统计学基础 I : 数理统计 Assignment 4

姓名: 雍崔扬
学号: 21307140051
习题: E1.30, E1.39, E1.40(3)(5)(6), E2.24(4)(8)(10), 补充题 3

Problem 1 (习题 1.30)

设 $X = (X_1, \dots, X_n)$ 为取自双参数指数分布的简单随机样本,
分布函数为 $F(x) = 1 - \exp\left(-\frac{x-\mu}{\sigma}\right) I_{(\mu,\infty)}(x), \quad (\mu \in \mathbb{R}, \sigma > 0)$

(1) 试证明: 若 $\begin{cases} Y_1 = \frac{n(X_{(1)}-\mu)}{\sigma} \\ Y_r = \frac{(n+1-r)(X_{(r)}-X_{(r-1)})}{\sigma}, 2 \leq r \leq n \end{cases}$
则 $Y_1, \dots, Y_n$ 相互独立，且具有相同分布密度 $f_Y(y) = e^{-y} I_{(0,\infty)}(y)$

- **Lemma:**

> **定理** $1.3.4$: **(次序统计量的联合概率密度函数, 数理统计讲义 命题** $1.4.7$)
> 设 $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ 为对应于简单随机样本 $X = (X_1, X_2, \dots, X_n)$ 的次序统计量,
> (我们可以看作存在映射关系 $T(X_1, X_2, \dots, X_n) = (X_{(1)}, X_{(2)}, \dots, X_{(n)})$)
> 总体分布具有分布函数 F 和概率密度函数 f.
> 则对于任意 $\begin{cases} 1 \leq r \leq n \\ 1 \leq j_1 < j_2 < \cdots < j_r \leq n \end{cases}$
> $(X_{(j_1)}, X_{(j_2)}, \dots, X_{(j_r)})$ 具有联合概率密度函数:
>
> $$\begin{aligned} & f_{X_{(j_1)},X_{(j_2)},\dots,X_{(j_r)}}(y_{j_1}, y_{j_2}, \dots, y_{j_r}) \\ & = \frac{n!}{(j_1-1)!(j_2-j_1-1)!\dots(j_r-j_{r-1}-1)!(n-j_r)!} \\ & \quad \times [F(y_{j_1})]^{j_1-1}[F(y_{j_2})-F(y_{j_1})]^{j_2-j_1-1}\cdots[F(y_{j_r})-F(y_{j_{r-1}})]^{j_r-j_{r-1}-1}[1-F(y_{j_r})]^{n-j_r} \\ & \quad \times f(y_{j_1})f(y_{j_2})\cdots f(y_{j_r}) \\ & \quad \times I(y_{j_1} < y_{j_2} < \cdots < y_{j_r}) \end{aligned}$$
>
> - **推论** 1: **(全体次序统计量的联合概率密度函数)**
>   全体次序统计量 $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ 的联合概率密度函数为:
>
> $$f_{X_{(1)},X_{(2)},\dots,X_{(n)}}(y_1, y_2, \dots, y_n) = n! \prod_{i=1}^n f(y_i) I(y_1 < y_2 < \cdots < y_n)$$
>
>   如果不将其视为推论的话，那么上式的得到是由于:
>   - ① 次序统计量 $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ 的取值是 $y_1, y_2, \dots, y_n$,
>   当且仅当 $X_1, X_2, \dots, X_n$ 的取值是 $y_1, y_2, \dots, y_n$ 的 $n!$ 个排列中的任意一个.
>   - ② 对于 $1, 2, \dots, n$ 的任意排列 $i_1, i_2, \dots, i_n$,
> $X_1, X_2, \dots, X_n$ 的取值为 $y_{i_1}, y_{i_2}, \dots, y_{i_n}$ 的概率密度都是 $\prod_{j=1}^n f(y_{i_j}) = \prod_{j=1}^n f(y_j)$

Solution:
根据引理可知，$(X_{(1)}, \dots, X_{(n)})$ 具有联合概率密度:

$$\begin{aligned} f_{X_{(1)},\dots,X_{(n)}}(x_1, \dots, x_n) &= n! \prod_{i=1}^n f(x_i) I(\mu < x_1 < \cdots < x_n) \\ &= n! \prod_{i=1}^n \frac{1}{\sigma} \exp\left(-\frac{x_i-\mu}{\sigma}\right) I(\mu < x_1 < \cdots < x_n) \\ &= \frac{n!}{\sigma^n} \exp\left\{-\frac{1}{\sigma}\sum_{i=1}^n (x_i-\mu)\right\} I(\mu < x_1 < \cdots < x_n) \end{aligned}$$

记 $Z_r = \frac{X_{(r)}-\mu}{\sigma} \quad (1 \leq r \leq n)$
则 $\begin{cases} Y_1 = \frac{n(X_{(1)}-\mu)}{\sigma} \\ Y_r = \frac{(n+1-r)(X_{(r)}-X_{(r-1)})}{\sigma}, 2 \leq r \leq n \end{cases}$ 化为 $\begin{cases} Y_1 = nZ_1 \\ Y_r = (n+1-r)(Z_r - Z_{r-1}), 2 \leq r \leq n \end{cases}$
且 $(Z_1, \dots, Z_n)$ 具有联合概率密度:

$$
\begin{aligned}
f_{Z_1,\ldots,Z_n}(z_1,\ldots,z_n) &= |\det(\sigma I_n)| \cdot f_{X_{(1)},\ldots,X_{(n)}}(\sigma z_1+\mu,\ldots,\sigma z_n+\mu) \\
&= \sigma^n \cdot \frac{n!}{\sigma^n} \exp\left\{-\frac{1}{\sigma}\sum_{i=1}^n(\sigma z_i+\mu-\mu)\right\} I(\mu<\sigma z_1+\mu<\cdots<\sigma z_n+\mu) \\
&= n!\exp\left\{-\sum_{i=1}^n z_i\right\} I(0<z_1<\cdots<z_n)
\end{aligned}
$$

注意到 $\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} n & & & \\ -(n-1) & n-1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix}$

记线性变换矩阵为 J，

容易看出 $\det(J) = \prod_{r=1}^n (n+1-r) = n!$ 且逆矩阵为

$$
J^{-1} = \begin{bmatrix} \frac{1}{n} & & & & \\ \frac{1}{n} & \frac{1}{n-1} & & & \\ \vdots & \vdots & \ddots & & \\ \frac{1}{n} & \frac{1}{n-1} & \cdots & \frac{1}{2} & \\ \frac{1}{n} & \frac{1}{n-1} & \cdots & \frac{1}{2} & 1 \end{bmatrix}
$$

说明 $Z_r = \sum_{i=1}^r \frac{1}{n+1-i} Y_i$, $1 \le r \le n$

因此 $(Y_1,\ldots,Y_n)$ 具有联合概率密度:

$$
\begin{aligned}
f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n) &= |\det(J)|^{-1} \cdot f_{Z_1,\ldots,Z_n}\left(\frac{1}{n}y_1, \frac{1}{n}y_1+\frac{1}{n-1}y_2, \ldots, \frac{1}{n}y_1+\frac{1}{n-1}y_2+\cdots+\frac{1}{1}y_n\right) \\
&= \frac{1}{n!}\cdot n!\exp\left\{-\sum_{r=1}^n\sum_{i=1}^r \frac{1}{n+1-i}y_i\right\} I(0<\frac{1}{n}y_1<\frac{1}{n}y_1+\frac{1}{n-1}y_2<\cdots<\sum_{i=1}^n \frac{1}{n+1-i}y_i) \\
&= \exp\left\{-\sum_{i=1}^n y_i\right\} I(y_i>0 \text{ for all } i=1,\ldots,n) \\
&= \prod_{i=1}^n \exp(-y_i) I(y_i>0)
\end{aligned}
$$

得证 $Y_1,\ldots,Y_n$ 相互独立，且具有相同分布密度 $f_Y(y) = \exp(-y)I(y>0) = e^{-y}I_{(0,\infty)}(y)$

也就是说 $Y_1,\ldots,Y_n \overset{\text{i.i.d.}}{\sim} \exp(1) = \text{Gamma}(1,1)$

---

(2) 给定 $r \le n$，求截尾样本 $(X_{(1)},\ldots,X_{(r)})$ 的联合概率密度，

并证明 $\begin{cases} \sum_{i=r+1}^n X_{(i)} - (n-r)X_{(r)} \overset{\text{d}}{=} \frac{\sigma}{2}\chi^2(2(n-r)) \\ \sum_{i=1}^r X_{(i)} - n\mu + (n-r)X_{(r)} \overset{\text{d}}{=} \frac{\sigma}{2}\chi^2(2r) \end{cases}$

Solution:

根据引理可知，$(X_{(1)},\ldots,X_{(r)})$ 的联合概率密度为:

$$
\begin{aligned}
& f_{X_{(1)},\ldots,X_{(r)}}(x_1,\ldots,x_r) \\
&= \frac{n!}{1!\cdots 1!(n-r)!}(1-F(x_r))^{n-r} f(x_1)\cdots f(x_r) I(\mu<x_1<\cdots<x_r) \\
&= \frac{n!}{(n-r)!}\left[1-1+\exp\left(-\frac{x_r-\mu}{\sigma}\right)I(x_r>\mu)\right]^{n-r}\prod_{i=1}^r \frac{1}{\sigma}\exp\left(-\frac{x_i-\mu}{\sigma}\right) I(\mu<x_1<\cdots<x_r) \\
&= \frac{n!}{(n-r)!}\exp\left\{-(n-r)\frac{x_r-\mu}{\sigma}\right\}\prod_{i=1}^r \frac{1}{\sigma}\exp\left(-\frac{x_i-\mu}{\sigma}\right) I(\mu<x_1<\cdots<x_r)
\end{aligned}
$$

根据第 (1) 问的记号，我们有 $Z_r = \frac{X_{(r)}-\mu}{\sigma}$ $(1 \le r \le n)$

则 $\begin{cases} Y_1 = \frac{n(X_{(1)}-\mu)}{\sigma} \\ Y_r = \frac{(n+1-r)(X_{(r)}-X_{(r-1)})}{\sigma}, 2 \le r \le n \end{cases}$ 化为 $\begin{cases} Y_1 = nZ_1 \\ Y_r = (n+1-r)(Z_r - Z_{r-1}), 2 \le r \le n \end{cases}$

其逆变换为 $Z_r = \sum_{i=1}^r \frac{1}{n+1-r} Y_i$, $1 \le r \le n$

因此我们知道:

要证明 $\begin{cases} \sum_{i=r+1}^n X_{(i)} - (n-r)X_{(r)} \overset{\mathrm{d}}{=} \frac{\sigma}{2}\chi^2(2(n-r)) \\ \sum_{i=1}^r X_{(i)} - n\mu + (n-r)X_{(r)} \overset{\mathrm{d}}{=} \frac{\sigma}{2}\chi^2(2r) \end{cases}$
等价于证明 $\begin{cases} \sum_{i=r+1}^n Z_i - (n-r)Z_r \overset{\mathrm{d}}{=} \frac{1}{2}\chi^2(2(n-r)) \\ \sum_{i=1}^r Z_i + (n-r)Z_r \overset{\mathrm{d}}{=} \frac{1}{2}\chi^2(2r) \end{cases}$
也就等价于证明 $\begin{cases} \sum_{i=r+1}^n Y_i \overset{\mathrm{d}}{=} \frac{1}{2}\chi^2(2(n-r)) \\ \sum_{i=1}^r Y_i \overset{\mathrm{d}}{=} \frac{1}{2}\chi^2(2r) \end{cases}$
根据第 (1) 问的 $Y_1,\dots,Y_n$ 的独立性，
也就等价于证明 $\begin{cases} 2\sum_{i=r+1}^n Y_i \overset{\mathrm{d}}{=} \chi^2(2(n-r)) = \mathrm{Gamma}(n-r,\frac{1}{2}) \\ 2\sum_{i=1}^r Y_i \overset{\mathrm{d}}{=} \chi^2(2r) = \mathrm{Gamma}(r,\frac{1}{2}) \end{cases}$
也就等价于证明 $2Y_1, 2Y_2, \dots, 2Y_n \overset{\text{i.i.d.}}{\sim} \mathrm{Gamma}(1,\frac{1}{2})$

根据第 (1) 问的 $Y_1,\dots,Y_n \overset{\text{i.i.d.}}{\sim} \exp(1) = \mathrm{Gamma}(1,1)$
我们很容易证明这一点:

- 对于 $Y \sim \mathrm{Gamma}(1,1)$，$2Y$ 的概率密度为:

$$\begin{aligned} f_{2Y}(y) &= \frac{1}{2}\cdot f_Y(y/2)I(y/2>0) \\ &= \frac{1}{2}e^{-\frac{1}{2}y}I(y>0) \\ &= \mathrm{P}\left\{\mathrm{Gamma}\left(1,\frac{1}{2}\right)=y\right\} \end{aligned}$$

  因此 $2Y \sim \mathrm{Gamma}(1,\frac{1}{2})$.

因此我们有 $2Y_1, 2Y_2, \dots, 2Y_n \overset{\text{i.i.d.}}{\sim} \mathrm{Gamma}(1,\frac{1}{2})$
根据前文的推导，命题得证.

Problem 2 (习题 1.39)

对于下列正态总体的简单随机样本 $X=(X_1,\dots,X_n)$，
写出参数的统计量，并利用因子化定理证明其为充分统计量:

- > **定理** $1.3.5$: **(因子化定理, 数理统计讲义 命题** $1.5.6$**)**
  > 设样本的可能分布族为 $\mathscr{F}_X = \{f_X(x;\theta):\theta\in\Theta\}$
  > 其中 $f_X(x;\theta)$ 为分布密度或离散的概率分布，
  > 则统计量 $T=T(X)$ 为分布族 $\mathscr{F}_X$ 参数 θ 的**充分统计量**的**充要条件**是:
  > 对于任意 $\theta\in\Theta$，$f_X(x;\theta)$ 都可分解为 $g(T(x);\theta)\cdot h(x)$，
  > 其中 $h(x)$ 是与 θ 无关的**非负函数**.
  >
  > - 根据下面的必要性证明，
  >   我们可以知道 $g(T(x);\theta)$ 可以是 $f_T(T(x);\theta)$，
  >   而 $h(x)$ 可以是 $\mathrm{P}\{X=x|T(X)=T(x)\}$
  >   (由于 $T(X)$ 此时是充分统计量，故这个条件概率与 θ 无关)
  >   但实际应用时可能相差一些因式 (例如常数).

(1) $\{N(\mu,1):\mu\in\mathbb{R}\}$

Solution:
样本均值 $\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i$ 是参数 μ 的充分统计量.

利用因子化定理证明:
记 $x=(x_1,\dots,x_n)$，则我们有:

$$
\begin{aligned}
\mathrm{P}\{X=x\} &= \mathrm{P}\{X_1=x_1,\ldots,X_n=x_n\} \\
&= \prod_{i=1}^n \mathrm{P}\{N(\mu,1)=x_i\} \\
&= \prod_{i=1}^n (2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x_i-\mu)^2\right\} \\
&= (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^n (x_i-\mu)^2\right\} \\
&= (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2\right]\right\} \\
&= (2\pi)^{-n/2} \exp\left\{-\frac{n}{2}(\bar{x}-\mu)^2\right\} \exp\left\{-\frac{1}{2}\sum_{i=1}^n (x_i-\bar{x})^2\right\}
\end{aligned}
$$

考虑统计量 $T(X)=\overline{X}$,

记 $\begin{cases} g(T(X);\mu)=g(\overline{X};\mu)=(2\pi)^{-n/2}\exp\{-\frac{n}{2}(\bar{x}-\mu)^2\} \\ h(x)=\exp\{-\frac{1}{2}\sum_{i=1}^n (x_i-\bar{x})^2\} \end{cases}$

根据因子化定理我们知道，$\overline{X}$ 是参数 μ 的充分统计量.

---

(2) $\{N(0,\sigma^2):\sigma^2>0\}$

Solution:

记已修偏的样本方差为 $S^2=\frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X})^2$

$T(X)=(\overline{X},S^2)$ 是参数 σ^2 的充分统计量.

利用因子化定理证明:

记 $x=(x_1,\ldots,x_n)$，则我们有:

$$
\begin{aligned}
\mathrm{P}\{X=x\} &= \mathrm{P}\{X_1=x_1,\ldots,X_n=x_n\} \\
&= \prod_{i=1}^n \mathrm{P}\{N(0,\sigma^2)=x_i\} \\
&= \prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}x_i^2\right\} \\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right\} \\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (x_i-\bar{x})^2 + n\bar{x}^2\right]\right\} \\
&= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{(n-1)s^2+n\bar{x}^2}{2\sigma^2}\right\}
\end{aligned}
$$

其中 $\begin{cases} s^2=\frac{1}{n-1}\sum_{i=1}^n (x_i-\bar{x})^2 \\ \bar{x}=\frac{1}{n}\sum_{i=1}^n x_i \end{cases}$

考虑统计量 $T=(\overline{X},S^2)$,

记 $\begin{cases} g(T(x);\sigma^2)=g(\bar{x},s^2;\sigma^2)=(2\pi\sigma^2)^{-n/2}\exp\{-\frac{(n-1)s^2+n\bar{x}^2}{2\sigma^2}\} \\ h(x)\equiv 1 \end{cases}$

根据因子化定理我们知道，$T=(\overline{X},S^2)$ 是参数 σ^2的充分统计量.

Problem 3 (习题 1.40(3)(5)(6))

记 $X=(X_1,\ldots,X_n)$ 为取自具有下列总体分布的样本.

求参数的充分统计量.

(若有两个未知参数，则分别讨论当参数有一个未知或两个都未知时的充分统计量)

(1) $\{p(x;\theta_1,\theta_2)=\frac{1}{\theta_2-\theta_1}I_{(\theta_1,\theta_2)}(x):\theta_1<\theta_2\in\mathbb{R}\}$ (均匀分布族)

Solution:

记 $x=(x_1,\ldots,x_n)$，则我们有:

$$
\begin{aligned}
\mathrm{P}\{X=x\} &= \mathrm{P}\{X_1=x_1,\ldots,X_n=x_n\}\\
&= \prod_{i=1}^{n} \mathrm{P}\{\mathrm{Uniform}(\theta_1,\theta_2)=x_i\}\\
&= \prod_{i=1}^{n} \frac{1}{\theta_2-\theta_1} I(\theta_1<x_i<\theta_2)\\
&= \frac{1}{(\theta_2-\theta_1)^n} I(\min\{x_i\}>\theta_1) I(\max\{x_i\}<\theta_2)
\end{aligned}
$$

- **Case 1: (θ_1 已知，θ_2 未知)**
  考虑统计量 $T=X_{(n)}=\max\{x_i\}$
  记 $\begin{cases} g(T(x);\theta_2)=g(\max\{x_i\};\theta_2)=\frac{1}{(\theta_2-\theta_1)^n} I(\max\{x_i\}<\theta_2) \\ h(x)=I(\min\{x_i\}>\theta_1) \end{cases}$
  根据因子化定理我们知道，$T=T(X)=X_{(n)}=\max\{X_i\}$ 是参数 θ_2 的充分统计量.
- **Case 2: (θ_1 未知，θ_2 已知)**
  考虑统计量 $T=X_{(1)}=\min\{x_i\}$
  记 $\begin{cases} g(T(x);\theta_1)=g(\min\{x_i\};\theta_1)=\frac{1}{(\theta_2-\theta_1)^n} I(\min\{x_i\}>\theta_1) \\ h(x)=I(\max\{x_i\}<\theta_2) \end{cases}$
  根据因子化定理我们知道，$T=T(X)=X_{(1)}=\min\{X_i\}$ 是参数 θ_1 的充分统计量.
- **Case 3: (θ_1 和 θ_2 均未知)**
  考虑统计量 $T=(X_{(1)},X_{(n)})=(\min\{x_i\},\max\{x_i\})$
  记 $\begin{cases} g(T(x);\theta_1,\theta_2)=g(\min\{x_i\},\max\{x_i\};\theta_1,\theta_2)=\frac{1}{(\theta_2-\theta_1)^n} I(\min\{x_i\}>\theta_1) I(\max\{x_i\}<\theta_2) \\ h(x)\equiv 1 \end{cases}$
  根据因子化定理我们知道，
  $T=T(X)=(X_{(1)},X_{(n)})=(\min\{x_i\},\max\{x_i\})$ 是参数 (θ_1,θ_2) 的充分统计量.

---

(2) $\{f(x;p,\theta)=(1-p)p^{x-\theta},\ x=\theta,\theta+1,\ldots,\ p\in(0,1)\}$ (带位移的几何分布族)

Solution:
记 $x=(x_1,\ldots,x_n)$，则我们有:

$$
\begin{aligned}
\mathrm{P}\{X=x\} &= \mathrm{P}\{X_1=x_1,\ldots,X_n=x_n\}\\
&= \prod_{i=1}^{n} (1-p)p^{x_i-\theta} I(x_i=\theta,\theta+1,\ldots)\\
&= (1-p)^n p^{\sum_{i=1}^n x_i-n\theta} I(x_i=\theta,\theta+1,\ldots \text{for all } i=1,\ldots,n)\\
&= (1-p)^n p^{n\bar{x}-n\theta} I(\min\{x_i\}\ge\theta) I(x_i\in\mathbb{Z} \text{ for all } i=1,\ldots,n)
\end{aligned}
$$

- **Case 1: (p 已知，θ 未知)**
  考虑统计量 $T=T(X)=X_{(1)}=\min\{X_i\}$
  记 $\begin{cases} g(T(x);\theta)=g(\min\{x_i\};\theta)=p^{n\bar{x}-n\theta} I(\min\{x_i\}\ge\theta) \\ h(x)=(1-p)^n I(x_i\in\mathbb{Z} \text{ for all } i=1,\ldots,n) \end{cases}$
  根据因子化定理我们知道，$T=T(X)=X_{(1)}=\min\{X_i\}$ 是参数 θ 的充分统计量.
- **Case 2: (p 未知，θ 已知)**
  考虑统计量 $T=T(X)=\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i$
  记 $\begin{cases} g(T(x);\theta)=g(\bar{x};\theta)=(1-p)^n p^{n\bar{x}-n\theta} \\ h(x)=I(x_i\in\mathbb{Z} \text{ for all } i=1,\ldots,n) I(\min\{x_i\}\ge\theta) \end{cases}$
  根据因子化定理我们知道，$T=T(X)=\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i$ 是参数 p 的充分统计量.
- **Case 3: (p 和 θ 均未知)**
  考虑统计量 $T=T(X)=(\overline{X},X_{(1)})=(\frac{1}{n}\sum_{i=1}^n X_i,\min\{X_i\})$
  记 $\begin{cases} g(T(x);\theta)=g(\bar{x},\min\{x_i\};\theta)=(1-p)^n p^{n\bar{x}-n\theta} I(\min\{x_i\}\ge\theta) \\ h(x)=I(x_i\in\mathbb{Z} \text{ for all } i=1,\ldots,n) \end{cases}$
  根据因子化定理我们知道，
  $T=T(X)=(\overline{X},X_{(1)})=(\frac{1}{n}\sum_{i=1}^n X_i,\min\{X_i\})$ 是参数 (p,θ) 的充分统计量.

---

(3) $\{f(x;\theta)=\frac{\theta a^\theta}{x^{\theta+1}} I_{(a,\infty)}(x):\theta>0\}$ (Pareto 分布族，a 为已知常数)

Solution:
记 $x=(x_1,\ldots,x_n)$，则我们有:

$$
\begin{aligned}
\mathrm{P}\{X=x\} &= \mathrm{P}\{X_1=x_1,\ldots,X_n=x_n\} \\
&= \prod_{i=1}^{n} \frac{\theta a^{\theta}}{x^{\theta+1}} I_{(a,\infty)}(x_i) \\
&= \theta^n a^{n\theta} \left(\prod_{i=1}^{n} x_i^{-1}\right)^{\theta} \left(\prod_{i=1}^{n} x_i^{-1}\right) I(\min\{x_i\}>a) \\
&= \theta^n a^{n\theta} \exp\left\{-\theta \sum_{i=1}^{n} \log(x_i)\right\} \left(\prod_{i=1}^{n} x_i^{-1}\right) I(\min\{x_i\}>a)
\end{aligned}
$$

考虑统计量 $T(X) = -\log\left(\prod_{i=1}^{n} x_i^{-1}\right) = \sum_{i=1}^{n} \log(x_i)$

记 $\begin{cases} g(T(x);\theta) = g(\sum_{i=1}^{n}\log(x_i);\theta) = \theta^n a^{n\theta} \exp\{-\theta \sum_{i=1}^{n} \log(x_i)\} \\ h(x) = (\prod_{i=1}^{n} x_i^{-1}) I(\min\{x_i\}>a) \end{cases}$

根据因子化定理我们知道,

$T(X) = -\log\left(\prod_{i=1}^{n} x_i^{-1}\right) = \sum_{i=1}^{n} \log(x_i)$ 是参数 θ 的充分统计量.

Problem 4 (习题 2.24(4)(8)(10))

判别下列分布族的完备性:

(1) 均匀分布族 $\{\mathrm{Uniform}(\theta,\theta+1):\theta\in\mathbb{R}\}$

- **Solution:**
  均匀分布族 $\{\mathrm{Uniform}(\theta,\theta+1):\theta\in\mathbb{R}\}$ 是完备的.

  证明:
  设随机变量 X 的分布族为 $\{\mathrm{Uniform}(\theta,\theta+1):\theta\in\mathbb{R}\}$
  任意给定 X 的函数 ϕ,
  假设有 $\mathrm{E}_\theta[\phi(X)] = \int_\theta^{\theta+1} \phi(x)\cdot 1\mathrm{d}x = 0 \ \ (\forall\,\theta\in\mathbb{R})$ 成立.
  根据测度论的结论,
  如果函数 $\phi(\cdot)$ 在每一个长度为 1 的区间上的积分都为 0,
  我们可以推断 $\phi(\cdot)$ 几乎处处为 0 (即仅在某个零测度集合上不为 0)
  也就是说, 我们可以推断出 $\mathrm{P}_\theta\{\phi(X)=0\}=1 \ \ (\forall\,\theta\in\mathbb{R})$
  因此均匀分布族 $\{\mathrm{Uniform}(\theta,\theta+1):\theta\in\mathbb{R}\}$ 是完备的.

(2) 双参数 (但 $\lambda=1$) 的指数分布族 $\{p(x;\mu)=e^{-(x-\mu)}I_{(\mu,\infty)}(x):\mu\in\mathbb{R}\}$

- **Solution:**
  指数分布族 $\{p(x;\mu)=e^{-(x-\mu)}I_{(\mu,\infty)}(x):\mu\in\mathbb{R}\}$ 是完备的.

  证明:
  设随机变量 X 的分布族为 $\{p(x;\mu)=e^{-(x-\mu)}I_{(\mu,\infty)}(x):\mu\in\mathbb{R}\}$
  任意给定 X 的函数 ϕ,
  假设有 $\mathrm{E}_\mu[\phi(X)] = \int_\mu^{\infty} \phi(x)\cdot e^{-(x-\mu)}\mathrm{d}x = 0 \ \ (\forall\,\mu\in\mathbb{R})$ 成立.
  根据测度论的结论,
  上述条件意味着 $\phi(x)e^{-(x-\mu)}$ 几乎处处为 0,
  也就是说, 我们可以推断出 $\mathrm{P}_\mu\{\phi(X)e^{-(X-\mu)}=0\} = \mathrm{P}_\mu\{\phi(X)=0\}=1 \ \ (\forall\,\mu\in\mathbb{R})$
  因此指数分布族 $\{p(x;\mu)=e^{-(x-\mu)}I_{(\mu,\infty)}(x):\mu\in\mathbb{R}\}$ 是完备的.

(3) Cauchy 分布族 $\{p(x;\theta)=[\pi\theta(1+\frac{x^2}{\theta^2})]^{-1}:\theta>0\}$

- **Solution:**
  Cauchy 分布族 $\{p(x;\theta)=[\pi\theta(1+\frac{x^2}{\theta^2})]^{-1}:\theta>0\}$ 不具有完备性.

  反例:
  设随机变量 X 的分布族为 $\{p(x;\theta)=[\pi\theta(1+\frac{x^2}{\theta^2})]^{-1}:\theta>0\}$
  考虑 X 的函数 $\phi(x)=\mathrm{sgn}(x)$
  它显然不处处为 0, 但我们可以验证:

  $$\mathrm{E}_\theta[\phi(X)] = \mathrm{E}_\theta[\mathrm{sgn}(X)] = \int_{-\infty}^{\infty} \mathrm{sgn}(x)\cdot\left[\pi\theta\left(1+\frac{x^2}{\theta^2}\right)\right]^{-1}\mathrm{d}x = 0 \ \ (\forall\,\theta>0)$$

  (奇函数在对称区间上的积分为 0)

Problem 5 (补充题 3)

设 $X=(X_1,\dots,X_n)$ 为取自下列分布族的简单随机样本.
请找出参数 θ 的充分统计量，并检查其是否完备.

(1) $\{N(\theta,\theta):\theta>0\}$ (scaled normal distribution)
(the distribution does not fit into the typical framework of "curved" normal distributions
because the relationship between the variance and the mean is linear)

- **Lemma:**

> **定理 $1.3.6$: (数理统计讲义 命题 $2.2.28$, 指数族分布完备统计量的充分条件)**
> 设随机变量 X 的分布族为**指数分布族** $\{p_X(x;\theta)=C(\theta)\exp[\sum_{j=1}^k Q_j(\theta)T_j(x)]h(x):\theta\in\Theta\}$
> 若集合 $\{(Q_1(\theta),\dots,Q_k(\theta)):\theta\in\Theta\}$ 包含 $\mathbb{R}^k$ 中的一个 k 维邻域,
> 则 $T=(T_1(X),\dots,T_k(X))$ 为完备的统计量.

Solution:
设 $X=(X_1,\dots,X_n)$ 为
取自总体分布为**正态分布族** $\{N(\theta,\theta):\theta>0\}$ 的简单随机样本.
记 $x=(x_1,\dots,x_n)$，则我们有:

$$
\begin{aligned}
p_X(x;\theta)&=\mathrm{P}_\theta\{X=x\}\\
&=\mathrm{P}_\theta\{X_1=x_1,\dots,X_n=x_n\}\\
&=\prod_{i=1}^n \mathrm{P}\{N(\theta,\theta)=x_i\}\\
&=\prod_{i=1}^n (2\pi\theta)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\theta}(x_i-\theta)^2\right\}\\
&=(2\pi\theta)^{-n/2}\exp\left\{-\frac{1}{2\theta}\sum_{i=1}^n (x_i-\theta)^2\right\}\\
&=(2\pi\theta)^{-n/2}\exp\left\{-\frac{1}{2\theta}\sum_{i=1}^n x_i^2+\sum_{i=1}^n x_i-\frac{n\theta}{2}\right\}\\
&=(2\pi\theta)^{-n/2}\exp\left(-\frac{n\theta}{2}\right)\exp\left\{-\frac{1}{2\theta}\sum_{i=1}^n x_i^2\right\}\exp\left\{\sum_{i=1}^n x_i\right\}
\end{aligned}
$$

考虑统计量 $T=\sum_{i=1}^n X_i^2$,
记 $\begin{cases} g(T;\theta)=(2\pi\theta)^{-n/2}\exp\{-\frac{1}{2\theta}T-\frac{n\theta}{2}\} \\ h(X)=\exp\{\sum_{i=1}^n X_i\} \end{cases}$
根据因子化定理我们知道，$T=\sum_{i=1}^n X_i^2$ 是参数 θ 的**充分统计量**.

记 $\begin{cases} C(\theta)=(2\pi\theta)^{-n/2}\exp\left(-\frac{n\theta}{2}\right) \\ Q(\theta)=-\frac{1}{2\theta} \\ T(X)=\sum_{i=1}^n X_i^2 \\ h(X)=\exp\{\sum_{i=1}^n X_i\} \end{cases}$ 将 $p_X(x;\theta)$ 写成指数分布族形式.
我们发现 $\{Q(\theta):\theta>0\}$ 显然包含 $\mathbb{R}$ 的某个 1 维邻域,
根据引理可知 $T=\sum_{i=1}^n X_i^2$ 是参数 θ 的**完备统计量**.

综上所述，$T=\sum_{i=1}^n X_i^2$ 是参数 θ 的**充分完备统计量**.

---

(2) $\{N(\theta,\theta^2):\theta\in\mathbb{R}\}$ (curved normal distribution)
(A curved normal distribution refers to a situation
where the variance of the distribution is a non-linear function of the mean)

- **Reference:**
  - normal distribution - Sufficient statistic for $N(\theta,\theta^2)$ - Mathematics Stack Exchange
  - real analysis - Completeness of X with normal distribution $N(\theta,\theta^2)$ with equal mean and standard deviation: Integral of scale of a function - Mathematics Stack Exchange
  - Complete statistic for Normal Distribution $\mathcal{N}(\mu,\mu^2)$ - Mathematics Stack Exchange
- **Lemma:**

> **正态总体的样本均值与样本方差的联合分布 (S. Ross 命题 2.5):**
> 若 $X = (X_1, \dots, X_n)^T$ 为取自 $\mathrm{N}(\mu, \sigma^2)$ 的简单随机样本，样本量为 n，
> 则对于样本均值 $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ 和样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ 有 $\begin{cases} \overline{X} \perp S^2 \\ \overline{X} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n}) \\ S^2 \sim \sigma^2 \frac{\chi^2(n-1)}{n-1} \end{cases}$ 成立.

Solution:
设 $X = (X_1, \dots, X_n)$ 为
取自总体分布为**正态分布族** $\{N(\theta, \theta^2) : \theta > 0\}$ 的简单随机样本.
记 $x = (x_1, \dots, x_n)$，则我们有:

$$\begin{aligned} \mathrm{P}\{X = x\} &= \mathrm{P}\{X_1 = x_1, \dots, X_n = x_n\} \\ &= \prod_{i=1}^n \mathrm{P}\{N(\theta, \theta^2) = x_i\} \\ &= \prod_{i=1}^n (2\pi\theta^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\theta^2}(x_i - \theta)^2\right\} \\ &= (2\pi\theta^2)^{-n/2} \exp\left\{-\frac{1}{2\theta^2}\sum_{i=1}^n (x_i - \theta)^2\right\} \\ &= (2\pi\theta^2)^{-n/2} \exp\left\{-\frac{1}{2\theta^2}\left[\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \theta)^2\right]\right\} \\ &= (2\pi\theta^2)^{-n/2} \exp\left(-\frac{n-1}{2\theta^2}s^2\right) \exp\left\{-\frac{n}{2\theta^2}(\bar{x} - \theta)^2\right\} \end{aligned}$$

其中 $\begin{cases} s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2 \\ \bar{x} = \frac{1}{n}\sum_{i=1}^n x_i \end{cases}$

考虑统计量 $T = (\overline{X}, S^2) = (\frac{1}{n}\sum_{i=1}^n X_i, \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2)$,
记 $\begin{cases} g(T(x); \theta) = g(\bar{x}, s^2; \theta) = (2\pi\theta^2)^{-n/2} \exp\{-\frac{n-1}{2\theta^2}s^2\} \exp\{-\frac{n}{2\theta^2}(\bar{x} - \theta)^2\} \\ h(x) \equiv 1 \end{cases}$
根据因子化定理我们知道，$T = (\overline{X}, S^2)$ 是参数 θ 的**充分统计量**.

通过反例说明完备性不成立:
根据引理可知 $T = (\overline{X}, S^2)$ 的分布族为 $\{(N(\theta, \frac{\theta^2}{n}), \theta^2 \frac{\chi^2(n-1)}{n-1}) : \theta > 0\}$
定义 T 的函数 $\phi(T) = \phi(\overline{X}, S^2) = \overline{X}^2 - \frac{n+1}{n}S^2$,
则对于任意 $\theta > 0$ 都有 $\mathrm{E}_\theta[\phi(T)] = \mathrm{E}_\theta[\overline{X}^2 - \frac{n+1}{n}S^2] = (\frac{\theta^2}{n} + \theta^2) - \frac{n+1}{n}\theta^2 = 0$ 成立,
但显然 $\phi(t)$ 在 $\mathbb{R} \times \mathbb{R}_{++}$ 上并非几乎处处为零.

The End