Statistical Theory Problem set 6

Bohao Tang

Problem 1:

(i) $P(X_i, Y_{i,1}, Y_{i,2}; \theta) = P(Y_{i,1}|X_i,\theta)\cdot P(Y_{i,2}|X_i,\theta)\cdot P(X_i|\theta)$

$$= e^{-\beta_0 \mu_0(X_i)} \frac{[\beta_0\mu_0(X_i)]^{Y_{i,1}}}{Y_{i,1}!} \cdot e^{-\mu_0(X_i)} \frac{[\mu_0(X_i)]^{Y_{i,2}}}{Y_{i,2}!} \cdot P_0(X_i)$$

(ii) $pr(Y_{i,1}+Y_{i,2}|X_i,\theta)$, since Poisson have additivity, and give X_i $Y_{i,1}$, $Y_{i,2}$ are independent fallow Poisson distribution

$$\Rightarrow Pr(Y_{i,1}+Y_{i,2}|X_i,\theta) = \text{Poisson}[(\beta_0+1)\mu_0(X_i)](Y_{i,1}+Y_{i,2})$$

$$= e^{-(\beta_0+1)\mu_0(X_i)} \frac{[(\beta_0+1)\mu_0(X_i)]^{Y_{i,1}+Y_{i,2}}}{(Y_{i,1}+Y_{i,2})!}$$

(iii) Given X_i, θ, $Y_{i,1}$ $Y_{i,2}$ are independent Poisson variable

therefore Given $X_i, \theta, Y_{i,1}+Y_{i,2}$, $Y_{i,1}$ is Binomial variable

conditionally $Y_{i,1} \sim \text{Binomial}\left(Y_{i,1}+Y_{i,2}, \frac{\beta_0\mu_0(X_i)}{\beta_0\mu_0(X_i)+\mu_0(X_i)}\right) = \text{Binomial}\left(Y_{i,1}+Y_{i,2}, \frac{\beta_0}{\beta_0+1}\right)$

$$\Rightarrow Pr(Y_{i,1}|Y_{i,1}+Y_{i,2}, X_i, \theta) = \binom{Y_{i,1}+Y_{i,2}}{Y_{i,1}} \left[\frac{\beta_0\mu_0(X_i)}{(\beta_0+1)\mu_0(X_i)}\right]^{Y_{i,1}} \left[\frac{\mu_0(X_i)}{(\beta_0+1)\mu_0(X_i)}\right]^{Y_{i,2}}$$

$$= \binom{Y_{i,1}+Y_{i,2}}{Y_{i,1}} \left(\frac{\beta_0}{\beta_0+1}\right)^{Y_{i,1}} \left(\frac{1}{\beta_0+1}\right)^{Y_{i,2}}.$$

(iv) $P = \frac{\beta_0\mu_0(X_i)}{(\beta_0+1)\mu_0(X_i)} = \frac{\beta_0}{\beta_0+1} \Rightarrow \beta_0 = \frac{P}{1-P}$

and MLE for P is $\frac{\sum Y_{i,1}}{\sum Y_{i,1}+\sum Y_{i,2}}$ $\Rightarrow$ MLE for β_0 is $\frac{\frac{\sum Y_{i,1}}{\sum(Y_{i,1}+Y_{i,2})}}{\sum Y_{i,2}/\sum(Y_{i,1}+Y_{i,2})}$

$$= \frac{\sum_i Y_{i,1}}{\sum_i Y_{i,2}}.$$

(V) See the code in appendix, The Function is Poisson.ratio. exactCI.(- - -)
It has a parameter "interval" to decide whether to form a interval all return all posible β

Here we first do simulation for $\beta = 0,1,2 \cdots,10$, each sampling 10000 samples and ~~do~~ return all possible β, we have β can $= 2,3,4,5$

Then do simulation for $\beta = 1, 1.1, 1.2, 1.3 \cdots, 6.0$, each sampling 100000 and return all β
we have β can $= 1.1 \rightarrow 5.1$

Since β actually forms a interval well, we lastly do simullation for $\beta = seq(1, 5.2, by=0.01)$
each sampling 1000000 and return a interval of β

we have $\beta \in (1.10, 5.12)$ So the 95% CI is $(1.10, 5.12)$ error up to 0.02
(notice that the result is somehow stable).

Problem 2:

(i)
$$\log Pr(X_i, Y_{i,1}, Y_{i,2}) = \log Pr(Y_{i,1}|X_i,\theta) + \log Pr(Y_{i,2}|X_i,\theta) + \log Pr(X_i|\theta)$$
$$= -\beta\mu(X_i) - \mu(X_i) + \log P(X_i) - \log Y_{i,1}! - \log Y_{i,2}!$$
$$+ Y_{i,1}\log\beta\mu(X_i) + Y_{i,2}\log\mu(X_i)$$

(ii)

(a)
$$\psi_{P,S}(D_i,\theta_0) = \frac{\partial}{\partial v_2}\log Pr(D_i;\theta)\Big|_{\theta=\theta_0} = \frac{\frac{\partial P}{\partial v_2}(x_i, v_2^*)}{P(x_i, v_2^*)}$$

(b)
$$\psi_{\mu,S}(D_i,\theta_0) = \frac{\partial}{\partial v_1}\log Pr(D_i;\theta)\Big|_{\theta=\theta_0} = -\beta\frac{\partial\mu}{\partial v_1}(x_i,v_1^*) - \frac{\partial\mu}{\partial v_1}(x_i,v_1^*)$$
$$+ Y_{i,1}\frac{\frac{\partial\mu}{\partial v_1}(x_i,v_1^*)}{\mu(x_i,v_1^*)} + Y_{i,2}\frac{\frac{\partial\mu}{\partial v_1}(x_i,v_1^*)}{\mu(x_i,v_1^*)}$$

(c)
$$\psi_{\beta,S}(D_i,\theta_0) = \frac{\partial}{\partial\beta}\log Pr(D_i;\theta)\Big|_{\theta=\theta_0} = -\mu(x_i,v_1^*) + \frac{Y_{i,1}}{\beta}$$
$$= -\mu_0(x_i) + \frac{Y_{i,1}}{\beta}$$

We can remove S because in the full model we also have parameter β
and its score in full model is also $-\mu_0(x_i) + \frac{Y_{i,1}}{\beta}$.

(iii)

(a) Since $E_{\theta_0} \psi'_{P,S}(D_i,\theta_0) = 0$, obviously all function in Λ_P have mean 0 under θ_0

On the other side For any function ~~under~~ $f(X_i,\theta_0)$ which $E_{\theta_0} f(X_i,\theta_0)=0$ we can build a submodel as.

$$p(X_i,\nu_2) = p_0(X_i)\left[1+(\nu_2-\nu_2^*)\, f(X_i,\theta_0)\right]$$

this is a likelihood because $E_{\theta_0} f(X_i,\theta_0)=0 \Rightarrow \int p_0(X_i)(\nu_2-\nu_2^*) f(X_i,\theta_0)\,dx_i = 0$

$\Rightarrow \int p(X_i,\nu_2)\,dX_i = 1$

Then the score is $\frac{\partial}{\partial \nu_2} \log p(X_i,\nu_2)\Big|_{\nu_2=\nu_2^*} = f(X_i,\theta_0)$

therefore $\Lambda_P = \{$all function of X_i and θ_0, that have mean 0 at $\theta_0\}$

(b) $$\psi_{\mu,s}(D_i,\theta_0) = (Y_{i,1}+Y_{i,2})\frac{\partial \mu}{\partial \nu_1}(X_i,\nu_1^*)\Big/\mu_0(X_i) - (\beta+1)\frac{\partial \mu}{\partial \nu_1}(X_i,\nu_1^*)$$

$$= \frac{\frac{\partial \mu}{\partial \nu_1}(X_i,\nu_1^*)}{\mu_0(X_i)}\left[(Y_{i,1}+Y_{i,2})-(\beta+1)\mu_0(X_i)\right]$$

~~Since~~. Since $Y_{i,1}, Y_{i,2}, \beta, \mu_0(X_i)$ are all given

$\text{span}\{\psi'_{\mu,s}(D_i,\theta_0)\} = \left[(Y_{i,1}+Y_{i,2})-(\beta+1)\mu_0(X_i)\right] \text{span}\left\{\frac{\partial \mu}{\partial \nu_1}(X_i,\nu_1^*)\Big/\mu_0(X_i)\right\}$

For all $g(X_i,\theta_0)$ just let submodel be $\mu(X_i,\nu_1) = e^{[(\nu_1-\nu_1^*)\, g(X_i,\theta_0)]} \cdot \mu_0(X_i)$

Then $\frac{\partial \mu}{\partial \nu_1}(X_i,\nu_1^*)/\mu_0(X_i) = g(X_i,\theta_0)$

$\Rightarrow \text{span}\{\psi_{\mu,s}(D_i,\theta_0)\} = \{g(X_i,\theta_0)\cdot[(Y_{i,1}+Y_{i,2}) - (1+\beta)\mu_0(X_i)], \text{all } g(X_i,\theta_0)\}$

~~Here X_i only have 15 value, therefore $g(X_i,\theta_0)$ will always be bounded~~

~~therefore choose ν_1: $|\nu_1-\nu_1^*|$ small enough $\mu(X_i,\nu_1)$ will be a mean function~~

Here the $\mu(X_i,\nu_1)$ is positive therefore do be a mean ~~(>0)~~ function (iv).

(a) $\psi_\beta(D_i,\theta_0) = \dfrac{Y_{i,1}}{\beta} - \mu_0(X_i)$

$\forall$ function $f(x_i,\theta_0) \in \Lambda_p$

$$E_{\theta_0} \psi_\beta(D_i,\theta_0) f(x_i,\theta_0) = E\left[E\left[\left(\frac{Y_{i,1}}{\beta} - \mu_0(X_i)\right]\cdot f(x_i,\theta_0)\right] \middle| X_i\right]$$
$$= E\left\{f(x_i,\theta_0)\, E\left[\frac{Y_{i,1}}{\beta} - \mu_0(X_i) \middle| X_i\right]\right\}$$
$$= E\left[f(x_i,\theta_0)\cdot 0\right] = 0$$

therefore $\psi_\beta(D_i,\theta_0) \perp \Lambda_p \Rightarrow$ proj $\psi_\beta(D_i,\theta_0)$ on $\Lambda_p = 0$

(b) Since $\psi_\beta(D_i,\theta_0)$ is $\perp \Lambda_p$

therefore proj of $\psi_\beta(D_i,\theta_0)$ on $\Lambda_p + \Lambda_\mu$ is the same as proj of $\psi_\beta(D_i,\theta_0)$ on Λ_μ.

therefore the efficient score for $\beta = \psi_\beta - \psi_\beta$ proj on $\Lambda_p + \Lambda_\mu$
$= \psi_\beta -$ proj ψ_β on Λ_μ.

(C) from (b) we know that $\psi(D_i,\theta_0) = \psi_\beta$ − some function in Λ_μ

$$\Rightarrow \psi(D_i,\theta_0) = \left[\frac{Y_{1,i}}{\beta} - \mu_0(X_i)\right] - \left[(Y_{i,1}+Y_{i,2}) - (1+\beta_0)\mu_0(X_i)\right] g^*(X_i,\theta_0)$$

Then since $\psi(D_i,\theta_0)$ is the residul of projection, we have $\psi(D_i,\theta_0) \perp \Lambda_\mu$

$$\Rightarrow E_{\theta_0}\left[\psi(D_i,\theta_0)\cdot\{(Y_{i,1}+Y_{i,2}) - (1+\beta_0)\mu_0(X_i)\}\cdot \check{g}(X_i,\theta_0)\right] = 0. \quad \forall \check{g}$$

(v) Do the expectation $E_{\theta_0}(\cdot)$ by $E[E(\cdot \mid X_i)]$

Then we have $E_{\theta_0}\left[\overset{\mu_0(X_i)}{\cancel{\beta_0 \mu_0^2(X_i)}}\, \check{g}(X_i,\theta_0) - g^*(X_i,\theta_0)\,\check{g}(X_i,\theta_0)\,(1+\beta_0)\mu_0(X_i)\right] = 0$

$$\Rightarrow E_{\theta_0}\left[\check{g}(X_i,\theta_0)\left[\underset{\mu_0(X_i)}{\cancel{\beta_0\mu_0^2(X_i)}} - g^*(1+\beta_0)\mu_0(X_i)\right]\right] = 0 \Rightarrow g^* = \cancel{\frac{\beta_0}{1+\beta_0}\mu_0(X_i)}\ \frac{1}{1+\beta_0}$$

(vi) Based on (v), the efficient score for β is

$$\psi(D_i, \theta_0) = \left[\frac{Y_{i,1}}{\beta_0} - \mu_0(X_i)\right] - \cancel{(Y_{i,1}+Y_{i,2})\frac{\beta_0}{1+\beta_0}\mu_0(X_i) + \beta_0\mu_0^2(X_i)}$$
$$- (Y_{i,1}+Y_{i,2})\frac{1}{1+\beta_0} + \mu_0(X_i)$$
$$= \frac{Y_{i,1}}{\beta_0} - (Y_{i,1}+Y_{i,2})\frac{1}{1+\beta_0}$$

Then solve efficient score function

$$\sum \psi(D_i, \hat{\theta}_0) = 0 \Rightarrow \frac{\sum Y_{i,1}}{\beta_0} - \frac{\sum(Y_{i,1}+Y_{i,2})}{1+\beta_0} = 0$$
$$\Rightarrow \hat{\beta}_0 = \frac{\sum Y_{i,1}}{\sum Y_{i,2}} = \text{MLE}$$

So that conditional MLE is somehow efficient in general sense that it archieves C-R bound for the whole semi-parametric model.