习题5.1

补充题：设总体 X 的概率分布为

$$P\{X=0\}=p_0，P\{X=1\}=p_1，P\{X=2\}=p_2,$$

其中 $p_i>0$ 且 $p_0+p_1+p_2=1$。求样本 $X_1, X_2, \ldots, X_n$ 的联合质量函数

进一步，设总体 X 的概率分布为

$$P\{X=a_1\}=p_1，P\{X=a_2\}=p_2，\ldots，P\{X=a_m\}=p_m$$

其中 $a_1, a_2, \ldots, a_m$ 互不相同，$p_i>0$ 且 $p_1+p_2+\ldots+p_m=1$ 。

样本 $X_1, X_2, \ldots, X_n$ 的联合质量函数又如何呢？请都用两种方法给出结论。

解：方法一：三点分布

$$p(x_1, x_2, \cdots, x_n) = p_0^{n_0} p_1^{n_1} p_2^{n_2}$$

其中 n_0, n_1, n_2 分别是样本 $X_1, X_2, \ldots, X_n$ 中取 0, 1, 2 的样品个数。

多点分布

$$p(x_1, x_2, \cdots, x_n) = p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}$$

其中 $n_1, n_2, \ldots, n_m$ 分别是 $X_1, X_2, \ldots, X_n$ 中取 $a_1, a_2, \ldots, a_m$的样品个数

方法二：三点分布：总体 X 的质量函数为

$$p(x)=p_1^{?}\cdot p_2^{?}\cdot p_3^{?}$$

$x=0$	1	0	0
$x=1$	0	1	0
$x=2$	0	0	1
	$\frac{1}{2}(x-1)(x-2)$	$-x(x-2)$	$\frac{1}{2}x(x-1)$

$$p(x)=p_1^{\frac{1}{2}(x-1)(x-2)}\cdot p_2^{-x(x-2)}\cdot p_3^{\frac{1}{2}x(x-1)}$$

样本联合质量函数为 $p(x_1,x_2,\cdots,x_n)=p_1^{\frac{1}{2}\sum_{i=1}^{n}(x_i-1)(x_i-2)}\cdot p_2^{-\sum_{i=1}^{n}x_i(x_i-2)}\cdot p_3^{\frac{1}{2}\sum_{i=1}^{n}x_i(x_i-1)}$

多点分布：总体 X 的质量函数为

$$p(x) = p_1^{?} \cdot p_2^{?} \cdot \mathrm{L} \cdot p_m^{?}$$

$x = a_1$	1	0	……	0
$x = a_2$	0	1	……	0
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$x = a_m$	0	0	……	1

$$\frac{(x-a_2)\dots(x-a_m)}{(a_1-a_2)\dots(a_1-a_m)}$$

一般地，$p(x)$ 中 p_j 的次数为 $$\frac{(x-a_1)\dots(x-a_{j-1})(x-a_{j+1})\dots(x-a_m)}{(a_j-a_1)\dots(a_j-a_{j-1})(a_j-a_{j+1})\dots(a_j-a_m)}$$

p_j 的次数 $\dfrac{(x-a_1)\cdots(x-a_{j-1})(x-a_{j+1})\cdots(x-a_m)}{(a_j-a_1)\cdots(a_j-a_{j-1})(a_j-a_{j+1})\cdots(a_j-a_m)}$

$$=\prod_{\substack{k=1\\k\neq j}}^{m}\frac{x-a_k}{a_j-a_k} \qquad j=1,2,\ldots,m$$

这称为拉格朗日插值函数。总体 X 的质量函数为

$$p(x)=p_1^{\prod\limits_{\substack{k=1\\k\neq 1}}^{m}\frac{x-a_k}{a_1-a_k}}\,p_2^{\prod\limits_{\substack{k=1\\k\neq 2}}^{m}\frac{x-a_k}{a_2-a_k}}\cdots\,p_m^{\prod\limits_{\substack{k=1\\k\neq m}}^{m}\frac{x-a_k}{a_m-a_k}}=\prod_{j=1}^{m}p_j^{\prod\limits_{\substack{k=1\\k\neq j}}^{m}\frac{x-a_k}{a_m-a_k}}$$

样本联合质量函数为

$$p(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}p(x_i)=\prod_{j=1}^{m}p_j^{\sum\limits_{i=1}^{n}\prod\limits_{\substack{k=1\\k\neq j}}^{m}\frac{x_i-a_k}{a_m-a_k}}$$

习题5.3

第9题 设总体二阶矩存在，$X_1, X_2, \ldots, X_n$ 是样本，证明 $X_i - \bar{X}$ 与 $X_j - \bar{X}$ $(i \neq j)$ 的相关系数为 $-(n-1)^{-1}$

思路：计算相关系数需要计算它们的协方差以及各自的方差，
利用协方差与方差的性质展开后，会涉及计算

$$\mathrm{Cov}(X_i, X_j), \quad \mathrm{Cov}(X_i, \bar{X}), \quad \mathrm{Cov}(\bar{X}, \bar{X})$$

先计算这几个式子，再计算题中两项的协方差与方差，最后就得到相关系数。

证明：记$\mathrm{Var}(X)=\sigma^2$，可知 $X_1, X_2, \ldots, X_n$ 相互独立，且 $\mathrm{Var}(X_i)=\sigma^2$，

则 $\mathrm{Cov}(X_i, X_j)=0,\ (i \neq j)$

$$\mathrm{Cov}(X_i, \bar{X})=\mathrm{Cov}\left(X_i, \frac{1}{n}(\underline{X_1+\cdots}+X_i\underline{+\cdots+X_n})\right)=\mathrm{Cov}\left(X_i, \frac{1}{n}X_i\right)=\frac{\sigma^2}{n}$$

$$\mathrm{Cov}(\bar{X}, \bar{X})=\frac{\sigma^2}{n}$$

按照协方差、方差的性质计算、化简，则

$$\mathrm{Cov}(X_i-\bar{X}, X_j-\bar{X})=\mathrm{Cov}(X_i, X_j)-\mathrm{Cov}(X_i, \bar{X})-\mathrm{Cov}(\bar{X}, X_j)+\mathrm{Cov}(\bar{X}, \bar{X})=-\frac{\sigma^2}{n}$$

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

故相关系数

$$\mathrm{Corr}(X_i-\bar{X},X_j-\bar{X})=\frac{\mathrm{Cov}(X_i-\bar{X},X_j-\bar{X})}{\sqrt{\mathrm{Var}(X_i-\bar{X})}\sqrt{\mathrm{Var}(X_j-\bar{X})}}$$

$$=\frac{-\frac{1}{n}\sigma^2}{\sqrt{\frac{n-1}{n}\sigma^2}\sqrt{\frac{n-1}{n}\sigma^2}}=-\frac{1}{n-1}$$

第10题 设 $X_1, X_2, \ldots, X_n$为一个样本，$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 是样本方差，试证：

$$\frac{1}{n(n-1)}\sum_{i<j}(X_i - X_j)^2 = S^2$$

思路：这里的关键是求和号为二重求和，对两个变量 i,j 同时做要求，这就相当于二重积分是对两个变量 x,y 同时做要求

对于二重积分我们知道需要化为二次积分计算，这里的二重求和也需要化为二次求和，即对 i,j 分开分别求和。因

$$\sum_{i<j}(X_i - X_j)^2 = \sum_{i>j}(X_i - X_j)^2 \qquad \sum_{i=j}(X_i - X_j)^2 = 0$$

有

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - X_j)^2 = \sum_{i<j}(X_i - X_j)^2 + \sum_{i>j}(X_i - X_j)^2 + \sum_{i=j}(X_i - X_j)^2 = 2\sum_{i<j}(X_i - X_j)^2$$

证明：因 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right)$

则 $$\sum_{i<j}(X_i-X_j)^2=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-X_j)^2=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i^2+X_j^2-2X_iX_j)$$

$$=\frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}X_i^2+\sum_{i=1}^{n}\sum_{j=1}^{n}X_j^2-2\sum_{i=1}^{n}\sum_{j=1}^{n}X_iX_j\right)$$

$$=\frac{1}{2}\left(2\sum_{i=1}^{n}\sum_{j=1}^{n}X_i^2-2\sum_{i=1}^{n}X_i\sum_{j=1}^{n}X_j\right)$$

$$=\sum_{i=1}^{n}nX_i^2-n\bar{X}\cdot n\bar{X}=n\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right)=n(n-1)S^2$$

故 $$\frac{1}{n(n-1)}\sum_{i<j}(X_i-X_j)^2=S^2$$

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

思路：利用最大与最小顺序统计量的分布函数

$$F_n(x)=P\{X_{(n)}\le x\}=P\{X_1\le x,X_2\le x,\dots,X_n\le x\}=\boldsymbol{[F(x)]^n}$$

$$\begin{aligned}F_1(x)&=P\{X_{(1)}\le x\}=1-P\{X_{(1)}>x\}\\&=1-P\{X_1>x,X_2>x,\dots,X_n>x\}=\boldsymbol{1-[1-F(x)]^n}\end{aligned}$$

先根据总体 X 的概率分布列通过求和得到 X 的分布函数 $F(x)$，求得 $X_{(n)},X_{(1)}$ 的分布函数 $F_n(x),F_1(x)$

再通过求差得到 $X_{(n)},X_{(1)}$ 的概率分布列。

解：因 X 的分布列为 $P\{X=k\}=pq^{k-1},\ k=1,2,\dots$

则分布函数值为

$$F(k)=P\{X\le k\}=\sum_{j=1}^{k}pq^{j-1}=\frac{p(1-q^k)}{1-q}=1-q^k,\quad k=1,2,\dots$$

得到 $X_{(n)},X_{(1)}$ 的分布函数值

$$F_n(k)=P\{X_{(n)}\le k\}=F(k)^n=(1-q^k)^n,\ k=1,2,\dots$$

$$F_1(k)=P\{X_{(1)}\le k\}=1-[1-F(k)]^n=1-(q^k)^n=1-q^{nk},k=1,2,\dots$$

故 $X_{(n)},X_{(1)}$ 的概率分布列为

$$P\{X_{(n)}=k\}=P\{X_{(n)}\le k\}-P\{X_{(n)}\le k-1\}=(1-q^k)^n-(1-q^{k-1})^n,\ k=1,2,\dots$$

$$P\{X_{(1)}=k\}=P\{X_{(1)}\le k\}-P\{X_{(1)}\le k-1\}=q^{n(k-1)}-q^{nk},\ k=1,2,\dots$$

第24题　设 $X_1, X_2, \ldots, X_{16}$ 是来自 $N(8, 4)$ 的样本，试求下列概率。

（1）$P\{X_{(16)} > 10\}$

（2）$P\{X_{(1)} > 5\}$

思路：利用最大与最小顺序统计量的分布函数

$$F_n(x) = [F(x)]^n,\quad F_1(x) = 1 - [1 - F(x)]^n$$

而总体 X 服从正态分布，其分布函数 $F(x)$ 可标准化化为标准正态分布函数

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

再查附表2得到相应概率

解：$X_1, X_2, \ldots, X_{16}$ 是来自 $N(8, 4)$ 的样本，$\mu = 8$，$\sigma = 2$, $n = 16$

（1）$X_{(16)}$ 为最大顺序统计量，

$$P\{X_{(16)} > 10\} = 1 - F_{16}(10) = 1 - F(10)^{16} = 1 - \Phi\left(\frac{10-8}{2}\right)^{16}$$

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

（2）$X_{(1)}$ 为最小顺序统计量

$$P\{X_{(1)} > 5\} = 1 - F_1(5) = [1 - F(5)]^{16} = \left[1 - \Phi\left(\frac{5-8}{2}\right)\right]^{16}$$

$$= [1 - \Phi(-1.5)]^{16} = 0.9332^{16} = 0.3308$$

第33题 （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。

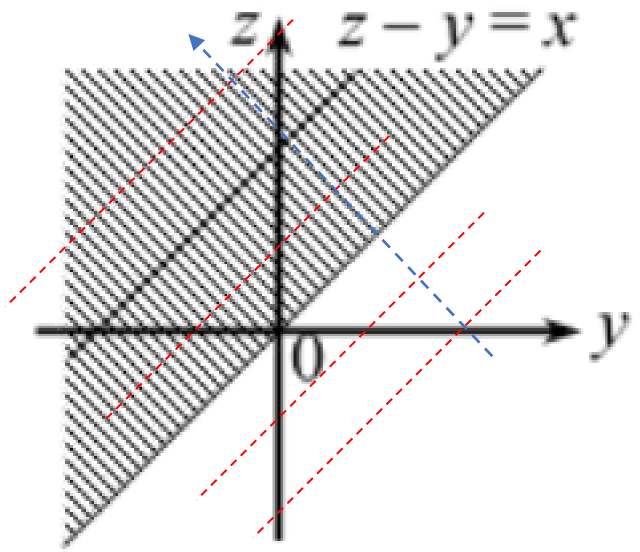
注：第（1）问应添上 $x>0$ 的要求.

思路：先根据两个顺序统计量的联合密度写出 $(X_{(1)}, X_{(n)})$ 的联合密度函数 $p_{1n}(y,z)$，再根据概率论第三章二维随机变量函数的分布，求得极差 $R_n = X_{(n)} - X_{(1)}$ 的分布函数。

$$p_{1n}(y,z) = n(n-1)[F(z) - F(y)]^{n-2} p(y)p(z)I_{y\le z}$$

解：（1）在 yOz 平面上作曲线族 $z-y=x$，（x 从$-\infty$ 到 $+\infty$）

分布函数法得 $R_n=X_{(n)}-X_{(1)}$ 的分布函数



当 $x>0$ 时，

$$F_{R_n}(x)=P\{R_n=X_{(n)}-X_{(1)}\le x\}$$

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

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

$$=\int_{-\infty}^{+\infty}dy\cdot np(y)\cdot[F(z)-F(y)]^{n-1}\Big|_{y}^{y+x}$$

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

（2）因指数分布 $Exp(\lambda)$ 的密度函数与分布函数分别为

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

故 $R_n=X_{(n)}-X_{(1)}$ 的分布函数当 $x>0$ 时为

$$\begin{aligned}F_{R_n}(x)&=n\int_{-\infty}^{+\infty}[F(y+x)-F(y)]^{n-1}p(y)dy\\&=n\int_{0}^{+\infty}[(1-e^{-\lambda(y+x)})-(1-e^{-\lambda y})]^{n-1}\cdot\lambda e^{-\lambda y}dy\\&=n\int_{0}^{+\infty}(e^{-\lambda y})^{n-1}(1-e^{-\lambda x})^{n-1}\cdot(-1)d(e^{-\lambda y})\\&=n(1-e^{-\lambda x})^{n-1}\cdot\left(-\frac{1}{n}\right)(e^{-\lambda y})^{n}\Big|_{0}^{+\infty}\\&=(1-e^{-\lambda x})^{n-1},\quad x>0\end{aligned}$$

习题5.4

第9题　设 X_1, X_2 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。

思路：分子分母都是正态变量平方，即 χ^2 变量之商，考虑用 F 分布

因 X_1, X_2 都服从 $N(0, \sigma^2)$ 且相互独立，从而 X_1, X_2 服从二维正态分布，那么 $(X_1 + X_2, X_1 - X_2)$ 也服从二维正态分布。

证明 $X_1 + X_2$ 与 $X_1 - X_2$ 协方差等于 0，二者不相关从而独立

再分别进行标准化后求商。

证明：因 X_1, X_2 都服从 $N(0, \sigma^2)$ 且相互独立，从而 X_1, X_2 服从二维正态分布，那么 (X_1+X_2, X_1-X_2) 也服从二维正态分布。因

$$\begin{aligned}\mathrm{Cov}(X_1+X_2, X_1-X_2) &= \mathrm{Var}(X_1) - \mathrm{Cov}(X_1, X_2) + \mathrm{Cov}(X_2, X_1) - \mathrm{Var}(X_2)\\ &= 0\end{aligned}$$

则 X_1+X_2 与 X_1-X_2 不相关从而独立，且

$$X_1+X_2 \sim N(0, 2\sigma^2)，X_1-X_2 \sim N(0, 2\sigma^2)$$

即

$$\frac{X_1+X_2}{\sqrt{2}\sigma} \sim N(0,1) \qquad \frac{X_1-X_2}{\sqrt{2}\sigma} \sim N(0,1)$$

故

$$\frac{\left(\dfrac{X_1+X_2}{\sqrt{2}\sigma}\right)^2}{\left(\dfrac{X_1-X_2}{\sqrt{2}\sigma}\right)^2} = \left(\frac{X_1+X_2}{X_1-X_2}\right)^2 \sim F(1,1)$$

第10题 设总体为 $N(0, 1)$，X_1, X_2 为样本，试求常数 k，使得

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

思路：分子分母都是 χ^2 变量，即 χ^2 变量之商，考虑用 F 分布。

但分子分母不独立，不能用 $F(1, 2)$。

要使得分子分母独立，可利用第9题的结论，构造一个服从 $F(1, 1)$ 的 F 变量，再利用它们的函数关系得出结论。

解：根据第9题的结论可知

$$F=\left(\frac{X_1+X_2}{X_1-X_2}\right)^2 \sim F(1,1) \qquad \frac{1}{F}\sim F(1,1)$$

则

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

$$=P\left\{\frac{(X_1-X_2)^2}{(X_1+X_2)^2}<\frac{1}{k}-1\right\}=P\left\{\frac{1}{F}<\frac{1}{k}-1\right\}=0.05$$

故 $$\frac{1}{k}-1=f_{0.05}(1,1)=\frac{1}{f_{0.95}(1,1)}=\frac{1}{161.45}$$

可得 $$k=\frac{161.45}{162.45}=0.9938$$

第12题 设 $X_1, X_2, \ldots, X_n, X_{n+1}$ 是来自 $N(\mu, \sigma^2)$ 的样本，

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

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

思路：分子为正态变量，分母为 χ 变量，并且分子分母独立，标准化后，根据 t 分布构成计算常数 c。

解： 因 $X_{n+1} \sim N(\mu, \sigma^2)$ ，$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$，且相互独立，有

$$X_{n+1} - \bar{X}_n \sim N\left(0, \frac{n+1}{n}\sigma^2\right)$$

标准化

$$U = \frac{X_{n+1} - \bar{X}_n}{\sqrt{\frac{n+1}{n}}\sigma} \sim N(0, 1)$$

且

$$\chi^2 = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$$

因 S_n^2 与 $X_{n+1}-\bar{X}_n$ 独立，可知 U 与 χ^2 独立

从而根据 t 分布的构成可得

$$T=\frac{U}{\sqrt{\frac{\chi^2}{n-1}}}=\frac{\frac{X_{n+1}-\bar{X}_n}{\sigma\sqrt{\frac{n+1}{n}}}}{\sqrt{\frac{S_n^2}{\sigma^2}}}=\sqrt{\frac{n}{n+1}}\cdot\frac{X_{n+1}-\bar{X}_n}{S_n}\sim t(n-1)$$

故 $c=\sqrt{\frac{n}{n+1}}$

自由度为 $n-1$

第20题 设 $X_1, X_2, \ldots, X_n$ 是来自正态分布 $N(\mu, \sigma^2)$ 的一个样本，

$S_n^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$是样本方差，试求满足 $P\left\{\dfrac{S_n^2}{\sigma^2} \le 1.5\right\} \ge 0.95$

的最小 n 值。

解：因 $\chi^2 = \dfrac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$，有

$$P\left\{\frac{S_n^2}{\sigma^2} \le 1.5\right\} = P\left\{\chi^2 = \frac{(n-1)S_n^2}{\sigma^2} \le 1.5(n-1)\right\} \ge 0.95$$

则

$$1.5(n-1) \ge \chi^2_{0.95}(n-1) \qquad \frac{\chi^2_{0.95}(n-1)}{n-1} \le 1.5$$

k = n-1	20	21	22	23	24	25	**26**	27	28	29	30
0.95分位数	31.41	32.671	33.924	35.172	36.415	37.652	**38.89**	40.113	41.337	42.557	43.773
比值	1.5705	1.5557	1.5420	1.5292	1.5173	1.5061	**1.4956**	1.4857	1.4763	1.4675	1.4591

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

故 $n-1=26$，即 n 至少为 27。

第21题 设 $X_1, X_2, \ldots, X_n$ 独立同分布服从 $N(\mu, \sigma^2)$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

记 $\xi = \dfrac{X_1 - \bar{X}}{S}$ 。试找出 ξ 与 t 分布的联系，因而定出 ξ 的密度函数

思路：分子正态变量，分母 χ 变量，用 t 分布。但分子分母不独立

对于分母 S，利用定理2的证明方法二，将二次型 $f = \sum_{i=1}^{n}(X_i - \bar{X})^2$ 通过正交变换 $\vec{X} = C\vec{Y}$ 化为标准形 $Y_2^2 + Y_3^2 + \mathrm{L} \ + Y_n^2$

此方法中，只给定了 Y_1 的具体表达式，其它的 Y 变量并没有给定

本题中考虑将分子作为 Y_2，再仿照 t 分布那里的一个例题处理

解：因 $\bar{X}=\frac{1}{n}(X_1+X_2+\cdots+X_n)=\frac{1}{n}(1,1,\cdots,1)\vec{X}$

$$X_1-\bar{X}=X_1-\frac{1}{n}(X_1+X_2+\cdots+X_n)=\frac{1}{n}(n-1,-1,\cdots,-1)\vec{X}$$

分别取 α_1 与 α_2 为上面得到的两个 n 维向量（写成列向量）单位化之后的结果，即

$$\alpha_1=\frac{1}{\sqrt{n}}\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}\qquad\qquad \alpha_2=\frac{1}{\sqrt{n(n-1)}}\begin{pmatrix}n-1\\-1\\\vdots\\-1\end{pmatrix}$$

可得 $\sqrt{n}\bar{X}=\frac{1}{\sqrt{n}}(1,1,\cdots,1)\vec{X}$

$$=\alpha_1^T\vec{X}$$

$$\sqrt{\frac{n}{n-1}}(X_1-\bar{X})=\frac{1}{\sqrt{n(n-1)}}(n-1,-1,\cdots,-1)\vec{X}$$

$$=\alpha_2^T\vec{X}$$

注意到取 α_1 与 α_2 都是单位向量且正交，将 α_1, α_2 扩充为 R^n 中一组标准正交基 $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n$

再令 $C=(\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n)$，$C$ 为正交矩阵，作正交变换 $\vec{X}=C\vec{Y}$

有

$$\vec{Y}=\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{pmatrix}=C^T\vec{X}=\begin{pmatrix} \alpha_1^T \\ \alpha_2^T \\ \alpha_3^T \\ \vdots \\ \alpha_n^T \end{pmatrix}\vec{X}=\begin{pmatrix} \alpha_1^T\vec{X} \\ \alpha_2^T\vec{X} \\ \alpha_3^T\vec{X} \\ \vdots \\ \alpha_n^T\vec{X} \end{pmatrix}$$

可得

$$Y_i=\alpha_i^T\vec{X}, \quad i=1,2,3,\cdots,n$$

特别是 $Y_1 = \alpha_1^T \vec{X} = \sqrt{n}\bar{X}$ $\qquad Y_2 = \alpha_2^T \vec{X} = \sqrt{\dfrac{n}{n-1}}(X_1 - \bar{X})$

可化二次型为标准形

$$(n-1)S^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = Y_2^2 + Y_3^2 + \cdots + Y_n^2$$

由引理知 $Y_1, Y_2, Y_3, \ldots, Y_n$ 相互独立都服从方差同为 σ^2 的正态分布

再求期望，因 $Y_i = \alpha_i^T \vec{X},\quad i = 1, 2, 3, \cdots, n$

可得

$$E(Y_i) = \alpha_i^T E(\vec{X}) = \alpha_i^T \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix} = \alpha_i^T \cdot \mu\sqrt{n}\alpha_1 = \mu\sqrt{n} \cdot \alpha_i^T \alpha_1$$

即 $E(Y_1)=\mu\sqrt{n}$，$E(Y_2)=E(Y_3)=\ldots=E(Y_n)=0$

则$Y_1, Y_2, Y_3, \ldots, Y_n$相互独立，且$Y_1\sim N(\mu\sqrt{n},\sigma^2)$，$Y_2, Y_3, \ldots, Y_n\sim N(0, \sigma^2)$

因

$$Y_2=\sqrt{\frac{n}{n-1}}(X_1-\bar{X}) \qquad (n-1)S^2=Y_2^2+Y_3^2+\cdots+Y_n^2$$

可得

$$\xi=\frac{X_1-\bar{X}}{S}=\frac{\sqrt{\frac{n-1}{n}}Y_2}{\sqrt{\frac{1}{n-1}(Y_2^2+Y_3^2+\cdots+Y_n^2)}}$$

$$=\frac{n-1}{\sqrt{n}}\cdot\frac{Y_2}{\sqrt{Y_2^2+Y_3^2+\cdots+Y_n^2}}$$

因 $\dfrac{Y_2}{\sigma} \sim N(0,1)$ $\quad \left(\dfrac{Y_3}{\sigma}\right)^2 + \cdots + \left(\dfrac{Y_n}{\sigma}\right)^2 = \dfrac{Y_3^2 + \cdots + Y_n^2}{\sigma^2} \sim \chi^2(n-2)$

且相互独立，根据 t 分布的构成可知

$$T = \frac{\dfrac{Y_2}{\sigma}}{\sqrt{\dfrac{Y_3^2 + \cdots + Y_n^2}{(n-2)\sigma^2}}}$$

$$= \sqrt{n-2}\,\frac{Y_2}{\sqrt{Y_3^2 + \cdots + Y_n^2}} \sim t(n-2)$$

可得 $\xi=\dfrac{n-1}{\sqrt{n}}\cdot\dfrac{Y_2}{\sqrt{Y_2^2+Y_3^2+\cdots+Y_n^2}}$

$$=\frac{n-1}{\sqrt{n}}\cdot\frac{\dfrac{Y_2}{\sqrt{Y_3^2+\cdots+Y_n^2}}}{\sqrt{\dfrac{Y_2^2}{Y_3^2+\cdots+Y_n^2}+1}}=\frac{n-1}{\sqrt{n}}\cdot\frac{\dfrac{T}{\sqrt{n-2}}}{\sqrt{\dfrac{T^2}{n-2}+1}}$$

即 $T\sim t(n-2)$，且

$$\xi=\frac{n-1}{\sqrt{n}}\cdot\frac{T}{\sqrt{T^2+n-2}}$$

进一步可根据随机变量函数的分布由 T 的分布求得 ξ 的分布