrac{1}{k}\right\}=P\left\{\dfrac{(X_1-X_2)^2}{(X_1+X_2)^2}<\dfrac{1}{k}-1\right\}=0.05$，

得 $P\left\{\dfrac{(X_1-X_2)^2}{(X_1+X_2)^2}\geq\dfrac{1-k}{k}\right\}=0.95$，即 $P\left\{\dfrac{(X_1+X_2)^2}{(X_1-X_2)^2}\leq\dfrac{k}{1-k}\right\}=0.95$，

故 $\dfrac{k}{1-k}=F_{0.95}(1,1)$，$k=\dfrac{F_{0.95}(1,1)}{1+F_{0.95}(1,1)}=\dfrac{161.45}{1+161.45}=0.9938$.

注：此题 $\dfrac{(X_1+X_2)^2}{2}\sim\chi^2(1)$，$\dfrac{(X_1+X_2)^2+(X_1-X_2)^2}{2}\sim\chi^2(2)$，

但 $\dfrac{\dfrac{(X_1+X_2)^2}{2}}{\dfrac{(X_1+X_2)^2+(X_1-X_2)^2}{2}\Big/2}=\dfrac{2(X_1+X_2)^2}{(X_1+X_2)^2+(X_1-X_2)^2}$ 并不服从 $F(1, 2)$，因为二者不独立.

11．设 $x_1, \cdots, x_n$ 是来自 $N(\mu_1, \sigma^2)$的样本，$y_1, \cdots, y_m$ 是来自 $N(\mu_2, \sigma^2)$的样本，c, d 是任意两个不为 0 的常数，证明 $t=\dfrac{c(\bar{x}-\mu_1)+d(\bar{y}-\mu_2)}{s_w\sqrt{\dfrac{c^2}{n}+\dfrac{d^2}{m}}}\sim t(n+m-2)$，其中 $s_w^2=\dfrac{(n-1)S_x^2+(m-1)S_y^2}{n+m-2}$.

解：因 $\overline{X}\sim N(\mu_1,\dfrac{\sigma^2}{n})$，$\overline{Y}\sim N(\mu_2,\dfrac{\sigma^2}{m})$，有 $c(\overline{X}-\mu_1)+d(\overline{Y}-\mu_2)\sim N(0,\dfrac{c^2\sigma^2}{n}+\dfrac{d^2\sigma^2}{m})$，

则 $\dfrac{c(\overline{X}-\mu_1)+d(\overline{Y}-\mu_2)}{\sigma\sqrt{\dfrac{c^2}{n}+\dfrac{d^2}{m}}}\sim N(0,1)$，

又因 $\dfrac{(n-1)S_x^2}{\sigma^2}=\dfrac{\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}\sim\chi^2(n-1)$，$\dfrac{(m-1)S_y^2}{\sigma^2}=\dfrac{\sum\limits_{j=1}^{m}(Y_j-\overline{Y})^2}{\sigma^2}\sim\chi^2(m-1)$，且相互独立，

则 $\dfrac{(n-1)S_x^2+(m-1)S_y^2}{\sigma^2}\sim\chi^2(n+m-2)$，且与 $c(\overline{X}-\mu_1)+d(\overline{Y}-\mu_2)$ 相互独立，

故由 t 分布定义知

$$\frac{\dfrac{c(\overline{X}-\mu_1)+d(\overline{Y}-\mu_2)}{\sigma\sqrt{\dfrac{c^2}{n}+\dfrac{d^2}{m}}}}{\sqrt{\dfrac{(n-1)S_x^2+(m-1)S_y^2}{\sigma^2}\Big/(n+m-2)}}=\frac{c(\overline{X}-\mu_1)+d(\overline{Y}-\mu_2)}{\sqrt{\dfrac{(n-1)S_x^2+(m-1)S_y^2}{n+m-2}}\cdot\sqrt{\dfrac{c^2}{n}+\dfrac{d^2}{m}}}\sim t(n+m-2),$$

注：t 分布结构为 $T=\dfrac{X}{\sqrt{Y/n}}\sim t(n)$，其中 $X\sim N(0, 1)$，$Y\sim\chi^2(n)$，且 X 与 Y 相互独立.

12．设 $x_1, \cdots, x_n, x_{n+1}$ 是来自 $N(\mu, \sigma^2)$的样本，$\bar{x}=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i$，$s_n^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i-\bar{x}_n)^2$，试求常数 c，使得 $t_c=c\dfrac{x_{n+1}-\bar{x}_n}{s_n}$ 服从 t 分布，并指出分布的自由度.

解：因 $\overline{X}_n \sim N(\mu, \dfrac{\sigma^2}{n})$，$X_{n+1} \sim N(\mu, \sigma^2)$，有 $X_{n+1} - \overline{X}_n \sim N(0, \sigma^2 + \dfrac{\sigma^2}{n})$，

即 $\dfrac{X_{n+1} - \overline{X}_n}{\sigma\sqrt{\dfrac{n+1}{n}}} \sim N(0,1)$，又因 $\dfrac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$，且与 $X_{n+1} - \overline{X}_n$ 相互独立，

则由 t 分布定义知 $\dfrac{\dfrac{X_{n+1} - \overline{X}_n}{\sigma\sqrt{\dfrac{n+1}{n}}}}{\sqrt{\dfrac{(n-1)S_n^2}{\sigma^2}\Big/(n-1)}} = \sqrt{\dfrac{n}{n+1}}\dfrac{X_{n+1} - \overline{X}_n}{S_n} \sim t(n-1)$，

故当 $c = \sqrt{\dfrac{n}{n+1}}$ 时，$c\dfrac{X_{n+1} - \overline{X}_n}{S_n}$ 服从自由度为 $n-1$ 的 t 分布．

13．设从两个方差相等的正态总体中分别抽取容量为 15, 20 的样本，其样本方差分别为 s_1^2，s_2^2，试求 $P\{S_1^2 / S_2^2 > 2\}$．

解：因 $\dfrac{(n_1-1)S_1^2}{\sigma^2} = \dfrac{14S_1^2}{\sigma^2} \sim \chi^2(14)$，$\dfrac{(n_2-1)S_2^2}{\sigma^2} = \dfrac{19S_2^2}{\sigma^2} \sim \chi^2(19)$，且相互独立，

则由 F 分布定义知 $\dfrac{\dfrac{14S_1^2}{\sigma^2}\Big/14}{\dfrac{19S_2^2}{\sigma^2}\Big/19} = \dfrac{S_1^2}{S_2^2} \sim F(14,19)$，

故 $P\{S_1^2 / S_2^2 > 2\} = \int_2^{+\infty} p_{F(14,19)}(x)\mathrm{d}x = 1 - \int_0^2 p_{F(14,19)}(x)\mathrm{d}x = 0.0798$．

注：最后一步的积分利用 MATLAB 计算，命令窗口输入：1-fcdf(2,14,19)
这里 fcdf(x, n, m) 表示自由度为 n, m 的 F 分布在点 x 处的分布函数值．

14．设 $X_1, X_2, \cdots, X_{15}$ 是总体 $N(0, \sigma^2)$的一个样本，求

$$Y = \frac{X_1^2 + X_2^2 + \cdots + X_{10}^2}{2(X_{11}^2 + X_{12}^2 + \cdots + X_{15}^2)}$$

的分布．

解：因 $X_1, X_2, \cdots, X_{15}$ 相互独立，且 $X_i \sim N(0, \sigma^2)$，有 $\dfrac{X_i}{\sigma} \sim N(0,1)$，$i = 1, 2, \cdots, 15$，

则由χ^2 分布的构成可知 $\dfrac{X_1^2 + X_2^2 + \cdots + X_{10}^2}{\sigma^2} \sim \chi^2(10)$，$\dfrac{X_{11}^2 + X_{12}^2 + \cdots + X_{15}^2}{\sigma^2} \sim \chi^2(5)$，且相互独立，

故由 F 分布的构成可知 $Y = \dfrac{X_1^2 + X_2^2 + \cdots + X_{10}^2}{2(X_{11}^2 + X_{12}^2 + \cdots + X_{15}^2)} = \dfrac{\dfrac{X_1^2 + X_2^2 + \cdots + X_{10}^2}{\sigma^2}\Big/10}{\dfrac{X_{11}^2 + X_{12}^2 + \cdots + X_{15}^2}{\sigma^2}\Big/5} \sim F(10,5)$．

15．设$(X_1, X_2, \cdots, X_{17})$是来自正态分布 $N(\mu, \sigma^2)$的一个样本，$\overline{X}$ 与 S^2 分别是样本均值与样本方差．求 k，使得 $P\{\overline{X} > \mu + kS\} = 0.95$．

解：因$(X_1, X_2, \cdots, X_{17})$是来自正态分布$N(\mu, \sigma^2)$的一个样本，$n = 17$，有$\dfrac{\overline{X}-\mu}{S/\sqrt{17}} \sim t(16)$，

则$P\{\overline{X} > \mu + kS\} = P\left\{\dfrac{\overline{X}-\mu}{S/\sqrt{17}} > \sqrt{17}k\right\} = 0.95$，即$\sqrt{17}k = -t_{0.95}(16) = -1.7459$，

故$k = -0.4234$.

16．设总体X服从$N(\mu, \sigma^2)$，$\sigma^2 > 0$，从该总体中抽取简单随机样本$X_1, X_2, \cdots, X_{2n}$ $(n \ge 1)$，其样本均值$\overline{X} = \dfrac{1}{2n}\sum\limits_{i=1}^{2n} X_i$，求统计量$Y = \sum\limits_{i=1}^{n}(X_i + X_{n+i} - 2\overline{X})^2$的数学期望.

解：因$E(X_i) = \mu$，$\mathrm{Var}(X_i) = \sigma^2$，$E(\overline{X}) = \dfrac{1}{2n}\sum\limits_{i=1}^{2n} E(X_i) = \mu$，$\mathrm{Var}(\overline{X}) = \dfrac{1}{4n^2}\sum\limits_{i=1}^{2n}\mathrm{Var}(X_i) = \dfrac{\sigma^2}{2n}$，

且$Y = \sum\limits_{i=1}^{n}[(X_i^2 + X_{n+i}^2 + 2X_iX_{n+i}) - 4\overline{X}(X_i + X_{n+i}) + 4\overline{X}^2]$

$$= \sum_{i=1}^{n}(X_i^2 + X_{n+i}^2) + 2\sum_{i=1}^{n}X_iX_{n+i} - 4\overline{X}\sum_{i=1}^{n}(X_i + X_{n+i}) + 4n\overline{X}^2$$

$$= \sum_{i=1}^{2n}X_i^2 + 2\sum_{i=1}^{n}X_iX_{n+i} - 4\overline{X}\cdot 2n\overline{X} + 4n\overline{X}^2 = \sum_{i=1}^{2n}X_i^2 + 2\sum_{i=1}^{n}X_iX_{n+i} - 4n\overline{X}^2,$$

故$E(Y) = \sum\limits_{i=1}^{2n}E(X_i^2) + 2\sum\limits_{i=1}^{n}E(X_iX_{n+i}) - 4nE(\overline{X}^2)$

$$= \sum_{i=1}^{2n}[\mathrm{Var}(X_i) + E(X_i)^2] + 2\sum_{i=1}^{n}E(X_i)E(X_{n+i}) - 4n[\mathrm{Var}(\overline{X}) + E(\overline{X})^2]$$

$$= 2n(\sigma^2 + \mu^2) + 2n\mu^2 - 4n\left(\frac{\sigma^2}{2n} + \mu^2\right) = 2(n-1)\sigma^2.$$

17．证明：若随机变量$T \sim t(k)$，则对$r < k$有

$$E(T^r) = \begin{cases} 0, & r\text{ 为奇数;} \\ \dfrac{k^{\frac{r}{2}}\Gamma\left(\dfrac{r+1}{2}\right)\Gamma\left(\dfrac{k-r}{2}\right)}{\sqrt{\pi}\Gamma\left(\dfrac{k}{2}\right)}, & r\text{ 为偶数.} \end{cases}$$

并由此写出$E(T)$，$\mathrm{Var}(T)$.

证：因$T \sim t(k)$，有T的密度函数为

$$p(x) = \frac{\Gamma\left(\dfrac{k+1}{2}\right)}{\sqrt{k\pi}\Gamma\left(\dfrac{k}{2}\right)}\left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}},\quad -\infty < x < +\infty,$$

则 $E(T^r)=\dfrac{\Gamma\left(\dfrac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\dfrac{k}{2}\right)}\int_{-\infty}^{+\infty}x^r\left(1+\dfrac{x^2}{k}\right)^{-\frac{k+1}{2}}dx$，

因当 $x\to\infty$时，$x^r\left(1+\dfrac{x^2}{k}\right)^{-\frac{k+1}{2}}\sim x^r\cdot\left(\dfrac{x^2}{k}\right)^{-\frac{k+1}{2}}=k^{\frac{k+1}{2}}x^r\cdot x^{-(k+1)}=\dfrac{k^{\frac{k+1}{2}}}{x^{k-r+1}}$，

则对 $r<k$，有反常积分 $\int_{-\infty}^{+\infty}x^r\left(1+\dfrac{x^2}{k}\right)^{-\frac{k+1}{2}}dx$ 收敛，即 $E(T^r)$存在，

当 r 为奇数时，$x^r\left(1+\dfrac{x^2}{k}\right)^{-\frac{k+1}{2}}$ 为奇函数，有 $\int_{-\infty}^{+\infty}x^r\left(1+\dfrac{x^2}{k}\right)^{-\frac{k+1}{2}}dx=0$，即 $E(T^r)=0$，

当 r 为偶数时，$x^r\left(1+\dfrac{x^2}{k}\right)^{-\frac{k+1}{2}}$ 为偶函数，有 $\int_{-\infty}^{+\infty}x^r\left(1+\dfrac{x^2}{k}\right)^{-\frac{k+1}{2}}dx=2\int_{0}^{+\infty}x^r\left(1+\dfrac{x^2}{k}\right)^{-\frac{k+1}{2}}dx$，

令 $t=\left(1+\dfrac{x^2}{k}\right)^{-1}$，有 $x=k^{\frac{1}{2}}\left(\dfrac{1}{t}-1\right)^{\frac{1}{2}}=k^{\frac{1}{2}}t^{-\frac{1}{2}}(1-t)^{\frac{1}{2}}$，$dx=\dfrac{k^{\frac{1}{2}}}{2}\left(\dfrac{1}{t}-1\right)^{-\frac{1}{2}}\cdot\left(-\dfrac{1}{t^2}\right)dt=-\dfrac{k^{\frac{1}{2}}}{2}t^{-\frac{3}{2}}(1-t)^{-\frac{1}{2}}dt$，

且当 $x=0$ 时，$t=1$；当 $x\to+\infty$时，$t\to0$，

则 $\int_{-\infty}^{+\infty}x^r\left(1+\dfrac{x^2}{k}\right)^{-\frac{k+1}{2}}dx=2\int_{1}^{0}k^{\frac{r}{2}}t^{-\frac{r}{2}}(1-t)^{\frac{r}{2}}\cdot t^{\frac{k+1}{2}}\cdot(-1)\dfrac{k^{\frac{1}{2}}}{2}t^{-\frac{3}{2}}(1-t)^{-\frac{1}{2}}dt=k^{\frac{r+1}{2}}\int_{0}^{1}t^{\frac{k-r-2}{2}}(1-t)^{\frac{r-1}{2}}dt$

$$=k^{\frac{r+1}{2}}B\left(\frac{k-r}{2},\frac{r+1}{2}\right)=k^{\frac{r+1}{2}}\frac{\Gamma\left(\frac{k-r}{2}\right)\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{k+1}{2}\right)},$$

故 $E(T^r)=\dfrac{\Gamma\left(\dfrac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\dfrac{k}{2}\right)}\cdot k^{\frac{r+1}{2}}\dfrac{\Gamma\left(\dfrac{k-r}{2}\right)\Gamma\left(\dfrac{r+1}{2}\right)}{\Gamma\left(\dfrac{k+1}{2}\right)}=\dfrac{k^{\frac{r}{2}}\Gamma\left(\dfrac{r+1}{2}\right)\Gamma\left(\dfrac{k-r}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\dfrac{k}{2}\right)}$；

取 $r=1$，r 为奇数，当 $k=1$ 时，$E(T)$不存在；当 $k>1$ 时，$E(T)=0$；
取 $r=2$，r 为偶数，
故当 $k\le2$ 时，$E(T^2)$不存在，即 $\mathrm{Var}(T)$不存在；

当 $k>2$ 时，$\mathrm{Var}(T)=E(T^2)=\dfrac{k\Gamma\left(\dfrac{3}{2}\right)\Gamma\left(\dfrac{k-2}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\dfrac{k}{2}\right)}=\dfrac{k\cdot\dfrac{1}{2}\Gamma\left(\dfrac{1}{2}\right)\cdot\Gamma\left(\dfrac{k-2}{2}\right)}{\sqrt{\pi}\cdot\dfrac{k-2}{2}\Gamma\left(\dfrac{k-2}{2}\right)}=\dfrac{k}{k-2}$.

18．证明：若随机变量 $F\sim F(k,m)$，则当 $-\dfrac{k}{2}<r<\dfrac{m}{2}$ 时，有

$$E(F^r)=\frac{m^r\Gamma\left(\frac{k}{2}+r\right)\Gamma\left(\frac{m}{2}-r\right)}{k^r\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)},$$

由此写出 $E(F)$，$\mathrm{Var}(F)$.

证：因 $F \sim F(k, m)$，有 F 的密度函数为

$$p(x)=\frac{\Gamma\left(\frac{k+m}{2}\right)\left(\frac{k}{m}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}x^{\frac{k}{2}-1}\left(1+\frac{k}{m}x\right)^{-\frac{k+m}{2}}，\ x>0，$$

则 $E(F^r)=\dfrac{\Gamma\left(\frac{k+m}{2}\right)\left(\frac{k}{m}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}\displaystyle\int_0^{+\infty}x^r\cdot x^{\frac{k}{2}-1}\left(1+\frac{k}{m}x\right)^{-\frac{k+m}{2}}dx=\dfrac{\Gamma\left(\frac{k+m}{2}\right)\left(\frac{k}{m}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}\displaystyle\int_0^{+\infty}x^{\frac{k}{2}+r-1}\left(1+\frac{k}{m}x\right)^{-\frac{k+m}{2}}dx$，

因当 $x\to 0$ 时，$x^{\frac{k}{2}+r-1}\left(1+\frac{k}{m}x\right)^{-\frac{k+m}{2}}\sim x^{\frac{k}{2}+r-1}$；当 $x\to\infty$时，$x^{\frac{k}{2}+r-1}\left(1+\frac{k}{m}x\right)^{-\frac{k+m}{2}}\sim\left(\frac{k}{m}\right)^{-\frac{k+m}{2}}x^{-\frac{m}{2}+r-1}$，

则当 $\frac{k}{2}+r-1>-1$ 且 $-\frac{m}{2}+r-1<-1$ 时，即 $-\frac{k}{2}<r<\frac{m}{2}$，反常积分 $\displaystyle\int_0^{+\infty}x^{\frac{k}{2}+r-1}\left(1+\frac{k}{m}x\right)^{-\frac{k+m}{2}}dx$ 收敛，

令 $t=\left(1+\frac{k}{m}x\right)^{-1}$，有 $x=\frac{m}{k}\left(\frac{1}{t}-1\right)$，$dx=\frac{m}{k}\cdot\left(-\frac{1}{t^2}\right)dt$，

且当 $x=0$ 时，$t=1$；当 $x\to+\infty$时，$t\to 0$，

则 $\displaystyle\int_0^{+\infty}x^{\frac{k}{2}+r-1}\left(1+\frac{k}{m}x\right)^{-\frac{k+m}{2}}dx=\int_1^0\left(\frac{m}{k}\right)^{\frac{k}{2}+r-1}\left(\frac{1-t}{t}\right)^{\frac{k}{2}+r-1}\cdot t^{\frac{k+m}{2}}\cdot\frac{m}{k}\left(-\frac{1}{t^2}\right)dt=\left(\frac{m}{k}\right)^{\frac{k}{2}+r}\int_0^1 t^{\frac{m}{2}-r-1}(1-t)^{\frac{k}{2}+r-1}dt$

$$=\left(\frac{m}{k}\right)^{\frac{k}{2}+r}B\left(\frac{m}{2}-r,\frac{k}{2}+r\right)=\left(\frac{m}{k}\right)^{\frac{k}{2}+r}\frac{\Gamma\left(\frac{m}{2}-r\right)\Gamma\left(\frac{k}{2}+r\right)}{\Gamma\left(\frac{m+k}{2}\right)}，$$

故 $E(F^r)=\dfrac{\Gamma\left(\frac{k+m}{2}\right)\left(\frac{k}{m}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}\cdot\left(\frac{m}{k}\right)^{\frac{k}{2}+r}\dfrac{\Gamma\left(\frac{m}{2}-r\right)\Gamma\left(\frac{k}{2}+r\right)}{\Gamma\left(\frac{m+k}{2}\right)}=\dfrac{m^r\Gamma\left(\frac{k}{2}+r\right)\Gamma\left(\frac{m}{2}-r\right)}{k^r\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}$；

取 $r=1$，

当 $m\le 2$ 时，$E(F)$不存在；

当 $m>2$ 时，$E(F)=\dfrac{m\Gamma\left(\frac{k}{2}+1\right)\Gamma\left(\frac{m}{2}-1\right)}{k\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}=\dfrac{m\cdot\frac{k}{2}\Gamma\left(\frac{k}{2}\right)\cdot\Gamma\left(\frac{m}{2}-1\right)}{k\Gamma\left(\frac{k}{2}\right)\cdot\left(\frac{m}{2}-1\right)\Gamma\left(\frac{m}{2}-1\right)}=\dfrac{m}{m-2}$；

取 $r=2$，

当 $m\le 4$ 时，$E(F^2)$不存在，即 $\mathrm{Var}(F)$不存在；

当 $m>4$ 时，$E(F^2)=\dfrac{m^2\Gamma\left(\frac{k}{2}+2\right)\Gamma\left(\frac{m}{2}-2\right)}{k^2\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)}=\dfrac{m^2\cdot\left(\frac{k}{2}+1\right)\frac{k}{2}\Gamma\left(\frac{k}{2}\right)\cdot\Gamma\left(\frac{m}{2}-2\right)}{k^2\Gamma\left(\frac{k}{2}\right)\cdot\left(\frac{m}{2}-1\right)\left(\frac{m}{2}-2\right)\Gamma\left(\frac{m}{2}-2\right)}=\dfrac{m^2(k+2)}{k(m-2)(m-4)}$，

故 $\mathrm{Var}(F)=E(F^2)-[E(F)]^2=\frac{m^2(k+2)}{k(m-2)(m-4)}-\left(\frac{m}{m-2}\right)^2=\frac{2m^2(m+k-2)}{k(m-2)^2(m-4)}$.

19．设 $X_1, X_2, \cdots, X_n$ 是来自某连续总体的一个样本．该总体的分布函数 $F(x)$是连续严格单增函数，证明：统计量 $T=-2\sum_{i=1}^{n}\ln F(X_i)$ 服从$\chi^2(2n)$.

证：因 $Y_i=-2\ln F(X_i)$的分布函数：

$$F_Y(y)=P\{-2\ln F(X_i)\le y\}=P\{X_i\ge F^{-1}(\mathrm{e}^{-\frac{y}{2}})\}=1-F[F^{-1}(\mathrm{e}^{-\frac{y}{2}})]=1-\mathrm{e}^{-\frac{y}{2}},\ y>0,$$

则 $Y_i=-2\ln F(X_i)$服从指数分布 $Exp\left(\frac{1}{2}\right)$，也就是服从自由度为 2 的$\chi^2$分布$\chi^2(2)$，

因 $X_1, X_2, \cdots, X_n$ 相互独立，有 $Y_1, Y_2, \cdots, Y_n$ 相互独立，

故由χ^2分布的可加性知 $T=-2\sum_{i=1}^{n}\ln F(X_i)$ 服从$\chi^2(2n)$.

20．设 $X_1, X_2, \cdots, X_n$ 是来自正态分布 $N(\mu, \sigma^2)$的一个样本，$S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 是样本方差，试求满足 $P\left\{\frac{S_n^2}{\sigma^2}\le 1.5\right\}\ge 0.95$ 的最小 n 值.

解：因 $\frac{(n-1)S_n^2}{\sigma^2}\sim\chi^2(n-1)$，有 $P\left\{\frac{S_n^2}{\sigma^2}\le 1.5\right\}=P\left\{\frac{(n-1)S_n^2}{\sigma^2}\le 1.5(n-1)\right\}\ge 0.95$

则 $1.5(n-1)\ge\chi_{0.95}^2(n-1)$，即 $1.5\ge\frac{\chi_{0.95}^2(n-1)}{n-1}$，

因 $\frac{\chi_{0.95}^2(k)}{k}$ 单调下降，且 $\frac{\chi_{0.95}^2(25)}{25}=1.5061$，$\frac{\chi_{0.95}^2(26)}{26}=1.4956$，

故 $n-1\ge 26$，即 n 至少为 27.

21．设 $X_1, X_2, \cdots, X_n$ 独立同分布服从 $N(\mu, \sigma^2)$，$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$，$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$，记 $\xi=\frac{X_1-\overline{X}}{S}$. 试找出$\xi$与 t 分布的联系，因而定出ξ的密度函数（提示：作正交变换 $Y_1=\sqrt{n}\overline{X}$，$Y_2=\sqrt{\frac{n}{n-1}}(X_1-\overline{X})$，$Y_i=\sum_{j=1}^{n}c_{ij}X_j$，$j=3, \cdots, n$）.

解：因 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i=(X_1,\ X_2,\ \cdots,\ X_n)\cdot\frac{1}{n}(1,1,\cdots,1)^T$，

$$X_1-\overline{X}=\frac{n-1}{n}X_1-\frac{1}{n}\sum_{i=2}^{n}X_i=(X_1,\ X_2,\ \cdots,\ X_n)\cdot\frac{1}{n}(n-1,-1,\cdots,-1)^T,$$

且向量 $\alpha_1=\dfrac{1}{\sqrt{n}}(1,1,\cdots,1)^T$， $\alpha_2=\dfrac{1}{\sqrt{n(n-1)}}(n-1,-1,\cdots,-1)^T$ 正交并都是单位向量，

将单位向量α_1, α_2扩充为 n 维向量空间的一组标准正交基$\alpha_1, \alpha_2, \alpha_3, \cdots, \alpha_n$，

令 $C=(\alpha_1, \alpha_2, \alpha_3, \cdots, \alpha_n)$，$C$ 为正交阵，设 $(Y_1, Y_2, \cdots, Y_n)^T=C^T(X_1, X_2, \cdots, X_n)^T$，即$\vec{Y}=C^T\vec{X}$，

因 X_1, X_2, X_3， $\cdots$ ，X_n 相互独立且都服从方差同为σ^2的正态分布，
可知 $Y_1, Y_2, Y_3, \cdots, Y_n$ 相互独立且都服从方差同为σ^2的正态分布，

当 $i\ge 2$ 时， $E(Y_i)=E(\alpha_i^T\vec{X})=\alpha_i^T(\mu,\mu,\cdots,\mu)^T=\alpha_i^T\cdot\mu\cdot\sqrt{n}\alpha_1=0$，

则 $Y_2, Y_3, \cdots, Y_n$ 相互独立且都服从正态分布 $N(0,\sigma^2)$，即 $\dfrac{Y_i}{\sigma}\sim N(0,1)$，$i=2, 3, \ldots, n$，

因$\displaystyle\sum_{i=1}^n Y_i^2=\vec{Y}^T\vec{Y}=\vec{X}^TCC^T\vec{X}=\vec{X}^TE\vec{X}=\sum_{i=1}^n X_i^2$，

且$Y_1=\alpha_1^T\vec{X}=\dfrac{1}{\sqrt{n}}(X_1+X_2+\cdots+X_n)=\sqrt{n}\overline{X}$，

$$Y_2=\alpha_2^T\vec{X}=\frac{1}{\sqrt{n(n-1)}}[(n-1)X_1-X_2-\cdots-X_n]=\sqrt{\frac{n}{n-1}}(X_1-\overline{X})=\sqrt{\frac{n}{n-1}}S\xi,$$

则$\displaystyle(n-1)S^2=\sum_{i=1}^n(X_i-\overline{X})^2=\sum_{i=1}^n X_i^2-n\overline{X}^2=\sum_{i=1}^n Y_i^2-Y_1^2=\sum_{i=2}^n Y_i^2$，有$\displaystyle(n-1)S^2-Y_2^2=\sum_{i=3}^n Y_i^2$，

即$\displaystyle\frac{Y_2}{\sigma}\sim N(0,1)$， $\displaystyle\frac{(n-1)S^2-Y_2^2}{\sigma^2}=\sum_{i=3}^n\left(\frac{Y_i}{\sigma}\right)^2\sim\chi^2(n-2)$，且相互独立，

$$故T=\frac{\dfrac{Y_2}{\sigma}}{\sqrt{\dfrac{(n-1)S^2-Y_2^2}{\sigma^2}\Big/(n-2)}}=\frac{\sqrt{n-2}\cdot\sqrt{\dfrac{n}{n-1}}S\xi}{\sqrt{(n-1)S^2-\dfrac{n}{n-1}S^2\xi^2}}=\frac{\sqrt{n(n-2)}\xi}{\sqrt{(n-1)^2-n\xi^2}}\sim t(n-2).$$

22．设 $X_1, X_2, \cdots, X_m$ 相互独立，X_i 服从$\chi^2(n_i)$，$i=1, 2, \cdots, m$．令$U_1=\dfrac{X_1}{X_1+X_2}$， $U_2=\dfrac{X_1+X_2}{X_1+X_2+X_3}$， $\cdots$，

$U_{m-1}=\dfrac{X_1+\cdots+X_{m-1}}{X_1+\cdots+X_m}$．证明：$U_1, \cdots, U_{m-1}$ 相互独立，且 U_i 服从 $Be\left(\dfrac{n_1+\cdots+n_i}{2},\dfrac{n_{i+1}}{2}\right)$，$i=1, \cdots, m-1$，

（提示：令 $U_m=X_1+\cdots+X_m$，作变换 $X_1=U_1\cdots U_m$，$X_2=U_2\cdots U_m-U_1\cdots U_m$，$\cdots$，$X_m=U_m-U_{m-1}U_m$）.

证：因 $X_1, X_2, \cdots, X_m$ 相互独立，X_i 服从$\chi^2(n_i)$，$i=1, 2, \cdots, m$，

则$(X_1, X_2, \cdots, X_m)$的联合密度函数为

$$p_X(x_1,x_2,\cdots,x_m)=\prod_{i=1}^m\frac{\left(\frac{1}{2}\right)^{\frac{n_i}{2}}}{\Gamma\left(\frac{n_i}{2}\right)}x_i^{\frac{n_i}{2}-1}\mathrm{e}^{-\frac{x_i}{2}}\mathrm{I}_{x_i>0}=\frac{\left(\frac{1}{2}\right)^{\frac{1}{2}\sum_{i=1}^m n_i}}{\prod_{i=1}^m\Gamma\left(\frac{n_i}{2}\right)}\prod_{i=1}^m x_i^{\frac{n_i}{2}-1}\mathrm{e}^{-\frac{1}{2}\sum_{i=1}^m x_i}\mathrm{I}_{x_1,x_2,\cdots,x_m>0},$$

因$U_1=\dfrac{X_1}{X_1+X_2}$，$U_2=\dfrac{X_1+X_2}{X_1+X_2+X_3}$，…，$U_{m-1}=\dfrac{X_1+\cdots+X_{m-1}}{X_1+\cdots+X_m}$，且$X_i>0$，$i=1, 2, \cdots, m$，

则$0<U_i<1$，$i=1, 2, \cdots, m-1$，$U_m>0$，
令$U_m=X_1+\cdots+X_m$，有$X_1=U_1\cdots U_m$，$X_2=U_2\cdots U_m-U_1\cdots U_m$，…，$X_m=U_m-U_{m-1}U_m$，
设$Y_1=U_1\cdots U_m$，$Y_2=U_2\cdots U_m$，…，$Y_{m-1}=U_{m-1}U_m$，$Y_m=U_m$，
有$X_1=Y_1$，$X_2=Y_2-Y_1$，…，$X_m=Y_m-Y_{m-1}$，
则$(X_1, X_2, \cdots, X_m)$关于$(U_1, U_2, \cdots, U_m)$的雅可比行列式为

$$J=\left|\frac{\partial(x_1,x_2,\cdots,x_m)}{\partial(u_1,u_2,\cdots,u_m)}\right|=\left|\frac{\partial(x_1,x_2,\cdots,x_m)}{\partial(y_1,y_2,\cdots,y_m)}\right|\cdot\left|\frac{\partial(y_1,y_2,\cdots,y_m)}{\partial(u_1,u_2,\cdots,u_m)}\right|$$

$$=\begin{vmatrix}1&0&0&\cdots&0&0\\-1&1&0&\cdots&0&0\\0&-1&1&\cdots&0&0\\\cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\0&0&0&\cdots&1&0\\0&0&0&\cdots&-1&1\end{vmatrix}\cdot\begin{vmatrix}u_2\cdots u_m&u_1u_3\cdots u_m&u_1u_2u_4\cdots u_m&\cdots&u_1\cdots u_{m-2}u_m&u_1\cdots u_{m-1}\\0&u_3\cdots u_m&u_2u_4\cdots u_m&\cdots&u_2\cdots u_{m-2}u_m&u_2\cdots u_{m-1}\\0&0&u_4\cdots u_m&\cdots&u_3\cdots u_{m-2}u_m&u_3\cdots u_{m-1}\\\cdots&\cdots&\cdots&\cdots&\cdots&\cdots\\0&0&0&\cdots&u_m&0\\0&0&0&\cdots&0&1\end{vmatrix}$$

$$=u_2u_3^2\cdots u_m^{m-1},$$

可得$(U_1, U_2, \cdots, U_m)$的联合密度函数为

$$p_U(u_1,u_2,\cdots,u_m)$$

$$=\frac{\left(\frac{1}{2}\right)^{\frac{1}{2}\sum_{i=1}^{m}n_i}}{\prod_{i=1}^{m}\Gamma\left(\frac{n_i}{2}\right)}(u_1u_2\cdots u_m)^{\frac{n_1}{2}-1}\cdot\prod_{i=2}^{m}[(1-u_{i-1})u_i\cdots u_m)]^{\frac{n_i}{2}-1}\mathrm{e}^{-\frac{u_m}{2}}\mathrm{I}_{0<u_1,u_2,\cdots,u_{m-1}<1,\,u_m>0}\cdot u_2u_3^2\cdots u_m^{m-1}$$

$$=\frac{\left(\frac{1}{2}\right)^{\frac{1}{2}\sum_{i=1}^{m}n_i}}{\prod_{i=1}^{m}\Gamma\left(\frac{n_i}{2}\right)}u_1^{\frac{n_1}{2}-1}(1-u_1)^{\frac{n_2}{2}-1}\cdot u_2^{\frac{n_1+n_2}{2}-1}(1-u_2)^{\frac{n_3}{2}-1}\cdots u_{m-1}^{\frac{n_1+n_2+\cdots+n_{m-1}}{2}-1}(1-u_{m-1})^{\frac{n_m}{2}-1}$$

$$\cdot u_m^{\frac{n_1+n_2+\cdots+n_m}{2}-1}\mathrm{e}^{-\frac{u_m}{2}}\mathrm{I}_{0<u_1,u_2,\cdots,u_{m-1}<1,\,u_m>0},$$

$$=\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}u_1^{\frac{n_1}{2}-1}(1-u_1)^{\frac{n_2}{2}-1}\mathrm{I}_{0<u_1<1}\cdot\frac{\Gamma\left(\frac{n_1+n_2+n_3}{2}\right)}{\Gamma\left(\frac{n_1+n_2}{2}\right)\Gamma\left(\frac{n_3}{2}\right)}u_2^{\frac{n_1+n_2}{2}-1}(1-u_2)^{\frac{n_3}{2}-1}\mathrm{I}_{0<u_2<1}$$

$$\cdots\frac{\Gamma\left(\frac{n_1+n_2+\cdots+n_m}{2}\right)}{\Gamma\left(\frac{n_1+n_2+\cdots+n_{m-1}}{2}\right)\Gamma\left(\frac{n_m}{2}\right)}u_{m-1}^{\frac{n_1+n_2+\cdots+n_{m-1}}{2}-1}(1-u_{m-1})^{\frac{n_m}{2}-1}\mathrm{I}_{0<u_{m-1}<1}$$

$$\cdot\frac{\left(\frac{1}{2}\right)^{\frac{n_1+n_2+\cdots+n_m}{2}}}{\Gamma\left(\frac{n_1+n_2+\cdots+n_m}{2}\right)}u_m^{\frac{n_1+n_2+\cdots+n_m}{2}-1}\mathrm{e}^{-\frac{u_m}{2}}\mathrm{I}_{u_m>0}$$

由于$(U_1, U_2, \cdots, U_m)$的联合密度函数 $p_U(u_1, u_2, \cdots, u_m)$可分离变量，

故 $U_1, U_2, \cdots, U_m$ 相互独立，且 U_i 服从 $Be\left(\dfrac{n_1+\cdots+n_i}{2}, \dfrac{n_{i+1}}{2}\right)$，$i = 1, \cdots, m-1$；$U_m$ 服从$\chi^2(n_1 + \cdots + n_m)$.

习题 5.5

1．设 $X_1, \cdots, X_n$ 是来自几何分布 $P\{X = x\} = \theta(1-\theta)^x, x = 0, 1, 2, \cdots$的样本，证明$T=\sum_{i=1}^{n} X_i$ 是充分统计量.

证：方法一：根据充分统计量的定义

样本联合概率函数

$$p(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} \theta(1-\theta)^{x_i} = \theta^n (1-\theta)^{\sum_{i=1}^{n} x_i},$$

因 $X_i + 1$ 的概率函数为 $P\{X_i + 1 = x\} = \theta(1-\theta)^x$，$x = 1, 2, \cdots$，即服从几何分布 $Ge(\theta)$，$i = 1, 2, \cdots, n$，

则根据几何分布与负二项分布的关系可知$\sum_{i=1}^{n}(X_i+1) = T + n$服从负二项分布 $Nb(n, \theta)$，即概率函数为

$$P\{T+n=k\} = \binom{k-1}{n-1} \theta^n (1-\theta)^{k-n}, \quad k = n, n+1, n+2, \cdots,$$

即$T=\sum_{i=1}^{n} X_i$ 的概率函数为 $p_T(t;\theta) = \binom{t+n-1}{n-1} \theta^n (1-\theta)^t, \quad t = 0, 1, 2, \cdots,$

可得在 $T = t$ 时，即$t = \sum_{i=1}^{n} x_i$，$X_1, X_2, \cdots, X_n$ 的条件概率函数为

$$p(x_1, x_2, \cdots, x_n; \theta \mid T = t) = \frac{p(x_1, x_2, \cdots, x_n; \theta\}}{p_T(t;\theta)} = \frac{\theta^n (1-\theta)^{\sum_{i=1}^{n} x_i}}{\binom{t+n-1}{n-1} \theta^n (1-\theta)^t} = \frac{1}{\binom{t+n-1}{n-1}},$$

这与参数θ 无关，

故根据充分统计量的定义可知$T=\sum_{i=1}^{n} X_i$ 是θ 的充分统计量.

方法二：根据因子分解定理

样本联合概率函数

$$p(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} \theta(1-\theta)^{x_i} = \theta^n (1-\theta)^{\sum_{i=1}^{n} x_i},$$

因$T=\sum_{i=1}^{n} X_i$，有$t=\sum_{i=1}^{n} x_i$，即 $p(x_1, x_2, \cdots, x_n; \theta) = \theta^n (1-\theta)^t$，

取 $g(t; \theta) = \theta^n (1-\theta)^t$，$h(x_1, x_2, \cdots, x_n) = 1$ 与参数θ 无关，

故根据因子分解定理可知 $T=\sum_{i=1}^{n} X_i$ 是 θ 的充分统计量.

2． 设 $X_1,\ \cdots, X_n$ 是来自泊松分布 $P(\lambda)$ 的样本，证明 $T=\sum_{i=1}^{n} X_i$ 是充分统计量.

证：方法一：根据充分统计量的定义

样本联合概率函数

$$p(x_1, x_2, \cdots, x_n; \lambda)=\prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} \mathrm{e}^{-\lambda}=\frac{\lambda^{\sum_{i=1}^{n} x_i}}{x_1! x_2! \cdots x_n!} \mathrm{e}^{-n\lambda}=\lambda^{\sum_{i=1}^{n} x_i} \mathrm{e}^{-n\lambda} \cdot \frac{1}{x_1! x_2! \cdots x_n!},$$

根据泊松分布的可加性可知 $T=\sum_{i=1}^{n} X_i$ 服从泊松分布 $P(n\lambda)$，即概率函数为

$$p_T(t; \lambda)=\frac{(n\lambda)^t}{t!} \mathrm{e}^{-n\lambda}, \quad t=0,1,2, \cdots,$$

可得在 $T=t$ 时，即 $t=\sum_{i=1}^{n} x_i$，$X_1, X_2,\ \cdots, X_n$ 的条件概率函数为

$$p(x_1, x_2, \cdots, x_n; \theta \mid T=t)=\frac{p(x_1, x_2, \cdots, x_n; \theta\}}{p_T(t; \theta)}=\frac{\lambda^{\sum_{i=1}^{n} x_i} \mathrm{e}^{-n\lambda} \cdot \frac{1}{x_1! x_2! \cdots x_n!}}{\frac{n^t \lambda^t}{t!} \mathrm{e}^{-n\lambda}}=\frac{t!}{n^t \cdot x_1! x_2! \cdots x_n!},$$

这与参数 λ 无关，

故根据充分统计量的定义可知 $T=\sum_{i=1}^{n} X_i$ 是 λ 的充分统计量.

方法二：根据因子分解定理

样本联合概率函数

$$p(x_1, x_2, \cdots, x_n; \lambda)=\prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} \mathrm{e}^{-\lambda}=\frac{\lambda^{\sum_{i=1}^{n} x_i}}{x_1! x_2! \cdots x_n!} \mathrm{e}^{-n\lambda}=\lambda^{\sum_{i=1}^{n} x_i} \mathrm{e}^{-n\lambda} \cdot \frac{1}{x_1! x_2! \cdots x_n!},$$

因 $T=\sum_{i=1}^{n} X_i$，有 $t=\sum_{i=1}^{n} x_i$，即 $p(x_1, x_2, \cdots, x_n; \lambda)=\lambda^t \mathrm{e}^{-n\lambda} \cdot \dfrac{1}{x_1! x_2! \cdots x_n!}$,

取 $g(t; \lambda)=\lambda^t \mathrm{e}^{-n\lambda}$，$h(x_1, x_2, \cdots, x_n)=\dfrac{1}{x_1! x_2! \cdots x_n!}$ 与参数 λ 无关，

故根据因子分解定理可知 $T=\sum_{i=1}^{n} X_i$ 是 λ 的充分统计量.

3． 设总体为如下离散型分布，

X	a_1	a_2	$\cdots$	a_k
P	p_1	p_2	$\cdots$	p_k

$X_1, \cdots, X_n$ 是来自该总体的样本，

（1）证明次序统计量 $(X_{(1)}, \cdots, X_{(n)})$ 是充分统计量.

（2）以 n_j 表示 $X_1, \cdots, X_n$ 中等于 a_j 的个数，证明 $(n_1, \cdots, n_k)$ 是充分统计量.

证：设样本 $(X_1, X_2, \cdots, X_n)$ 中有 n_1 个 a_1，n_2 个 a_2，…，n_k 个 a_k，

显然次序统计量 $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ 中同样有 n_1 个 a_1，n_2 个 a_2，…，n_k 个 a_k，

样本联合概率函数

$$p(x_1, x_2, \cdots, x_n; p_1, p_2, \cdots, p_k) = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k},$$

（2）因 $T_2 = (n_1, \cdots, n_k)$，取 $g(n_1, n_2, \cdots, n_k; p_1, p_2, \cdots, p_k) = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$，$h(x_1, x_2, \cdots, x_n) = 1$，

故根据因子分解定理可知 $T_2 = (n_1, n_2, \cdots, n_k)$ 是 $(p_1, p_2, \cdots, p_k)$ 的充分统计量；

（1）因 $T_1 = (X_{(1)}, X_{(2)}, \cdots, X_{(n)})$，显然$(n_1, n_2, \cdots, n_k)$ 与 $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ 一一对应，

故由第（2）小题结论知 $T_1 = (X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ 是 $(p_1, p_2, \cdots, p_k)$ 的充分统计量.

4． 设 $X_1, \cdots, X_n$ 是来自正态分布 $N(\mu, 1)$ 的样本，证明 $T = \sum_{i=1}^{n} X_i$ 是充分统计量

证：方法一：根据充分统计量的定义

样本联合密度函数

$$p(x_1, x_2, \cdots, x_n; \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(x_i-\mu)^2}{2}} = \frac{1}{(\sqrt{2\pi})^n} \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n}(x_i^2 - 2\mu x_i + \mu^2)} = \frac{1}{(\sqrt{2\pi})^n} \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2 + \mu \sum_{i=1}^{n} x_i - \frac{1}{2} n\mu^2},$$

根据正态分布的可加性可知 $T = \sum_{i=1}^{n} X_i$ 服从正态分布 $N(n\mu, n)$，即密度函数为

$$p_T(t) = \frac{1}{\sqrt{2\pi} \cdot \sqrt{n}} \mathrm{e}^{-\frac{(t-n\mu)^2}{2n}} = \frac{1}{\sqrt{2\pi} \cdot \sqrt{n}} \mathrm{e}^{-\frac{t^2}{2n} + \mu t - \frac{1}{2} n\mu^2},$$

可得在 $T = t$ 时，即 $t = \sum_{i=1}^{n} x_i$，$X_1, X_2, \cdots, X_n$ 的条件概率函数为

$$p(x_1, x_2, \cdots, x_n; \mu \mid T = t) = \frac{p(x_1, x_2, \cdots, x_n; \mu)}{p_T(t)}$$

$$= \frac{\dfrac{1}{(\sqrt{2\pi})^n} \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2 + \mu \sum_{i=1}^{n} x_i - \frac{1}{2} n\mu^2}}{\dfrac{1}{\sqrt{2\pi} \cdot \sqrt{n}} \mathrm{e}^{-\frac{t^2}{2n} + \mu t - \frac{1}{2} n\mu^2}} = \frac{\sqrt{n}}{(\sqrt{2\pi})^{n-1}} \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2 + \frac{t^2}{2n}} = \frac{\sqrt{n}}{(\sqrt{2\pi})^{n-1}} \mathrm{e}^{-\frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)},$$

这与参数 μ 无关，

故根据充分统计量的定义可知 $T = \sum_{i=1}^{n} X_i$ 是 μ 的充分统计量.

方法二：根据因子分解定理

样本联合密度函数

$$p(x_1,x_2,\cdots,x_n;\mu)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(x_i-\mu)^2}{2}}=\frac{1}{(\sqrt{2\pi})^n}\mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n}(x_i^2-2\mu x_i+\mu^2)}=\frac{1}{(\sqrt{2\pi})^n}\mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n}x_i^2+\mu\sum_{i=1}^{n}x_i-\frac{1}{2}n\mu^2},$$

因$T=\sum_{i=1}^{n}X_i$，有$t=\sum_{i=1}^{n}x_i$，即 $p(x_1,x_2,\cdots,x_n;\mu)=\frac{1}{(\sqrt{2\pi})^n}\mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n}x_i^2+\mu t-\frac{1}{2}n\mu^2}=\frac{1}{(\sqrt{2\pi})^n}\mathrm{e}^{\mu t-\frac{1}{2}n\mu^2}\cdot\mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n}x_i^2}$，

取$g(t;\mu)=\frac{1}{(\sqrt{2\pi})^n}\mathrm{e}^{\mu t-\frac{1}{2}n\mu^2}$，$h(x_1,x_2,\cdots,x_n)=\mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n}x_i^2}$与参数$\mu$无关，

故根据因子分解定理可知$T=\sum_{i=1}^{n}X_i$是μ的充分统计量.

5． 设$X_1,\cdots,X_n$是来自$p(x;\theta)=\theta x^{\theta-1}$，$0<x<1$，$\theta>0$的样本，试给出一个充分统计量.

解：样本联合密度函数

$$p(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^{n}\theta x_i^{\theta-1}\mathrm{I}_{0<x_i<1}=\theta^n(x_1x_2\cdots x_n)^{\theta-1}\mathrm{I}_{0<x_1,x_2,\cdots,x_n<1},$$

令$T=X_1X_2\cdots X_n$，有$t=x_1x_2\cdots x_n$，即$p(x_1,x_2,\cdots,x_n;\theta)=\theta^n t^{\theta-1}\mathrm{I}_{0<x_1,x_2,\cdots,x_n<1}$，

取$g(t;\theta)=\theta^n t^{\theta-1}$，$h(x_1,x_2,\cdots,x_n)=\mathrm{I}_{0<x_1,x_2,\cdots,x_n<1}$与参数$\theta$无关，

故根据因子分解定理可知$T=X_1X_2\cdots X_n$是θ的充分统计量.

6． 设$X_1,\cdots,X_n$是来自韦布尔分布$p(x;\theta)=mx^{m-1}\theta^{-m}\mathrm{e}^{-(x/\theta)^m}$，$x>0$，$\theta>0$的样本（$m>0$已知），试给出一个充分统计量.

解：样本联合密度函数

$$p(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^{n}mx_i^{m-1}\theta^{-m}\mathrm{e}^{-(x_i/\theta)^m}\mathrm{I}_{x_i>0}=m^n(x_1x_2\cdots x_n)^{m-1}\theta^{-mn}\mathrm{e}^{-\sum_{i=1}^{n}(x_i/\theta)^m}\mathrm{I}_{x_1,x_2,\cdots,x_n>0}$$

$$=\theta^{-mn}\mathrm{e}^{-\frac{1}{\theta^m}\sum_{i=1}^{n}x_i^m}\cdot m^n(x_1x_2\cdots x_n)^{m-1}\mathrm{I}_{x_1,x_2,\cdots,x_n>0},$$

令$T=\sum_{i=1}^{n}X_i^m$，有$t=\sum_{i=1}^{n}x_i^m$，即$p(x_1,x_2,\cdots,x_n;\theta)=\theta^{-mn}\mathrm{e}^{-\frac{1}{\theta^m}t}\cdot m^n(x_1x_2\cdots x_n)^{m-1}\mathrm{I}_{x_1,x_2,\cdots,x_n>0}$，

取$g(t;\theta)=\theta^{-mn}\mathrm{e}^{-\frac{1}{\theta^m}t}$，$h(x_1,x_2,\cdots,x_n)=m^n(x_1x_2\cdots x_n)^{m-1}\mathrm{I}_{x_1,x_2,\cdots,x_n>0}$与参数$\theta$无关，

故根据因子分解定理知$T=\sum_{i=1}^{n}X_i^m$是θ的充分统计量.

7． 设$X_1,\cdots,X_n$是来 Pareto 分布$p(x;\theta)=\theta a^\theta x^{-(\theta+1)}$，$x>a$，$\theta>0$的样本（$a>0$已知），试给出一个充分统计量.

解：样本联合密度函数

$$p(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^{n}\theta a^\theta x_i^{-(\theta+1)}\mathrm{I}_{x_i>a}=\theta^n a^{n\theta}(x_1x_2\cdots x_n)^{-(\theta+1)}\mathrm{I}_{x_1,x_2,\cdots,x_n>a},$$

令 $T = X_1X_2\cdots X_n$，有 $t = x_1x_2\cdots x_n$，即 $p(x_1, x_2, \cdots, x_n; \theta) = \theta^n a^{n\theta} t^{-(\theta+1)} \mathrm{I}_{x_1, x_2, \cdots, x_n > a}$，

取 $g(t; \theta) = \theta^n a^{n\theta} t^{-(\theta+1)}$，$h(x_1, x_2, \cdots, x_n) = \mathrm{I}_{x_1, x_2, \cdots, x_n > a}$ 与参数 θ 无关，

故根据因子分解定理知 $T = X_1X_2\cdots X_n$ 是 θ 的充分统计量.

8．设 $X_1, \cdots, X_n$ 是来自 Laplace 分布 $p(x;\theta) = \dfrac{1}{2\theta} e^{-|x|/\theta}$，$\theta > 0$ 的样本，试给出一个充分统计量.

解：样本联合密度函数

$$p(x_1, x_2, \cdots, x_n; \mu) = \prod_{i=1}^n \frac{1}{2\theta} \mathrm{e}^{-\frac{|x_i|}{\theta}} = \frac{1}{(2\theta)^n} \mathrm{e}^{-\frac{1}{\theta}\sum\limits_{i=1}^n |x_i|},$$

令 $T = \sum\limits_{i=1}^n |X_i|$，有 $t = \sum\limits_{i=1}^n |x_i|$，即 $p(x_1, x_2, \cdots, x_n; \mu) = \dfrac{1}{(2\theta)^n} \mathrm{e}^{-\frac{1}{\theta}t}$，

取 $g(t; \theta) = \dfrac{1}{(2\theta)^n} \mathrm{e}^{-\frac{1}{\theta}t}$，$h(x_1, x_2, \cdots, x_n) = 1$ 与参数 θ 无关，

故根据因子分解定理知 $T = \sum\limits_{i=1}^n |X_i|$ 是 θ 的充分统计量.

9．设 $X_1, \cdots, X_n$ 独立同分布，X_1 服从以下分布，求相应的充分统计量：

（1）负二项分布 $X_1 \sim p(x_1;\theta) = \dbinom{x_1 + r - 1}{r - 1} \theta^r (1-\theta)^{x_1}$，$x_1 = 0, 1, 2, \cdots$，$r$ 已知；

（2）离散均匀分布 $X_1 \sim p(x_1; m) = \dfrac{1}{m}$，$x_1 = 1, 2, \cdots, m$，$m$ 未知；

（3）对数正态分布 $X_1 \sim p(x_1; \mu, \sigma) = \dfrac{1}{\sqrt{2\pi}\sigma x_1} \exp\left\{-\dfrac{1}{2\sigma^2}(\ln x_1 - \mu)^2\right\}$，$x_1 > 0$；

（4）瑞利（Rayleigh）分布 $X_1 \sim p(x_1; \mu, \sigma) = 2\lambda x_1 \mathrm{e}^{-\lambda x_1^2} \cdot \mathrm{I}_{x_1 \ge 0}$.

注：第（4）小题有误，密度函数应为 $p(x_1; \lambda)$，即参数应为 λ，而不是 μ, σ.

解：（1）样本联合密度函数为

$$p(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^n \binom{x_i + r - 1}{r - 1} \theta^r (1-\theta)^{x_i} = \theta^{nr} (1-\theta)^{\sum\limits_{i=1}^n x_i} \cdot \prod_{i=1}^n \binom{x_i + r - 1}{r - 1},$$

令 $T = \sum\limits_{i=1}^n X_i$，有 $t = \sum\limits_{i=1}^n x_i$，即 $p(x_1, x_2, \cdots, x_n; \theta) = \theta^{nr}(1-\theta)^t \cdot \prod\limits_{i=1}^n \dbinom{x_i + r - 1}{r - 1}$，

取 $g(t; \theta) = \theta^{nr}(1-\theta)^t$，$h(x_1, x_2, \cdots, x_n) = \prod\limits_{i=1}^n \dbinom{x_i + r - 1}{r - 1}$ 与参数 θ 无关，

故根据因子分解定理知 $T = \sum\limits_{i=1}^n X_i$ 是参数 θ 的充分统计量；

（2）样本联合密度函数为

$$p(x_1, x_2, \cdots, x_n; m) = \prod_{i=1}^{n} \frac{1}{m} \cdot \mathrm{I}_{1 \le x_i \le m,\, x_i\text{为整数}} = \frac{1}{m^n} \cdot \mathrm{I}_{1 \le x_1, x_2, \cdots, x_n \le m,\, x_1, x_2, \cdots, x_n\text{为整数}} ,$$

$$= \frac{1}{m^n} \cdot \mathrm{I}_{1 \le x_{(1)} \le x_{(n)} \le m,\, x_1, x_2, \cdots, x_n\text{为整数}} = \frac{1}{m^n} \cdot \mathrm{I}_{x_{(n)} \le m} \cdot \mathrm{I}_{x_{(1)} \ge 1,\, x_1, x_2, \cdots, x_n\text{为整数}} ,$$

令 $T = X_{(n)} = \max\limits_{1 \le i \le n}\{X_i\}$，有 $t = x_{(n)}$，即 $p(x_1, x_2, \cdots, x_n; m) = \dfrac{1}{m^n} \cdot \mathrm{I}_{t \le m} \cdot \mathrm{I}_{x_{(1)} \ge 1,\, x_1, x_2, \cdots, x_n\text{为整数}}$，

取 $g(t; m) = \dfrac{1}{m^n} \cdot \mathrm{I}_{t \le m}$，$h(x_1, x_2, \cdots, x_n) = \mathrm{I}_{x_{(1)} \ge 1,\, x_1, x_2, \cdots, x_n\text{为整数}}$ 与参数 m 无关，

故根据因子分解定理知 $T = X_{(n)}$是参数 m 的充分统计量；

（3）样本联合密度函数为

$$p(x_1, x_2, \cdots, x_n; \mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma x_i} \exp\left\{-\frac{1}{2\sigma^2}(\ln x_i - \mu)^2\right\} \cdot \mathrm{I}_{x_i > 0}$$

$$= \frac{1}{(\sqrt{2\pi}\sigma)^n x_1 x_2 \cdots x_n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln^2 x_i - 2\mu \ln x_i + \mu^2)\right\} \cdot \mathrm{I}_{x_1, x_2, \cdots, x_n > 0}$$

$$= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}\ln^2 x_i - 2\mu\sum_{i=1}^{n}\ln x_i + n\mu^2\right)\right\} \cdot \frac{1}{x_1 x_2 \cdots x_n} \mathrm{I}_{x_1, x_2, \cdots, x_n > 0} ,$$

令 $T_1 = \sum\limits_{i=1}^{n} \ln X_i$，$T_2 = \sum\limits_{i=1}^{n} \ln^2 X_i$，有 $t_1 = \sum\limits_{i=1}^{n} \ln x_i$，$t_2 = \sum\limits_{i=1}^{n} \ln^2 x_i$，

则 $p(x_1, x_2, \cdots, x_n; \mu, \sigma) = \dfrac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\dfrac{1}{2\sigma^2}(t_2 - 2\mu t_1 + n\mu^2)\right\} \cdot \dfrac{1}{x_1 x_2 \cdots x_n} \cdot \mathrm{I}_{x_1, x_2, \cdots, x_n > 0}$，

取 $g(t; \mu, \sigma) = \dfrac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\dfrac{1}{2\sigma^2}(t_2 - 2\mu t_1 + n\mu^2)\right\}$，

$h(x_1, x_2, \cdots, x_n) = \dfrac{1}{x_1 x_2 \cdots x_n} \cdot \mathrm{I}_{x_1, x_2, \cdots, x_n > 0}$ 与参数μ, σ 无关，

故根据因子分解定理知 $(T_1, T_2) = \left(\sum\limits_{i=1}^{n} \ln X_i, \sum\limits_{i=1}^{n} \ln^2 X_i\right)$是参数$(\mu, \sigma)$的充分统计量；

（4）样本联合密度函数为

$$p(x_1, x_2, \cdots, x_n; \lambda) = \prod_{i=1}^{n} 2\lambda x_i \,\mathrm{e}^{-\lambda x_i^2} \cdot \mathrm{I}_{x_i > 0} = 2^n \lambda^n x_1 x_2 \cdots x_n \,\mathrm{e}^{-\lambda \sum_{i=1}^{n} x_i^2} \cdot \mathrm{I}_{x_1, x_2, \cdots, x_n > 0} ,$$

令 $T = \sum\limits_{i=1}^{n} X_i^2$，有 $t = \sum\limits_{i=1}^{n} x_i^2$，即 $p(x_1, x_2, \cdots, x_n; \lambda) = 2^n \lambda^n \mathrm{e}^{-\lambda t} \cdot x_1 x_2 \cdots x_n \mathrm{I}_{x_1, x_2, \cdots, x_n > 0}$，

取 $g(t; \lambda) = 2^n \lambda^n \mathrm{e}^{-\lambda t}$，$h(x_1, x_2, \cdots, x_n) = x_1 x_2 \cdots x_n \cdot \mathrm{I}_{x_1, x_2, \cdots, x_n > 0}$ 与参数λ无关，

故根据因子分解定理知 $T = \sum\limits_{i=1}^{n} X_i^2$ 是参数λ的充分统计量.

10．设 $X_1, \cdots, X_n$ 是来自正态分布 $N(\mu, \sigma^2)$的样本.

（1）在μ 已知时给出σ^2 的一个充分统计量；

（2）在σ^2 已知时给出μ 的一个充分统计量.

解：因总体密度函数为

$$p(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

则样本联合密度函数为

$$p(x_1,x_2,\cdots,x_n;\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}}=\frac{1}{(\sqrt{2\pi}\sigma)^n}\mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2},$$

（1）在μ 已知时，令 $T_1=\sum_{i=1}^{n}(X_i-\mu)^2$ ，有 $t=\sum_{i=1}^{n}(x_i-\mu)^2$ ，即 $p(x_1,x_2,\cdots,x_n;\sigma^2)=\frac{1}{(\sqrt{2\pi}\sigma)^n}\mathrm{e}^{-\frac{t}{2\sigma^2}}$ ，

取 $g(t;\sigma^2)=\frac{1}{(\sqrt{2\pi}\sigma)^n}\mathrm{e}^{-\frac{t}{2\sigma^2}}$ ，$h(x_1, x_2, \cdots, x_n) = 1$ 与参数σ^2 无关，

故根据因子分解定理知 $T_1=\sum_{i=1}^{n}(X_i-\mu)^2$ 是参数σ^2 的充分统计量；

（2）在σ^2 已知时，

$$p(x_1,x_2,\cdots,x_n;\mu)=\frac{1}{(\sqrt{2\pi}\sigma)^n}\mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2-2\mu x_i+\mu^2)}=\frac{1}{(\sqrt{2\pi}\sigma)^n}\mathrm{e}^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}x_i^2-2\mu\sum_{i=1}^{n}x_i+n\mu^2\right)}$$

$$=\frac{1}{(\sqrt{2\pi}\sigma)^n}\mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2-2\mu x_i+\mu^2)}=\frac{1}{(\sqrt{2\pi}\sigma)^n}\mathrm{e}^{\frac{\mu}{\sigma^2}\sum_{i=1}^{n}x_i}\cdot\mathrm{e}^{-\frac{n\mu^2}{2\sigma^2}}\cdot\mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2},$$

令 $T_2=\sum_{i=1}^{n}X_i$ ，有 $t=\sum_{i=1}^{n}x_i$ ，即 $p(x_1,x_2,\cdots,x_n;\mu)=\frac{1}{(\sqrt{2\pi}\sigma)^n}\mathrm{e}^{\frac{\mu}{\sigma^2}t}\cdot\mathrm{e}^{-\frac{n\mu^2}{2\sigma^2}}\cdot\mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2}$ ，

取 $g(t;\mu)=\frac{1}{(\sqrt{2\pi}\sigma)^n}\mathrm{e}^{\frac{\mu}{\sigma^2}t}\cdot\mathrm{e}^{-\frac{n\mu^2}{2\sigma^2}}$ ，$h(x_1,x_2,\cdots,x_n)=\mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2}$ 与参数μ 无关，

故根据因子分解定理知 $T_2=\sum_{i=1}^{n}X_i$ 是参数μ 的充分统计量.

11．设 $X_1, \cdots, X_n$ 是来自均匀分布 $U(\theta_1, \theta_2)$ 的样本，试给出一个充分统计量.

解：样本联合密度函数

$$p(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^{n}\frac{1}{\theta_2-\theta_1}I_{\theta_1<x_i<\theta_2}=\frac{1}{(\theta_2-\theta_1)^n}I_{\theta_1<x_1,\,x_2,\,\cdots,\,x_n<\theta_2}=\frac{1}{(\theta_2-\theta_1)^n}I_{\theta_1<x_{(1)}\le x_{(n)}<\theta_2},$$

令 $(T_1, T_2) = (X_{(1)}, X_{(n)})$，有$(t_1, t_2) = (x_{(1)}, x_{(n)})$，即 $p(x_1,x_2,\cdots,x_n;\theta)=\frac{1}{(\theta_2-\theta_1)^n}I_{\theta_1<t_1\le t_2<\theta_2}$ ，

取 $g(t_1,t_2;\theta_1,\theta_2)=\frac{1}{(\theta_2-\theta_1)^n}I_{\theta_1<t_1\le t_2<\theta_2}$ ，$h(x_1, x_2, \cdots, x_n) = 1$ 与参数θ_1, θ_2 无关，

故根据因子分解定理知$(T_1, T_2) = (X_{(1)}, X_{(n)})$是$(\theta_1, \theta_2)$的充分统计量.

12．设 $X_1, \cdots, X_n$ 是来自均匀分布 $U(\theta, 2\theta)$ ，$\theta > 0$ 的样本，试给出充分统计量．
解：样本联合密度函数

$$p(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} \frac{1}{\theta} I_{\theta < x_i < 2\theta} = \frac{1}{\theta^n} I_{\theta < x_1, x_2, \cdots, x_n < 2\theta} = \frac{1}{\theta^n} I_{\theta < x_{(1)} \le x_{(n)} < 2\theta},$$

令 $(T_1, T_2) = (X_{(1)}, X_{(n)})$，有$(t_1, t_2) = (x_{(1)}, x_{(n)})$，即 $p(x_1, x_2, \cdots, x_n; \theta) = \dfrac{1}{\theta^n} I_{\theta < t_1 \le t_2 < 2\theta}$

取 $g(t_1, t_2; \theta) = \dfrac{1}{\theta^n} I_{\theta < t_1 \le t_2 < 2\theta}$，$h(x_1, x_2, \cdots, x_n) = 1$ 与参数θ无关，

故根据因子分解定理知$(T_1, T_2) = (X_{(1)}, X_{(n)})$是$\theta$ 的充分统计量．

13．设 $X_1, \cdots, X_n$ 来自伽玛分布族$\{Ga(\alpha, \lambda) \mid \alpha > 0, \lambda > 0\}$的一个样本，寻求$(\alpha, \lambda)$的充分统计量．
解：总体 X 的密度函数为

$$p(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\lambda x} \mathrm{I}_{x>0},$$

样本联合密度函数为

$$p(x_1, x_2, \cdots, x_n; \alpha, \lambda) = \prod_{i=1}^{n} \frac{\lambda^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} \mathrm{e}^{-\lambda x_i} \mathrm{I}_{x_i>0} = \frac{\lambda^{n\alpha}}{\Gamma(\alpha)^n} (x_1 x_2 \cdots x_n)^{\alpha-1} \mathrm{e}^{-\lambda \sum\limits_{i=1}^{n} x_i} \mathrm{I}_{x_1, x_2, \cdots, x_n > 0},$$

令 $(T_1, T_2) = \left(X_1 X_2 \cdots X_n, \sum\limits_{i=1}^{n} X_i \right)$，有 $(t_1, t_2) = \left(x_1 x_2 \cdots x_n, \sum\limits_{i=1}^{n} x_i \right)$，

则 $p(x_1, x_2, \cdots, x_n; \alpha, \lambda) = \dfrac{\lambda^{n\alpha}}{\Gamma(\alpha)^n} t_1^{\alpha-1} \mathrm{e}^{-\lambda t_2} \mathrm{I}_{x_1, x_2, \cdots, x_n > 0}$，

取 $g(t_1, t_2; \alpha, \lambda) = \dfrac{\lambda^{n\alpha}}{\Gamma(\alpha)^n} t_1^{\alpha-1} \mathrm{e}^{-\lambda t_2}$，$h(x_1, x_2, \cdots, x_n) = \mathrm{I}_{x_1, x_2, \cdots, x_n > 0}$ 与参数α, λ无关，

故 $(T_1, T_2) = \left(X_1 X_2 \cdots X_n, \sum\limits_{i=1}^{n} X_i \right)$是参数$(\alpha, \lambda)$的充分统计量．

14．设 $X_1, \cdots, X_n$ 是来自贝塔分布族$\{Be(a, b) \mid a > 0, b > 0\}$的一个样本，寻求$(a, b)$的充分统计量．
解：总体 X 的密度函数为

$$p(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \mathrm{I}_{0<x<1},$$

样本联合密度函数

$$p(x_1, x_2, \cdots, x_n; a, b) = \prod_{i=1}^{n} p(x_i; a, b) = \prod_{i=1}^{n} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x_i^{a-1} (1-x_i)^{b-1} \mathrm{I}_{0<x_i<1}$$

$$= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right]^n \left(\prod_{i=1}^{n} x_i \right)^{a-1} \left[\prod_{i=1}^{n} (1-x_i) \right]^{b-1} \mathrm{I}_{0<x_1, x_2, \cdots, x_n<1},$$

令 $(T_1, T_2) = \left(\prod\limits_{i=1}^{n} X_i, \prod\limits_{i=1}^{n} (1-X_i) \right)$，有 $(t_1, t_2) = \left(\prod\limits_{i=1}^{n} x_i, \prod\limits_{i=1}^{n} (1-x_i) \right)$，

则 $p(x_1, x_2, \cdots, x_n; a, b) = \left[\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]^n t_1^{a-1} t_2^{b-1} \cdot \mathrm{I}_{0<x_1, x_2, \cdots, x_n<1}$，

取 $g(t_1, t_2; a, b) = \left[\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right]^n t_1^{a-1} t_2^{b-1}$， $h(x_1, x_2, \cdots, x_n) = \mathrm{I}_{0<x_1, x_2, \cdots, x_n<1}$ 与参数 a, b 无关，

故根据因子分解定理知 $(T_1, T_2) = \left(\prod_{i=1}^n X_i, \prod_{i=1}^n (1-X_i)\right)$ 是 a, b 的充分统计量.

15．若 $X = (X_1, \cdots, X_n)$ 为从分布族 $f(x;\theta) = C(\theta)\exp\left\{\sum_{i=1}^k Q_i(\theta) T_i(x)\right\} h(x)$ 中抽取的简单样本，试证

$$T(X) = \left(\sum_{j=1}^n T_1(X_j), \cdots, \sum_{j=1}^n T_k(X_j)\right)$$

为充分统计量.

证：样本联合密度函数为

$$p(x_1, x_2, \cdots, x_n; \theta) = \prod_{j=1}^n C(\theta)\exp\left\{\sum_{i=1}^k Q_i(\theta) T_i(x_j)\right\} h(x_j)$$

$$= C(\theta)^n \exp\left\{\sum_{j=1}^n \sum_{i=1}^k Q_i(\theta) T_i(x_j)\right\} \cdot \prod_{j=1}^n h(x_j) = C(\theta)^n \exp\left\{\sum_{i=1}^k Q_i(\theta) \sum_{j=1}^n T_i(x_j)\right\} \cdot \prod_{j=1}^n h(x_j),$$

因 $T(X) = \left(\sum_{j=1}^n T_1(X_j), \cdots, \sum_{j=1}^n T_k(X_j)\right)$，有 $T(x) = (t_1, \cdots, t_k) = \left(\sum_{j=1}^n T_1(x_j), \cdots, \sum_{j=1}^n T_k(x_j)\right)$，

则 $p(x_1, x_2, \cdots, x_n; \theta) = C(\theta)^n \exp\left\{\sum_{i=1}^k Q_i(\theta) t_i\right\} \cdot \prod_{j=1}^n h(x_j)$，

取 $g(T(x);\theta) = C(\theta)^n \exp\left\{\sum_{i=1}^k Q_i(\theta) t_i\right\}$， $h(x_1, x_2, \cdots, x_n) = \prod_{j=1}^n h(x_j)$ 与参数 θ 无关，

故 $T(X) = \left(\sum_{j=1}^n T_1(X_j), \cdots, \sum_{j=1}^n T_k(X_j)\right)$ 为参数 θ 的充分统计量.

16．设 $X_1, \cdots, X_n$ 是来自正态总体 $N(\mu, \sigma_1^2)$ 的样本，$Y_1, \cdots, Y_m$ 是来自另一正态总体 $N(\mu, \sigma_2^2)$ 的样本，这两个样本相互独立，试给出 $(\mu, \sigma_1^2, \sigma_2^2)$ 的充分统计量.

解：两个总体的密度函数分别为

$$p_X(x; \mu, \sigma_1^2) = \frac{1}{\sqrt{2\pi}\sigma_1} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma_1^2}}, \quad p_Y(y; \mu, \sigma_2^2) = \frac{1}{\sqrt{2\pi}\sigma_2} \mathrm{e}^{-\frac{(y-\mu)^2}{2\sigma_2^2}},$$

全部样本的联合密度函数为

$$p(x_1,\cdots,x_n,y_1,\cdots,y_m;\mu,\sigma_1^2,\sigma_2^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma_1^2}}\cdot\prod_{j=1}^{m}\frac{1}{\sqrt{2\pi}\sigma_2}\mathrm{e}^{-\frac{(y_j-\mu)^2}{2\sigma_2^2}}$$

$$=\frac{1}{(\sqrt{2\pi})^{n+m}\sigma_1^n\sigma_2^m}\mathrm{e}^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(x_i^2-2\mu x_i+\mu^2)-\frac{1}{2\sigma_2^2}\sum_{j=1}^{m}(y_j^2-2\mu y_j+\mu^2)}$$

$$=\frac{1}{(\sqrt{2\pi})^{n+m}\sigma_1^n\sigma_2^m}\mathrm{e}^{-\frac{1}{2\sigma_1^2}\left(\sum_{i=1}^{n}x_i^2-2\mu\sum_{i=1}^{n}x_i+n\mu^2\right)-\frac{1}{2\sigma_2^2}\left(\sum_{j=1}^{m}y_j^2-2\mu\sum_{j=1}^{m}y_j+m\mu^2\right)},$$

令 $(T_1,T_2,T_3,T_4)=\left(\sum_{i=1}^{n}X_i,\sum_{j=1}^{m}Y_j,\sum_{i=1}^{n}X_i^2,\sum_{j=1}^{m}Y_j^2\right)$，有 $(t_1,t_2,t_3,t_4)=\left(\sum_{i=1}^{n}x_i,\sum_{j=1}^{m}y_j,\sum_{i=1}^{n}x_i^2,\sum_{j=1}^{m}y_j^2\right)$，

则 $p(x_1,\cdots,x_n,y_1,\cdots,y_m;\mu,\sigma_1^2,\sigma_2^2)=\dfrac{1}{(\sqrt{2\pi})^{n+m}\sigma_1^n\sigma_2^m}\mathrm{e}^{-\frac{1}{2\sigma_1^2}\left(t_2-2\mu t_1+n\mu^2\right)-\frac{1}{2\sigma_2^2}\left(t_4-2\mu t_3+m\mu^2\right)}$，

取 $g(t_1,t_2,t_3,t_4;\mu,\sigma_1^2,\sigma_2^2)=\dfrac{1}{(\sqrt{2\pi})^{n+m}\sigma_1^n\sigma_2^m}\mathrm{e}^{-\frac{1}{2\sigma_1^2}\left(t_2-2\mu t_1+n\mu^2\right)-\frac{1}{2\sigma_2^2}\left(t_4-2\mu t_3+m\mu^2\right)}$，

$h(x_1,\cdots,x_n,y_1,\cdots,y_m)=1$ 与参数 $\mu,\sigma_1^2,\sigma_2^2$ 无关，

故 $(T_1,T_2,T_3,T_4)=\left(\sum_{i=1}^{n}X_i,\sum_{j=1}^{m}Y_j,\sum_{i=1}^{n}X_i^2,\sum_{j=1}^{m}Y_j^2\right)$ 是参数 $(\mu,\sigma_1^2,\sigma_2^2)$ 的充分统计量.

17．设 $\begin{pmatrix}X_i\\Y_i\end{pmatrix}$，$i=1,\cdots,n$ 是来自正态分布族

$$\left\{N\left(\begin{pmatrix}\theta_1\\\theta_2\end{pmatrix},\begin{pmatrix}\sigma_1^2&\rho\sigma_1\sigma_2\\\rho\sigma_1\sigma_2&\sigma_2^2\end{pmatrix}\right),\quad -\infty<\theta_1,\theta_2<+\infty,\ \sigma_1,\sigma_2>0,\ |\rho|\le 1\right\}$$

的一个二维样本，寻求$(\mu_1,\sigma_1,\mu_2,\sigma_2,\rho)$的充分统计量.

注：此题有误，应改为寻求$(\theta_1,\sigma_1,\theta_2,\sigma_2,\rho)$的充分统计量.

解：总体密度函数为

$$p(x,y;\theta_1,\sigma_1,\theta_2,\sigma_2,\rho)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\theta_1)^2}{\sigma_1^2}-2\rho\frac{(x-\theta_1)(y-\theta_2)}{\sigma_1\sigma_2}+\frac{(y-\theta_2)^2}{\sigma_2^2}\right]},$$

样本联合密度函数为

$$p(x_1,y_1,\cdots,x_n,y_n;\theta_1,\sigma_1,\theta_2,\sigma_2,\rho)=\prod_{i=1}^{n}\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_i-\theta_1)^2}{\sigma_1^2}-2\rho\frac{(x_i-\theta_1)(y_i-\theta_2)}{\sigma_1\sigma_2}+\frac{(y_i-\theta_2)^2}{\sigma_2^2}\right]}$$

$$=\frac{1}{(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^n}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\frac{1}{\sigma_1^2}\sum_{i=1}^{n}(x_i^2-2\theta_1x_i+\theta_1^2)-\frac{2\rho}{\sigma_1\sigma_2}\sum_{i=1}^{n}(x_iy_i-\theta_2x_i-\theta_1y_i+\theta_1\theta_2)+\frac{1}{\sigma_2^2}\sum_{i=1}^{n}(y_i^2-2\theta_2y_i+\theta_2^2)\right]}$$

$$=\frac{1}{(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^n}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\frac{1}{\sigma_1^2}\left(\sum_{i=1}^{n}x_i^2-2\theta_1\sum_{i=1}^{n}x_i+n\theta_1^2\right)-\frac{2\rho}{\sigma_1\sigma_2}\left(\sum_{i=1}^{n}x_iy_i-\theta_2\sum_{i=1}^{n}x_i-\theta_1\sum_{i=1}^{n}y_i+n\theta_1\theta_2\right)+\frac{1}{\sigma_2^2}\left(\sum_{i=1}^{n}y_i^2-2\theta_2\sum_{i=1}^{n}y_i+n\theta_2^2\right)\right]},$$

令 $(T_1,T_2,T_3,T_4,T_5)=\left(\sum\limits_{i=1}^n X_i,\sum\limits_{i=1}^n Y_i,\sum\limits_{i=1}^n X_i^2,\sum\limits_{i=1}^n Y_i^2,\sum\limits_{i=1}^n X_iY_i\right)$,

有 $(t_1,t_2,t_3,t_4,t_5)=\left(\sum\limits_{i=1}^n x_i,\sum\limits_{i=1}^n y_i,\sum\limits_{i=1}^n x_i^2,\sum\limits_{i=1}^n y_i^2,\sum\limits_{i=1}^n x_iy_i\right)$,

则 $p(x_1, y_1, \cdots, x_n, y_n;\ \theta_1, \sigma_1, \theta_2, \sigma_2, \rho)$

$$=\frac{1}{(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^n}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\frac{1}{\sigma_1^2}(t_3-2\theta_1t_1+n\theta_1^2)-\frac{2\rho}{\sigma_1\sigma_2}(t_5-\theta_2t_1-\theta_1t_2+n\theta_1\theta_2)+\frac{1}{\sigma_2^2}(t_4-2\theta_2t_2+n\theta_2^2)\right]},$$

取 $g(t_1, t_2, t_3, t_4, t_5;\ \theta_1, \sigma_1, \theta_2, \sigma_2, \rho)$

$$=\frac{1}{(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^n}\mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\frac{1}{\sigma_1^2}(t_3-2\theta_1t_1+n\theta_1^2)-\frac{2\rho}{\sigma_1\sigma_2}(t_5-\theta_2t_1-\theta_1t_2+n\theta_1\theta_2)+\frac{1}{\sigma_2^2}(t_4-2\theta_2t_2+n\theta_2^2)\right]},$$

$h(x_1, y_1, \cdots, x_n, y_n) = 1$ 与参数 $\theta_1, \sigma_1, \theta_2, \sigma_2, \rho$ 无关，

故 $(T_1,T_2,T_3,T_4,T_5)=\left(\sum\limits_{i=1}^n X_i,\sum\limits_{i=1}^n Y_i,\sum\limits_{i=1}^n X_i^2,\sum\limits_{i=1}^n Y_i^2,\sum\limits_{i=1}^n X_iY_i\right)$是参数$(\theta_1, \sigma_1, \theta_2, \sigma_2, \rho)$的充分统计量.

18．设二维随机变量 $X=\begin{pmatrix}X_1\\X_2\end{pmatrix}$ 服从二元正态分布，其均值向量为零向量，协方差阵为

$$\begin{pmatrix}\sigma^2+r^2 & \sigma^2-r^2\\ \sigma^2-r^2 & \sigma^2+r^2\end{pmatrix},\quad \sigma>0, r>0\ .$$

证明：二维统计量 $T = ((X_1 + X_2)^2, (X_1 - X_2)^2)$是该二元正态分布族的充分统计量.

注：此题有误，应改为 $T=\left(\sum\limits_{i=1}^n(X_{1i}+X_{2i})^2,\sum\limits_{i=1}^n(X_{1i}-X_{2i})^2\right)$.

证：因二元正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 的均值向量为 $\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}$，协方差阵为 $\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$,

则$\mu_1 = \mu_2 = 0$，$\sigma_1^2=\sigma_2^2=\sigma^2+r^2$，$\rho\sigma_1\sigma_2 = \sigma^2 - r^2$，有 $\rho=\dfrac{\sigma^2-r^2}{\sigma^2+r^2}$，$1-\rho^2=\dfrac{4\sigma^2r^2}{(\sigma^2+r^2)^2}$,

可得

$$-\frac{1}{2(1-\rho)^2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]$$

$$=-\frac{1}{2}\frac{(\sigma^2+r^2)^2}{4\sigma^2r^2}\left(\frac{x_1^2}{\sigma^2+r^2}-2\frac{\sigma^2-r^2}{\sigma^2+r^2}\cdot\frac{x_1x_2}{\sigma^2+r^2}+\frac{x_2^2}{\sigma^2+r^2}\right)$$

$$=-\frac{1}{8\sigma^2r^2}[(\sigma^2+r^2)x_1^2-2(\sigma^2-r^2)x_1x_2+(\sigma^2+r^2)x_2^2]$$

$$=-\frac{1}{8\sigma^2r^2}[\sigma^2(x_1-x_2)^2+r^2(x_1+x_2)^2],$$

即总体密度函数为

$$p(x_1,x_2;\sigma,r)=\frac{1}{4\pi\sigma r}\mathrm{e}^{-\frac{1}{8\sigma^2r^2}[\sigma^2(x_1-x_2)^2+r^2(x_1+x_2)^2]},$$

样本联合密度函数为

$$p(x_{11}, x_{21}, \cdots, x_{1n}, x_{2n}; \sigma, r) = \prod_{i=1}^{n} \frac{1}{4\pi\sigma r} e^{-\frac{1}{8\sigma^2 r^2}[\sigma^2 (x_{1i}-x_{2i})^2 + r^2 (x_{1i}+x_{2i})^2]}$$

$$= \frac{1}{(4\pi\sigma r)^n} e^{-\frac{1}{8\sigma^2 r^2}\left[\sigma^2 \sum_{i=1}^{n}(x_{1i}-x_{2i})^2 + r^2 \sum_{i=1}^{n}(x_{1i}+x_{2i})^2\right]},$$

令 $T = \left(\sum_{i=1}^{n}(X_{1i}+X_{2i})^2, \sum_{i=1}^{n}(X_{1i}-X_{2i})^2\right)$，有 $t = (t_1, t_2) = \left(\sum_{i=1}^{n}(x_{1i}+x_{2i})^2, \sum_{i=1}^{n}(x_{1i}-x_{2i})^2\right)$，

则 $p(x_{11}, x_{21}, \cdots, x_{1n}, x_{2n}; \sigma, r) = \frac{1}{(4\pi\sigma r)^n} e^{-\frac{1}{8\sigma^2 r^2}(\sigma^2 t_2 + r^2 t_1)}$，

取 $g(t_1, t_2; \sigma, r) = \frac{1}{(4\pi\sigma r)^n} e^{-\frac{1}{8\sigma^2 r^2}(\sigma^2 t_2 + r^2 t_1)}$，$h(x_{11}, x_{21}, \cdots, x_{1n}, x_{2n}) = 1$ 与参数σ, r 无关，

故$T = \left(\sum_{i=1}^{n}(X_{1i}+X_{2i})^2, \sum_{i=1}^{n}(X_{1i}-X_{2i})^2\right)$是参数$(\sigma, r)$的充分统计量.

19．设 $X_1, \cdots, X_n$ 是来自两参数指数分布 $p(x;\theta,\mu) = \frac{1}{\theta} e^{-(x-\mu)/\theta}$，$x > \mu$，$\theta > 0$ 的样本，证明 $(\bar{x}, x_{(1)})$ 是充分统计量.

解：样本联合密度函数

$$p(x_1, x_2, \cdots, x_n; \theta, \mu) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{x_i-\mu}{\theta}} \mathrm{I}_{x_i>\mu} = \frac{1}{\theta^n} e^{-\frac{\sum_{i=1}^{n} x_i - n\mu}{\theta}} \mathrm{I}_{x_1, x_2, \cdots, x_n > \mu} = \frac{1}{\theta^n} e^{-\frac{n\bar{x}-n\mu}{\theta}} \mathrm{I}_{x_{(1)}>\mu},$$

令 $(T_1, T_2) = (\bar{X}, X_{(1)})$，有 $(t_1, t_2) = (\bar{x}, x_{(1)})$，即 $p(x_1, x_2, \cdots, x_n; \theta, \mu) = \frac{1}{\theta^n} e^{-\frac{nt_1-n\mu}{\theta}} \mathrm{I}_{t_2>\mu}$，

取 $g(t_1, t_2; \theta, \mu) = \frac{1}{\theta^n} e^{-\frac{nt_1-n\mu}{\theta}} I_{t_2>\mu}$，$h(x_1, x_2, \cdots, x_n) = 1$ 与参数θ, μ 无关，

故根据因子分解定理知 $(T_1, T_2) = (\bar{X}, X_{(1)})$ 是参数(θ, μ)的充分统计量.

20．设 $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$，$i = 1, \cdots, n$，诸 Y_i 独立，$x_1, \cdots, x_n$ 是已知常数，证明$\left(\sum_{i=1}^{n} Y_i, \sum_{i=1}^{n} x_i Y_i, \sum_{i=1}^{n} Y_i^2\right)$是充分统计量.

解：联合密度函数

$$p(y_1, y_2, \cdots, y_n; \beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^2}} = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2}$$

$$= \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} y_i^2 - 2\beta_0 \sum_{i=1}^{n} y_i - 2\beta_1 \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n}(\beta_0+\beta_1 x_i)^2\right]},$$

令 $(T_1, T_2, T_3) = (\sum_{i=1}^{n} Y_i, \sum_{i=1}^{n} x_i Y_i, \sum_{i=1}^{n} Y_i^2)$，有 $(t_1, t_2, t_3) = (\sum_{i=1}^{n} y_i, \sum_{i=1}^{n} x_i y_i, \sum_{i=1}^{n} y_i^2)$，

则 $p(y_1, y_2, \cdots, y_n; \beta_0, \beta_1, \sigma^2) = \dfrac{1}{(\sqrt{2\pi}\sigma)^n} \mathrm{e}^{-\frac{1}{2\sigma^2}\left[t_3 - 2\beta_0 t_1 - 2\beta_1 t_2 + \sum_{i=1}^{n}(\beta_0 + \beta_1 x_i)^2\right]}$,

取 $g(T_1, T_2, T_3; \beta_0, \beta_1, \sigma^2) = \dfrac{1}{(\sqrt{2\pi}\sigma)^n} \mathrm{e}^{-\frac{1}{2\sigma^2}\left[t_3 - 2\beta_0 t_1 - 2\beta_1 t_2 + \sum_{i=1}^{n}(\beta_0 + \beta_1 x_i)^2\right]}$,

$h(y_1, y_2, \cdots, y_n) = 1$ 与参数 β_0, β_1, σ^2 无关,

故根据因子分解定理知 $(T_1, T_2, T_3) = (\sum_{i=1}^{n} Y_i, \sum_{i=1}^{n} x_i Y_i, \sum_{i=1}^{n} Y_i^2)$ 是参数$(\beta_0, \beta_1, \sigma^2)$的充分统计量.