统计学基础 I : 数理统计 Assignment 3

姓名: 雍崔扬
学号: 21307140051
习题: E1.29, E1.31(2)(3), E1.34(1)(2), E1.37(3)(7)(9), 补充题2

Problem 1 (习题 1.29)

设 $X=(X_1,\dots,X_n)$ 为取自以分布函数 $F(x)$ 为分布的简单随机样本.
证明 $\mathrm{P}\{X_{(j)}\le x\}=\sum_{k=j}^n\binom{n}{k}F(x)^k[1-F(x)]^{n-k}$

- **证明:**

$$
\begin{aligned}
\mathrm{P}\{X_{(j)}\le x\} &= \sum_{k=j}^n \mathrm{P}\{X_1,\dots,X_n\text{中有 }k\text{ 个属于}(-\infty,x],\text{有 }n-k\text{ 个属于 }(x,\infty)\} \\
&= \sum_{k=j}^n\left\{\binom{n}{k}\prod_{i=1}^k \mathrm{P}\{X_i\le x\}\prod_{i=k+1}^n \mathrm{P}\{X_i>x\}\right\} \\
&= \sum_{k=j}^n\binom{n}{k}F(x)^k[1-F(x)]^{n-k}
\end{aligned}
$$

命题得证.

Problem 2 (习题 1.31 (2)(3))

设 $X_1,\dots,X_n$ 为取自具有连续分布函数 F 总体的样本，$X_{(1)}\le\cdots\le X_{(n)}$ 为顺序统计量.
试证明:

(1) $F(X_{(k)})\overset{\mathrm{d}}{=}\mathrm{Beta}(k,n+1-k)$
证明:

- **引理 1: (数理统计讲义, 命题 $1.4.7$ 推论)**
第 k 个次序统计量 $X_{(k)}$ 的概率密度函数为:

$$
f_{X_{(k)}}(x)=\frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}f(x)[1-F(x)]^{n-k}
$$

- **应用引理 1:**
简单起见，假设 F 严格单调递增，存在逆函数 F^{-1}.
对于任意 $y\in[0,1]$，我们记 $x=F^{-1}(y)$，则有:

$$
\begin{aligned}
f_{F(X_{(k)})}(y) &= f_{X_{(k)}}(x)\cdot\left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|^{-1} \\
&= \frac{n!}{(k-1)!(n-k)!}F(x)^{k-1}f(x)[1-F(x)]^{n-k}\cdot f(x)^{-1} \\
&= \frac{n!}{(k-1)!(n-k)!}F(x)^{k-1}[1-F(x)]^{n-k} \\
&= \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n+1-k)}y^{k-1}(1-y)^{(n+1-k)-1} \\
&= \mathrm{P}\{\mathrm{Beta}(k,n+1-k)=y\}
\end{aligned}
$$

说明 $F(X_{(k)})\overset{\mathrm{d}}{=}\mathrm{Beta}(k,n+1-k)$
命题得证.

---

(2) 当 $1\le i<j\le n$ 时，有 $F(X_{(j)})-F(X_{(i)})\overset{\mathrm{d}}{=}\mathrm{Beta}(j-i,n+1-(j-i))$ 成立.
证明:

- **引理 2: (数理统计讲义, 命题 $1.4.7$ 推论)**
对于任意 $1\le i<j\le n$，$(X_{(i)},X_{(j)})$ 的联合概率密度函数为:

$$
f_{X_{(i)},X_{(j)}}(x,y)=\frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(x)]^{i-1}[F(y)-F(x)]^{j-i-1}[1-F(y)]^{n-j}f(x)f(y)I(x<y)
$$

- **应用引理 2:**

记 $\begin{cases} Y_1 = F(X_{(i)}) \\ Y_2 = F(X_{(j)}) \\ Z_1 = g_1(Y_1, Y_2) = Y_1 = F(X_{(i)}) \\ Z_2 = g_2(Y_1, Y_2) = Y_2 - Y_1 = F(X_{(j)}) - F(X_{(i)}) \end{cases}$ 则有 $\begin{cases} Y_1 = h_1(Z_1, Z_2) = Z_1 \\ Y_2 = h_2(Z_1, Z_2) = Z_1 + Z_2 \end{cases}$

简单起见，假设 F 严格单调递增，存在逆函数 F^{-1}.

对于任意 $y_1, y_2 \in [0,1]$，我们记 $\begin{cases} x_1 = F^{-1}(y_1) \\ x_2 = F^{-1}(y_2) \end{cases}$，则我们有:

$$
\begin{aligned}
& f_{Y_1,Y_2}(y_1, y_2) \\
& = f_{X_{(i)},X_{(j)}}(x_1, x_2) \cdot \left| \frac{\partial y_1}{\partial x_1} \frac{\partial y_2}{\partial x_2} \right|^{-1} \\
& = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} y_1^{i-1} (y_2 - y_1)^{j-i-1} (1-y_2)^{n-j} f(x_1) f(x_2) I(0 \le y_1 < y_2 \le 1) \cdot f(x_1)^{-1} f(x_2)^{-1} \\
& = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} y_1^{i-1} (y_2 - y_1)^{j-i-1} (1-y_2)^{n-j} I(0 \le y_1 < y_2 \le 1)
\end{aligned}
$$

且有:

$$
J(y_1, y_2) = \begin{vmatrix} \frac{\partial g_1}{\partial y_1} & \frac{\partial g_1}{\partial y_2} \\ \frac{\partial g_2}{\partial y_1} & \frac{\partial g_2}{\partial y_2} \end{vmatrix} \equiv \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} \equiv 1
$$

因此有:

$$
\begin{aligned}
f_{Z_1,Z_2}(z_1, z_2) & = f_{Y_1,Y_2}(h_1(z_1, z_2), h_2(z_1, z_2)) \cdot |J(h_1(z_1, z_2), h_2(z_1, z_2))|^{-1} \\
& = f_{Y_1,Y_2}(z_1, z_1 + z_2) \cdot 1 \\
& = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} z_1^{i-1} [(z_1 + z_2) - z_1]^{j-i-1} [1 - (z_1 + z_2)]^{n-j} I(0 \le z_1 < z_1 + z_2 \le 1) \\
& = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} z_1^{i-1} z_2^{j-i-1} (1 - z_1 - z_2)^{n-j} I(0 \le z_1 < z_1 + z_2 \le 1)
\end{aligned}
$$

于是有:

$$
\begin{aligned}
& f_{Z_2}(z_2) \\
& = \int_0^1 f_{Z_1,Z_2}(z_1, z_2) \mathrm{d}z_1 \\
& = \int_0^1 \frac{n!}{(i-1)!(j-i-1)!(n-j)!} z_1^{i-1} z_2^{j-i-1} (1 - z_1 - z_2)^{n-j} I(0 \le z_1 < z_1 + z_2 \le 1) \mathrm{d}z_1 \\
& = \frac{n!}{(j-i-1)!(n-j+i)!} z_2^{j-i-1} (1-z_2)^{n-j+i-1} I(0 < z_2 \le 1) \int_0^{1-z_2} \frac{(n-j+i)!}{(i-1)!(n-j)!} (\frac{z_1}{1-z_2})^{i-1} (1 - \frac{z_1}{1-z_2})^{n-j} \mathrm{d}z_1 \\
& = \frac{n!}{(j-i-1)!(n-j+i)!} z_2^{j-i-1} (1-z_2)^{n-j+i-1} I(0 < z_2 \le 1) \int_0^{1-z_2} \mathrm{P}\left\{ \mathrm{Beta}(i, n+1-j) = \frac{z_1}{1-z_2} \right\} \mathrm{d}z_1 \\
& = \frac{n!}{(j-i-1)!(n-j+i)!} z_2^{j-i-1} (1-z_2)^{n-j+i-1} I(0 < z_2 \le 1) \cdot (1-z_2) \int_0^1 \mathrm{P}\{\mathrm{Beta}(i, n+1-j) = t\} \mathrm{d}t \\
& \quad (\text{denote } t := \frac{z_1}{1-z_2}) \\
& = \frac{n!}{(j-i-1)!(n-j+i)!} z_2^{j-i-1} (1-z_2)^{n-j+i} I(0 < z_2 \le 1) \cdot 1 \\
& = \frac{\Gamma(n+1)}{\Gamma(j-i)\Gamma(n+1-j+i)} z_2^{j-i-1} (1-z_2)^{n-j+i} I(0 < z_2 \le 1) \\
& = \mathrm{P}\{\mathrm{Beta}(j-i, n+1-(j-i)) = z_2\}
\end{aligned}
$$

因此 $F(X_{(j)}) - F(X_{(i)}) = Z_2 \overset{\mathrm{d}}{=} \mathrm{Beta}(j-i, n+1-(j-i))$

命题得证.

Problem 3 (习题 1.34 (1)(2))

设 F 为分布函数，

定义**分位函数** (quantile function) 为 $F^{-1}(p) = \inf\{x : F(x) \ge p\} \quad (p \in [0,1])$

试证明:

(1) $F(F^{-1}(p)_-) \le p \le F(F^{-1}(p))$，即 $F^{-1}(p)$ 是 F 的 p 分位数.

证明:

- 对于任意 $\varepsilon > 0$，我们都有 $F(F^{-1}(p) + \varepsilon) \geq p$
  由于 F 是右连续的，因此 $F(F^{-1}(p)) = \lim_{\varepsilon\to 0^+} F(F^{-1}(p) + \varepsilon) \geq p$
- 对于任意 $\varepsilon > 0$，我们都有 $F(F^{-1}(p) - \varepsilon) < p$ (否则与 $F^{-1}(p)$ 的定义相矛盾)
  因此我们有 $F(F^{-1}(p)_-) = F(\lim_{\varepsilon\to 0^+}\{F^{-1}(p) - \varepsilon\}) \leq p$ 成立.
  (取极限，严格不等号变为非严格不等号)

命题得证.

(2) $F^{-1}(p) \leq x \ \Leftrightarrow \ F(x) \geq p$
证明:

- 证明充分性 $(\Rightarrow)$:
  若 $F^{-1}(p) \leq x$，则根据 F 的单调性和 (1) 的结论有:
  $$F(x) \geq F(F^{-1}(p)) \geq p$$
- 证明必要性 $(\Leftarrow)$:
  若 $F(x) \geq p$，则根据定义可知 $F^{-1}(p) = \inf\{x : F(x) \geq p\} \leq x$

命题得证.

Problem 4 (习题 1.37 (3)(7)(9))

验证以下分布族是否为指数型分布族:

(1) Poisson 双参数指数分布族 $\{p(x;\lambda,\mu) = \frac{1}{\lambda}\exp\left(-\frac{x-\mu}{\lambda}\right) I_{(\mu,\infty)}(x) : \lambda > 0, \mu \in \mathbb{R}\}$

- **判断:** 不是指数族.
- **证明:**
  其支撑集 (μ,∞) 与参数 μ 有关，所以它不是指数族.
  这是因为指数族 $\{p(x;\theta) = C(\theta)\exp\left(\sum_{i=1}^k Q_i(\theta)T_i(x)\right)h(x) : \theta \in \Theta\}$
  的支撑集 $\{x : p(x;\theta) > 0\} = \{x : h(x) > 0\}$ 与参数 θ 无关.

(2) $\{p(x;\theta) = \exp[-2\log(\theta) + \log(2x)] I_{(0,\theta)}(x) : \theta > 0\}$

- **判断:** 不是指数族.
- **证明:**
  其支撑集 $(0,\theta)$ 与参数 θ 有关，所以它不是指数族.

(3) $\{p(x;\theta) = \frac{2(x+\theta)}{1+2\theta} I_{(0,1)}(x) : \theta > 0\}$

- **判断:** 不是指数族.
- **证明:**
  - **引理: ($\log(1+u)$ 的 Taylor 展开式)**
    $$\log(1+u) = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{u^n}{n}$$
    等式右边的级数在 $(-1,1]$ 收敛，在 $(1,\infty)$ 上发散.
  - **应用引理:**
    $$\begin{aligned}
    p(x;\theta) &= \frac{2(x+\theta)}{1+2\theta} I_{(0,1)}(x) \\
    &= \frac{2}{1+2\theta}\exp\{\log(x+\theta)\} I_{(0,1)}(x) \\
    &= \frac{2}{1+2\theta}\exp\left\{\log(\theta) + \log\left(1+\frac{x}{\theta}\right)\right\} I_{(0,1)}(x) \\
    &= \frac{2}{1+2\theta}\exp\left\{\log(\theta) + \sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n}\left(\frac{x}{\theta}\right)^n\right\} I_{(0,1)}(x) \\
    &= \frac{2}{1+2\theta}\exp\left\{\log(\theta) + \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n\theta^n}x^n\right\} I_{(0,1)}(x)
    \end{aligned}$$

取 $\begin{cases} C(\theta)=\frac{2}{1+2\theta} \\ k=\infty \\ Q_0(\theta)=\log(\theta) \\ T_0(x)=1 \\ Q_i(\theta)=\frac{(-1)^{i+1}}{n\theta^n} \quad (i=1,2,\dots) \\ T_i(x)=x^i \quad (i=1,2,\dots) \\ h(x)=I_{(0,1)}(x) \end{cases}$ 似乎可以构造指数族的形式.

但是传统的指数族的标准形式中，k 一般是一个有限的正整数.
而且级数 $\sum_{n=1}^{\infty}(-1)^{n+1}\frac{1}{n}(\frac{x}{\theta})^n\}$ 在 $x\in(0,1)$ 上不一定收敛，
任意给定 $\theta\in(0,\infty)$ 时，级数的收敛域为 $(-\theta,\theta]$
因此当 $\theta\in(0,1)$ 时，级数在 $x\in(0,1)$ 上不总是收敛.
综上所述，我们认为 $\{p(x;\theta)=\frac{2(x+\theta)}{1+2\theta}I_{(0,1)}(x):\theta>0\}$ 不是一个指数族.

- **反证法:**
  假设该分布族是一个一维指数族
  (若能表示为有限维指数族，因只有一个参数 θ 可变，必能写成一维指数族的形式)

$$p(x;\theta)=C(\theta)\exp\left(Q(\theta)T(x)\right)h(x)=\frac{2(x+\theta)}{1+2\theta}I_{(0,1)}(x)$$

  对于任意 $x\in(0,1)$ 和 $\theta>0$，我们有

$$C(\theta)\exp\left(Q(\theta)T(x)\right)h(x)=\frac{2(x+\theta)}{1+2\theta}$$

  因此存在 $A(\theta),B(\theta)$ 和 $D(x)$ 使得

$$\log(x+\theta)=A(\theta)+B(\theta)D(x)\quad(\forall\, x\in(0,1),\theta>0)$$

  左右两式对 x 求导，得到

$$\frac{1}{x+\theta}=B(\theta)D'(x)$$

  对于任意给定的 $x\in(0,1)$ 和不同的 $\theta_1,\theta_2>0$，我们有

$$\frac{x+\theta_2}{x+\theta_1}=\frac{B(\theta_1)D'(x)}{B(\theta_2)D'(x)}=\frac{B(\theta_1)}{B(\theta_2)}$$

  但左式依赖于 x，而右式不依赖于 x，产生矛盾.
  因此该指数族不是一维 (亦即有限维) 指数族.

Problem 5 (补充题 2)

设随机变量 X 均值有限，$0<p<1$
证明随机变量 X 的 p-分位数 ζ_p 满足 $\zeta_p=\text{argmin}_a\mathbb{E}[p\cdot(X-a)_+-(1-p)\cdot(X-a)_-]$
其中 $(c)_+=\max(c,0)$, $(c)_-=\min(c,0)$

证明:
为简单起见，我们首先假设 X 是连续随机变量，具有累积分布函数 F 和概率密度函数 f.
我们定义目标函数为:

$$\begin{aligned}L(a)&:=\mathbb{E}[p\cdot(X-a)_+-(1-p)\cdot(X-a)_-]\\&=p\int_a^{\infty}(x-a)f(x)\mathrm{d}x-(1-p)\int_{-\infty}^{a}(x-a)f(x)\mathrm{d}x\end{aligned}$$

- **引理: (Leibniz 积分规则)**
  变限积分 $\int_{g(a)}^{h(a)}F(x,a)\mathrm{d}x$ 关于 a 的导数为:

$$\frac{\mathrm{d}}{\mathrm{d}a}\int_{g(a)}^{h(a)}F(x,a)\mathrm{d}x=F(h(a),a)h'(a)-F(g(a),a)g'(a)+\int_{g(a)}^{h(a)}\frac{\partial}{\partial a}F(x,a)\mathrm{d}x$$

- **应用引理:**
  我们知道 $\begin{cases}\frac{\mathrm{d}}{\mathrm{d}a}\int_a^{\infty}(x-a)f(x)\mathrm{d}x=-\int_a^{\infty}f(x)\mathrm{d}x=F(a)-1\\ \frac{\mathrm{d}}{\mathrm{d}a}\int_{-\infty}^{a}(x-a)f(x)\mathrm{d}x=-\int_{-\infty}^{a}f(x)\mathrm{d}x=-F(a)\end{cases}$
  于是我们有 $\frac{\partial}{\partial a}L(a)=p(F(a)-1)+(1-p)F(a)=F(a)-p$
  记其**零点集**为 $\mathcal{A}^\star=\{a:F(a)=p\}$ (显然不为空集)

根据 F 的单调性可知 $\frac{\mathrm{d}L}{\mathrm{d}a}$ 关于 a (非严格) 单调递增，
因此目标函数 $L(a)$ 在 $a \in \mathcal{A}^\star$ 处取得最小值.
于是有:

$$\begin{aligned}\zeta_p &= \underset{a}{\operatorname{argmin}}\ L(a) \\ &= \inf\{a : a \in \mathcal{A}^\star\} \\ &= \inf\{a : F(a) = p\} \\ &= \inf\{a : F(a) \geq p\} \\ &= F^{-1}(p)\end{aligned}$$

根据本次作业第三题的结论 (其中 $F^{-1}(p) = \inf\{a : F(a) \geq p\}$ 为**分位函数**),
我们知道 ζ_p 满足 $F(\lim_{\varepsilon \to 0^+}\{\zeta_p - \varepsilon\}) \leq p \leq F(\zeta_p)$，即 ζ_p 是 p-分位数.

对于更一般的情况，我们记 $p_X(a) = \mathrm{P}\{X = a\}$
定义目标函数为:

$$\begin{aligned}L(a) &= \mathbb{E}[p \cdot (X-a)_+ - (1-p) \cdot (X-a)_-] \\ &= p \sum_{x > a : p_X(x) > 0} (x-a) p_X(x) - (1-p) \sum_{x \leq a : p_X(x) > 0} (x-a) p_X(x)\end{aligned}$$

根据 $\frac{\mathrm{d}}{\mathrm{d}a}(x-a)p_X(x) = -p_X(x)$ 可知，
当 a 从 a^- 到 a^+ 穿过某个离散点 x 时，
$\sum_{x>a}(x-a)p_X(x)$ 的贡献的变化率近似为 $-\sum_{x>a} p_X(x) = F(a) - 1$
$\sum_{x \leq a}(x-a)p_X(x)$ 的贡献的变化率近似为 $-\sum_{x \leq a} p_X(x) = -F(a)$

根据这种变化的单调性我们可以推断:
问题的任意最优解 ζ_p 是 X 满足 $p \cdot (F(a) - 1) - (1-p) \cdot (-F(a)) \geq 0$ 的所有取值的下确界，
即 $\zeta_p = \inf\{a : F(a) \geq p\}$
和连续情形类似地，我们知道 ζ_p 是 p-分位数.

The End