HOMEWORK 3

2021/10/13

[Wei] 2.47. 样本联合 p.d.f 为

$$
\begin{aligned}
f(\mathbf{x},\mathbf{y};a,b,\sigma) &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{m+n} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}(X_i-a)^2+\sum_{i=1}^{n}(Y_i-b)^2\right]\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{m+n} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum X_i^2+\sum Y_i^2-2ma\bar{X}-2nb\bar{Y}+ma^2+nb^2\right)\right\}.
\end{aligned}
$$

由因子分解定理知，$\left(\sum X_i^2+\sum Y_i^2,\bar{X},\bar{Y}\right)$ 是充分统计量。样本联合 p.d.f. 是指数族；令 $\eta := \left(-\frac{1}{2\sigma^2},\frac{ma}{\sigma^2},\frac{nb}{\sigma^2}\right) \in \Theta^*$，这里自然参数空间 $\Theta^* = \mathbf{R}_- \times \mathbf{R} \times \mathbf{R}$ 作为 $\mathbf{R}^3$ 的子集有内点，由课本定理 2.8.1，$\left(\sum X_i^2+\sum Y_i^2,\bar{X},\bar{Y}\right)$ 是完全统计量。注意到 $(n+m-2)S^2 = \sum X_i^2+\sum Y_i^2-m\bar{X}^2-n\bar{Y}^2$，故从 $\left(\sum X_i^2+\sum Y_i^2,\bar{X},\bar{Y}\right)$ 到 $\left(S^2,\bar{X},\bar{Y}\right)$ 有 1-1 映射，故 $\left(S^2,\bar{X},\bar{Y}\right)$ 也是充分完全统计量。 □

Remark: *两个说明指数族的完全统计量的定理的条件和结论形式不太一样. 其中课本定理 2.8.1 是写出样本联合 p.d.f.，验证自然参数空间有内点；《统计推断》定理 6.2.25 是写出总体 p.d.f.，验证参数空间有内点，但这里 pdf 指数上的参数部分形如 $w(\theta_j)$，注意这里是同一个 w，不同的 θ_j，也就是说，这里本质还是化到某种自然参数来做。建议大家书写时统一化成自然参数形式*

定理 2.8.1 设样本 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ 的概率函数

$$f(\boldsymbol{x},\boldsymbol{\theta}) = C(\boldsymbol{\theta})\exp\left\{\sum_{i=1}^{k}\theta_i T_i(\boldsymbol{x})\right\}h(\boldsymbol{x}),\quad \boldsymbol{\theta}=(\theta_1,\cdots,\theta_k)\in\Theta^*$$

为指数族的自然形式. 令 $T(\boldsymbol{X})=(T_1(\boldsymbol{X}),\cdots,T_k(\boldsymbol{X}))$, 若自然参数空间 Θ^* 作为 R_k 的子集有内点, 则 $T(\boldsymbol{X})$ 是完全统计量.

Date: 2021/10/27.

Thanks for Weiyu Li who is with the School of the Gifted Young, University of Science and Technology of China. Corresponding Email: liweiyu@mail.ustc.edu.cn.

Theorem 6.2.25 (Complete statistics in the exponential family) *Let $X_1, \dots, X_n$ be iid observations from an exponential family with pdf or pmf of the form*

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{j=1}^{k} w(\theta_j) t_j(x)\right), \tag{6.2.7}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$. Then the statistic

$$T(\mathbf{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \sum_{i=1}^{n} t_2(X_i), \dots, \sum_{i=1}^{n} t_k(X_i)\right)$$

is complete as long as the parameter space Θ contains an open set in $\Re^k$.

[Wei] 2.48.

$$f(\mathbf{x};\theta) = \left(\frac{1}{2\theta}\right)^n \exp\left\{-\frac{\sum_{i=1}^n |x_i|}{\theta}\right\}.$$

令 $\eta := -\frac{1}{\theta} \in \Theta^*$, 自然参数空间 $\Theta^* = \mathbf{R}_-$ 在 $\mathbf{R}$ 中有内点. 由因子分解定理及定理 2.8.1 知 $T = \sum_{i=1}^n |X_i|$ 是充分完全统计量. □

[Wei] 2.49.

$$f(\mathbf{x};\theta) = \exp\left\{n\theta - \sum_{i=1}^{n} x_i\right\} I_{(\theta,+\infty)}(x_{(1)}).$$

由因子分解定理, $T = X_{(1)}$ 是充分统计量.

$$\begin{aligned}
\because\quad & f_T(t) = ne^{-n(t-\theta)} I_{(\theta,+\infty)}(t) \\
\therefore\quad & E_\theta(\phi(T)) = \int_\theta^{+\infty} \phi(t) ne^{-n(t-\theta)} dt, \\
\therefore\quad & \int_\theta^{+\infty} \phi(t) e^{-nt} dt = 0
\end{aligned}$$

对上式关于 θ 求导，得 $\phi(\theta)e^{-n\theta} = 0$, 故 $\phi \overset{a.s.}{=} 0$. 由定义知 T 是完全统计量. □

2021/10/15

[Wei] 2.50.

$$f(\mathbf{x};\theta) = \left(\frac{1}{\theta}\right)^n I\left(-\theta/2 < x_{(1)} \le x_{(n)} < \theta/2\right).$$

由因子分解定理知 $T = (X_{(1)}, X_{(n)})$ 是充分统计量.。由于 $Y_i := \frac{X_i}{\theta} \overset{i.i.d.}{\sim} U(-\frac{1}{2}, \frac{1}{2})$, 故 $\frac{X_{(n)}}{X_{(1)}} = \frac{Y_{(n)}}{Y_{(1)}}$ 是辅助统计量. 故 T 不是完全统计量. □

Remark: Actually the statistic $\frac{X_{(n)}}{X_{(1)}}$ is not well-defined. Some tricks such as $\frac{X_{(n)}+1}{X_{(1)}+1}$ can fix this problem.

[Wei] 2.51.

$$f(\mathbf{x};\theta)=\left(\frac{1}{\theta}\right)^n I\left(\theta<x_{(1)}\leq x_{(n)}<2\theta\right).$$

由因子分解定理知 $T=(X_{(1)},X_{(n)})$ 是充分统计量.。由于 $Y_i:=\frac{X_i}{\theta}\overset{i.i.d.}{\sim}U(1,2)$, 故 $\frac{X_{(n)}}{X_{(1)}}=\frac{Y_{(n)}}{Y_{(1)}}$ 是辅助统计量. 故 T 不是完全统计量. □

[Wei] 2.52. 样本联合 p.d.f. 为

$$f(\mathbf{x};\lambda,\mu)=\lambda^{-n}\exp\left\{\frac{n\mu-\sum_i x_{(i)}}{\lambda}\right\}I_{[x_{(1)}>\mu]}$$

由因子分解定理知 $(X_{(1)},\sum_i X_{(i)})$ 是 (λ,μ) 的充分统计量。

任意固定 λ, 由样本联合 p.d.f. 知 $X_{(1)}$ 是 μ 的充分统计量。且 $X_{(1)}$ 的 p.d.f. 为

$$f_1(x|\mu)=\frac{n}{\lambda}\exp\{-\frac{n(x-\mu)}{\lambda}\}\mathbf{1}(x>\mu)$$

若 $E\phi(X_{(1)})=0$, 即

$$\frac{n}{\lambda}\int_{\mu}^{+\infty}\phi(x)e^{-n(x-\mu)/\lambda}=0$$

则 $\int_{\mu}^{+\infty}\phi(x)e^{-nx/\lambda}=0$, 关于 μ 求导整理得 $\phi(\mu)=0,\forall\mu\in\mathbb{R}$. 由定义知 $X_{(1)}$ 关于 μ 是完全统计量.

令 $Y_i=X_i-\mu$, 其分布与 μ 无关。则 $\sum_i X_i-X_{(1)}=\sum_i Y_i-Y_{(1)}$ 关于 μ 是辅助统计量. 而 $X_{(1)}$ 是 μ 的充分完全统计量，由 **Basu** 定理, λ 固定时，两统计量独立.

又由 λ 的任意性知 $X_{(1)}$ 与 $\sum_{i=1}^n(X_i-X_{(1)})$ 独立. □

Remark1:多参数情况下，应指明统计量关于哪个参数是充分完全统计量。如本题中任意固定 λ 后，$X_{(1)}$ 是 μ 的充分完全统计量

Remark2:*Basu* 定理是充分完全统计量和辅助统计量独立，不少同学只验证了完全性

Remark3:称 V 是 θ 的有界完全统计量是若任意有界可测函数 ϕ 使得 $\mathbb{E}_\theta[\phi(V)]=0$ 必有 $\phi=0,\quad a.s.\quad\theta$; 不是指统计量本身有界

[Wei] 2.53. Fix σ^2, from the factorization and that the natural parameter space is $\mathbb{R}$ (which has inner point), $\bar{X}$ is a **sufficient complete** statistic for a. Let $Y_i=X_i-a$, then $X_{(n)}-X_{(1)}=Y_{(n)}-Y_{(1)}$ is **auxiliary**. Derive independence from **Basu** theorem. According to the **arbitrariness** of σ^2, the two statistics are independent if σ^2 is a parameter. □

[Wei] 2.54. Hint:

From the **factorization** and that the **natural parameter space** is

$$\left\{\left(\frac{na}{\sigma_1^2}, \frac{nb}{\sigma_2^2}, -\frac{1}{2\sigma_1^2}, -\frac{1}{2\sigma_2^2}\right)\right\} = \mathbb{R}^2 \times \mathbb{R}_-^2$$

(which has inner point in $\mathbb{R}^4$), $T(\mathbf{X}, \mathbf{Y})$ is **sufficient complete** for $(a, b, \sigma_1^2, \sigma_2^2)$. Rescaling $\tilde{X}_i = \frac{X_i - a}{\sigma_1} \sim N(0,1)$, $\tilde{Y}_i = \frac{Y_i - b}{\sigma_2} \sim N(0,1)$ to find that r is **auxiliary**. Derive independence from **Basu** theorem.

Assignment:. X_1, X_2 独立同分布于 $N(0, \sigma^2)$, 用 Basu 定理证明统计量 $\frac{X_1}{X_2}$ 和 $\sqrt{X_1^2 + X_2^2}$ 独立。

令 $Y_i = X_i/\sigma$, 则 Y_1, Y_2 同分布于 $N(0,1)$.$\frac{X_1}{X_2} = \frac{Y_1}{Y_2}$，其分布与 σ^2 无关，故 $\frac{X_1}{X_2}$ 是辅助统计量。

(X_1, X_2) 的联合 p.d.f. 为

$$f(x_1, x_2|\sigma^2) = \frac{1}{2\pi\sigma^2} \exp\{-\frac{1}{2\sigma^2}(\sqrt{x_1^2 + x_2^2})^2\}$$

它是指数族，且由因子分解定理知 $\sqrt{X_1^2 + X_2^2}$ 是充分统计量。令 $\eta = -\frac{1}{2\sigma^2}$, 则自然参数空间 $\Theta^* = (-\infty, 0)$ 作为 $\mathbb{R}$ 的子集有内点，故 $X_1^2 + X_2^2$ 是完全统计量，从而 $\sqrt{X_1^2 + X_2^2}$ 也是完全统计量。

由 Basu 定理，$\sqrt{X_1^2 + X_2^2}$ 和 $\frac{X_1}{X_2}$ 独立。 $\square$