Stats 300A HW6 Solutions

Song Mei

November 15, 2018

Problem 1

(a)

Since $\boldsymbol{T}(\boldsymbol{x})$ is a sufficient statistics of statistical model $\mathcal{P}_1$, by definition, the conditional distribution of $[\boldsymbol{X}|\boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t}]$ under each $\mathsf{P}_\theta \in \mathcal{P}_1$ are the same. Since $\mathcal{P}_0 \subseteq \mathcal{P}_1$, the conditional distribution of $[\boldsymbol{X}|\boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t}]$ under each $\mathsf{P}_\theta \in \mathcal{P}_0$ are the same. By definition, $\boldsymbol{T}(\boldsymbol{x})$ is a sufficient statistics of statistical model $\mathcal{P}_0$.

(b)

For any g such that $\mathsf{E}_{\boldsymbol{\theta}}[g(\boldsymbol{T}(\boldsymbol{X}))] = 0$ for $\theta \in \mathcal{P}_1$, by the completeness of $\boldsymbol{T}$ w.r.t. $\mathcal{P}_0$, we have $\mathsf{P}_{\boldsymbol{\theta}}(g(\boldsymbol{T}(\boldsymbol{X})) = 0) = 1$ for $\boldsymbol{\theta} \in \mathcal{P}_0$. By the null set property of $\mathcal{P}_1$ and $\mathcal{P}_0$, we have $\mathsf{P}_{\boldsymbol{\theta}}(g(\boldsymbol{T}(\boldsymbol{X})) = 0) = 1$ for $\boldsymbol{\theta} \in \mathcal{P}_1$. That is, $\boldsymbol{T}$ is complete sufficient w.r.t. $\mathcal{P}_1$.

(c)

Consider the exponential family

$$\mathcal{P}_0 = \Big\{ p_{\boldsymbol{\theta}}(x) = c(\boldsymbol{\theta}) \exp\Big(\theta_1 \sum_{i=1}^n x_i + \ldots + \theta_n \sum_{i=1}^n x_i^n - \sum_{i=1}^n x_i^{2n} \Big) : \boldsymbol{\theta} \in \mathbb{R}^n \Big\}.$$

It is easy to see that $\mathcal{P}_0 \subseteq \mathcal{P}$. By the property of exponential family, the complete sufficient statistics for $\mathcal{P}_0$ gives $T_0(\boldsymbol{x}) = (\sum_{i=1}^n x_i, \ldots, \sum_{i=1}^n x_i^n)$. Denote $T(\boldsymbol{x}) = (x_{(1)}, \ldots, x_{(n)})$. By the hints, for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have $T_0(\boldsymbol{x}) = T_0(\boldsymbol{y})$ if and only if $T(\boldsymbol{x}) = T(\boldsymbol{y})$. Hence T is also a complete sufficient statistics for $\mathcal{P}_0$. Since all distributions in $\mathcal{P}$ has densities, it satisfies the null set property. By problem (b), T is a complete sufficient statistics for $\mathcal{P}$.

(d)

Consider estimator

$$A(\boldsymbol{X}) = \sum_{i=1}^n \mathbf{1}\{x_{(i)} \le x\}/n.$$

This estimator is a function of the order statistics. It is also unbiased for $\int_{-\infty}^x f(t)\mathrm{d}t$. So it is UMVU.

Problem 2

(a)

Let $\boldsymbol{x} \sim \mathrm{Binom}(2, \theta)$. For any estimator $\hat{\theta}(\boldsymbol{x})$, we have

$$\mathsf{E}_\theta[\hat{\theta}(\boldsymbol{x})] = \mathsf{E}[\hat{\theta}(0)](1-\theta)^2 + \mathsf{E}[\hat{\theta}(1)]2\theta(1-\theta) + \mathsf{E}[\hat{\theta}(2)]\theta^2.$$

For any fixed numbers $\mathsf{E}[\hat\theta(i)]$ for $i = 0, 1, 2$, $\mathsf{E}_\theta[\hat\theta(\boldsymbol{x})]$ is a degree 2 polynomial of θ, and it cannot be identical to a degree 3 polynomial.

Let $\boldsymbol{x} \sim \mathrm{Binom}(3, \theta)$. For any deterministic estimator $\hat\theta(\boldsymbol{x})$, we have

$$\mathsf{E}_\theta[\hat\theta(\boldsymbol{x})] = \hat\theta(0)(1-\theta)^3 + \hat\theta(1)3\theta(1-\theta)^2 + \hat\theta(2)3\theta^2(1-\theta) + \hat\theta(3)\theta^3.$$

Setting $\hat\theta(3) = 1$ and $\hat\theta(0) = \hat\theta(1) = \hat\theta(2) = 0$, we get an unbiased estimator for θ^3.

(b)

An unbiased estimator is

$$A(\boldsymbol{x}) = \mathbf{1}\{x_1 = x_2 = x_3 = 1\}.$$

Using the conditioning mechanism, we have

$$\begin{aligned}
A_\star(t) =& \mathsf{E}_\theta[\mathbf{1}\{X_1 = X_2 = X_3 = 1\} | T(\boldsymbol{X}) = t] = \mathsf{P}_\theta(X_1 = X_2 = X_3 = 1 | T(\boldsymbol{X}) = t)\\
=& \mathsf{P}_\theta\Big(X_1 = X_2 = X_3 = 1, \sum_{i=4}^n X_i = t-3\Big) / \mathsf{P}_\theta\Big(\sum_{i=1}^n X_i = t\Big)\\
=& \theta^3 \binom{n-3}{t-3}\theta^{t-3}(1-\theta)^{n-t}\mathbf{1}\{t \ge 3\} / \Big[\binom{n}{t}\theta^t(1-\theta)^{n-t}\Big]\\
=& t(t-1)(t-2)/[n(n-1)(n-2)].
\end{aligned}$$

Hence the UMVU gives $A_\star(T(\boldsymbol{x})) = T(\boldsymbol{x})(T(\boldsymbol{x})-1)(T(\boldsymbol{x})-2)/[n(n-1)(n-2)]$.

Problem 3

(a)

The log-likelihood function gives

$$\begin{aligned}
\ell(\boldsymbol{\theta}) =& \log \prod_{i=1}^n \Big[\mathsf{P}(Y_i = y_i | \boldsymbol{X}_i = \boldsymbol{x}_i)\mathsf{p}_{\boldsymbol{X}}(\boldsymbol{x}_i)\Big]\\
=& \sum_{i=1}^n \log \frac{\exp\{y_i\langle\boldsymbol{\theta}, \boldsymbol{x}_i\rangle\}}{1 + \exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}_i\rangle\}}\mathsf{p}_{\boldsymbol{X}}(\boldsymbol{x}_i)\\
=& \sum_{i=1}^n \Big[y_i\langle\boldsymbol{\theta}, \boldsymbol{x}_i\rangle - \log(1 + \exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}_i\rangle\}) + \log \mathsf{p}_{\boldsymbol{X}}(\boldsymbol{x}_i)\Big].
\end{aligned}$$

Hence

$$\dot\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \Big[y_i - \exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle\}/(1 + \exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle\})\Big]\boldsymbol{x}_i.$$

The Fisher information matrix gives

$$\begin{aligned}
I_F(\boldsymbol{\theta}) =& n\mathsf{E}_{\boldsymbol{\theta}}\{[y - \exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle\}/(1 + \exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle\})]^2\boldsymbol{x}\boldsymbol{x}^\mathsf{T}\}\\
=& n\mathsf{E}_{\boldsymbol{x}}\mathsf{E}_{\boldsymbol{\theta}}\{[y - \exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle\}/(1 + \exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle\})]^2\boldsymbol{x}\boldsymbol{x}^\mathsf{T} | \boldsymbol{x}\}\\
=& n\mathsf{E}_{\boldsymbol{x}}\{\{[\exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle\}/(1 + \exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle\})]^2/(1 + \exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle\})\\
&+ [1/(1 + \exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle\})]^2\exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle\}/(1 + \exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle\})\}\boldsymbol{x}\boldsymbol{x}^\mathsf{T}\}\\
=& n\mathsf{E}_{\boldsymbol{x}}\{[\exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle\}/(1 + \exp\{\langle\boldsymbol{\theta}, \boldsymbol{x}\rangle\})^2]\boldsymbol{x}\boldsymbol{x}^\mathsf{T}\}.
\end{aligned}$$

(b)

Let U be the orthogonal transform such that $U\boldsymbol{\theta} = \theta \boldsymbol{e}_1$ where $\theta = \|\boldsymbol{\theta}\|_2$. Let $\boldsymbol{z} \equiv U\boldsymbol{x}$. Since $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, we have $\boldsymbol{z} = U\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. Note $\langle \boldsymbol{\theta}, \boldsymbol{x} \rangle = \theta z_1$, we have

$$I_F(\boldsymbol{\theta}) = n\mathsf{E}_{\boldsymbol{x}}\{[\exp\{\langle \boldsymbol{\theta}, \boldsymbol{x}\rangle\}/(1+\exp\{\langle \boldsymbol{\theta}, \boldsymbol{x}\rangle\})^2]\boldsymbol{x}\boldsymbol{x}^\mathsf{T}\} = nU^\mathsf{T}\mathsf{E}_{\boldsymbol{z}}\{[\exp\{\theta z_1\}/(1+\exp\{\theta z_1\})^2]\boldsymbol{z}\boldsymbol{z}^\mathsf{T}\}U.$$

Denote

$$S(\theta) \equiv \mathsf{E}_{\boldsymbol{z}}\{[\exp\{\theta z_1\}/(1+\exp\{\theta z_1\})^2]\boldsymbol{z}\boldsymbol{z}^\mathsf{T}\}.$$

Then $I_F(\boldsymbol{\theta}) = nU^\mathsf{T} S(\theta) U$.

The off-diagonal elements of $S(\theta)$ must be 0, because z_i and z_j for $i \neq j$ are mean zero independent Gaussians. The first diagonal element of $S(\theta)$ gives

$$d_1(\theta) \equiv S(\theta)_{11} = \mathsf{E}_{G\sim\mathcal{N}(0,1)}\{[\exp\{\theta G\}/(1+\exp\{\theta G\})^2]G^2\}.$$

The other diagonal elements of $S(\theta)$ gives

$$d_2(\theta) \equiv S(\theta)_{22} = \cdots = S(\theta)_{nn} = \mathsf{E}_{G\sim\mathcal{N}(0,1)}\{[\exp\{\theta G\}/(1+\exp\{\theta G\})^2]\}.$$

As a result, we have

$$\begin{aligned} I_F(\boldsymbol{\theta}) =& nU^\mathsf{T} S(\theta) U = nU^\mathsf{T}[d_2(\theta)\mathbf{I} + (d_1(\theta) - d_2(\theta))\boldsymbol{e}_1\boldsymbol{e}_1^\mathsf{T}]U = n[d_2(\theta)\mathbf{I} + (d_1(\theta) - d_2(\theta))\boldsymbol{\theta}\boldsymbol{\theta}^\mathsf{T}/\theta^2] \\ =& c_0(\|\boldsymbol{\theta}\|_2)\mathbf{I}_d + c_1(\|\boldsymbol{\theta}\|_2)\boldsymbol{\theta}\boldsymbol{\theta}^\mathsf{T}, \end{aligned}$$

where

$$\begin{aligned} c_0(\theta) =& n\mathsf{E}_{G\sim\mathcal{N}(0,1)}\{[\exp\{\theta G\}/(1+\exp\{\theta G\})^2]\}, \\ c_1(\theta) =& n\mathsf{E}_{G\sim\mathcal{N}(0,1)}\{[\exp\{\theta G\}/(1+\exp\{\theta G\})^2](G^2-1)\}/\theta^2. \end{aligned}$$

(c)

Let $\boldsymbol{y} \equiv \Sigma^{-1/2}\boldsymbol{x}$. Since $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, we have $\boldsymbol{y} = \Sigma^{-1/2}\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. Let U be the orthogonal transform such that $U\Sigma^{1/2}\boldsymbol{\theta} = \theta\boldsymbol{e}_1$, where $\theta = \|\Sigma^{1/2}\boldsymbol{\theta}\|_2$. Let $\boldsymbol{z} = U\boldsymbol{y} = U\Sigma^{-1/2}\boldsymbol{x}$, then $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. We have (note $\langle \boldsymbol{\theta}, \boldsymbol{x}\rangle = \theta z_1$, where $\theta = \|\Sigma^{1/2}\boldsymbol{\theta}\|_2$)

$$\begin{aligned} I_F(\boldsymbol{\theta}) =& n\Sigma^{1/2}U^\mathsf{T}\mathsf{E}_{\boldsymbol{z}}\{[\exp\{\theta z_1\}/(1+\exp\{\theta z_1\})^2]\boldsymbol{z}\boldsymbol{z}^\mathsf{T}\}U\Sigma^{1/2} = n\Sigma^{1/2}U^\mathsf{T}S(\theta)U\Sigma^{1/2}, \\ =& n\Sigma^{1/2}U^\mathsf{T}[d_2(\theta)\mathbf{I} + (d_1(\theta) - d_2(\theta))]U\Sigma^{1/2} = n[d_2(\theta)\Sigma + (d_1(\theta) - d_2(\theta))\Sigma\boldsymbol{\theta}\boldsymbol{\theta}^\mathsf{T}\Sigma/\theta^2] \\ =& c_0(\|\Sigma^{1/2}\boldsymbol{\theta}\|_2)\Sigma + c_1(\|\Sigma^{1/2}\boldsymbol{\theta}\|_2)\Sigma\boldsymbol{\theta}\boldsymbol{\theta}^\mathsf{T}\Sigma, \end{aligned}$$

where

$$\begin{aligned} c_0 =& n\mathsf{E}_{G\sim\mathcal{N}(0,1)}\{[\exp\{\theta G\}/(1+\exp\{\theta G\})^2]\}, \\ c_1 =& n\mathsf{E}_{G\sim\mathcal{N}(0,1)}\{[\exp\{\theta G\}/(1+\exp\{\theta G\})^2](G^2-1)\}/\theta^2. \end{aligned}$$