SM.

AIC: $-2\sum\log + 2\dim(\Theta)$

BIC: $-2\sum\log + \dim(\Theta)\log n$

Mallow C_p: $\|Y-\hat{\mu}\|^2 + 2p\sigma^2$.

F-test: $\dfrac{\|(I-H)Y\|^2}{\|(H_0-H)Y\|^2} \sim F_{n-p, p-p_0}$.

Expo family: $f(\underline{y};\underline{\theta}) = e^{\theta^T T(y) - K(\underline{\theta})} f_0(\underline{y})$.

Ex. Binomial $f(y;\pi) = \binom{n}{y}\pi^y(1-\pi)^{n-y} = e^{y\log\frac{\pi}{1-\pi} + n\log(1-\pi)}\binom{n}{y}$

$\theta = \log\frac{\pi}{1-\pi}$. Expit function $\pi(\theta) = \frac{e^\theta}{1+e^\theta}$.

$K(\theta) = n\log(1+e^\theta)$.

Multinomial: $f(y;\pi) = \sim \exp\{\sum_{l=1}^{L-1} y_l \log\left(\frac{\pi_l}{1-\sum_{l=1}^{L-1}\pi_l}\right) + n\log(1-\sum_{l=1}^{L-1}\pi_l)\}$.

θ: $(L-1)$dim, $\theta_l = \log\frac{\pi_l}{\pi_L} = \log\frac{\pi_l}{1-\sum_{l=1}^{L-1}\pi_l}$, $\pi_l(\theta) = \frac{e^{\theta_l}}{\sum_{l=1}^{L} e^{\theta_l}}$

$K(\theta) = \log(\sum_{l=1}^{L} e^{\theta_l}) = -\log(1-\sum_{l=1}^{L-1}\pi_l)$.

$\mu(\theta) = K'(\theta)$, $V(\theta) = K''(\theta)$; $\mu'(\theta) = K''(\theta) = V(\theta) > 0$.

$\mu(\theta) = \mathbb{E}_\theta(Y)$, $V(\theta) = Var_\theta(Y)$

Write inverse of $\mu(\theta)$ is $\theta(\mu)$. Then $\theta'(\mu) = \frac{1}{V(\theta)}$.

cumulant: $\frac{1}{\sigma^2}(K(\theta + t\sigma^2) - K(\theta))$

$l(\theta) = n\{\theta\bar{Y} - K(\theta)\} + \text{const.}$

$$i^{(n)}(\theta) = Var(l'(\theta)) = \mathbb{E}(-l''(\theta)).$$

$$= nK''(\theta) = nV(\theta).$$

Score: $U(\theta) = l'(\theta) \cancel{= nK''(\theta) = nV(\theta)} = n(\bar{Y} - K'(\theta)) = n\{\bar{Y} - \mu(\theta)\}.$

$\Rightarrow i^{(n)}(\theta) = nK''(\theta) = nV(\theta).$

θ

Asymptotic: $\sqrt{n}(\hat{\eta} - \eta) \xrightarrow{d} N(0, \tau^2)$

$\sqrt{n}\{g(\hat{\eta}) - g(\eta)\} \xrightarrow{d} N(0, \tau^2 g'(\eta)^2).$

$$D(\theta_1, \theta_2) = 2\mathbb{E}_{\theta_1}\left\{\log \frac{f(Y;\theta_1)}{f(Y;\theta_2)}\right\} = 2\{(\theta_1 - \theta_2)\mu_1 - K(\theta_1) + K(\theta_2)\}.$$

$\mu(\theta, \sigma^2) = K'(\theta)$; $V(\theta, \sigma^2) = \sigma^2 K''(\theta).$

GLM.

(i)

(ii) $\eta_i = x_i^T \beta$

(iii) $\mu_i = \mathbb{E}(Y_i | X_i)$; $\eta_i = g(\mu_i)$

link function

Poisson regression: $\log \mu_{i,j} = \beta_i + r_j$. $y_{i,j} \overset{iid}{\sim} Po(\mu_{i,j})$.

$\theta_i = \log(\mu_i)$, $K(\theta) = e^\theta$, $g(\mu) = \theta(\mu) = \log(\mu)$

$\mu_i = e^{x_i^T\beta} = \prod_{j=1}^{p} (e^{\beta_j})^{x_{ij}}.$

Ex

Fisher update: $I(\beta) = \mathrm{Var}(U(\beta;Y))$

$$= \mathbb{E}_{\beta}\{-\nabla^2 l(\beta;Y)\}.$$

$$U(\beta;Y) = \nabla l(\beta;Y).$$

$$\hat{\beta}^{(m)} = \hat{\beta}^{(m-1)} + i^{-1}(\hat{\beta}^{(m-1)})\,U(\hat{\beta}^{(m-1)})$$

2022. P2, S1.

(b) $f(\bar{Y};\theta) = \prod f(y_i;\theta)$

$$= \prod e^{\theta y_i - k(\theta)} f_0(y_i)$$

$$= e^{\theta \sum y_i - nk(\theta)} \prod f_0(y_i)$$

$$\cancel{= e^{n\theta(\frac{\sum y_i}{n})}}$$

$$= e^{n(\theta \bar{y} - k(\theta))} \prod f_0(y_i)$$

$$\cancel{= e} = e^{\theta' \bar{y} - k(\frac{\theta'}{n}) \cdot n} \prod f_0(y_i)$$

$$\cancel{\sum y_i - n k'(\theta) = 0.}$$

$$\boxed{k'(\theta) = \mu(\theta)}.$$

$$= \frac{\sum y_i}{n}.$$

(b) $\mathbb{E}(\|Y-\hat{\mu}_\lambda\|^2)$

$= \mathbb{E}(\|\mu+\varepsilon - H_\lambda(\mu+\varepsilon)\|^2)$

$= \mathbb{E}(\|\mu(I-H_\lambda)\|^2) + \mathbb{E}(\|(I-H_\lambda)\varepsilon\|^2)$.

$\cancel{\varepsilon^T(I-H_\lambda)^T(I-H_\lambda)\varepsilon}$

$= \|(I-H_\lambda)\mu\|^2 + \{n - 2tr(H_\lambda) + tr(H_\lambda^2)\}\sigma^2$

(c) $\mathbb{E}(\|Y^* - \hat{\mu}_\lambda\|^2)$

$= \mathbb{E}(\|\mu+\varepsilon^* - H_\lambda(\mu+\varepsilon)\|^2)$

$= \mathbb{E}(\|(I-H_\lambda)\mu + \varepsilon^* - H_\lambda\varepsilon\|^2)$

$= \|(I-H_\lambda)\mu\|^2 + n\sigma^2 + tr(H_\lambda^2)\sigma^2$.

$= \mathbb{E}(\|Y-\hat{\mu}_\lambda\|^2) + 2\sigma^2 tr(H_\lambda)$. ✓

(d) $X = I$: $H_\lambda = (I+\lambda I)^{-1} = \frac{1}{\lambda+1} I$.

$\mathbb{E}(\|Y^* - \hat{\mu}_\lambda\|^2) = \|\frac{1}{\lambda+1}\mu\|^2 + n\sigma^2 + \frac{1}{(\lambda+1)^2}\sigma^2$

$= \left(\frac{\lambda}{\lambda+1}\right)^2\|\mu\|^2 + n\sigma^2 + \frac{1}{(\lambda+1)^2}\sigma^2$

min when $\lambda = \frac{\sigma^2}{\mu_i}$:

min $= \mathbb{E} = \frac{\mu^2(\sigma^2+\sigma^4)}{(\sigma^2+\mu)^2} + n\sigma^2$.

(c) $Y^* = \mu + \varepsilon^*$

$\mathbb{E}(\|Y^* - \hat{\mu}_\lambda\|^2)$

$= \mathbb{E}(\|\mu + \varepsilon^* -$

$\lambda = \frac{\sigma^2}{\mu}$.

$\left(\frac{\frac{\sigma^2}{\mu}}{\frac{\sigma^2}{\mu}+1}\right)^2$

$= \left(\frac{\sigma^2}{\sigma^2+\mu}\right)^2 \mu^2$

$+ \frac{1}{(\frac{\sigma^2}{\mu}+1)^2}\sigma^2$

$\frac{\mu^2\sigma^2}{(\sigma^2+\mu)^2}$

$+ 2\sigma^2 tr(H_\lambda)$

(a) ez.

(b) ez

(c) $Y^* = \mu + \varepsilon^*$. $\varepsilon^* \sim N(0, \sigma^2 I) \perp Y$.

$\mathbb{E}(\|Y^* - \hat{\mu}_\lambda\|^2 \cancel{+ 2\sigma^2 tr(H_\lambda)})$

$= \mathbb{E}(\|\mu + \varepsilon^* - H_\lambda(\mu+\varepsilon)\|^2)$

$= \mathbb{E}(\|(I-H_\lambda)\mu + \varepsilon^* - H_\lambda\varepsilon\|^2)$

$= \|(I-H_\lambda)\mu\|^2 + \sigma^2 I + \sigma^2$

$\frac{2\lambda\mu}{\cancel{(1+\lambda)}} = \frac{2\sigma^2}{\cancel{(1+\lambda)}}$

$\lambda = \frac{\sigma^2}{\mu}$.

$\mu = (X^TX)^{-1}X^TY$

$= Y$.

$(1+\frac{1}{\lambda})^{-2}$. $(\lambda+1)^{-2}$

minimize $\left(\frac{1}{1+\frac{1}{\lambda}}\right)^2 \mu^2 + \frac{1}{(\lambda+1)^2}\sigma^2$

$\to (-2)\cdot(-\lambda^{-2})(1+\lambda^{-1})^{-3}\mu^2 \cancel{+ (1+\lambda)^{-3}\sigma^2}$

$+ (-2)(\lambda+1)^{-3}\sigma^2$.

$= \frac{2\lambda^{-2}\mu}{(1+\frac{1}{\lambda})^3} - \frac{2}{(\lambda+1)^3}\sigma^2 = 0$

$$\hat{\sigma}^2 = \frac{\|Y - X\hat\beta\|^2}{n}.$$

$$\frac{\|Y - X\hat\beta\|^2}{n-p}$$

$$\hat{\sigma}^2 = \frac{\sigma^2}{n-p}\chi^2_{n-p}.$$

$$\hat\beta - \beta \qquad (0, (X^TX)^{-1})$$

$$\hat\sigma$$

$$\sigma = \sqrt{\frac{\sigma^2}{n-p}\chi^2_{n-p}}$$

$$\frac{\hat\beta - \beta}{\sqrt{(X^TX)^{-1}\hat\sigma^2}} \sim \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-p}}{n-p}}}$$

$$\sim t_{n-p}.$$

$$\mathbb{P}\left(t_{n-p}\left(\frac{\alpha}{2}\right) \le \frac{\hat\beta - \beta}{\sqrt{X^TX^{-1}\hat\sigma^2}} \le t_{n-p}\left(1-\frac{\alpha}{2}\right)\right) = 1-\alpha.$$

$$\Rightarrow \beta \in [\hat\beta \sim].$$

SM. 2021

P1, 5I, 5J.

$$f(x,\mu) = \frac{1}{\sqrt{2\pi x^3}} \exp\left[-\frac{(x-\mu)^2}{2\mu^2 x}\right].$$

exp-family: $f(y;\theta)$

$$= e^{\theta T(y) - K(\theta)} f_0(y).$$

$\cancel{f(x;\theta}$ $f(x;\theta)$

$$= e^{\theta T(x) - K(\theta)} f_0(x).$$

$$\frac{1}{\sqrt{2\pi x^3}} e^{-\frac{x^2-2\mu x+\mu^2}{2\mu^2 x}}$$

$$= \frac{1}{\sqrt{2\pi x^3}} e^{-\frac{1}{2x}} e^{-\frac{x}{2\mu^2}} e^{\frac{1}{\mu}}.$$

So $T(x) = x$. $\theta = -\frac{1}{2\mu^2}$.

$$K(\theta) = \frac{-1}{\mu} = \cancel{-2\theta}\ -\sqrt{-2\theta}.$$

$$f_0(x) = \frac{1}{\sqrt{2\pi x^3}} e^{-\frac{1}{2x}}. \qquad \cancel{(-2\theta)}^{\frac{1}{2}}$$

$$\cancel{\mu = K'(\theta) = (\tfrac{1}{2})(-2)(-2\theta)^{-\frac{1}{2}} = -\frac{1}{\sqrt{-2\theta}} =}$$

$$\mu = K'(\theta) = -(-2)\left(\tfrac{1}{2}\right)(-2\theta)^{-\frac{1}{2}} = \frac{1}{\sqrt{-2\theta}} = \mu.$$

$$V = K''(\theta) = \frac{d}{d\theta}\left((-2\theta)^{-\frac{1}{2}}\right) = \left(-\tfrac{1}{2}\right)(-2)(-2\theta)^{-\frac{3}{2}} = \mu^3.$$

SM 2021.

P3, SI.

$$\hat{\beta} = (X^TX)^{-1}X^TY.$$

$$f(\beta; \sigma^2) = \cancel{e^{Y-X^T\beta}} \prod \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_i - X_i^T\beta)^2}{2\sigma^2}}$$

$$\frac{d\|Y-X^T\beta\|^2}{d\beta} = (Y-X^T\beta)^T(Y-X^T\beta).$$

$$= 2X^TY - 2(X^TX)\beta$$

$$\Rightarrow \hat{\beta} = \sim$$

$$\hat{\sigma}^2 = \frac{\|Y - X^T\hat{\beta}\|^2}{n}.$$

$$\|Y - X^T\beta\|^2$$

$$= \|[I - X(X^TX)^{-1}X^T]Y\|^2$$

$$= \|[I - X(X^TX)^{-1}X^T]\varepsilon\|^2.$$

$$\sim \chi^2_{n-p} \cdot \sigma^2.$$

$$\text{so } \hat{\sigma}^2 \sim \frac{\chi^2_{n-p}}{n}\sigma^2$$

$$\frac{\hat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2_{n-p}}{n}.$$

$$\mathbb{P}\left(\frac{\chi^2_{n-p}(\frac{\alpha}{2})}{n} \le \frac{\hat{\sigma}^2}{\sigma^2} \le \sim\right) = 1-\alpha \sim.$$

(b) $\cancel{\quad}\ \mathbb{E}(\|Y-\hat{\mu}_\lambda\|^2)$

$=\mathbb{E}(\|(I-H_\lambda)Y\|^2)$

$=\mathbb{E}\|(I-H_\lambda)(\mu+\varepsilon)\|^2$

$=\cancel{\mathbb{E}}\ \|(I-H_\lambda)\mu\|^2+\mathbb{E}[(I-H_\lambda)^T(I-H_\lambda)\varepsilon\|^2$

$+\ \mathbb{E}[\cdots\cdots\varepsilon]$ $\uparrow$ $=0.$

$=\|(I-H_\lambda)\mu\|^2+\{n-2\mathrm{tr}(H_\lambda)+\mathrm{tr}(H_\lambda^2)\}\sigma^2.$

(c) ✓

(d) ...

Stats Modelling Revision

Xuanang Shawn Chen

To: Bethany Heath

crsid: bah40

SM 2020.

P3, 5I, 5J. either define notation or used notation that has been used previously, i.e. $\ell(\theta)$

(a) $\tilde{F} = \prod_{i=1}^{n} \log f_{x_i}(Y_i, \theta)$

$= \sum_{i=1}^{n} \log f_{x_i}(Y_i;\theta)$. ✓

$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ F(Y,\theta)$. ✓

$AIC = -2\sum_{i=1}^{n} \log f(Y_i;\hat{\theta}) + 2\dim(\Theta)$

$= -2F(Y,\hat{\theta}) + 2\dim(\Theta)$. ✓

(b) ~~AIC =~~ $\dim(\Theta) = \dim(\beta, \sigma^2) = p+1$. ✓✓

$\hat{\sigma}^2 = \frac{\|Y - X\hat{\beta}\|^2}{n}$ ✓

$L = \left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}\right)^{\frac{n}{2}} e^{-\frac{\|Y-X\hat{\beta}\|^2}{2\hat{\sigma}^2}}$

$= e^{-\frac{n}{2}} \left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}\right)^{-\frac{n}{2}}$.

$F = -\frac{n}{2} - \frac{n}{2}\log(2\pi\hat{\sigma}^2)$. ✓

$AIC = -2\left(-\frac{n}{2} - \frac{n}{2}\log(2\pi\hat{\sigma}^2)\right) + 2(p+1)$

$= n(1 + \log(2\pi\hat{\sigma}^2)) + 2(p+1)$. ✓✓

(c) $\frac{\hat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2_{n-p}}{n}$ ✓✓ So $\log(2\pi\hat{\sigma}^2)$

$\sim \log\left(2\pi\frac{\sigma^2\chi^2_{n-p}}{n}\right) = \log(\chi^2_{n-p}) + \log\left(\frac{2\pi\sigma^2}{n}\right)$.

Thus $AIC \sim n\left(1 + \log(\chi^2_{n-p}) + \log\left(\frac{2\pi\sigma^2}{n}\right)\right) + 2(p+1)$

$= n\log(\chi^2_{n-p}) + n\left(\log\left(\frac{2\pi\sigma^2}{n}\right) + 1\right) + 2(p+1)$. ✓

SM 2020

P1, SII, 13J.

(i) Kilometres is not categorical variable.
its values has numerical meaning which we can regress on. ↳ and it naturally has ordering

(ii) Idk what the plot is.
But maybe lm(log(Pay) ~ as.numeric(Kilo) + Brands + Bonus) ?
↳ is box-Cox will go though.

(iii) 95% is 2 standard deviation.

$$\cancel{\text{so ratio}} \sim [e^{0.057-2\times0.032}, e^{0.57+2\times0.032}]$$

independent by ~~deparate~~ proof in class [

$$\frac{\hat{\beta}-\beta}{\hat{\sigma}} \sim \frac{N(0,(X^TX)^{-1})}{\sqrt{\frac{\chi^2_{n-p}}{n-p}}}$$

nice

$\hat{\sigma} = 0.032$

$\hat{\beta} = 0.057$.

Ratio is $\exp\left[\hat{\beta} \pm \hat{\sigma}\sqrt{(X^TX)^{-1}_{Kilo}}\ t_{284-11}(0.025)\right]$ will go though.

(iv) idk. logistic regression? $g = \log\frac{\mu}{1-\mu}$.

SM. 2020.

P2, 5I, 5J.

(a) Binomial regression

what link function is being used?

Hasn't been changed

$\therefore$ is canonical link

$Y_i \sim \frac{1}{n_i} Bin(n_i, \mu_i)$. ✓ logit $p_i = \mu + \alpha \, height_i + \beta \, cig_i$

$$f(y; n, \mu) = \binom{n}{ny} \mu^{ny} (1-\mu)^{n(1-y)}$$

$$= \exp\left\{ \frac{1}{n^{-1}} \left(y \log \frac{\mu}{1-\mu} + \log(1-\mu) \right) \right\} \binom{n}{ny}$$

$\sigma^2 = 1$, $w = \frac{1}{n}$, $\theta = \log\left\{\frac{\mu}{1-\mu}\right)$. $K(\theta) = \log(1+e^{\theta})$.

~~(b)~~ $\cancel{\eta = x\beta = 2 \cdot 20 \cdot 0.02313 = 0.92}$.

$\cancel{\mu = \frac{e^{\theta}}{1+e^{\theta}} = \frac{e^{0.92}}{1+e^{0.92}}}$ ~~chance of having HD, given no height info.~~

(b) η-change $= x'\beta_{cigs} = 2 \cdot 20 \cdot 0.02313 = 0.92$

$$\mu = \frac{e^{0.92+\eta_0}}{1+e^{0.92+\eta_0}}$$ ✓ chance of having HD,

η_0 is ~~other~~ regressor on height & intercept.

(c) Let $Y^* \sim Poisson(e^{\eta})$. $Y = \mathbb{1}\{Y^* > 0\}$

Then ~~$E\mu =$~~ $\mu = \mathbb{E}(Y) = \mathbb{P}(Y^* > 0) = \mathbb{P}(\varepsilon > \eta)$

$= 1 - F(-\eta)$ ✓

$1 - \mu = \mathbb{P}(Y=0) = \mathbb{P}(Y^*=0) = e^{-e^{\eta}}$.

$\hookrightarrow$ want you to use the hint & the fact F is symmetric to say

link is $\eta = \log(-\log(1-\mu))$.

$F(x) = 1 - F(-x)$

SM 2020. P4. SI. 5J.

P4. SI. 5J.

(a) (i) If set 5% as significance value,

then covar1, covar2 ~~can't~~ aren't significantly explain response

as p-value > 0.05

good

(ii) Both variables together explain response

as F-test p value < 0.05

value significant

(iii) Both not explain response.

combined with (ii) to suggest not one variable does not explain the response

(b)

48   1

47   1

$$\frac{\|(P_0 - P_I)Y\|^2}{\|(I-P)Y\|^2/48}$$

$$\frac{\|(P_1 - P_0)Y\|^2}{\|(I-P)Y\|^2/48}.$$

$\mathbb{P}(Z_2 \geq F)$.

$$= \frac{133.68}{30.817/48}$$

$$= \frac{0.014}{30.817/48}.$$

$$\mathbb{P}\left(\frac{0.014}{30.817/48} \geq F_{1,48}\right)$$

0.886 from above

SM 2020.

P1, SI, SJ.

(a) Score function $U(\beta) = \nabla l(\beta)$

$U_i(\beta) = \frac{\partial}{\partial \beta_i} l(\beta)$ ✓

$i(\beta) \doteq Var_\beta(u(\beta))$

Fisher info $I^n(\beta) = \cancel{\nabla^2 l(\beta)}\ -\mathbb{E}[H]$ ✓

$= -\mathbb{E}[\nabla^2 l(\beta)]$

[where is the first bit of the question? - what is the iterative Fisher Scoring algorithm]

(b) Let $X' = W^{\frac{1}{2}} X$, $Y' = W^{\frac{1}{2}} Y$ ✓ nice

~~since~~ ($W^{\frac{1}{2}}$ defined as W is positive entries)

$\underset{b \in \mathbb{R}^p}{\operatorname{argmin}} \left(\sum_{i=1}^n W_{ii} (Y_i - X_i^T b)^2 \right)$ ✓

$= \underset{b \in \mathbb{R}^p}{\operatorname{argmin}} \left(\sum_{i=1}^n (W_{ii}^{\frac{1}{2}} Y_i - W_{ii}^{\frac{1}{2}} X_i^T b)^2 \right)$ ✓

$= \underset{b \in \mathbb{R}^p}{\operatorname{argmin}} \| Y' - X' b \|^2$

$= (X'^T X')^{-1} X'^T Y'$

$= (X^T W X)^{-1} X^T W Y$. ✓

(c) Newton-Raphson: $\beta^{(t)} = \beta^{(t-1)} - \{H(\beta^{(t-1)})\}^{-1} U(\beta^{(t-1)})$

Fisher scoring: $\beta^{(t)} = \beta^{(t-1)} + I(\beta^{(t-1)})^{-1} U(\beta^{(t-1)})$ ✓

$\beta_j^{(t)} = \beta_j^{(t-1)} + \left(\sum_{i=1}^n \frac{X_{ij} X_{ik}}{a_i \sigma^2 V(\mu_i) \{g'(\mu_i)\}^2} \right)^{-1} \left(\sum_{i=1}^n \frac{(Y_i - \mu_i) X_{ik}}{a_i \sigma^2 V(\mu_i) g'(\mu_i)} \right)$

$= \beta^{(t-1)} + (X^T W X)^{-1} X^T W R \big|_j$

where $W = \operatorname{diag}\left(\frac{1}{a_i \sigma^2 V(\mu_i) \{g'(\mu_i)\}^2}, \dots \right)$, $R_i = (Y_i - \mu_i) g'(\mu_i)$

$U(\beta^{(t-1)}) = \frac{1}{\sigma^2} X^T W R$.

Note that

$$\beta^{(t-1)} = (X^TWX)^{-1}X^TWX\beta^{(t-1)}$$
$$= (X^TWX)^{-1}X^TW\eta^{(t-1)}.$$

So $$\beta^{(t)} = (X^TWX)^{-1}X^TW(\underbrace{\eta^{(t-1)}+R}_{Y})$$

$$= \underset{b\in\mathbb{R}^p}{\arg\min}\left\{\sum_{i=1}^{n} W_{ii}(Y_i - x_i^Tb)^2\right\}.$$

Can be interpreted a least squares operation with weights matrix $W^{(r-1)}$, which gets updated in every iteration because $W^{(r-1)}$ depends on $\hat{\beta}^{(r-1)}$.

SM 299.

P4, 5I, 5J.

$$Y = X\beta + \varepsilon. \qquad \hat\beta - \beta \sim N(0, \sigma^2 (X^TX)^{-1})$$

$\cancel{Y^* = x^* \hat\beta + \varepsilon}$

$$Y^* = (x^*)^T\beta + \varepsilon.$$

$$\hat{Y}^* = (x^*)^T \hat\beta.$$

$$Y^* - \hat{Y}^* = (x^*)^T(\beta - \hat\beta) + \varepsilon$$

$$\sim N(0, (x^*)^T(X^TX)^{-1}(x^*)^T\sigma^2 + \sigma^2 I)$$

Note: $\cancel{\frac{\hat\sigma^2(n-p)}{\sigma^2}}$ $\frac{\hat\sigma^2}{\sigma^2} \sim \frac{\chi^2_{n-p}}{n-p}$

$$\frac{Y^* - \hat{Y}^*}{\sqrt{\hat\sigma^2} \cdot \sqrt{(x^*)^T(X^TX)^{-1}(x^*)^T + I}} \sim \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-p}}{n-p}}} = t_{n-p}.$$

Thus $(1-\alpha)$ CI for Y^* is

$\cancel{[x^*(X^TX)^{-1}(x^*)^T}$

$$\left[(x^*)^T\hat\beta \pm \sqrt{\hat\sigma^2 (x^*)(X^TX)^{-1}(x^*)^T + I} \; t_{n-p}\left(\frac{\alpha}{2}\right) \right]$$

where $\hat\beta = (X^TX)^{T}X^T$. $\hat\sigma^2 = \frac{\|Y - X\hat\beta\|^2}{n}$.

SM 2019

P1, SI, 5J.

Exponential dispersion family:

A collection of density functions of the form

$$f(y;\theta,\sigma^2) = e^{\frac{\theta y - K(\theta)}{\sigma^2}} f_0(y;\sigma^2).$$

$$f(y;\alpha,\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} \quad \text{for } y>0.$$

$$= e^{-\lambda y + \alpha\log\lambda} \cdot y^{\alpha-1} \frac{1}{\Gamma(\alpha)}.$$

~~Take~~

Guess: $\theta = -\frac{\lambda}{\alpha}$, $\sigma^2 = \frac{1}{\alpha}$. $\left[\alpha = \frac{1}{\sigma^2}, \lambda = -\frac{\theta}{\sigma^2}.\right]$

$$f(y;\alpha,\lambda) = e^{\frac{-\frac{\lambda}{\alpha}y - [-\log\frac{\lambda}{\alpha}]}{\frac{1}{\alpha}}} \cdot \alpha^\alpha y^{\alpha-1} \cdot \frac{1}{\Gamma(\alpha)}$$

$$= e^{\frac{\theta y - (-\log\theta)}{\sigma^2}} \cdot e^{\frac{1}{\sigma^2}\log(\frac{1}{\sigma^2})} y^{\frac{1}{\sigma^2}-1} \cdot \frac{1}{\Gamma(\frac{1}{\sigma^2})}.$$

$$K(\theta) = -\log(-\theta)$$

$$\mu = K'(\theta) = -\frac{1}{\theta} = \frac{\alpha}{\lambda}$$

$$V = \sigma^2 K''(\theta) = \sigma^2 \theta^{-2} = \frac{1}{\alpha} \cdot \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}.$$

cumulant generating function given by

$$\frac{1}{\sigma^2}\left(k(\theta + t\sigma^2) - k(\theta)\right)$$

SM 2019. P2. SI.

5J.

(a) model 1: Participants with strong/more histories have more episodes to work.

Model 2: Strong ~~history~~ more episodes

Mode 3: Strong history tend to be more distant miles to work

↳ It wants you to ~~give~~ talk about other correlates fixing history on the same.

(b) Those being advised/incentiled and more distant to work will cycle more often than those not distant and advised.

~~Not~~ Not advisable, as more distant more episodes.

Also $\mathbb{E}[\widehat{\text{episodes}} \mid \text{Incentive yes}] > \mathbb{E}[\text{episodes}]$.

SM 2019. P3.5I.

5J. (a) $l(\beta) = \log L(\beta)$.

Score function ~~$U(\beta) = \frac{d}{d\beta}\log L(\beta)$~~.

$$U(\beta) = \nabla l(\beta).$$

H = Hessian = $\nabla^2 l(\beta)$

$$H_{ij} = \frac{\partial^2}{\partial\beta_i\partial\beta_j} l(\beta).$$

Fisher information matrix = $I^n(\beta) = -\mathbb{E}[H(\beta)]$ ✓

(b) $\underbrace{f(Y;\theta,\sigma^2)}$ $= \cancel{\log} \prod_{i=1}^{n} e^{\frac{\theta_i y_i - K(\theta_i)}{\sigma^2}} f_0(y_i,\sigma^2)$

$$= \sum_{i=1}^{n} \frac{\theta_i y_i - K(\theta_i)}{\sigma^2} + \sum_{i=0}^{n} f_0(y_i,\sigma^2)$$

be v. careful with notation – it is asking for likelihood.

asked to give log likelihood of a glm – state standard form

Where $\theta_i = \theta(\mu_i) = \theta(g^{-1}(x_i^T\beta))$.

(c) For canonical link, $g(\mu) = \theta(\mu)$. $= \eta$ ~~$g(\mu_i)$~~

$g'(\mu_i) = \theta'(\mu_i) = \frac{1}{V(\theta(\mu_i))} = \frac{1}{V(\mu_i)}$. by inverse function thm.

$$\frac{\partial l}{\partial\beta_r} = \frac{\partial}{\partial\beta_r} \sum_{i=1}^{n} \frac{\theta(g^{-1}(x_i^T\beta))y_i - K(\theta(g^{-1}(x_i^T\beta)))}{\sigma^2}$$

$$= \sum_{i=1}^{n} \left[\theta'(\mu_i) x_{ir} y_i \frac{1}{g'(\mu_i)} - K'(\theta(\mu_i))\frac{1}{g'(\mu_i)} x_{ir}\theta'(\mu_i)\right]\sigma^{-2}$$

$$= \sum_{i=1}^{n} \frac{(Y_i-\mu_i)x_{ir}}{\sigma^2 V(\theta(\mu_i)) g'(\mu_i)} = \sum_{i=1}^{n} \frac{(Y_i-\mu_i)x_{ir}}{\sigma^2}$$

Since:

$K'(\theta(\mu_i)) = \mu_i$

$\theta'(\mu_i) = \frac{1}{V(\theta)}$.

$V(\theta) = K''(\theta) = \mu'(\theta)$.

(c) continue.

Note: shorter version:

for canonical link, $\theta(\mu)=g(\mu)=\eta=X^T\beta$

$$g'(\mu_i)=\theta'(\mu_i)=\frac{1}{V(\theta(\mu_i))}=\frac{1}{V(\mu_i)}.$$

$$K'(\theta(\mu_i))=\mu_i.$$

$$\cancel{\theta'(\mu_i)}\ V(\theta)=K''(\theta)=\mu'(\theta).$$

can do it simple by keeping it in form, i.e.

by doing transpose $k'(x_i^T\beta)=\theta^{-1}(x_i^T\beta)$

+ leaving it in this form.

$$\frac{\partial L}{\partial \beta_r}=\frac{\partial}{\partial \beta_r}\sum_{i=1}^{n}\frac{X_i^T\beta y_i-K(X_i^T\beta)}{\sigma^2}$$

$$=\sum_{i=1}^{n}\frac{(Y_i-\mu_i)X_{ir}}{\sigma^2}.$$

$$\frac{\partial L}{\partial \beta_r\partial \beta_k}=\sum_{i=1}^{n}\frac{X_{ir}}{\sigma^2}\cdot\left(-\frac{X_{ik}}{g'(\mu_i)}\right)$$

$$=-\sum_{i=1}^{n}\frac{X_{ir}X_{ik}}{\sigma^2 g'(\mu_i)}$$

$$=-\sum_{i=1}^{n}\frac{X_{ir}X_{ik}}{\sigma^2}V(\mu_i).$$

since $\mu_i=g^{-1}(X_i^T\beta)$.

$$\frac{\partial \mu_i}{\partial \beta_k}=\frac{X_{ik}}{g'(X_i^T\beta)}=\frac{X_{ik}}{g'(\mu_i)}$$

Thus $I^n(\beta)=-\mathbb{E}(H)$

$$=\sigma^{-2}X^TWX,$$

$$\cancel{W=V(\mu_i)=\frac{1}{g'(\mu_i)}}$$

$$W=\operatorname{diag}(V(\mu_1), V(\mu_2),\dots)$$

$$=\operatorname{diag}\left(\frac{1}{g'(\mu_1)},\frac{1}{g'(\mu_2)},\dots\right).$$

SM 2019.

P1, SV, 13J.

State what you are fitting, i.e. $y_{ij} \overset{ind}{\sim} Po(\mu_{ij})$
canonical link $\log \mu_{ij} = \beta_i + \gamma_j$
+ identifiability of corner point constraint $\beta_1 = 0$

(a). $f(y_i; \mu_i, \sigma^2) = e^{\frac{1}{\sigma^2}(y_i \log \mu_i - \mu_i)} f_0(y_i; \sigma^2)$.

$\theta_i = \log(\mu_i), K(\theta) = e^{\theta}$. $g(\mu) = \theta(\mu) = \log(\mu)$

(ii) $\mu_i = e^{x_i^T \beta} = \prod_{j=1}^{p} (e^{\beta_j})^{x_{ij}}$.

(b) $L(\text{multinomial})$
$= \prod_{i=1}^{I} \frac{n_i!}{y_{1i}! \cdots y_{Ji}!} p_{1i}^{y_{1i}} \cdots p_{Ji}^{y_{Ji}}$

$\left(\sum y_{1i} + \cdots + y_{Ji} = n_i \right.$.

$l(\text{multi})$
$= \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij} \log p_{ij} + \text{const}$.

$p_{ij} = \frac{e^{x_{ij}^T \beta}}{\sum_{j=1}^{J} e^{x_{ij}^T \beta}}$

$= \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij} x_{ij}^T \beta - \sum_{i=1}^{I} n_i \log \sum_{j=1}^{J} e^{x_{ij}^T \beta}$.

$L(\text{Poisson}) = \prod_{i=1}^{I} \prod_{j=1}^{J} \frac{\mu_{ij}^{y_{ij}} e^{-\mu_{ij}}}{y_{ij}}$

$\mu_{ij} = \exp(\alpha_i + x_{ij}^T \beta)$

$l(\text{Poisson}) = \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij} \log \mu_{ij} - \mu_{ij} + \text{const}$.

$= \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij} (\alpha_i + x_{ij}^T \beta) - \exp(\alpha_i + x_{ij}^T \beta) + \text{const}$.

$= \sum_{i=1}^{I} \left(n_i \alpha_i + \sum_{j=1}^{J} y_{ij} x_{ij}^T \beta - e^{\alpha_i} \sum_{j=1}^{J} e^{x_{ij}^T \beta} \right)$

$= \sum_{i=1}^{I} \left(\sum_{j=1}^{J} y_{ij} x_{ij}^T \beta - n_i \log \sum_{j=1}^{J} e^{x_{ij}^T \beta} \right)$
$+ \sum_{i=1}^{I} n_i \delta_i - e^{\delta_i}$

Let $e^{\alpha_i} = \frac{e^{\delta_i}}{\sum_{j=1}^{J} e^{x_{ij}^T \beta}}$,
$\alpha_i = \delta_i - \log \sum_{j=1}^{J} e^{x_{ij}^T \beta}$

$= l(\text{multi}; \beta) + l(\delta)$.

Thus $\hat{\beta}_{mle}$ is the same.

Mult: $\hat{\mathbb{E}}[Y_{ij}] = n_i \hat{p}_{ij} = n_i \dfrac{e^{x_{ij}^T\hat{\beta}}}{\sum_{j=1}^{J} e^{x_{ij}^T\hat{\beta}}}$ //

$e^{\hat{\delta}_i + \hat{\gamma}_j}$

Poisson: $\hat{\mathbb{E}}[Y_{ij}] = \hat{\mu}_{ij} = e^{\hat{\alpha}_i} e^{x_{ij}^T\hat{\beta}}$ /

$$= e^{\hat{\delta}_i} \frac{e^{x_{ij}^T\hat{\beta}}}{\sum_{j=1}^{J} e^{x_{ij}^T\hat{\beta}}}$$ //

$$\frac{\partial l(\text{Poisson})}{\partial \delta_i} = n_i - e^{\delta_i}. \Rightarrow \hat{\mathbb{E}}[Y_{ij}]_{multi} = \hat{\mathbb{E}}[Y_{ij}]_{Poisson}$$

(c). ~~Null~~ $H_0: \pi_{jk} = \pi_j^A \pi_k^B$

$(\Leftrightarrow H_0: \log \mu_{jk} = \alpha + \beta_j^A + \beta_k^B)$

Test $H_1: \log \mu_{jk} = \alpha + \beta_j^A + \beta_k^B + \beta_{jk}^{AB}$.

Test statistic: $D_+(Y, \hat{\mu}) - D_+(Y, \hat{\mu}_0)$.

$\sim \chi^2_{p-p_0}$ /

(~~strong~~) assuming small dispersion asymptotics hold [i.e. expectation of each entry is large]

Which values are you using?

~~257-4=253.~~

$\chi^2_{9.4}(0.01) < 257 - 4.$ //

$\Rightarrow$ reject H_0. accept H_1.

(d) ~~T-t~~ Student-t test on β_{price},

$\hookrightarrow$ Deviance statistic compares any deviation from homogeneity assumption of first hypothesis. If better i.e.-cheaper, more expensive, can fit a more extensive alternate model

More powerful test as ~~Fitted value~~ ~~the~~ H_1 is not that large degrees of freedom

SM 2019.

P4, SII, 13J.

(a)

$$f = \prod_{i,j} \left(\frac{e^{\theta_{i,j}}}{1+e^{\theta_{i,j}}}\right)^{y_{i,j}} \left(\frac{1}{1+e^{\theta_{i,j}}}\right)^{1-y_{i,j}}$$

remember by definition $y_{i,j}$ can only take 2 values → 0 or 1.

(b) Nested: $\theta = X\beta$

model(ii) < model(iii) < model (i).

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_C \end{pmatrix} \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_C \end{pmatrix} \qquad \beta = \begin{pmatrix} \beta_{1,1} \\ \beta_{1,2} \\ \vdots \\ \beta_{1,C} \\ \beta_{2,1} \\ \vdots \\ \beta_{C,C} \end{pmatrix}$$

X $m^2 \times C$.

Each row, (i,j)th row, ith column, jth column is 1

Can view $\beta_{i,j}$ as β_i.

Compare Model (i) to (ii): By Wilk's Thm

$$D_+(Y, \hat{\mu}) - D_+(Y, \hat{\mu}_0) \xrightarrow{d} \chi^2_{p-p_0} = \chi^2_{C^2-(C+1)}$$

$$= 2\{\sup_{H_0 \cup H_1} l(\beta; Y) - \sup_{H_0} l(\beta; Y)\}.$$

$\frac{(C^2+2C)}{2}.$

(c)

~~$M_i = \sum_{j\in\{1,2,\dots,m\}} \mathbb{1}\{y_{i,j}=1\}$~~

$$M_i = \sum_{t\in\{1,2,\dots,m\}} \mathbb{1}\{y_{i,j}=1\}\mathbb{1}\{z_j \neq k\}$$

Idk what's this Q asking → will go through in repo

(d)

$$\frac{f(\beta;y)}{f(\beta,y')} = \prod_{i,j}\left(\frac{e^{\beta_{z_i}+\beta_{z_j}+\beta_0\delta_{z_i,z_j}}}{1+e^{\beta_{z_i}+\beta_{z_j}+\beta_0\delta_{z_i,z_j}}}\right)^{y_{i,j}-y'_{i,j}}\left(\frac{1}{1+e^{\beta_{z_i}+\beta_{z_j}+\beta_0\delta_{z_i,z_j}}}\right)^{1-y_{i,j}-(1-y'_{i,j})}$$

$$= \prod_{i,j}\left(e^{\beta_{z_i}+\beta_{z_j}+\beta_0\delta_{z_i,z_j}}\right)^{y_{i,j}-y'_{i,j}}$$

$$\log\frac{f(\beta;y)}{f(\beta,y')} = \sum_{i,j}(y_{i,j}-y'_{i,j})(\beta_{z_i}+\beta_{z_j}+\beta_0\delta_{z_i,z_j})$$

$$= 2\sum_{i=1}^{c}\beta_i\left(\sum_{j:z_j=i}\sum_{k=1}^{m}y_{j,k}\right) - 2\sum_{i=1}^{c}\beta_i\left(\sum_{j:z_j=i}\sum_{k=1}^{m}y'_{j,k}\right) + \beta_0\left(\sum_{i\neq j}y_{i,j}-\sum_{i\neq j}y'_{i,j}\right).$$

$$\therefore \text{Minimal } T = t(y) = \begin{pmatrix} \sum_{j:z_j=1}\sum_{k=1}^{m}y_{j,k} \\ \sum_{j:z_j=2}\sum_{k=1}^{m}y_{j,k} \\ \vdots \\ \sum_{j:z_j=c}\sum_{k=1}^{m}y_{j,k} \\ \sum_{i\neq j}y_{i,j} \end{pmatrix}$$

Stats Modelling 2018.

P2, SI.

5J. $$F = \frac{\|(P-P_0)Y\|^2 / p-p_0}{\|(I-P)Y\|^2 / n-p} \sim F_{p-p_0, n-p} \checkmark$$

where $P = X(X^TX)^{-1}X^T$, $P_0 = X_0(X_0^TX_0)^{-1}X_0^T$,

$X = \begin{pmatrix} X_0 \\ X_1 \end{pmatrix}^T$, X_0 is first p_0 columns of X.

~~Then~~ ~~Note:~~ $$\cancel{\frac{\sup_{\beta,\sigma} L(\beta,\sigma^2)}{\sup_{\beta,\sigma,\ \beta_{p_0+1}=\cdots=\beta_p=0} L(\beta,\sigma^2)} = \frac{\exp\{-\frac{n}{2}\log\hat{\sigma}^2 - \frac{n}{2} + \text{const}\}}{\exp\{-\frac{n}{2}\log\tilde{\sigma}^2 - \frac{n}{2} + \text{const}\}} =}$$

Generalized likelihood ratio test is

$$2\log\Lambda = 2\{\max_{\beta\in\mathbb{R}^p, \sigma^2>0} l(\beta,\sigma^2) - \max_{\substack{\beta\in\mathbb{R}^p, \sigma^2>0 \\ \beta_{p_0+1}=\cdots=\beta_p=0}} l(\beta,\sigma^2)\}$$

$$= 2\left[-\frac{n}{2}\log\hat{\sigma}^2 - \frac{n}{2}\right] - 2\left[-\frac{n}{2}\log\tilde{\sigma}^2 - \frac{n}{2}\right]$$

$$= n\left\{\log\frac{\|(I-P_0)Y\|^2}{\|(I-P)Y\|^2}\right\}. \checkmark$$

Note that $(I-P_0)(P-P_0)=0$, $P-P_0$ is orthogonal matrix,

$$\|(I-P_0)Y\|^2 = \|(I-P)Y\|^2 + \|(P-P_0)Y\|^2.$$

So $2\log\Lambda$ is monotone in $\frac{\|(I-P_0)Y\|^2}{\|(I-P)Y\|^2}$

$$= 1 + \frac{\|(P-P_0)Y\|^2}{\|(I-P)Y\|^2}.$$

$$= 1 + F\text{-stats}.$$

nice

$Y = X\beta + \sigma^2 \varepsilon$. $\varepsilon \sim N(0, I)$. $X: n \times p$.

$400/4 = 100$.

$60 + 45$

$2\alpha + 3\beta$

$\Downarrow$

$8\alpha + 12\beta$.

F-stats:
$$\frac{\frac{\|(P-P_0)Y\|^2}{p-p_0}}{\frac{\|(I-P)Y\|^2}{n-p}} \sim F_{p-p_0, n-p}.$$

$X = (X_0 \ X_1)$

$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$.

σ^2

$e^{-\frac{\|Y-}{2\sigma^2}}$

$$Y - (X^TX)^{-1}X^TY$$
$$= (I-P)Y.$$

$\hat{\sigma}^2 = \frac{\|Y - X_i^T\beta\|^2}{n}$.

$-\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\|Y - X_i^T\beta)^2 + \text{const}$.

$-\frac{n}{2} - \frac{n}{2}\log(\hat{\sigma}^2)$

$-\frac{n}{2}\log\left(\frac{\|(I-P)Y\|^2}{n}\right) + \text{const}$.

SM 2018. P4, SII. 13J.

(a) No. As $\sum_{\text{teams}} \mathbb{I}\{\text{win}\} - \mathbb{I}\{\text{lose}\} = 0$.

So intercept should be 0. //

as $\mathbb{P}(\text{win}|X)$ [?].

(b) No.

By looking at expression,

$\beta_1 + \beta_2 + \beta_{\{1,2\}}$ play as a whole term,

as they always come together.

So only $\widehat{\beta_1 + \beta_2 + \beta_{\{1,2\}}}$ mle matters //

(c) By delta method,

$$\sqrt{n}\{g(\hat{\eta}) - g(\eta)\} \xrightarrow{d} N(0, \tau^2 g'(\eta)^2).$$

$$\tilde{\theta} \sim N\left(\theta, \frac{1}{i^{(n)}(\theta)}\right) \ //$$

Looking for g which is a transformation.

SM 2018. P1. S$\underline{\text{V}}$.

(a) $F_i(t) = \int_0^t f_i(s)\,ds$

$= \int_0^t h_i(s)(1-F_i(s))\,ds.$

$F_i(t)' = h_i(t)(1-F_i(t))$

$\Rightarrow \quad \log(1-F_i(t))' = -h_i(t). \quad \checkmark$

$\cancel{F_i(t)=1} \Rightarrow F_i(t) = 1 - \exp\left(-\int_0^t h_i(s)\,ds\right).$

as $\cancel{F_i(t)}$ $F_i(0)=0$,

(b). $h_i(t) = \lambda(t)\, e^{(\beta^T x_i)}.$

$f_i(t) = 1 - \exp\left(-\int_0^t h_i(s)\,ds\right) \cdot (-h_i(t))$

$= 1 + h_i(t) \exp\left(-\int_0^t h_i(s)\,ds\right)$

$= 1 + \lambda(t)\exp(\beta^T x_i)\, \exp\left(-\int_0^t \lambda(s)\exp(\beta^T x_i)\,ds\right)$

$f_i(Y_i, \beta) = 1 + \lambda(Y_i)\exp(\beta^T x_i)(1-Y_i)\, \exp\left(-\int_0^{Y_i} \lambda(s)\,ds\right)$

(c)

Start by writing out what the log likelihood is defined to be + work through from there.

SM 2018.

P4. S2, 5J.

Rate of weight loss with time: → Time effect in Control not significantly different to 0.

Time - estimate = -0.02602 3.

Yes. significant weight loss with time in control group:

Time : Grouptreatment −0.173515, p-value <2e-16.

I may not trust, as if model is good, residual against time should be normal distributed ~~across~~ with mean 0. ✓

But on plot, residual ~~has~~ seems not to be i.i.d. normal, and have more positive residual for mouse 1, 2 ; negative for 3.

Also for mouse 1, 2, ~~3~~ residual ↓ as time ↑.

~~So we~~ for mouse 3, residual increases as time increase.

So there are hidden factors not captured.