Tsinghua-Berkeley Shenzhen Institute
Inference and Information
Fall 2017

Homework 6

赵丰 March 15, 2018

- **Acknowledgments:** This coursework referes to wikipedia: `https://en.wikipedia.org`.
- **Collaborators:** I finish this coursework by myself.

I use `enumerate` to generate answers for each question:

6.1. (a)

$$\begin{aligned}\frac{p_{\underline{y}}(\underline{y};x)}{p_{t(t(\underline{y});x)}} &= \frac{x^{y_1+y_2}(1-x)^{2-y_1-y_2}}{\binom{2}{y_1+y_2}x^{y_1+y_2}(1-x)^{2-y_1-y_2}}\\ &= \frac{1}{\binom{2}{y_1+y_2}}\end{aligned}$$

上式与 x 无关，由 Fisher-Neyman 因子定理得 $t(\underline{y}) = y_1 + y_2$ 是充分统计量。

(b) $\hat{x}(\underline{y}) = y_1$ 是无偏估计量，MSE 等于 Bernoulli 分布的方差，因此 $\mathrm{MSE}_{\hat{x}}(x) = x(1-x)$

(c) i. **因为** t 是充分统计量，$p_{\underline{\mathsf{y}}|\mathsf{t}}(\underline{y}|\mathsf{t}=t)$ 与 x 无关。所以

$$\begin{aligned}\hat{x}'(t) &= \mathbb{E}[\hat{x}(\underline{\mathsf{y}})|\mathsf{t}=t]\\ &= \sum_{\underline{y}} \hat{x}(\underline{y}) p_{\underline{\mathsf{y}}|\mathsf{t}}(\underline{y}|\mathsf{t}=t)\end{aligned}$$

$\hat{x}'(t)$ 与 x 无关。

ii. 沿用 (a,b) 中的假设，由对称性 $\mathbb{E}[\mathsf{y}_1|\mathsf{t}=t] = \mathbb{E}[\mathsf{y}_2|\mathsf{t}=t] \Rightarrow \hat{x}'(\mathsf{t}) = \frac{\mathsf{t}}{2}$ 所以 $\hat{x}'(\mathsf{t})$ 是无偏的，$\mathrm{MSE}_{\hat{x}'}(x) = \frac{1}{2}x(1-x) \Rightarrow \gamma = \frac{1}{2}$

(d) i. 由 $\underline{\mathsf{t}}$ 是充分统计量可知。

ii.

$$\hat{x}'(\underline{t}) = \sum_{\underline{y}} \hat{x}(\underline{y}) p_{\underline{\mathsf{y}}|\underline{\mathsf{t}}}(\underline{y}|\underline{\mathsf{t}}=\underline{t})$$

由 $C(x,\hat{x})$ 关于 $\hat{x}$ 的凸性以及概率质量函数的归一化特性可得：

$$C(x,\hat{x}'(\underline{t})) \le \sum_{\underline{y}} C(x,\hat{x}(\underline{y})) p_{\underline{\mathsf{y}}|\underline{\mathsf{t}}}(\underline{y}|\underline{\mathsf{t}}=\underline{t})$$

$\Rightarrow$

$$\sum_{\underline{t}} C(x,\hat{x}'(\underline{t})) p_{\underline{\mathsf{t}}}(\underline{t}) \le \sum_{\underline{y},\underline{t}} C(x,\hat{x}(\underline{y})) p_{\underline{\mathsf{y}}|\underline{\mathsf{t}}}(\underline{y}|\underline{\mathsf{t}}=\underline{t}) p_{\underline{\mathsf{t}}}(\underline{t}) \tag{1}$$

因为 $p_{\mathsf{y}|\underline{\mathsf{t}}}(\underline{y}|\underline{\mathsf{t}}=\underline{t})p_{\underline{\mathsf{t}}}(\underline{t}) = p_{\underline{\mathsf{y}},\underline{\mathsf{t}}}(\mathsf{y},\underline{t}) = p_{\underline{\mathsf{y}}}(y)\cdot 1_{\underline{t}=\underline{t}(y)} \Rightarrow$
$\sum_{\underline{t}} p_{\underline{\mathsf{y}}|\mathsf{t}}(\underline{y}|\underline{\mathsf{t}}=\underline{t})p_{\underline{\mathsf{t}}}(\underline{t}) = p_{\underline{\mathsf{y}}}(\underline{y})$ 所以由(1)可得：

$$\mathbb{E}[C(x,\hat{x}'(\underline{\mathsf{t}}))] \le \sum_{\underline{y}} C(x,\hat{x}(\underline{y}))p_{\underline{\mathsf{y}}}(y) = \mathbb{E}[C(x,\hat{x}(\underline{\mathsf{y}}))]$$

6.2. (a)

$$\begin{aligned}\mathbb{E}_{\mathsf{x}|\mathsf{y}}[R(x,b_y)|\mathsf{y}=y] =& r(b_y)p_{\mathsf{x}|\mathsf{y}=y}(x=1) + r(1-b_y)p_{\mathsf{x}|\mathsf{y}=y}(x=0)\\ &\propto(1+\ln(b_y))p\,p_{\mathsf{y}|\mathsf{x}=1}(y) + (1+\ln(1-b_y))(1-p)p_{\mathsf{y}|\mathsf{x}=0}(y)\end{aligned}$$

令 $\frac{\partial R(x,b_y)}{\partial b_y}=0$，解得 $b_y = \frac{p\,p_{\mathsf{y}|\mathsf{x}=1}}{p\,p_{\mathsf{y}|\mathsf{x}=1}+(1-p)p_{\mathsf{y}|\mathsf{x}=1}} = P_{\mathsf{x}|\mathsf{y}}(1|\mathsf{y}=y)$
另解, 设 $\mathsf{z}\sim B(b_y)$：

$$\begin{aligned}\mathbb{E}_{\mathsf{x}|\mathsf{y}}[R(x,b_y)|\mathsf{y}=y] =& 1+\ln 2(P_{\mathsf{x}|\mathsf{y}}(1|\mathsf{y}=y)\log P_{\mathsf{z}}(1)+\\ &+P_{\mathsf{x}|\mathsf{y}}(0|\mathsf{y}=y)\log P_{\mathsf{z}}(0))\\ =& 1-\ln 2(H(\mathsf{x}|\mathsf{y}=y)+D(\mathsf{x}|\mathsf{y}=y||\mathsf{z}))\end{aligned}$$

由相对熵的非负性，当 $D(\mathsf{x}|\mathsf{y}=y||\mathsf{z}))$ 时回报的期望值最大，此时 $b_y = P_{\mathsf{x}|\mathsf{y}}(1|\mathsf{y}=y)$

(b)

$$\begin{aligned}R(x,b_y) =& r(b_y)p_{\mathsf{x}|\mathsf{y}=y}(x=1) + r(1-b_y)p_{\mathsf{x}|\mathsf{y}=y}(x=0)\\ &\propto b_y p\,p_{\mathsf{y}|\mathsf{x}=1}(y) + (1-b_y)(1-p)p_{\mathsf{y}|\mathsf{x}=0}(y)\end{aligned}$$

令 $\frac{\partial R(x,b_y)}{\partial b_y}=0$，解得

$$b_y = \begin{cases}0 & (1-p)p_{\mathsf{y}|\mathsf{x}=0}(y) > p\,p_{\mathsf{y}|\mathsf{x}=1}(y)\\ 1 & (1-p)p_{\mathsf{y}|\mathsf{x}=0}(y) \le p\,p_{\mathsf{y}|\mathsf{x}=1}(y)\end{cases} \tag{2}$$

(c) i. 是
ii. 否
iii. 与 p 的取值和分布 $p_{\mathsf{y}|\mathsf{x}}(\cdot|\cdot)$ 无关。

6.3. (a) $\mathbb{E}_p[\mathsf{y}]=t \Rightarrow p_1+2p_2=t$ 若 $\mathbb{E}_p[\mathsf{y}]=0 \Rightarrow p_0=1, p_1=p_2=0$
$\mathcal{L}_0=\{(1,0,0)\}$

(b) 当 $t=\frac{1}{2}$ 时，$\mathcal{L}_{\frac{1}{2}}$ 是如图 1所示的线段。

(c) 正交于 $\mathcal{L}_{\frac{1}{2}}$ 的指数族 $\mathcal{E}$ 上分布 q 可以写成如下的形式：

$$q(y) = p^*(y)e^{xy-\alpha(x)}, y=0,1,2 \tag{3}$$

其中 $p^*(y)\in\mathcal{L}_{\frac{1}{2}}$, 它满足如下的方程：

$$p^*(0)+p^*(1)+p^*(2) = 1 \tag{4}$$

$$p^*(1)+2p^*(2) = \frac{1}{2} \tag{5}$$

$$\tag{6}$$

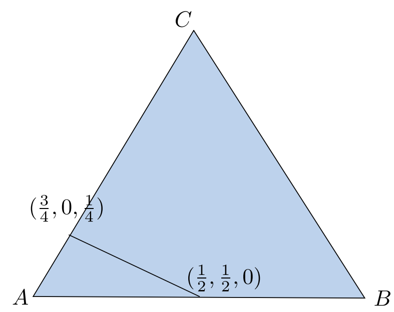

Figure 1: $\mathcal{L}_{\frac{1}{2}}$

利用 q 点在指数族上，有如下比例关系：

$$\frac{p^*(0)}{q(0)} = \frac{p^*(1)e^{x^*}}{q(1)} = \frac{p^*(2)e^{2x^*}}{q(2)}$$

由上面四个等式可以解出 $p^* = (\frac{2}{3}, \frac{1}{6}, \frac{1}{6})$ 在(3)式中，当 $x \to -\infty$ 时，由于 $q(0)$ 衰减最慢及概率的归一化条件可得：$q_{x\to-\infty} = (1, 0, 0)$, 同理可得 $q_{x\to\infty} = (0, 0, 1)$ 指数族 $\mathcal{E}$ 如图 2所示:

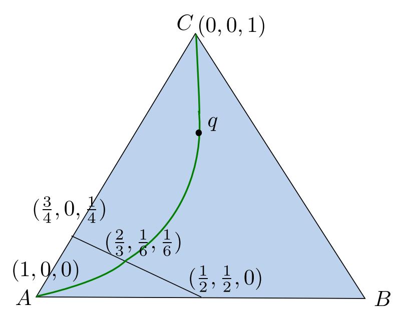

Figure 2: 正交于 $\mathcal{L}_{\frac{1}{2}}$ 的指数族 $\mathcal{E}$

(d) (c) 中已求出 $p^* = (\frac{2}{3}, \frac{1}{6}, \frac{1}{6})$

(e) 区域 $\mathcal{P}$ 如图 3所示:

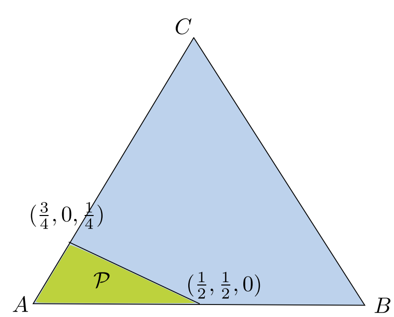

Figure 3: 区域 $\mathcal{P}$

(f) 若 p^* 在区域 $\mathcal{P}$ 内部取得, 则由偏导数为 0 的性质推出 $p^* = q$ 矛盾。因此只需分别考察 $D(\cdot||q)$ 在三角形区域 $\mathcal{P}$ 的三条边上的取值，得到

$$\arg\min_{p\in\mathcal{P}} D(p||q) = (\frac{2}{3}, \frac{1}{6}, \frac{1}{6})$$

与 (d) 的结果相同。

6.4. Thanks to 陆石, who gives me this template.