STA 512 HW8 Due 4/1 Yi Mi, ym154

1. (a) $L_n(\alpha, \beta) = \prod_{i=1}^{n} \frac{e^{(\alpha+\beta x_i)y_i}}{1+e^{\alpha+\beta x_i}}$

$$\ell_n(\alpha,\beta) = \sum_{i=1}^{n}\left[y_i(\alpha+\beta x_i) - \log(1+e^{\alpha+\beta x_i})\right] = \alpha\sum y_i + \beta\sum x_i y_i - \sum\log(1+e^{\alpha+\beta x_i})$$

The likelihood equations are

$$\frac{\partial \ell_n(\alpha,\beta)}{\partial\alpha} = \sum y_i - \sum\left[\frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}\right] = \sum y_i - \sum p_i = 0$$

$$\frac{\partial \ell_n(\alpha,\beta)}{\partial\beta} = \sum x_i y_i - \sum\left[\frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}\cdot x_i\right] = \sum x_i y_i - \sum p_i x_i = 0$$

(b) The second derivatives are

$$\frac{\partial^2}{\partial\alpha^2}\ell_n(\alpha,\beta) = -\sum\frac{e^{-(\alpha+\beta x_i)}}{[1+e^{-(\alpha+\beta x_i)}]^2}$$

$$\frac{\partial^2}{\partial\beta^2}\ell_n(\alpha,\beta) = -\sum\left[\frac{e^{-(\alpha+\beta x_i)}}{[1+e^{-(\alpha+\beta x_i)}]^2}\cdot x_i^2\right]$$

$$\frac{\partial^2}{\partial\alpha\partial\beta}\ell_n(\alpha,\beta) = -\sum\left[\frac{e^{-(\alpha+\beta x_i)}}{[1+e^{-(\alpha+\beta x_i)}]^2}\cdot x_i\right]$$

Let

$$\frac{e^{-(\alpha+\beta x_i)}}{[1+e^{-(\alpha+\beta x_i)}]^2} = q_i$$

Thus, the expected information is

$$I_n(\alpha,\beta) = -E\left(\frac{\partial^2}{\partial(\alpha,\beta)^2}\ell_n(\alpha,\beta)\right) = \begin{bmatrix}\sum q_i & \sum q_i x_i \\ \sum q_i x_i & \sum q_i x_i^2\end{bmatrix}$$

we then have

$$\left(\begin{bmatrix}\hat\alpha \\ \hat\beta\end{bmatrix} - \begin{bmatrix}\alpha \\ \beta\end{bmatrix}\right) \xrightarrow{d} N(0, I_n(a,b)^{-1})$$

The asy variance is $I_n^{-1}(\alpha,\beta)$. x_i exists in the second derivatives thus it's distribution relates to the asy variance.

(c) The observed information is

$$\hat{I}_n(\alpha, \beta) = -\frac{\partial^2}{\partial(\alpha,\beta)^2} \ell n(\hat{\alpha}, \hat{\beta}) = \begin{pmatrix} \Sigma \hat{q}_i & \Sigma \hat{q}_i x_i \\ \Sigma \hat{q}_i x_i & \Sigma \hat{q}_i x_i^2 \end{pmatrix}$$

The step of Newton-Raphson is as follows,

let $\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$

expand the derivative of the log-likelihood around θ^j

$$0 = \ell'(\hat{\theta}) = \ell'(\theta^j) + (\hat{\theta} - \theta^j)\ell''(\theta^j)$$

$$\Rightarrow \hat{\theta} \approx \theta^j - \frac{\ell'(\theta^j)}{\ell''(\theta^j)}$$

This suggests the following iterative schema.

$$\theta^{j+1} = \theta^j - \frac{\ell'(\theta^j)}{\ell''(\theta^j)}$$

Thus, the MLE $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ becomes

$$\theta^{j+1} = \theta^j - H^{-1}\ell'(\theta^j)$$

$\ell'(\theta^j)$ is the vector of the first derivatives

$$\ell'(\theta^j) = \begin{pmatrix} \Sigma y_i - \Sigma p_i \\ \Sigma x_i y_i - \Sigma p_i x_i \end{pmatrix}$$

H is the matrix of second derivatives.

$$H = \begin{pmatrix} \Sigma q_i & \Sigma q_i x_i \\ \Sigma q_i x_i & \Sigma q_i x_i^2 \end{pmatrix}$$

2. Let $Y_1 \dots Y_n \sim$ iid P_{θ_0} where $P_{\theta_0} \in \mathcal{P} = \{P_\theta : \theta \in \Theta\}$ and P_θ has density $p(y|\theta) = \theta y^{\theta-1}$ for $y \in (0,1)$

(a) $E(Y) = \int_0^1 y \cdot \theta \cdot y^{\theta-1} \, dy = \theta \int_0^1 y^{\theta} \, dy = \theta \frac{1}{\theta+1} y^{\theta+1} \Big|_0^1 = \frac{\theta}{\theta+1}$

$\mu = E(Y) = \frac{\theta}{\theta+1}$

then $\theta = \frac{\mu}{1-\mu}$

plug into the density, we get

$$p(y|\mu) = \frac{\mu}{1-\mu} \, y^{\frac{2\mu-1}{1-\mu}}$$

(b) $E(Y^2) = \int_0^1 y^2 \, \theta \cdot y^{\theta-1} \, dy = \theta \cdot \int_0^1 y^{\theta+1} \, dy = \theta \frac{1}{\theta+2} y^{\theta+2} \Big|_0^1 = \frac{\theta}{\theta+2}$

$Var(Y) = E(Y^2) - (EY)^2 = \frac{\theta}{\theta+2} - \left(\frac{\theta}{\theta+1}\right)^2 = \frac{\theta}{(\theta+2)(\theta+1)^2}$

$Var(\bar{Y}) = Var(Y)/n = \frac{\theta}{n(\theta+2)(\theta+1)^2} = \frac{\mu(1-\mu)^2}{n(2-\mu)}$

(c) For an unbiased estimator, the Cramér-Rao lower bound is

$Var(\hat{\mu}) \geq \frac{1}{I(\mu)}$, where $I(\mu)$ is the fisher information

Likelihood is

$L_n(\theta) = \prod_{i=1}^{n} \theta \cdot y^{\theta-1}$

$\ell_n(\theta) = \Sigma(\log\theta + (\theta-1)\log y_i) = n\log\theta + (n-1)\Sigma \log y_i$

The derivatives are

$$\frac{\partial l_n(\theta)}{\partial \theta} = \frac{n}{\theta} + \Sigma \log y_i$$

$$\frac{\partial^2 l_n(\theta)}{\partial \theta^2} = -\frac{n}{\theta^2}$$

The information is

$$I(\theta) = -E\left(\frac{\partial^2 l_n(\theta)}{\partial \theta^2}\right) = \frac{n}{\theta^2}$$

Thus. the lower Bond for θ is

$$Var(\hat{\theta}) \geq 1/I(\theta) = \frac{\theta^2}{n}$$

The derivative of μ to θ is

$$\mu = \frac{\theta}{\theta+1}, \quad \mu'(\theta) = \frac{1}{(\theta+1)^2}$$

By Delta method

$$Var(\hat{\mu}) \geq \frac{\theta^2}{n} \frac{1}{(\theta+1)^4}$$

And we have

$$Var(\bar{Y}) = \frac{\theta}{n(\theta+2)(\theta+1)^2}$$

$$\frac{Var(\hat{\mu})}{Var(\bar{Y})} = \frac{\theta^2+2\theta}{\theta^2+2\theta+1} < 1$$

$$\Rightarrow Var(\hat{\mu}) < Var(\bar{Y})$$

(d) Since $\mu = \frac{\theta}{\theta+1}$

$$\hat{\mu}_{MLE} = \frac{\hat{\theta}_{MLE}}{\hat{\theta}_{MLE}+1}$$

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \frac{\partial \ell_n(\theta)}{\partial \theta} = -\frac{\sum \log y_i}{n}$$

$$\hat{\mu}_{MLE} = \frac{\sum \log y_i}{\sum \log y_i - n}$$

The asy variance of $\hat{\theta}$ is

$$\hat{Var}(\hat{\theta}_{MLE}) = \frac{\theta^2}{n}$$

By Delta Method

$$\hat{Var}(\hat{\mu}_{MLE}) = \frac{\theta^2}{n(\theta+1)^4}$$

By the results from the last question,

$$\hat{Var}(\hat{\mu}_{MLE}) < Var(\bar{y})$$

3. (a) Let $T = t(\underset{\sim}{Y})$ be the more general estimator that could be biased, and let $E(T) = \psi(\theta) = \int t(y) f(y;\theta)\, dy$

Score function is $S = \frac{\partial}{\partial\theta} \log f(Y;\theta)$

$$= \frac{1}{f(Y;\theta)} \frac{\partial}{\partial\theta} f(Y;\theta)$$

Since, $E(S) = \int \left[\frac{1}{f(y;\theta)} \frac{\partial}{\partial\theta} f(y;\theta)\right] \cdot f(y;\theta)\, dy$

$$= \int \frac{\partial}{\partial\theta} f(y;\theta)\, dy$$

$$= 0$$

thus,

$Cov(S,T) = E(ST) - E(S)E(T) = E(ST)$

$$= E\left(T \cdot \left[\frac{1}{f(Y;\theta)} \frac{\partial}{\partial\theta} f(Y;\theta)\right]\right)$$

$$= \int t(y) \left[\frac{1}{f(y;\theta)} \frac{\partial}{\partial\theta} f(y;\theta)\right] f(y;\theta)\, dy$$

$$= \frac{\partial}{\partial\theta} \int t(y) f(y;\theta)\, dy$$

$$= \psi'(\theta)$$

According to the Cauchy-Schwarz inequality,

$$\sqrt{Var(X) Var(Y)} \geq |cov(X,Y)|$$

We have

$$\sqrt{Var(S) Var(T)} \geq |cov(S,T)| = |\psi'(\theta)|$$

$$Var(T) \geq \frac{\psi'(\theta)^2}{Var(S)} = \frac{\psi'(\theta)^2}{I(\theta)}$$

(b) the posterior dist is

$$p(\mu \mid \underset{\sim}{y}) = p(\mu) \prod_{i=1}^{n} p(y \mid \mu) \times c$$

$$= \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2}\frac{\mu^2}{\tau^2}\right\} \times \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y-\mu)^2\right\} \times c$$

The likelihood is

$$L(\mu) = p(\mu \mid \underset{\sim}{y})$$

$$\ell(\mu) = -\frac{\mu^2}{2\tau^2} - \sum \frac{(y_i-\mu)^2}{2} + c = -\frac{\mu^2}{2\tau^2} - \frac{n}{2}\mu^2 + \mu \sum y_i + c$$

The derivates are

$$\frac{\partial \ell(\mu)}{\partial \mu} = -\frac{\mu}{\tau^2} - n\mu + \sum y_i$$

$$\frac{\partial^2 \ell(\mu)}{\partial \mu} = -\frac{1}{\tau^2} - n$$

The information is

$$I(\mu) = -E\left(-\frac{1}{\tau^2} - n\right) = \frac{1}{\tau^2} + n$$

And we know the posterior exp is

$$E(\hat{\mu}) = \frac{n\mu}{n + 1/\tau^2}$$

then $\frac{\partial E(\hat{\mu})}{\partial \mu} = \frac{n}{n + 1/\tau^2}$

Thus, the lower bound is

$$Var(\hat{\mu} \mid \underset{\sim}{y}) \geq \frac{n^2}{(n + 1/\tau^2)^3}$$

The frequentist variance is $Var(\hat{\mu} \mid \mu) = \tau^2$

$$\lim \frac{Var(\hat{\mu} \mid \underset{\sim}{y})}{Var(\hat{\mu} \mid \mu)} \to 0$$

the variance of $\hat{\mu}$ is smaller that the freq variance

4. (a) $Y_1 \cdots Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$

From the lecture note.

$\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$

thus,

$$\frac{\sqrt{n}(\bar{Y}-\mu)}{\sigma} \sim N(0,1)$$

(b) $y_i \sim N(\mu, \sigma^2)$

$\frac{y_i-\mu}{\sigma} \sim N(0,1)$

let $x_i = \frac{y_i-\mu}{\sigma} \sim N(0,1)$

by fact (a), we have

$$\Sigma(x_i-\bar{x})^2 \sim \chi^2_{n-1}$$

$\bar{x} = \frac{\Sigma x_i}{n} = \frac{\bar{y}-\mu}{\sigma}$

therefore,

$$\Sigma(x_i-\bar{x})^2 = \Sigma\left[\frac{y_i-\mu}{\sigma} - \frac{\bar{y}-\mu}{\sigma}\right]^2 = \frac{\Sigma(y_i-\bar{y})^2}{\sigma^2} \sim \chi^2_{n-1}$$

$s^2 = \frac{\Sigma(y_i-\bar{y})^2}{n-1} \Rightarrow \frac{(n-1)s^2}{\sigma^2} = \frac{\Sigma(y_i-\bar{y})^2}{\sigma^2}$

Then $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$

(c) $\begin{cases} \dfrac{\sqrt{n}(\bar{y}-\mu)}{\sigma} \sim N(0,1) \\ \dfrac{\sum(y_i-\bar{y})^2}{\sigma^2} \sim \chi^2_{n-1} \end{cases}$

by fact (b)

$$\frac{\frac{\sqrt{n}(\bar{y}-\mu)}{\sigma}}{\sqrt{\frac{\sum(y_i-\bar{y})^2}{(n-1)\sigma^2}}} = \frac{\sqrt{n}(\bar{y}-\mu)}{s} \sim t_{n-1}$$

(d) I'd like to construct a t-test.

reject H if

$$\left|\frac{\bar{y}-\mu_0}{s/\sqrt{n}}\right| > t_{1-\alpha/2}$$

the acceptance region is

$$A(\mu_0) = \left\{ \underset{\sim}{y} : \left|\frac{\bar{y}-\mu_0}{s/\sqrt{n}}\right| < t_{1-\alpha/2} \right\}$$