# 数理统计习题答案

**(Answers to the Exercises of Mathematical Statistics)**

**（第一版）**

王震宇

南 开 大 学
统 计 与 数 据 科 学 学 院

天　津

# 作 者 序

数理统计是统计学的一门核心基础课，本答案对应的习题来自于

-《数理统计》王兆军 邹长亮

- 部分补充习题

写下此答案的原因起源于大二上学期与王老师的一次面谈，他提到:“如果有人能把这本书习题做完，那将来一定能做好研究。”又如陈希孺院士在《高等数理统计》一书序言中写到“不做习题，犹如入宝山而空返”。或许本人也不太适合做学术研究，但总听到有高年级学长学姐抱怨习题过难，甚至有伯苓班的大佬直言只能完成 1/3 的习题，终归是想试着做完这本书。再者，往届流传的习题答案虽然排版精美，有一定的参考价值，但实在是错误较多，答案中也看不到思路所在，更无法体现统计学科的思想，故笔者写下此书，望给后来者些许帮助。

与先前答案不同的是，此书的所有答案均采用手写。一是时间实在匆忙，无法打成 LaTeX 文稿；二是笔者在做题过程中，总会有些“奇思妙想”、“易错点”，“统计思想的感悟”甚至“吐槽”。这些文字是有温度的，倘若用机械的 TeX 录入，终究是少了些韵味。

写下此答案，需要感谢的是王兆军老师与朱国钧学长。王老师的数理统计课程不仅硬核，更是在传达统计思想，时刻传述数学与统计的差异。当然，爱国思政教育也让我深受启发。而国钧学长不厌其烦地回答我一些“稀奇古怪”的问题，让我在对知识点有了更为深刻的了解的同时，也教我如何成为更好的助教。

由于笔者对数理统计的专业知识了解及理解有限，故在编写之中难免会有不妥之处，敬请批评指正。同时，由于时间所限，部分习题的答案暂时来不及整理，这将在第二版补上。

本书参考以下文献：

-《概率论与数理统计》陈希孺

-《数理统计学教程》陈希孺

-《高等数理统计学》陈希孺

-《数理统计引论》陈希孺

-《应用多元统计分析》高惠璇

-《数理统计》韦来生

-《数理统计》邵军

编者

2022 年 9 月于南开园

# 第 1 章　基本概念

# 第一章　基本概念

1. 设 $X_1,\ldots,X_n$ 为来自总体分布 $F(x)$ 的 iid 样本，记其经验分布函数为 $F_n(x)$，试证对于任意给定的 $x\in\mathbb{R}$，有

$$\mathbb{E}(F_n(x))=F(x),\quad Var(F_n(x))=\frac{1}{n}F(x)(1-F(x)).$$

**证明:** 由 $F_n(x)$ 的定义, 知 $F_n(x)=\frac{1}{n}\sum\limits_{i=1}^{n}I_{(x_I<x)}$. 不妨令 $Y_i=I_{(x_i<x)}$ ,$\forall x\in\mathbb{R}$.

故

$$\mathbb{P}\{Y_i=1\}=\mathbb{P}\{X_i<x\}=F(x),\ \mathbb{P}\{Y_i=0\}=1-F(x)$$

由二项分布定义，知 $Y_i\sim b(1,F(x))$.

而 $F_n(x)=\frac{1}{n}\sum\limits_{i=1}^{n}Y_i$. 根据 $X_1,\ldots,X_n\quad i.i.d.$，知 $Y_i,\ldots,Y_n\quad i.i.d.$

且

$$\mathbb{E}Y_i=F(x),\quad VarY_i=F(x)(1-F(x)).$$

因此

$$\mathbb{E}(F_n(x))=\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}Y_i\right]=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}Y_i=F(x)$$

$$Var(F_n(x))=Var(\frac{1}{n}\sum_{i=1}^{n}Y_i)=\frac{1}{n^2}\sum_{i=1}^{n}Var\ Y_i=\frac{1}{n}F(x)(1-F(x)).$$

2. 设 $X_1,\ldots,X_n$ 为来自总体分布 $F(x)$ 的 iid 样本，记其经验分布函数为 $F_n(x)$，试证对于任意给定的 $x\in\mathbb{R}$，有

$$F_n(x)\xrightarrow{P}F(x).$$

**证明：法 1:**

类似于第一章习题 1，我们定义 $F_n(x)=\frac{1}{n}\sum\limits_{i=1}^{n} I_{(X_i<x)}=\frac{1}{n}Y_i$，其中 $Y_i=I_{(X_i<x)}$.

对于给定的 $x$，$Y_i,\ldots,Y_n$ $i.i.d.\sim b(1,F(x))$，且

$$\mathbb{E}Y_i=F(x),\quad Var(Y_i)=F(x)(1-F(x))$$

根据中心极限定理知：

$$\frac{1}{n}\sum_{i=1}^{n}Y_i \xrightarrow{P} \mathbb{E}Y_i=F(x)$$

即

$$F_n(x)\xrightarrow{P}F(x)$$

**法 2:**

不妨直接由依概率收敛的定义入手，由 $\mathbb{E}F_n(x)=F(x)$. 根据 Chebyshev 不等式：对 $\forall\epsilon>0$

$$\mathbb{P}\{|F_n(x)-F(x)|\geq\epsilon\}\leq\frac{Var(F_n(x))}{\epsilon^2}=\frac{F(x)(1-F(x))}{n\epsilon^2}\to 0\ \ (n\to\infty)$$

故有

$$F_n(x)\xrightarrow{P}F(x)$$

3. 设 $X_1,\ldots,X_{n+1}$ 为来自总体 $X$ 的 iid 样本, 试证明

$${S_{n+1}^*}^2=\frac{n}{n+1}\left[S_n^{*2}+\frac{1}{n+1}(X_{n+1}-\bar{X}_n)^2\right].$$

**证明**：我们先证明一个常用的化简公式：

$$\begin{aligned}(n-1)S_n^2=nS_n^{*2}&=\sum_{i=1}^{n}(X_i-\bar{X})^2\\&=\sum_{i=1}^{n}(X_i^2-2X_i\bar{X}+\bar{X}^2)\\&=\sum_{i=1}^{n}X_i^2-2\bar{X}\sum_{i=1}^{n}X_i+n\bar{X}^2\\&=\sum_{i=1}^{n}X_i^2-2n\bar{X}^2+n\bar{X}^2\\&=\sum_{i=1}^{n}X_i^2-n\bar{X}^2\end{aligned}$$

在本题中:

$$
\begin{aligned}
S_{n+1}^{*2} &= \frac{1}{n+1}\sum_{i=1}^{n+1}(X_i - \bar{X}_{n+1})^2 \\
&= \frac{1}{n+1}(\sum_{i=1}^{n+1} X_i^2 - (n+1)\bar{X}_{n+1}^2) \\
&= \frac{1}{n+1}(\sum_{i=1}^{n+1} X_i^2 - \frac{(n\bar{X}_n + X_{n+1})^2}{n+1}) \\
&= \frac{1}{n+1}(\sum_{i=1}^{n+1} X_i^2 - \frac{n^2\bar{X}_n^2}{n+1} - \frac{X_{n+1}^2}{n+1} - \frac{2n\bar{X}_n X_{n+1}}{n+1}) \\
&= \frac{1}{n+1}(\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2 + \frac{n}{n+1}(\bar{X}_n^2 + X_{n+1}^2 - 2n\bar{X}_n X_{n+1})) \\
&= \frac{1}{n+1}(nS_n^{*2} + \frac{n}{n+1}(X_{n+1} - \bar{X}_n)^2) \\
&= \frac{n}{n+1}(S_n^{*2} + \frac{1}{n+1}(X_{n+1} - \bar{X}_n)^2)
\end{aligned}
$$

原式得证.

4. 设 $X_1, \ldots, X_n$ 为来自 $N(0, \sigma^2)$ 的 iid 样本，$A = (a_{ij})$ 是一 n 阶正交阵，$\mathbf{c} = (c_1, \ldots, c_n)^{\mathbf{T}}$ 为一 n 维常向量，记 $\mathbf{X} = (X_1, \ldots, X_n)^{\mathbf{T}}$，则 $\mathbf{Y} = A\mathbf{X} + \mathbf{c}$ 的各分量相互独立，且分别服从 $N(c_i, \sigma^2)$.

**证明：法 1：**

由 $X_1, \ldots, X_n \;\; i.i.d. \sim N(0, \sigma^2)$ 知

$$f(x_1, \ldots, x_n) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right\}$$

即

$$f(\boldsymbol{x}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{x}'\boldsymbol{x}\right\}$$

由 $y = Ax + c$，故根据密度变换公式知

$$f(\boldsymbol{y}) = (2\pi\sigma^2)^{-\frac{n}{2}}|A| \exp\left\{-\frac{(A'(\boldsymbol{y} - c))'(A'(\boldsymbol{y} - c))}{2\sigma^2}\right\}$$

考虑到 $A$ 是正交阵，故 $|A|=1, A'A=I$，则 $Y$ 的密度函数为

$$\begin{aligned} f(\boldsymbol{y}) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{(\boldsymbol{y}-c)'(\boldsymbol{y}-c)}{2\sigma^2}\right\} \\ &= \prod_{i=1}^{n} (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i-c_i)^2\right\} \end{aligned}$$

故 $Y_I,\ldots,Y_n$ 相互独立，但 $Y_i \sim N(c_i,\sigma^2)$.

**法 2:**

由多元统计分析知：$\boldsymbol{X} \sim N_n(\boldsymbol{0},\sigma^2 I_n)$, 故

$$A\boldsymbol{X}+\boldsymbol{c} \sim N_n(A\cdot\boldsymbol{0}+\boldsymbol{c},\sigma^2 A I_n A')$$

即

$$\boldsymbol{Y} = A\boldsymbol{X}+\boldsymbol{c} \sim N_n(\boldsymbol{c},\sigma^2 I)$$

因而 $Y_I,\ldots,Y_n$ 相互独立，但 $Y_i \sim N(c_i,\sigma^2)$.

5. 设 $X_1,\ldots,X_8$ 为来自 $N(0,1)$ 的 iid 样本，$Y = (X_1+\cdots+X_4)^2 + (X_5+\cdots+X_8)^2$，试求常数 c，使 $cY$ 服从 $\chi^2$ 分布.

**解:** 由 $X_1,\ldots,X_8$ $i.i.d. \sim N(0,1)$ 知 $Y_1 = X_1+\cdots+X_4 \sim N(0,4)$, $Y_2 = X_5+\cdots+X_8 \sim N(0,4)$, 且 $Y_1$ 与 $Y_2$ 独立.

**充分性:**

由 $\frac{Y_1}{2},\frac{Y_2}{2}$ $i.i.d. \sim N(0,1)$，故

$$(\frac{Y_1}{2})^2+(\frac{Y_2}{2})^2 \sim \chi^2(2)$$

即 $c=\frac{1}{4}$ 符合题意.

**必要性:**

由 $cY = cY_1^2+cY_2^2 \sim \chi^2$，而

$$\mathbb{E}\left[cY\right] = c\mathbb{E}\left[Y_1^2+Y_2^2\right] = 4c\cdot\mathbb{E}\left[\frac{Y_1^2}{4}+\frac{Y_2^2}{4}\right]$$

由 $\frac{Y_1^2}{4}+\frac{Y_2^2}{4} \sim \chi^2$ 知 $\mathbb{E}\left[cY\right] = 4c\cdot 2 = 8c$，而

$$Var(cY) = c^2 VarY = c^2 Var(Y_1^2+Y_2^2) = 16c^2 Var(\frac{Y_1^2}{4}+\frac{Y_2^2}{4}) = 64c^2$$

由 $\chi^2$ 分布的性质知：$64c^2 = 8c\cdot 2$，且 $c\neq 0$，即 $c=\frac{1}{4}$.

综上：$c=\frac{1}{4}$

6. 设 $X_1,\dots,X_8$ 为来自 $N(0,1)$ 的 iid 样本,$X=a(X_1-2X_2)^2+b(3X_3-24X_4)^2$，试求 $a,b$，使 $X$ 服从 $\chi^2$ 分布.

**解:** 由 $X_1,\dots,X_4\ \ i.i.d.\sim N(0,1)$

令 $Y_1=X_1-2X_2\sim N(0,5)$, $Y_2=3X_3-4X_4\sim N(0,25)$，且 $Y_1$ 与 $Y_2$ 独立.

由 $\frac{Y_1}{\sqrt{5}},\frac{Y_2}{5}\ \ i.i.d.\sim N(0,1)$ 知

①自由度为 2 时:

$$(\frac{Y_1}{\sqrt{5}})^2+(\frac{Y_2}{5})^2\sim\chi^2(2)$$

即

$$\frac{1}{5}Y_1^2+\frac{1}{25Y_2^2}\sim\chi^2(2)$$

故

$$a=\frac{1}{5},b=\frac{1}{25}.$$

②自由度为 1 时:

$$a=\frac{1}{5},b=0\text{ 或 }a=0,b=\frac{1}{25}.$$

7. 设 $X_1,\dots,X_n$ 为来自 $N(\mu,\sigma^2)$ 的 iid 样本，请问当 k 为多少时，

$$P\{\bar{X}>\mu+kS_n\}=0.95?$$

**解:** 由 $X_1,\dots,X_n\ \ i.i.d.\sim N(\mu,\sigma^2)$ 知 $\frac{\sqrt{n}(\bar{x}-\mu)}{S_n}\sim t(n-1)$.

故

$$\mathbb{P}\{\bar{x}>\mu+kS_n\}=\mathbb{P}\left\{\frac{\sqrt{n}(\bar{x}-\mu)}{S_n}>k\sqrt{n}\right\}=0.95$$

因而 $k\sqrt{n}=-t_{0.05}(n-1)$，即

$$k=-\frac{1}{\sqrt{n}}t_{0.05}(n-1).$$

8*. 设 $X_1,\dots,X_n$ 为来自总体 $X\sim F$ 的 iid 样本, 且 $EX=\mu,Var(X)=\sigma^2$. 记 $T_n=\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$，则在某些连续性假设下，证明 $T_n$ 的分布函数可展成如下形式:

$$\Phi(x)-\phi(x)\left[\frac{\gamma_1}{6\sqrt{n}}H_2(x)+\frac{1}{n}\left(\frac{\gamma_2}{24}H_3(x)+\frac{{\gamma_1}^2}{72}H_5(x)\right)+r_n\right]$$

其中 $\gamma_1, \gamma_2$ 分别为总体 $X$ 的偏度与峰度，$r_n = o(n^{-1})$，$H_r(\cdot)$ 为 r 阶 Hermite 多项式，齐定义如下：$\frac{\mathrm{d}^r(e^{-x^2/2})}{\mathrm{d}x^r} = (-1)^r H_r(x) e^{-x^2/2}$. 上述展开式被称为 Edgeworth 展开.

**证明:** 不失一般性，我们假设 $\mu = 0 \quad \sigma^2 = 1$，对于任意 $\mu, \sigma^2$，我们只需将 $x_j$ 替换为 $\frac{x_j - \mu}{\sigma}$ 即可.

由题意可知：$\gamma_1 = \mathbb{E}X_1^3$，$\gamma_2 = \mathbb{E}X_1^4 - 3$，$T_n = \frac{1}{\sqrt{n}} \sum\limits_{i=1}^{n} X_i$

不妨记 $\varphi_1(t) = \mathbb{E}\left[e^{itX_1/\sqrt{n}}\right]$ 为 $X_1/\sqrt{n}$ 的特征函数

将 $e^{itX_1/\sqrt{n}}$ 在 $t=0$ 处 Taylor 展开有

$$\begin{aligned}\varphi_1(t) &= \mathbb{E}\left[1 + \frac{itX_1}{\sqrt{n}} + \frac{(it)^2 X_1^2}{2n} + \frac{(it)^3 X_1^3}{6n\sqrt{n}} + \frac{(it)^4 X_1^4}{24n^2} + o\left(\frac{1}{n^2}\right)\right] \\ &= \left(1 - \frac{t^2}{2n}\right) + \frac{(it)^3\gamma_1}{6n\sqrt{n}} + \frac{(it)^4(\gamma_2+3)}{24n^2} + o\left(\frac{1}{n^2}\right)\end{aligned}$$

因而

$$\begin{aligned}\varphi_{T_n}(t) =& [\varphi_1(t)]^n \\ =& \left(1 - \frac{t^2}{2n}\right)^n + \left(1 - \frac{t^2}{2n}\right)^{n-1} \left[\frac{(it)^3\gamma_1}{6\sqrt{n}} + \frac{(it)^4(\gamma_2+3)}{24n}\right] \\ &+ \left(1 - \frac{t^2}{2n}\right)^{n-2} \frac{(n-1)(it)^6\gamma_1^2}{72n^2} + o\left(\frac{1}{n}\right)\end{aligned}$$

由于 $(1+\frac{a}{n})^{n-k} = e^a\left[1 - \frac{a(a+2k)}{2n}\right] + o(\frac{1}{n}) \qquad (*)$

故

$$\begin{aligned}\varphi_{T_n}(t) =& e^{-\frac{t^2}{2}}\left(1 - \frac{t^4}{8n}\right) + e^{-\frac{t^2}{2}}\left(1 - \frac{\frac{t^4}{4} - t^2}{2n}\right)\left[\frac{(it)^3\gamma_1}{6\sqrt{n}} + \frac{(it)^4(\gamma_2+3)}{24n}\right] \\ &+ e^{-\frac{t^2}{2}}\left(1 - \frac{\frac{t^4}{4} - 2t^2}{2n}\right)\frac{(n-1)(it)^6\gamma_1^2}{72n^2} + o\left(\frac{1}{n}\right) \\ =& e^{-\frac{t^2}{2}}\left[1 - \frac{t^4}{8n} + \frac{(it)^3\gamma_1}{6\sqrt{n}} + \frac{(it)^4(\gamma_2+3)}{24n} + \frac{n(it)^6\gamma_1^2}{72n^2}\right] + o\left(\frac{1}{n}\right) \\ =& e^{-\frac{t^2}{2}}\left[1 + \frac{(it)^3\gamma_1}{6\sqrt{n}} + \frac{(it)^4\gamma_2}{24n} + \frac{(it)^6\gamma_1^2}{72n}\right] + o\left(\frac{1}{n}\right)\end{aligned}$$

由逆转公式可知，$T_n$ 的 PDF $g(x)$ 满足

$$\begin{aligned}
g(x) =&\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\varphi_{T_n}(t)\,\mathrm{d}t \\
=&\frac{1}{2\pi}\left(\int_{-\infty}^{+\infty} e^{-itx}e^{-\frac{t^2}{2}}\,\mathrm{d}t + \frac{\gamma_1}{6\sqrt{n}}\int_{-\infty}^{+\infty} e^{-itx}e^{-\frac{t^2}{2}}(it)^3\,\mathrm{d}t\right. \\
&\left.+\frac{\gamma_2}{24n}\int_{-\infty}^{+\infty} e^{-itx}e^{-\frac{t^2}{2}}(it)^4\,\mathrm{d}t + \frac{\gamma_1^2}{72n}\int_{-\infty}^{+\infty} e^{-itx}e^{-\frac{t^2}{2}}(it)^6\,\mathrm{d}t\right) \\
&+o\left(\frac{1}{n}\right)
\end{aligned}$$

注意到，对 $\forall k \in \mathbb{N}^*$ 有

$$\begin{aligned}
\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}e^{-\frac{t^2}{2}}(it)^k\,\mathrm{d}t &= \frac{(-1)^k}{2\pi}\frac{\mathrm{d}^k}{\mathrm{d}x^k}\int_{-\infty}^{+\infty} e^{-itx}e^{-\frac{t^2}{2}}\,\mathrm{d}t \\
&= (-1)^k\frac{\mathrm{d}^k}{\mathrm{d}x^k}\phi(x) \\
&= H_k(x)\phi(x)
\end{aligned}$$

故 $g(x) = \phi(x)\left(1 + \frac{\gamma_1}{6\sqrt{n}}H_3(x) + \frac{\gamma_2}{24n}H_4(x) + \frac{\gamma_1^2}{72n}H_6(x)\right) + r_n$
又注意到 $\frac{\mathrm{d}^r(e^{-x^2/2})}{\mathrm{d}x^r} = (-1)^r H_r(x)e^{-x^2/2}$ 知

$$(-1)^r\frac{\mathrm{d}^r}{\mathrm{d}x^r}\phi(x) = H_r(x)\phi(x)$$

对等式两端求导可得 $\frac{\mathrm{d}}{\mathrm{d}x}\left[H_r(x)\phi(x)\right] = -H_{r+1}(x)\phi(x)$
因此 $T_n$ 的分布函数 $G(x)$ 满足：

$$\begin{aligned}
G(x) =&\Phi(x) + \frac{\gamma_1}{6\sqrt{n}}\left[-H_2(x)\phi(x)\right] + \frac{\gamma_2}{24n}\left[-H_3(x)\phi(x)\right] \\
&+\frac{\gamma_1^2}{72n}\left[-H_5(x)\phi(x)\right] + o\left(\frac{1}{n}\right) \\
=&\Phi(x) - \phi(x)\left[\frac{\gamma_1}{6\sqrt{n}}H_2(x) + \frac{\gamma_2}{24n}H_3(x) + \frac{\gamma_1^2}{72n}H_5(x) + r_n\right] \\
=&\Phi(x) - \phi(x)\left[\frac{\gamma_1}{6\sqrt{n}}H_2(x) + \frac{1}{n}\left(\frac{\gamma_2}{24}H_3(x) + \frac{\gamma_1^2}{72}H_5(x)\right) + r_n\right]
\end{aligned}$$

最后由 $0 < \phi(x) \le \phi(0)$，故 $r_n$ 与 $-\phi(x)r_n$ 等价
故有 $G(x) = \Phi(x) - \phi(x)\left[\frac{\gamma_1}{6\sqrt{n}}H_2(x) + \frac{1}{n}\left(\frac{\gamma_2}{24}H_3(x) + \frac{\gamma_1^2}{72}H_5(x)\right) + r_n\right]$

下面证明 (*) 式

$$\begin{aligned}\left(1+\frac{a}{n}\right)^{n-k} &= \exp\left\{(n-k)\log\left(1+\frac{a}{n}\right)\right\} \\ &= \exp\left\{(n-k)\left[\frac{a}{n}-\frac{a^2}{2n^2}+o\left(\frac{1}{n^2}\right)\right]\right\} \\ &= \exp\left\{a-\frac{ka}{n}-\frac{a^2}{2n}+o\left(\frac{1}{n}\right)\right\} \\ &= e^a\cdot\exp\left\{-\frac{a(a+2k)}{2n}+o\left(\frac{1}{n}\right)\right\} \\ &= e^a\left(1-\frac{a(a+2k)}{2n}\right)+o\left(\frac{1}{n}\right)\end{aligned}$$

得证.

9. 由上述 Edgeworth 展开，请验证：$T_n$ 的 $\alpha$ 分位数可以近似为

$$u_\alpha+\frac{\gamma_1}{6\sqrt{n}}({u_\alpha}^2-1)+\frac{\gamma_2}{24n}({u_\alpha}^3-3u_\alpha)-\frac{{\gamma_1}^2}{36n}(2{u_\alpha}^3-5u_\alpha)+r_n$$

其中 $u_\alpha$ 为标准正态的 $\alpha$ 分位数，且上述展式被称为 Fisher-Cornish 展开.

**证明**：由 Hermite 多项式定义知:

$$H_2(x)=x^2-1, H_3(x)=x^3-3x, H_5(x)=x^5-10x^3+15x.$$

令 $x=u_\alpha+\frac{r_1}{6\sqrt{n}}(u_\alpha^2-1)+\frac{r_2}{24n}(u_\alpha^3-3u_\alpha)+\frac{r_1^2}{36n}(2u_\alpha^3-5u_\alpha+r_n)$. 将 $\Phi(x)$ 在 $u_\alpha$ 点展开，则有

$$\begin{aligned}\Phi(x)=\alpha+\phi(u_\alpha)\frac{r_1}{6\sqrt{n}}(u_\alpha^2-1)+\phi(u_\alpha)\frac{r_2}{24n}(u_\alpha^3-3u_\alpha) \\ -\phi(u_\alpha)\frac{r_1^2}{72n}(u_\alpha^5+2u_\alpha^3-9u_\alpha)+o(n^{-1})\end{aligned}$$

而 $\phi(x)=\phi(u_\alpha)-\phi(u_\alpha)u_\alpha\frac{r_1}{6\sqrt{n}}+o(n^{-1})$，且

$$\begin{aligned}\frac{r_1}{6\sqrt{n}}H_2(x)+\frac{1}{n}(\frac{r_2}{24}H_3(x)+\frac{r_1^2}{72}H_5(x))=\frac{r_1}{6\sqrt{n}}(u_\alpha^2-1)+\frac{r_1^2}{18n}(u_\alpha^3-3u_\alpha) \\ +\frac{r_2}{24n}(u_\alpha^3-3u_\alpha)+\frac{r_1^2}{72n}(u_\alpha^5-10u_\alpha^3+15u_\alpha) \\ +o(n^{-1})\end{aligned}$$

代入上面结果知

$$\Phi(x)-\phi(x)\left[\frac{r_1}{6\sqrt{n}}H_2(x)+\frac{1}{n}(\frac{r_2}{24}H_3(x)+\frac{r_1^2}{72}H_5(x))\right]+r_n=\alpha$$

则由左式为分布函数的近似知，$x$ 是 $T_n$ 的 $\alpha$ 分位数.

10. $X_1,\ldots,X_n$ 为来自总体 $X\sim F$ 的 iid 样本，总体概率密度函数为 $f(x)$，则对于给定的 $1\le j\le k\le n$，统计量 $X_{(j)}$ 与 $X_{(k)}$ 的联合概率密度函数为

$$f_{j,k}(x,y)=\begin{cases}\frac{n!}{(j-1)!(k-j-1)!(n-k)!}F^{j-1}(x)[F(y)-F(x)]^{k-j-1}[1-F(y)]^{n-k}f(x)f(y) & ,x<y\\ 0 & ,否则.\end{cases}$$

**证明:**不妨将其化为多项分布,求解 $\mathbb{P}\left\{x\le X_{(j)}<x+\Delta x, y\le X_{(k)}<y+\Delta y\right\}$,即:

①在 $(-\infty,x)$ 上有 $j-1$ 个样本;
②在 $(x,x+\Delta x)$ 上有 1 个样本;
③在 $(x+\Delta x,y)$ 上有 $k-j-1$ 个样本;
④在 $(y,y+\Delta y)$ 上有 1 个样本;
⑤在 $(y+\Delta y,+\infty)$ 上有 $n-k$ 个样本. 故有

$$\begin{aligned}&\mathbb{P}\left\{x\le X_{(j)}<x+\Delta x, y\le X_{(k)}<y+\Delta y\right\}\\ =&\frac{n!}{(j-1)!1!(k-j-1)!1!(n-k)!}\left[F(x)\right]^{j-1}f(x)\Delta x\\ &\cdot\left[F(y)-F(x+\Delta x)\right]^{k-j-1}f(y)\Delta y\left[1-F(y+\Delta y)\right]^{n-k}.\end{aligned}$$

由 CDF 与 PDF 之间的关系可知:

$$\begin{aligned}&f_{j,k}(x,y)\\ =&\lim_{\substack{\Delta x\to 0\\ \Delta y\to 0}}\frac{\mathbb{P}\left\{x\le X_{(j)}<x+\Delta x, y\le X_{(k)}<y+\Delta y\right\}}{\Delta x\Delta y}\\ =&\begin{cases}\frac{n!}{(j-1)!(k-j-1)!(n-k)!}F^{j-1}(x)\left[F(y)-F(x)\right]^{k-j-1}f(y)\left[1-F(y)\right]^{n-k}f(x)f(y), & x<y,\\ 0, & 否则.\end{cases}\end{aligned}$$

11. 设 $X_1,\ldots,X_{n+1}$ 为来自 $N(\mu,\sigma^2)$ 的 iid 样本，试求统计量

$$T=\frac{X_{n+1}-\bar{X}_n}{S_n^*}\sqrt{\frac{n-1}{n+1}}$$

的抽样分布.

**解**：由 $X_1,\ldots,X_{n+1}\ \ i.i.d. \sim N(\mu,\sigma^2)$ 知
$X_{n+1}\sim N(\mu,\sigma^2)$, $\bar{X}_n\sim N(\mu,\frac{\sigma^2}{n})$, 且 $X_{n+1}$ 与 $\bar{X}_n$ 独立.
因此 $X_{n+1}-\bar{X}_n\sim N(0,\frac{n+1}{n}\sigma^2)$.
又 $nS_n^{*2}=(n+1)S_N^2\sim\sigma^2\chi^2(n-1)$, 且 $\bar{X}_n$ 与 $X_{n+1}$ 均与 $S_n^{*2}$ 独立,
故 $\bar{X}_n-X_{n+1}$ 与 $S_n^{*2}$ 独立.
由 t 分布的定义可知:

$$\frac{\frac{X_{n+1}-\bar{X}_n}{\sqrt{\frac{n+1}{n}}\sigma}}{\sqrt{\frac{nS_n^{*2}}{(n-1)\sigma^2}}}=\frac{X_{n+1}-\bar{X}_n}{S_n^*}\sqrt{\frac{n-1}{n+1}}\sim t(n-1)$$

12. 设 $X_1,\ldots,X_{m+n}$ 为来自 $N(0,\sigma^2)$ 的 iid 样本，试求统计量

$$Z=\sqrt{\frac{m}{n}}\frac{\sum_{i=1}^n X_i}{\sqrt{\sum_{i=n+1}^{n+m}X_i^2}},\ \ F=\frac{m}{n}\frac{\sum_{i=1}^n X_i^2}{\sum_{i=n+1}^{n+m}X_i^2}$$

的抽样分布.

**解**：由 $X_1,\ldots,X_{m+n}\ \ i.i.d. \sim N(0,\sigma^2)$ 知:

$$\sum_{i=1}^n X_i\sim N(0,n\sigma^2),\quad \sum_{i=n+1}^{n+m}(X_i/\sigma)^2\sim\chi^2(m)$$

且两者独立.
故由 t 分布定义可知:

$$\frac{\frac{1}{\sqrt{n}\sigma}\sum\limits_{i=1}^n X_i}{\sqrt{\sum\limits_{i=n+1}^{n+m}(X_i/\sigma)^2/m}}\sim t(m)$$

即 $Z\sim t(m)$.
同理可知:

$$\sum_{i=1}^n(X_i/\sigma)^2\sim\chi^2(n),\quad \sum_{i=n+1}^{n+m}(X_i/\sigma)^2\sim\chi^2(m)$$

且两者独立.

因而由 F 分布定义可知：

$$\frac{\sum\limits_{i=1}^{n}(\frac{X_i}{\sigma})^2/n}{\sum\limits_{i=n+1}^{n+m}(\frac{X_i}{\sigma})^2/m} \sim F(n,m)$$

即 $F \sim F(n,m)$.

13. 设随机变量 $X, Y$ 相互独立且均服从 $N(0,\sigma^2)$，则
(1) $X^2+Y^2$ 与 $X/\sqrt{X^2+Y^2}$ 相互独立；
(2) $\Theta = \sin^{-1}\frac{X}{\sqrt{X^2+Y^2}}$ 服从 $(-\pi/2, \pi/2)$ 上的均匀分布；
(3) $X/Y$ 服从 Cauchy 分布.

**证明：** (1) 由 $X, Y \quad i.i.d. \sim N(0,\sigma^2)$ 知

$$\begin{aligned} f(x,y) &= f_X(x)f_Y(y) \\ &= \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{y^2}{2\sigma^2}} \\ &= \frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \end{aligned}$$

不妨令 $\begin{cases} X = R\sin\Phi \\ Y = R\cos\Phi \end{cases}$，其中 $R \in [0,+\infty)$，$\Phi \in [0,2\pi)$.

因此欲证 $X^2+Y^2$ 与 $X/\sqrt{X^2+Y^2}$ 相互独立，即证 $R^2$ 与 $\sin\Phi$ 的独立性.

而

$$\begin{aligned} g(r,\theta) &= f(r\sin\phi, r\cos\phi)\left|\frac{D(x,y)}{D(r,\phi)}\right| \\ &= \frac{1}{2\pi\sigma^2}e^{-\frac{r^2}{2\sigma^2}}r \\ &= \frac{1}{2\pi}\frac{r}{\sigma^2}e^{-\frac{r^2}{2\sigma^2}} \end{aligned}$$

即

$$g_R(r) = \frac{r}{\sigma^2}e^{-\frac{r^2}{2\sigma^2}}, \quad g_\Phi(\phi) = \frac{1}{2\pi}, \quad g(r,\phi) = g_R(r)g_\Phi(\phi)$$

故 $R$ 与 $\Phi$ 独立 $\Rightarrow$ $R^2$ 与 $\sin\Phi$ 独立.

(2)

$$F_\Theta(\theta) = \mathbb{P}\{\Theta \le \theta\} = \mathbb{P}\left\{\sin^{-1}\sin\phi \le \theta\right\} = \mathbb{P}\{\sin\phi \le \sin\theta\}$$

注意到

$$\theta\in[0,\frac{\pi}{2})\text{时,}\quad F_\Theta(\theta)=\mathbb{P}\{0\le\Phi\le\theta\text{或}\pi-\theta\le\Phi\le2\pi\}=\frac{\pi+2\theta}{2\pi}$$
$$\theta\in(-\frac{\pi}{2},0)\text{时,}\quad F_{\hat\theta}(\theta)=\mathbb{P}\{\pi-\theta\le\Phi\le2\pi+\theta\}=\frac{\pi+2\theta}{2\pi}$$

则 $f_\Theta=F'_\Theta(\theta)=\frac{1}{\pi}$，$\theta\in(-\frac{\pi}{2},\frac{\pi}{2})$
故 $\hat\theta\sim U(-\frac{\pi}{2},\frac{\pi}{2})$

(3) 令 $C=\frac{X}{Y}=\tan\Theta$
则

$$\begin{aligned}F_C(x)&=\mathbb{P}\{C\le x\}\\&=\mathbb{P}\{\tan\Theta\le x\}\\&=\frac{1}{2\pi}\cdot2(\arctan x+\frac{\pi}{2})\\&=\frac{1}{2}+\frac{1}{\pi}\arctan x\end{aligned}$$

因此 $f_C(x)=\frac{1}{\pi}\cdot\frac{1}{1+x^2}$，$x\in(-\infty,+\infty)$
故 $C$ 服从 Cauchy 分布.

注：本题在划定 R 与 $\Phi$ 的定义域时，倘若为了追求第 2 问的方便而直接取 $R\in(-\infty,+\infty),\Phi\in(-\frac{\pi}{2},\frac{\pi}{2})$，则 $\frac{X}{\sqrt{X^2+Y^2}}=\frac{R\sin\theta}{|R|}=\sin\theta\cdot sgn(R)$，从而导致第一问的讨论更复杂

14. 设 $X_1,\ldots,X_4$ 为来自 $N(0,\sigma^2)$ 的 iid 样本，$U=\frac{\sqrt{3}n}{\sqrt{X_2^2+X_3^2+X_4^2}}$，试求 $EU,VarU$.

**解:** 由 $X_1,\ldots,X_4$ $i.i.d.\sim N(0,\sigma^2)$ 知
$X_1/\sigma\sim N(0,1)$，$(X_2^2+X_3^2+X_4^2)/\sigma^2\sim\chi^2(3)$ 且两者独立.
又由 t 分布的定义可知:

$$\frac{X_1/\sigma}{\sqrt{(X_2^2+X_3^2+X_4^2)/3\sigma^2}}\sim t(3),\quad\text{即 }U\sim t(3)$$

故 $\mathbb{E}U=0$，$VarU=\frac{n}{n-2}=3$.

15. 设 $X_1,\ldots,X_{n+1}$ 为来自 $N(\mu,\sigma^2)$ 的 iid 样本，记 $V_i=X_i-\frac{1}{n+1}\sum_{i=1}^{n+1}X_i,i=1,\ldots,n+1$. 试求 $V_i$ 的分布.

**解：** 由 $X_1,\ldots,X_{n+1}\ \ i.i.d.\sim N(\mu,\sigma^2)$,
又由

$$V_i=X_i-\frac{1}{n+1}\sum_{i=1}^{n+1}X_i=-\frac{1}{n+1}X_1-\frac{1}{n+1}X_2-\cdots+\frac{n}{n+1}X_i-\cdots-\frac{1}{n+1}X_{n+1}$$

可知，$V_i$ 服从正态分布.
而

$$\mathbb{E}V_i=\mathbb{E}\left[\frac{n}{n+1}X_i\right]-n\mathbb{E}\left[\frac{1}{n+1}X_1\right]=\frac{n}{n+1}\mu-n\cdot\mu\frac{1}{n+1}=0$$

$$VarV_i=\sigma^2\left[\frac{1}{(n+1)^2}\cdot n+\frac{n^2}{(n+1)^2}\right]=\frac{n}{n+1}\sigma^2$$

故 $V_i\sim N(0,\frac{n}{n+1}\sigma^2)$.

16. 设 $X_1,\ldots,X_m$ 为来自 $N(\mu_1,\sigma^2)$ 的 iid 样本,$Y_1,\ldots,Y_n$ 为来自 $N(\mu_2,\sigma^2)$ 的 iid 样本，且全样本独立，又设 $a,b$ 为两个常数，$Z=\frac{a(\bar{X}-\mu_1)+b(\bar{Y}-\mu_2)}{\sqrt{(m-1)S_{1m}^2+(n-1)S_{2n}^2}}$，试求常数 $c$，使得 $cZ$ 服从 $t$ 分布，并给出其自由度.

**解：** 由 $X_1,\ldots,X_m\ \ i.i.d.\sim N(\mu_1,\sigma^2)$, $\ Y_1,\ldots,Y_n\ \ i.i.d.\sim N(\mu_2,\sigma^2)$ 且全样本独立知
$\bar{X}\sim N(\mu_1,\frac{\sigma^2}{m})$, $\ \bar{Y}\sim N(\mu_2,\frac{\sigma^2}{n})$ 且两者独立
故 $a(\bar{X}-\mu_1)+b(\bar{Y}-\mu_2)\sim N(0,(\frac{a^2}{m}+\frac{b^2}{n}\sigma^2))$
对其标准化得 $a(\bar{X}-\mu_1)+b(\bar{Y}-\mu_2)/\sqrt{(\frac{a^2}{m}+\frac{b^2}{n})\sigma^2}\sim N(0,1)$
又 $(m-1)S_{1m}^2/\sigma^2\sim\chi^2(m-1)$, $\ (n-1)S_{2n}^2/\sigma^2\sim\chi^2(n-1)$ 且两者独立
因而由 $\chi^2$ 分布的可加性知

$$(m-1)S_{1m}^2+(n-1)S_{2n}^2/\sigma^2\sim\chi^2(m+n-2)$$

故由 t 分布定义有

$$\frac{\frac{a(\bar{X}-\mu_1)+b(\bar{Y}-\mu_2)}{\sqrt{(\frac{a^2}{m}+\frac{b^2}{n})\sigma^2}}}{\sqrt{\frac{(m-1)S_{1m}^2+(n-1)S_{2n}^2}{\sigma^2(m+n-2)}}}\sim t(m+n-2)$$

即 $\sqrt{\frac{m+n-2}{\frac{a^2}{m}+\frac{b^2}{n}}}Z\sim t(m+n-2)$
因此 $c=\sqrt{\frac{m+n-2}{\frac{a^2}{m}+\frac{b^2}{n}}}$，自由度为 $m+n-2$.

17. 设 $X_1,\ldots,X_n$ 为来自指数分布 $E(\lambda)$ 的 iid 样本，记 $T=2\lambda\sum_{i+1}^{n}X_i$，则证明 $T$ 服从 $\chi^2(2n)$.

**证明**：由题意知 $X_1,\ldots,X_n \quad i.i.d. \sim E(\lambda)$

即 $X_1,\ldots,X_n \quad i.i.d. \sim \Gamma(1,\lambda)$

由 $\Gamma$ 分布形状参数可加性知：$\sum\limits_{i=1}^{n}X_i \sim \Gamma(n,\lambda)$

又由 $\Gamma$ 分布刻度参数可乘性知：$2\lambda\sum\limits_{i=1}^{n}X_i \sim \Gamma(n,\frac{\lambda}{2\lambda})$

即 $T\sim\Gamma(\frac{2n}{2},\frac{1}{2})=\chi^2(2n)$.

18. 设随机变量 $X$ 服从自由度为 $n,m$ 的 $F$ 分布，试求 $Y=1/X$ 的分布，并试证明

$$Z=\frac{nX}{m+nX}\sim\beta(n/2,m/2).$$

**证明**：不妨设 $\xi\sim\chi^2(n)$，$\eta\sim\chi^2(m)$ 且两者独立

由 $X\sim F(n,m)$ 知，$X\overset{d}{=}\frac{\xi/n}{\eta/m}$

故 $Y=\frac{1}{X}\overset{d}{=}\frac{\eta/m}{\xi/n}\sim F(m,n)$

而 $Z=\frac{nX}{m+nX}\overset{d}{=}\frac{\xi}{\xi+\eta}$

由 $\Gamma$ 分布与 $\beta$ 分布的关系可知 $\frac{\xi}{\xi+\eta}\sim\beta(n/2,m/2)$

故 $Z\sim\beta(n/2,m/2)$.

19. 设 $X_1\sim\beta(a_1,b_1)$，$X_2\sim\beta(a_2,b_2)$，$a_1=a_2+b_2$，且 $X_1$ 与 $X_2$ 独立，则 $Y=X_1X_2\sim\beta(a_2,b_1+b_2)$.

**解**：由 $X_1\sim\beta(a_1,b_1)$，$X_2\sim\beta(a_2,b_2)$ 知

$$f_{X_1}(x_1)=\frac{\Gamma(a_1+b_1)}{\Gamma(a_1)\Gamma(b_1)}x_1^{a_1-1}(1-X_1)^{b_1-1},\ f_{X_2}(x_2)=\frac{\Gamma(a_2+b_2)}{\Gamma(a_2)\Gamma(b_2)}x_2^{a_2-1}(1-x_2)^{b_2-1}$$

又由 $X_1$ 与 $X_2$ 独立可知

$$f(x_1,x_2)=f_{X_1}(x_1)f_{X_2}(x_2)=\frac{\Gamma(a_2+b_2+b_1)}{\Gamma(b_1)\Gamma(a_2)\Gamma(b_2)}x_1^{a_1-1}(1-x_1)^{b_1-1}x_2^{a_2-1}(1-x_2)^{b_2-1}$$

由概率论知识，可得：

$$
\begin{aligned}
f_Y(y) &= \int_y^1 f(x_1, \frac{y}{x_1})\frac{1}{x_1}\,\mathrm{d}x_1 \\
&= \int_y^1 \frac{\Gamma(a_2+b_2+b_1)}{\Gamma(b_1)\Gamma(a_2)\Gamma(b_2)} x_1^{a_1-1}(1-x_1)^{b_1-1}(\frac{y}{x_1})^{a_2-1}(1-\frac{y}{x_1})^{b_2-1}\cdot\frac{1}{x_1}\,\mathrm{d}x_1 \\
&= y^{a_2-1}\int_y^1 \frac{\Gamma(a_2+b_2+b_1)}{\Gamma(b_1)\Gamma(a_2)\Gamma(b_2)}(1-x_1)^{b_1-1}(x_1-y)^{b_2-1}\,\mathrm{d}x_1 \\
&= y^{a_2-1}(1-y)^{b_1+b_2-1}\int_y^1 \frac{\Gamma(a_2+b_2+b_1)}{\Gamma(b_1)\Gamma(a_2)\Gamma(b_2)}(\frac{1-x_1}{1-y})^{b_1-1}(\frac{x_1-y}{1-y})^{b_2-1}\,\mathrm{d}(\frac{x_1-y}{1-y}) \\
&= y^{a_2-1}(1-y)^{b_1+b_2-1}\int_0^1 \frac{\Gamma(a_2+b_2+b_1)}{\Gamma(b_1)\Gamma(a_2)\Gamma(b_2)} t^{b_2-1}(1-t)^{b_1-1}\,\mathrm{d}t \\
&= y^{a_2-1}(1-y)^{b_1+b_2-1}\frac{\Gamma(a_2+b_2+b_1)}{\Gamma(b_1)\Gamma(a_2)\Gamma(b_2)}\cdot\frac{\Gamma(b_1)\Gamma(b_2)}{\Gamma(b_1+b_2)} \\
&= \frac{\Gamma(a_2+b_2+b_1)}{\Gamma(b_1)\Gamma(a_2)\Gamma(b_2)} y^{a_2-1}(1-y)^{b_1+b_2-1}
\end{aligned}
$$

故 $Y \sim \beta(a_2, b_1+b_2)$，得证.

20. 设 $X \sim N(0,1), T \sim t(n)$，则存在一个正数 $t_0$，使得

$$
P\{|T| \geq t_0\} \geq P\{|X| \geq t_0\}.
$$

**证明：**令

$$
\begin{aligned}
F(t) &= \mathbb{P}\{|T| \geq t\} - \mathbb{P}\{|X| \geq t\} \\
&= 2(1-F_T(t)) - 2(1-F_X(t)) \\
&= 2F_X(t) - 2F_T(t)
\end{aligned}
$$

由正态分布与 t 分布的对称性可知：

$$
F(0) = 1-1 = 0, \quad \lim_{t\to\infty} F(t) = 2-2 = 0
$$

又 $F'(t) = 2f_X(t) - 2f_T(t)$

根据 t 分布的厚尾性可知：$F'(0) = 2f_X(0) - 2f_T(0) > 0$

且 $\exists x_0 \in (0,+\infty)$, 在 $t \in (0, x_0)$ 上 $F'(t) > 0$, 在 $t \in (x_0, +\infty)$ 上 $F'(t) < 0$

因而对 $\forall t > 0$ 有 $F(t) > 0$.

21. 检验下列分布族是否是指数型分布族:

(1) Poisson 分布族;

(2) 双参数指数分布族;

(3) $\Gamma$ 分布族:$\Gamma(\alpha,\lambda)$;

(4) Cauchy 分布族:$\{f(x,\lambda)=\frac{\lambda}{\pi(\lambda^2+x^2)}:\lambda>0\}$.

**解**: (1) 对于 Poisson 分布族

$$f(x;\lambda)=\frac{\lambda^x}{x!}e^{-\lambda}=e^{-\lambda}\cdot\exp\{x\ln\lambda\}\frac{1}{x!}$$

即 $c(\lambda)=e^{-\lambda}$, $T_1(x)=x$, $c_1(\lambda)=\ln\lambda$, $h(x)=\frac{1}{x!}$

故 Poisson 分布族为指数型分布族.

(2) 对于双参数指数分布族

$$f(x;\lambda,\mu)=\lambda e^{-\lambda(x-\mu)}I_{(x>\mu)}$$

而 $f(x;\lambda,\mu)>0\Rightarrow x>\mu$,即支撑集与参数 $\mu$ 相关

故双参数指数分布族不是指数族

(3) 对于 $\Gamma$ 分布族

$$\begin{aligned}\Gamma(x;\alpha,\lambda)&=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}\\&=\frac{\lambda^\alpha}{\Gamma(\alpha)}e^{(\alpha-1)\ln x-\lambda x}\end{aligned}$$

则 $\theta=\binom{\alpha}{\lambda}$ $c(\theta)=\frac{\lambda^\alpha}{\Gamma(\alpha)}$, $T_1(x)=\ln x$, $T_2(x)=x$, $Q_1(\theta)=\alpha-1$, $Q_2(\theta)=-\lambda$, $h(x)=1$

故 $\Gamma$ 分布族为指数族.

(4) 由于 Cauchy 分布期望不存在,即一阶矩不存在,因此 Cauchy 分布族不是指数型分布族.

22. 设 $X_1,\dots,X_n$ 为来自 PDF 为

$$f(x,\theta)=\theta^{-1}exp\{-|x|/\theta\},x\in\mathbb{R},\theta>0$$

的总体的 iid 样本,证明 $T=\sum_{i+1}^n|X_i|$ 是 $\theta$ 的充分统计量.

**证明**: 由 $X_1,\dots,X_n$ $i.i.d.\sim f(x,\theta)$ 可知

$$f_\theta(x_1,\dots,x_n)=\frac{1}{2^n}\theta^{-n}\exp\left\{-\sum_{i=1}^n|x_i|/\theta\right\}$$

不妨令

$$g_\theta(T(x_1,\ldots,x_n))=\frac{1}{2^n}\cdot\frac{1}{\theta^n}\exp\left\{-\sum_{i=1}^n|x_i|/\theta\right\}$$

$$h(x_1,\ldots,x_n)=1$$

其中 $T(x_1,\ldots,x_n)=\sum\limits_{i=1}^n|x_i|$

故由因子分解定理可知：$T=\sum\limits_{i=1}^n|X_i|$ 为 $\theta$ 的充分统计量.

23. 设 $X_1,\ldots,X_n$ 是来自均匀分布族 $\{U(\theta_1,\theta_2):-\infty<\theta_1<\theta_2<\infty\}$ 的 iid 样本，证明:$(X_{(1)},X_{(2)})$ 是 $(\theta_1,\theta_2)$ 的充分统计量.

**证明：** 由 $X_1,\ldots,X_n\ \ i.i.d.\sim U(\theta_1,\theta_2)$ 知

$$f(x)=\frac{1}{\theta_2-\theta_1}I_{(\theta_1<x<\theta_2)}$$

因而

$$f_\theta(x_1,\ldots,x_n)=\frac{1}{(\theta_2-\theta_1)^n}I_{(\theta_1<X_{(1)}\le X_{(n)}\le\theta_2)}$$

令

$$g_\theta(x_1,\ldots,x_n)=\frac{1}{(\theta_2-\theta_1)^n}I_{(\theta_1<X_{(1)}\le X_{(n)}\le\theta_2)}$$

$$h(x_1,\ldots,x_n)=1$$

故由因子分解定理可知：$(X_{(1)},X_{(n)})$ 为 $(\theta_1,\theta_2)$ 的充分统计量.

24. 设 $Y_i\sim N(\alpha+\beta x_i,\sigma^2)$，$i=1,\ldots,n$，且诸 $Y_i$ 相互独立，其中 $\alpha,\beta\in\mathbb{R}$，$\sigma>0$ 是未知参数，$x_1,\ldots,x_n$ 为已知的常数. 试证明统计量

$$\left(\sum_{i=1}^n Y_i,\sum_{i=1}^n x_iY_i,\sum_{i=1}^n Y_i^2\right)$$

是 $(\alpha,\beta,\sigma^2)$ 的充分统计量.

**证明：** 由 $Y_1,\ldots,Y_n$ 相互独立且 $Y_i\sim N(\alpha+\beta x_i,\sigma^2)$ 知

$$f_{Y_i}(y_i)=(2\pi\sigma^2)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}(y_i-\alpha-\beta x_i)^2\right\}$$

因此

$$\begin{aligned} f_\theta(y_1,\ldots,y_n) &= (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i-\alpha-\beta x_i)^2\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n y_i^2+n\alpha^2+\beta^2\sum_{i=1}^n x_i^2\right.\right. \\ &\quad \left.\left.-2\alpha\sum_{i=1}^n y_i-2\beta\sum_{i=1}^n x_iy_i+2\alpha\beta\sum_{i=1}^n x_i\right]\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\left[n\alpha^2+\beta^2\sum_{i=1}^n x_i^2+2\alpha\beta\sum_{i=1}^n x_i\right]\right\} \\ &\quad \cdot\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n y_i^2-2\alpha\sum_{i=1}^n y_i-2\beta\sum_{i=1}^n x_iy_i\right]\right\} \end{aligned}$$

故 $T=\left(\sum\limits_{i=1}^n y_i^2,\sum\limits_{i=1}^n y_i,\sum\limits_{i=1}^n x_iy_i\right)$ 为 $(\alpha,\beta,\sigma^2)$ 的充分统计量.

25. 设 $X_1,\ldots,X_n$ 为来自 PDF 为

$$f(x,\theta)=\theta/x^2,\ 0<\theta<x<\infty$$

的总体的 iid 样本，试证明 $X_{(1)}$ 是充分统计量.

**证明：** 由 $X_1,X_2,\ldots,X_n\ \ i.i.d.\sim f(x,\theta)$ 知

$$\begin{aligned} f_\theta(x_1,x_2,\ldots,x_n) &= \frac{\theta^n}{x_1^2x_2^2\ldots x_n^2}I_{(x_1,x_2,\ldots,x_n>\theta)} \\ &= \theta^n(x_1x_2\ldots x_n)^{-2}I_{(X_{(1)}>\theta)} \end{aligned}$$

则有

$$h(x_1,\ldots,x_n)=(x_1,\ldots x_n)^{-2},\quad g_\theta(T(x_1,\ldots,x_n))=\theta^nI_{(X_{(1)}>\theta)}$$

因此 $X_{(1)}$ 为 $\theta$ 的充分统计量.

26. 设 $X_1,\ldots,X_n$ 为来自两点分布 $b(1,p)$ 的 iid 样本，试写出 $X_1,\ldots,X_n$ 的联合分布，并指出 $X_1+X_2,X_{(n)},X_n+2p,(X_n-X_1)^2$ 中哪些是统计量，为什么?

**解**：由 $X_i \sim b(1,p)$ 知

$$f)x_i = p^{x_i}(1-p)^{1-x_i} I_{(x_i \in \{0,1\})}$$

又由 $X_1,\ldots,X_n \quad i.i.d.$ 知

$$f(x_1,\ldots,x_n) = p^{\sum\limits_{i=1}^{n} x_i}(1-p)^{n-\sum\limits_{i=1}^{n} x_i} I_{(x_i \in \{0,1\}, i=1,2,\ldots,n)}$$

其中 $X_1+X_2, X_{(n)}, (X_n-X_1)^2$ 为统计量
当 $p$ 为已知数时，$X_n+2p$ 为统计量，反之则不是.

27. 设 $X_1,\ldots,X_n$ 为来自概率密度函数为

$$f(x) = \begin{cases} 2x, & 0<x<1, \\ 0, & \text{其他}, \end{cases}$$

的总体的 iid 样本，试求 $X_{(1)}, X_{(n)}$ 的 PDF.

**解**：由于 $F(x) = \begin{cases} 0, & x \le 0; \\ \int_0^x 2x \,\mathrm{d}x = x^2, & x \in (0,1); \\ 1, & x \ge 1. \end{cases}$
又

$$\begin{aligned} F_1(x) &= \mathbb{P}\left\{X_{(1)} \le x\right\} = 1-\mathbb{P}\left\{X_{(1)} > x\right\} = 1-\mathbb{P}\{X_1,\ldots,X_n > x\} \\ &= 1-(1-F(x))^n \end{aligned}$$

故当 $x \in (0,1)$ 时

$$f_1(x) = F_1'(x) = nf(x)(1-F(x))^{n-1} = 2nx(1-x^2)^{n-1}$$

又 $0 < X_{(1)} < 1$，因此 $f_1(x) = 2nx(1-x^2)^{n-1} I_{(0<x<1)}$
由于 $F_n(x) = \mathbb{P}\left\{X_{(n)} \le x\right\} = F^n(x)$
故当 $x \in (0,1)$ 时

$$f_n(x) = F_n'(x) = nf(x)F^{n-1}(x) = n \cdot 2x \cdot x^{2n-2} = 2nx^{2n-1}$$

同理，由 $0 < X_{(n)} < 1$ 可知

$$f_n(x) = 2nx^{2n-1} I_{(0<x<1)}$$

28. 设 $X_1, \ldots, X_n$ 为来自具有 PDF 为 $f(x)$ 的总体的 iid 样本，试证明：样本极差 $R = X(n) - X(1)$ 的 PDF 为

$$f_R(x) = n(n-1)\int_{-\infty}^{\infty}[F(t+x) - F(t)]^{n-2} f(x+2)f(t)\,\mathrm{d}t,$$

其中 $F$ 为总体的 CDF.

**证明：** 由第 1 章第 10 题可知

$$f(x_{(1)}, x_{(n)}) = f(x_{(1)})f(x_{(n)}) \cdot n(n-1)(F(x_{(n)}) - F(x_{(1)}))^{n-2}$$

则由卷积公式：

$$\begin{aligned} f_R(x) &= \int_{-\infty}^{+\infty} f(x_{(1)})f(x + x_{(1)})n(n-1)(F(x_{(1)} + x) - F(x_{(1)}))^{n-2}\,\mathrm{d}x_{(1)} \\ &\xlongequal{t = x_{(1)}} n(n-1)\int_{-\infty}^{+\infty} f(t)f(t+x)(F(t+x) - F(t))^{n-2}\,\mathrm{d}t \end{aligned}$$

得证.

29. 设 $X_1, \ldots, X_m$ 为来自 $N(\mu, \sigma_1^2)$ 的 iid 样本，$Y_1, \ldots, Y_n$ 为来自 $N(\mu, \sigma_2^2)$ 的 iid 样本，且全样本独立记 $\alpha_1 = S_{1m}^2/(S_{1m}^2 + S_{2n}^2), \alpha_2 = S_{2n}^2/(S_{1m}^2 + S_{2n}^2)$. 试求 $Z = \alpha_1\bar{X} + \alpha_2\bar{Y}$ 的期望.

**解：** 由 $X_1, \ldots, X_m \;\; i.i.d. \sim N(\mu, \sigma_1^2), \;\; Y_1, \ldots, Y_n \;\; i.i.d. \sim N(\mu, \sigma_2^2)$ 且全样本独立

根据 Fisher 定理可知：$\bar{X}, \bar{Y}, S_{1m}, S_{2n}$ 相互独立

因此 $\alpha_1 = \frac{S_{1m}^2}{S_{1m}^2 + S_{2n}^2}$ 与 $\bar{X}$ 独立，$\alpha_2 = \frac{S_{2n}^2}{S_{1m}^2 + S_{2n}^2}$ 与 $\bar{Y}$ 独立

故

$$\begin{aligned} \mathbb{E}Z &= \mathbb{E}\alpha_1 \cdot \mathbb{E}\bar{X} + \mathbb{E}\alpha_2 \cdot \mathbb{E}\bar{Y} \\ &= \mu(\mathbb{E}\alpha_1 + \mathbb{E}\alpha_2) \\ &= \mu\mathbb{E}(\alpha_1 + \alpha_2) \\ &= \mu \end{aligned}$$

30. 设 $X \sim B(n,p)$，对 $k \geq 1$，请证明

$$P\{X \geq k\} = \sum_{i=k}^{n} C_n^i p^i (1-p)^{n-i} = kC_n^k \int_0^p t^{k-1}(1-t)^{n-k}\,\mathrm{d}t.$$

**证明：** 不妨记

$$\begin{aligned} f(p) &= \sum_{i=k}^{n} C_n^i p^i (1-p)^{n-i} \\ g(p) &= kC_n^k \int_0^p t^{k-1}(1-t)^{n-k}\,\mathrm{d}t \end{aligned}$$

则有

$$\begin{aligned} f'(p) &= \sum_{i=k}^{n} \left[ iC_n^i p^{i-1}(1-p)^{n-i} - (n-i)C_n^i p^i (1-p)^{n-i-1} \right] \\ &= kC_n^k p^{k-1}(1-p)^{n-k} + \sum_{i=k+1}^{n} iC_n^i p^{i-1}(1-p)^{n-i} - \sum_{i=k}^{n}(n-i)C_n^i p^i (1-p)^{n-i-1} \\ &= kC_n^k p^{k-1}(1-p)^{n-k} + \sum_{i=k}^{n-1}(i+1)C_n^{i+1} p^i (1-p)^{n-i-1} - \sum_{i=k}^{n-1}(n-i)C_n^i p^i (1-p)^{n-i-1} \\ &= kC_n^k p^{k-1}(1-p)^{n-k} + \sum_{i=k}^{n-1} \left[ (i+1)C_n^{i+1} - (n-i)C_n^i \right] p^i (1-p)^{n-i-1} \\ &= kC_n^k p^{k-1}(1-p)^{n-k} \end{aligned}$$

又 $g'(p) = kC_n^k p^k (1-p)^{n-k}$

故 $f'(p) = g'(p)$

注意到 $f(0) = g(0) = 0$，因此 $f(p) = g(p)$ 恒成立

故原式得证.

31. 设 $X_1, \ldots, X_n$ 为来自总体分布为 $F$ 的 iid 样本，$X_{(1)} \leq \cdots \leq X_{(n)}$ 为其次序统计量，如果 $F(x)$ 连续，则

$$\mathbb{E}\left[F(X_{(i)})\right] = \frac{i}{n+1},\quad Var\left[F(X_{(i)})\right] = \frac{i(n+1-i)}{(n+1)^2(n+2)}.$$

**证明**：不妨设 $X_{(i)}$ 的 PDF 为 $f_i(x)$，则由

$$f_i(x) = \binom{n}{i-1}F^{i-1}(x)\binom{n-i+1}{1}f(x)(1-F(x))^{n-i}$$

因而

$$\begin{aligned}
\mathbb{E}\left[F(X_{(i)})\right] &= \int_{-\infty}^{+\infty} F(x)\binom{n}{i-1}F^{i-1}(x)\binom{n-i+1}{1}f(x)(1-F(x))^{n-i}\,\mathrm{d}x \\
&= \binom{n}{i-1}(n-i+1)\int_{-\infty}^{+\infty} F^{i}(x)(1-F(x))^{n-i}\,\mathrm{d}F(x) \\
&= \binom{n}{i-1}(n-i+1)\int_{0}^{1} t^{i}(1-t)^{n-i}\,\mathrm{d}t \\
&= \binom{n}{n-1}(n-i+1)\beta(i+1, n-i+1) \\
&= \frac{n!(n-i+1)}{(i-1)!(n-i+1)!}\frac{i!(n-i)!}{(n+1)!} \\
&= \frac{i}{n+1}
\end{aligned}$$

同理可知

$$\begin{aligned}
\mathbb{E}\left[F^2(X_{(i)})\right] &= (n-i+1)\binom{n}{i-1}\beta(i+2, n-i+1) \\
&= \frac{i(i+1)}{(n+1)(n+2)}
\end{aligned}$$

因而

$$\begin{aligned}
Var(F(X_{(i)})) &= \mathbb{E}\left[F^2(X_{(i)})\right] - \mathbb{E}\left[F(X_{(i)})\right]^2 \\
&= \frac{i(n+1-i)}{(n+1)^2(n+2)}
\end{aligned}$$

得证.

32. 设 $X_1, \ldots, X_n$ 是 $n$ 个相互独立的随机变量，且 $X_i \sim N(0, \sigma_i^2)$. 记

$$Y = \frac{\sum_{i=1}^{n} X_i/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2}, \quad Z = \sum_{i=1}^{n}\frac{(X_i - Y)^2}{\sigma_i^2},$$

试证明 $Z \sim \chi^2(n-1)$.

**证明**：由 $X_1, \ldots, X_n$ 相互独立且 $X_i \sim N(0, \sigma_i^2)$
不妨令 $T_i = \frac{X_i}{\sigma_i}$，则 $T_1, \ldots, T_n \ \ i.i.d. \sim N(0, 1)$

作正交变换 $\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \frac{\frac{1}{\sigma_1}}{\sqrt{\sum_{i=1}^n \frac{1}{\sigma_i^2}}} & \cdots & \frac{\frac{1}{\sigma_n}}{\sqrt{\sum_{i=1}^n \frac{1}{\sigma_i^2}}} \\ & & \\ & * & \end{pmatrix} \begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_n \end{pmatrix} \triangleq A\mathbf{T}$，其中 $A$ 为正交矩阵

因而 $Y_1 = \sum\limits_{i=1}^n \frac{\frac{T_i}{\sigma_i}}{\sqrt{\sum\limits_{i=1}^n \frac{1}{\sigma_i^2}}} = \sum\limits_{i=1}^n \frac{X_i/\sigma_i^2}{\sqrt{\sum\limits_{i=1}^n \frac{1}{\sigma_i^2}}} = \sqrt{\sum\limits_{i=1}^n \frac{1}{\sigma_i^2}} Y$

又

$$\begin{aligned}
Z &= \sum_{i=1}^n \frac{(X_i - Y)^2}{\sigma_i^2} = \sum_{i=1}^n (\frac{X_i}{\sigma_i} - \frac{Y}{\sigma_i})^2 \\
&= \sum_{i=1}^n \frac{X_i^2}{\sigma_i^2} + \sum_{i=1}^n \frac{1}{\sigma_i^2} \cdot Y^2 - 2Y \sum_{i=1}^n \frac{X_i}{\sigma_i^2} \\
&= \sum_{i=1}^n \frac{X_i^2}{\sigma_i^2} + \sum_{i=1}^n \frac{1}{\sigma_i^2} \cdot Y^2 - 2Y \cdot Y \sum_{i=1}^n \frac{1}{\sigma_i^2} \\
&= \sum_{i=1}^n \frac{X_i^2}{\sigma_i^2} + \sum_{i=1}^n \frac{1}{\sigma_i^2} \cdot Y^2 \\
&= \sum_{i=1}^n T_i^2 - (\sum_{i=1}^n \frac{1}{\sigma_i^2}) \cdot Y_1^2 / (\sum_{i=1}^n \frac{1}{\sigma_i^2})^2 \\
&= \sum_{i=1}^n T_i^2 - Y_1^2
\end{aligned}$$

由正交变换可知：$\sum\limits_{i=1}^n T_i^2 = \sum\limits_{i=1}^n Y_i^2$

因而 $Z = \sum\limits_{i=1}^n Y_i^2 - Y_1^2 = \sum\limits_{i=2}^n Y_i^2$

又考虑到 $T_1, \ldots, T_n \;\; i.i.d. \sim N(0,1)$

因此经过正交变换后 $Y_2, \ldots, Y_n \;\; i.i.d. \sim N(0,1)$

故 $Z \sim \chi^2(n-1)$，得证.

33*. 设 $X_1, \ldots, X_n$ 为来自正态总体 $N(\mu, \sigma^2)$ 的 iid 样本，记

$$\tau = \frac{X_1 - \bar{X}}{S} \sqrt{\frac{n}{n-1}},$$

试证明 $t = \frac{\tau\sqrt{n-2}}{\sqrt{n-1-\tau^2}} \sim t(n-2)$.

**证明：** 由 $\tau = \frac{X_1-\bar{X}}{S}\sqrt{\frac{n}{n-1}}$ 可知

$$\begin{aligned} t &= \frac{\frac{X_1-\bar{X}}{S}\sqrt{\frac{n}{n-1}}\sqrt{n-2}}{\sqrt{n-1-\frac{(X_1-\bar{X})^2}{S^2}\cdot\frac{n}{n-1}}} \\ &= \frac{\sqrt{n}(X_1-\bar{X})\sqrt{n-2}}{\sqrt{(n-1)^2S^2-n(X_1-\bar{X})^2}} \\ &= \frac{\sqrt{n(n-2)}(X_1-\bar{X})}{\sqrt{(n-1)\sum\limits_{i=1}^{n}(X_i-\bar{X})^2-n(X_1-\bar{X})^2}} \end{aligned}$$

注意到

$$\begin{aligned} &\sum_{i=1}^{n}(X_i-\bar{X})^2-\frac{n}{n-1}(X_1-\bar{X})^2 \\ =&\sum_{i=1}^{n}X_i^2-n\bar{X}^2-\frac{n}{n-1}X_1^2-\frac{n}{n-1}\bar{X}^2+\frac{2n}{n-1}X_1\bar{X} \\ =&\sum_{i=2}^{n}X_i^2-\frac{1}{n-1}X_1^2-\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i\right)^2+\frac{2}{n-1}X_1\left(\sum_{i=1}^{n}X_i\right) \\ =&\sum_{i=2}^{n}X_i^2-\frac{1}{n-1}\left[X_1^2+2X_1\left(\sum_{i=1}^{n}X_i\right)+\left(\sum_{i=1}^{n}X_i\right)^2\right] \\ =&\sum_{i=2}^{n}X_i^2-\frac{1}{n-1}\left(\sum_{i=2}^{n}X_i\right)^2 \\ =&\sum_{i=2}^{n}(X_i-\bar{X}')^2 \end{aligned}$$

其中 $\bar{X}'=\frac{1}{n-1}\sum\limits_{i=2}^{n}X_i$

因而 $\left(\sum\limits_{i=1}^{n}(X_i-\bar{X})^2-\frac{n}{n-1}(X_1-\bar{X})^2\right)\Big/\sigma^2\sim\chi^2(n-2)$

又 $X_1-\bar{X}\sim N(0,\frac{n-1}{n}\sigma^2)$ 则有 $\frac{X_1-\bar{X}}{\sqrt{\frac{n-1}{n}\sigma^2}}\sim N(0,1)$

根据 Fisher 定理可知 $\sum\limits_{i=2}^{n}(X_i-\bar{X}')^2$ 与 $\bar{X}'$ 相互独立

因而 $\frac{X_1-\bar{X}}{\sqrt{\frac{n-1}{n}\sigma^2}}\Big/\sqrt{\left(\sum\limits_{i=1}^{n}(X_i-\bar{X})^2-\frac{n}{n-1}(X_1-\bar{X})^2\right)\Big/n-2}\sim t(n-2)$

即 $t\sim t(n-2)$

注：由于一般情形下，两两独立并不能推得相互独立，因此我们给出上述独立性的第二种证明方法

不妨记 $\boldsymbol{x}=(X_1,\ldots,X_n)'$，则

$$\sum_{i=2}^{n}(X_i-\bar{X}')^2=\boldsymbol{x'}\left[\begin{pmatrix}0 & \\ & I_{n-1}\end{pmatrix}-\frac{1}{n-1}\begin{pmatrix}0\\ \boldsymbol{1_{n-1}}\end{pmatrix}\begin{pmatrix}0 & \boldsymbol{1_{n-1}}'\end{pmatrix}\right]\boldsymbol{x}\triangleq\boldsymbol{x'}A\boldsymbol{x}$$

$$X_1-\bar{X}=\left[\begin{pmatrix}1 & \boldsymbol{0_{n-1}}\end{pmatrix}-\frac{1}{n}\boldsymbol{1_n}'\right]\boldsymbol{x}$$

又由 $\boldsymbol{x}\sim N_n(0,I_n)$，$\boldsymbol{x'}A\boldsymbol{x}\sim\chi^2(n-1)$

且有 $Cov(AX,BX)=AB'=0$

根据 $\chi^2$ 分布与正态分布的独立性可知: $X_1-\bar{X}$ 与 $\sum\limits_{i=2}^{n}(X_i-\bar{X}')^2$ 相互独立.

34*. 设 $\begin{pmatrix}X_1\\ Y_1\end{pmatrix},\ldots,\begin{pmatrix}X_n\\ Y_n\end{pmatrix}$ 为来自二维正态 $N(\mu,\Sigma)$ 的 iid 样本，其中 $\mu=(\mu_1,\mu_2)^{\mathbf{T}}$，$\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$. 定义二样本间的相关系数为

$$r(X,Y)=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right]^{1/2}},$$

试证明: 如果 $\rho=0$，则统计量

$$t=\sqrt{n-2}\frac{r(X,Y)}{\sqrt{1-r^2(X,Y)}}\sim t(n-2).$$

**证明:** 对于给定的 $Y_1,\ldots,Y_n$，不妨对 $Y_1,\ldots,Y_n$ 中心标准化，即令 $Z_i=\frac{Y_i-\bar{Y}}{\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}$

注意到 $\sum_{i=1}^{n}\bar{X}Z_i=\bar{X}(\sum_{i=1}^{n}Z_i)=0$

故 $r(X,Y)=\frac{\sum_{i=1}^{n}X_iZ_i}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}}$，因此

$$\begin{aligned}t&=\sqrt{n-2}\frac{r(X,Y)}{\sqrt{1-r^2(X,Y)}}\\&=\frac{\sqrt{n-2}\frac{\sum_{i=1}^{n}X_iZ_i}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}}}{\sqrt{1-\frac{(\sum_{i=1}^{n}X_iZ_i)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}}}\\&=\frac{\sqrt{n-2}\sum_{i=1}^{n}X_iZ_i}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2-(\sum_{i=1}^{n}X_iZ_i)^2}}\end{aligned}$$

令 $\boldsymbol{x}=(X_1,\ldots,X_n)'$，$\boldsymbol{z}=(Z_1,\ldots,Z_n)'$

因此

$$t = \frac{\sqrt{n-2}\boldsymbol{z}'\boldsymbol{x}}{\sqrt{\boldsymbol{x}'(I-\frac{1}{n}\mathbf{1}_{\boldsymbol{n}}\mathbf{1}_{\boldsymbol{n}}')-\boldsymbol{x}'\boldsymbol{z}\boldsymbol{z}'\boldsymbol{x}}}$$

不妨记 $A = I - \frac{1}{n}\mathbf{1}_{\boldsymbol{n}}\mathbf{1}_{\boldsymbol{n}}' - \boldsymbol{z}\boldsymbol{z}'$

可验证 $A$ 为对称幂等矩阵，即 $A^2 = A$, $A' = A$

注意到 $\boldsymbol{z}'A = 0$，因此由正态分布与 $\chi^2$ 分布的独立性可知:

$\boldsymbol{z}'\boldsymbol{x}$ 与 $\boldsymbol{x}'(I-\frac{1}{n}\mathbf{1}_{\boldsymbol{n}}\mathbf{1}_{\boldsymbol{n}}')-\boldsymbol{x}'\boldsymbol{z}\boldsymbol{z}'\boldsymbol{x}$ 独立

又

$$\begin{aligned} rank(A) &= tr(A) \\ &= tr(I) - \frac{1}{n}tr(\mathbf{1}_{\boldsymbol{n}}\mathbf{1}_{\boldsymbol{n}}') - \frac{1}{n}tr(\boldsymbol{z}\boldsymbol{z}') \\ &= n - \frac{1}{n}\cdot n - \frac{1}{n}\cdot n \\ &= n-2 \end{aligned}$$

故 $\boldsymbol{x}'A\boldsymbol{x}/\sigma_1^2 \sim \chi^2(n-2)$

又 $\boldsymbol{x}/\sigma_1 \sim N_n(\mu_1/\sigma_1\mathbf{1}_{\boldsymbol{n}}, I)$

注意到，当 $Y_1,\ldots,Y_n$ 给定时

$$\mathbb{E}\left[\boldsymbol{z}'\boldsymbol{x}/\sigma_1\right] = \boldsymbol{z}'\mathbb{E}\left[\boldsymbol{x}/\sigma_1\right] = \mu_1/\sigma_1\boldsymbol{z}'\mathbf{1}_{\boldsymbol{n}} = 0$$

$$Var(\boldsymbol{z}'\boldsymbol{x}/\sigma_1) = \boldsymbol{z}'Cov(\boldsymbol{x}/\sigma_1)\boldsymbol{z} = \boldsymbol{z}'\boldsymbol{z} = 1$$

因此 $\boldsymbol{z}'\boldsymbol{x}/\sigma_1 \sim N(0,1)$

故 $t \sim t(n-2)$

又由于上述证明对于任意 $Y_1,\ldots,Y_n$ 均成立，因此 $t \sim t(n-2)$.

35. 证明服从非中心 $t$ 分布的随机变量的平方服从非中心 $F$ 分布.

**证明**: 不妨设 $T \sim t(n,\delta)$ $(\delta \neq 0)$

设 $\xi \sim N(\delta,1)$, $\eta \sim \chi^2(n)$ 且 $\xi$ 与 $\eta$ 相互独立

则 $T \overset{d}{=} \frac{\xi}{\sqrt{\eta/n}}$，因而 $T^2 \overset{d}{=} \frac{\xi^2}{\eta/n}$

注意到 $\xi^2 \sim \chi^2(1,\delta^2)$ 且与 $\eta$ 相互独立

因此 $\frac{\xi^2}{\eta/n} \sim F(1,n,\delta^2)$

即 $T^2$ 服从非中心 $F$ 分布.

36. 一个总体有 $N$ 个元素，其指标值分别为 $a_1 > a_2 > \cdots > a_N$. 指定自然数 $M < N$, $n < N$，并设 $m = nM/N$ 为整数. 在 $(a_{M+1},\ldots,a_N)$ 中不放

回的抽取 $n-m$ 个. 写出所得样本的分布.

**解:**

$$p(a_{i1},a_{i2},\ldots,a_{in})=\begin{cases}\frac{1}{\binom{M}{m}\binom{N-M}{n-m}}, & \#\{j:a_{ij}\le a_M\}=m;\\ 0, & \text{否则}.\end{cases}$$

37. 设 $X_1,\ldots,X_n$ 为来自区间 $(a,b)$ 上均匀分布的 IID 样本，$-\infty<a<b<\infty$. 证明对于任何 $a$ 和 $b$，$(X_{(i)}-X_{(1)})/(X_{(n)}-X_{(1)})$, $i=2,\ldots,n-1$ 独立于 $(X_{(1)},X_{(n)})$.

**证明:** 令 $Y_i=\frac{X_{(i)}-a}{b-a}$
由 $X_1,\ldots,X_n$ $i.i.d.\sim U(a,b)$ 可知 $Y_1,\ldots,Y_n$ $i.i.d.\sim U(0,1)$
而 $\frac{X_{(i)}-X_{(1)}}{X_{(n)}-X_{(1)}}=\frac{Y_{(i)}-Y_{(1)}}{Y_{(n)}-Y_{(1)}}$ 分布显然与参数 $a,b$ 无关
故 $\frac{X_{(i)}-X_{(1)}}{X_{(n)}-X_{(1)}}$ $(i=2,\ldots,n-1)$ 均为辅助统计量
又对于样本 $X_1,\ldots,X_n$

$$f(x_1,\ldots,x_n;a,b)=\frac{1}{(b-a)^n}I_{(a<X_{(1)}\le X_{(n)}<b)}$$

故 $T=(X_{(1)},X_{(n)})$ 为充分统计量
而统计量 $T$ 的完备性证明可见第二章 26 题
故 $T$ 为充分完备统计量，由 Basu 定理知
对于任何 a 和 b，$\frac{X_{(i)}-X_{(1)}}{X_{(n)}-X_{(1)}}$ $(i=2,\ldots,n-1)$ 独立于 $(X_{(1)},X_{(n)})$.

**定理:** (Basu 定理) 设 $\mathcal{F}=\{f(x,\theta),\ \theta\in\Theta\}$ 为一分布族，$\Theta$ 是参数空间. 样本 $\boldsymbol{X}=(X_1,\ldots,X_n)$ 是从分布族 $\mathcal{F}$ 中抽取的简单样本，设 $T(\boldsymbol{X})$ 是一由界完备统计量，且是充分统计量. 若 $r.v.$ $V(\boldsymbol{X})$ 的分布与 $\theta$ 无关，则对任何 $\theta\in\Theta$，$V(\boldsymbol{X})$ 与 $T(\boldsymbol{X})$ 独立.

38. 设 $X_1,\ldots,X_n$ 为来自总体 $P_\theta\in\{P_\theta:\theta\in\Theta\}$ 的 IID 样本. 在下列的几种情况下，找出一个与 $\theta\in\Theta$ 相同维数的充分统计量.

(1) $P_\theta$ 是泊松分布 $P(\theta)$, $\theta\in(0,\infty)$;

(2) $P_\theta$ 是负二项分布 $NB(\theta,r)$，且 $r$ 已知，$\theta\in(0,1)$;

(3) $P_\theta$ 是指数分布 $E(\theta)$, $\theta\in(0,\infty)$.

**解**: (1) 由 $X_1,\dots,X_n\ \ i.i.d.\sim P(\theta)$ 可知

$$f(x_i;\theta)=\frac{\theta^{x_i}}{x_i!}e^{-\theta}$$

故

$$\begin{aligned}f_\theta(x_1,\dots,x_n)&=\frac{\theta^{x_1+\cdots+x_n}}{x_1!\dots x_n!}e^{-n\theta}\\&=\theta^{\sum\limits_{i=1}^n x_i}e^{-n\theta}\cdot\frac{1}{x_1!\dots x_n!}\end{aligned}$$

由因子分解定理知 $T=\sum\limits_{i=1}^n X_i$ 为充分统计量.

(2) 由 $X_1,\dots,X_n\ \ i.i.d.\sim NB(\theta,r)$ 可知

$$f(x;\theta)=\tbinom{x+r-1}{r-1}\theta^r(1-\theta)^x$$

故

$$\begin{aligned}f_\theta(x_1,\dots,x_n)&=\tbinom{x_1+r-1}{r-1}\dots\tbinom{x_n+r-1}{r-1}\theta^{nr}(1-\theta)^{x_1+\cdots+x_n}\\&=\theta^{nr}(1-\theta)^{\sum\limits_{i=1}^n x_i}\tbinom{x_1+r-1}{r-1}\dots\tbinom{x_n+r-1}{r-1}\end{aligned}$$

由因子分解定理知 $T=\sum\limits_{i=1}^n X_i$ 为充分统计量.

(3) 由 $X_1,\dots,X_n\ \ i.i.d.\sim E(\theta)$ 可知

$$f_\theta(x)=\theta e^{-\theta x}$$

故

$$\begin{aligned}f_\theta(x_1,\dots,x_n)&=\theta^n e^{-\theta(x_1+\cdots+x_n)}\\&=\theta^n e^{-\theta\sum\limits_{i=1}^n x_i}\end{aligned}$$

由因子分解定理知 $T=\sum\limits_{i=1}^n X_i$ 为充分统计量.

39. 设 $X_1,\dots,X_n$ 为来自 $N(\theta,\theta^2)(\theta>0)$ 的 iid 样本. 问 $\bar{X}$ 是否仍为充分统计量?

**解**: 由 $X_1,\dots,X_n\ \ i.i.d.\sim N(\theta,\theta^2)\ \ (\theta>0)$ 知

$$f(x;\theta)=(2\pi\theta^2)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\theta^2}(x-\theta)^2\right\}$$

因此 $f_\theta(x_1,\dots,x_n)=(2\pi\theta^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\theta^2}\left[\sum\limits_{i=1}^n x_i^2-2\theta\sum\limits_{i=1}^n x_i+n\theta^2\right]\right\}$

又 $\bar{X}\sim N(\theta,\frac{\theta^2}{n})$

因而 $f_{\bar{X}}(t)=(2\pi\frac{\theta^2}{n})^{-\frac{1}{2}}\exp\left\{-\frac{n}{2\theta^2}(t-\theta)^2\right\}$

$$
\begin{aligned}
f_{\boldsymbol{x}|\bar{X}}(x_1,\dots,x_n|t)&=\frac{f(x_1,\dots,x_{n-1},nt-\sum_{i=1}^{n-1}x_i)}{f_{\bar{X}}(t)}\\
&=\frac{(2\pi\theta^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\theta^2}\left[\sum\limits_{i=1}^{n-1}x_i^2+(nt-\sum_{i=1}^{n-1}x_i)^2-2\theta nt+n\theta^2\right]\right\}}{(2\pi\frac{\theta^2}{n})^{-\frac{1}{2}}\exp\left\{-\frac{n}{2\theta^2}(t-\theta)^2\right\}}\\
&=n^{-\frac{1}{2}}(2\pi\theta^2)^{-\frac{n-1}{2}}\exp\left\{-\frac{1}{2\theta^2}\left[\sum_{i=1}^{n-1}x_i^2+(nt-\sum_{i=1}^{n-1}x_i)^2-nt^2\right]\right\}
\end{aligned}
$$

仍与参数 $\theta$ 相关

故 $\bar{X}$ 不是充分统计量.

40. 设 $X_1,\dots,X_n\sim N(0,\sigma^2)$，$\bar{X}$ 为样本均值，$\xi=f(X_1,\dots,X_n)$ 满足条件 $f(X_1+c,\dots,X_n+c)=f(X_1,\dots,X_n)$(对任何常数 $c$). 证明：$\xi$ 与 $\bar{X}$ 独立.

**法 1:**Basu 定理

由于 $\bar{X}$ 是正态分布下 $\mu$ 的充分完备统计量，但该分布中只含有参数 $\sigma^2$，因此，我们不妨加强所证结论，将 0 换成 $\mu$，人为引进参数.

即证明 $\xi$ 与 $\bar{X}$ 在 $X_1 dots,X_n\ \ i.i.d.\sim N(\mu,\sigma^2)$ 下独立.

现将 $\sigma^2$ 固定于 $\sigma_0^2$，则 $\bar{X}$ 是分布 $N(\mu,\sigma^2)$ 中参数 $\mu$ 的充分完备统计量，同时

$$
\begin{aligned}
f(X_1,X_2,\dots,X_n)&=f(X_1-\mu,X_2-\mu,\dots,X_n-\mu)\\
&\triangleq f(Z_1,Z_2,\dots,Z_n)
\end{aligned}
$$

而 $Z_1,Z_2,\dots,Z_n\ \ i.i.d.\sim N(0,\sigma_0^2)$ 与 $\mu$ 无关. 因此 $f(X_1,X_2,\dots,X_n)$ 的分布与 $\mu$ 无关，为辅助统计量.

由 Basu 定理, 知 $\bar{X}$ 与 $\xi$ 独立, 又由于 $\sigma^2$ 可以是任意的, 故对任意 $\mu\in\mathbb{R}$ 以及 $\sigma^2>0$，$\bar{X}$ 和 $\xi$ 都是独立的.

**法 2:** 正交变换

作正交变换 $\begin{pmatrix}Y_1\\ \vdots\\ Y_n\end{pmatrix}=\begin{pmatrix}\frac{1}{\sqrt{n}}&\cdots&\frac{1}{\sqrt{n}}\\ a_{21}&\cdots&a_{2n}\\ \vdots&&\vdots\\ a_{n1}&\cdots&a_{nn}\end{pmatrix}\begin{pmatrix}X_1\\ \vdots\\ X_n\end{pmatrix}$

因而 $Y_1,\ldots,Y_n$ 相互独立且 $Y_1=\sqrt{n}\bar{X}$，不妨再作逆变换

$$\begin{pmatrix}X_1\\ \vdots\\ X_n\end{pmatrix}=\begin{pmatrix}\frac{1}{\sqrt{n}} & a_{21} & \ldots & a_{n1}\\ \vdots & \vdots & & \vdots\\ \frac{1}{\sqrt{n}} & a_{2n} & \ldots & a_{nn}\end{pmatrix}\begin{pmatrix}Y_1\\ \vdots\\ Y_n\end{pmatrix}$$

即 $\begin{cases}X_1=\frac{1}{\sqrt{n}}Y_1+a_{21}Y_2+\cdots+a_{n1}Y_n\\ \qquad\qquad\cdots\\ X_n=\frac{1}{\sqrt{n}}Y_1+a_{2n}Y_2+\cdots+a_{nn}Y_n\end{cases}$

则 $\xi=f(X_1,\ldots,X_n)=f\begin{pmatrix}\frac{1}{\sqrt{n}}Y_1+a_{21}Y_2+\cdots+a_{n1}Y_n\\ \vdots\\ \frac{1}{\sqrt{n}}Y_1+a_{2n}Y_2+\cdots+a_{nn}Y_n\end{pmatrix}$

由 $f$ 的平移不变性可知:

$$\xi=\begin{pmatrix}a_{21}Y_2+\cdots+a_{n1}Y_n\\ \vdots\\ a_{2n}Y_2+\cdots+a_{nn}Y_n\end{pmatrix}$$

41. 非中心 $\chi$ 变量 $\xi=\sum_{i=1}^n(X_i+a_i^2)^2$, $X_1,\ldots,X_n\sim N(0,1)$, $a_1,\ldots,a_n$ 为常数. 证明: $\xi$ 的分布只依赖于 $n$ 和 $\delta=\sqrt{\sum_{i=1}^n a_i^2}$(提示: 作为正交变换，使 $Y_1=\sum_{j=1}^n a_jX_j/\delta$).

**证明:** 作正交变换

$$\begin{pmatrix}Y_1\\ Y_2\\ \vdots\\ Y_n\end{pmatrix}=\begin{pmatrix}\frac{a_1}{\delta} & \frac{a_2}{\delta} & \cdots & \cdots & \frac{a_n}{\delta}\\ & & & & \\ & & * & & \\ & & & & \end{pmatrix}=\begin{pmatrix}X_1\\ X_2\\ \vdots\\ X_n\end{pmatrix}$$

而

$$\begin{aligned}\xi&=\sum_{i=1}^n X_i^2+2\sum_{i=1}^n a_iX_i+\sum_{i=1}^n a_i^2\\ &=\sum_{i=1}^n Y_i+2\delta Y_1+\delta^2\end{aligned}$$

又 $X_1,\ldots,X_n\ \ i.i.d.\sim N(0,1)$

故正交变换得 $Y_1,\ldots,Y_n\ \ i.i.d.\sim N(0,1)$

则由 $\xi = \sum\limits_{i=1}^{n} Y_i^2 + 2\delta Y_1 + \delta^2$ 知 $\xi$ 的分布只依赖于 $n$ 和 $\delta$.

补充题

1. 设 $X_1, X_2, \ldots, X_n$ 为来自某总体 $F$ 的 i.i.d. 样本，且二阶矩有限，试证明样本方差 $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 的期望等于总体方差 $\sigma^2$，即 $ES_n^2 = \sigma^2$.

**证明：** 由 $X_1, \ldots, X_n$ $i.i.d.$，知

$$\mathbb{E}S_n^2 = \frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right] = \frac{n}{n-1}\left[\mathbb{E}X_1^2 - \mathbb{E}\bar{X}^2\right]$$

而

$$Var(\bar{X}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}$$

$$\mathbb{E}\bar{X} = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mathbb{E}X_1$$

故

$$\mathbb{E}\bar{X}^2 = Var\bar{X} + (\mathbb{E}\bar{X})^2 = \frac{\sigma^2}{n} + (\mathbb{E}X_1)^2$$

$$\mathbb{E}X_1^2 = VarX_1 + (\mathbb{E}X_1)^2 = \sigma^2 + (\mathbb{E}X_1)^2$$

因而

$$\begin{aligned}\mathbb{E}S_n^2 &= \frac{n}{n-1}\left[(\sigma^2 + (\mathbb{E}X_1)^2) - \left(\frac{\sigma^2}{n} + (\mathbb{E}X_1)^2\right)\right] \\ &= \frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2 \\ &= \sigma^2\end{aligned}$$

2. 证明：对于任意常数 $c, d$，有

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