Theory Exam Section II 2013

1). $y_{ij} \mid v_1, \ldots, v_m \sim Poi(\lambda_{ij}$

$i = 1, \ldots, m \; ; \; j = 1, 2$

$\log(\lambda_{ij}) = x_{ij}^T B + v_i$

(Poisson Mixed Effects Model)

x_{ij} $p \times 1$, $B_{p\times 1}$

$v_1, \ldots, v_m$ indep + identically distributed

$z_i = \exp(v_i)$

γ = coefficient of variation of z_i

$$= \frac{\sqrt{Var(z_i)}}{E(z_i)}$$

ⓐ Show that $Var(y_{ij} \mid x_{ij}) = \mu_{ij}(1 + \gamma^2 \mu_{ij})$

+ $Cov(y_{ij}, y_{ik} \mid x_{ij}, x_{ik}) = \gamma^2 \mu_{ij} \mu_{ik}$ for $j \neq k$

$\mu_{ij} = E[y_{ij} \mid x_{ij})$

(i) $Var(y_{ij} \mid x_{ij}) = E[Var(y_{ij} \mid x_{ij}, v_{ij})] + Var[E[y_{ij} \mid x_{ij}, v_{ij}]]$

$= E[\lambda_{ij}] + Var[\lambda_{ij}]$

$= E[\exp(x_{ij}^T B + v_i)] + Var[\exp(x_{ij}^T B + v_i)]$

$= \exp(x_{ij}^T B) \cdot \underbrace{E[\exp(v_i)]}_{E[z_i]} + \exp(x_{ij}^T B)^2 \, Var(\underbrace{\exp(v_i)}_{z_i})$

Notes

$\mu_{ij} = E[y_{ij} \mid x_{ij}] = E[E[y_{ij} \mid x_{ij}, v_i]]$

$= E[\lambda_{ij}] = E[\exp(x_{ij}^T B + v_i)]$

$= \exp(x_{ij}^T B) E[z_i] \iff E(z_i) = \mu_{ij} \exp(-x_{ij}^T B)$

$$\gamma = \frac{\sqrt{Var(z_i)}}{E(z_i)} = \frac{\sqrt{Var(\exp(v_i))}}{E[\exp(v_i)]}$$

$$\gamma^2 = \frac{Var(\exp(v_i))}{(E[\exp(v_i)])^2} = \frac{Var(\exp(v_i))}{(\mu_{ij}(\exp(x_{ij}^T B)^{-1}))^2} \longrightarrow$$

$$\left(\begin{aligned} y_{ij} \mid v_i, x_{ij} &\sim Poi(\lambda_{ij}) \\ &\equiv Poi(\exp(x_{ij}^T B + v_i)) \end{aligned} \right)$$

$$\Rightarrow \quad Var(\overbrace{\exp(v_i)}^{z_i}) = \gamma^2 \mu_{ij}^2 \exp(-2 x_{ij}^T B)$$

And

$$E[z_i] = \exp(-x_{ij}^T B)\, \mu_{ij}$$

$$\begin{aligned} & E[\lambda_{ij}] + Var[\lambda_{ij}] \\ &= E[\exp(x_{ij}^T B) z_i] + \exp(2x_{ij}^T B) Var(z_i) \\ &= \exp(x_{ij}^T B) E[z_i] + \exp(2x_{ij}^T B) Var(z_i) \\ &= \exp(\cancel{x_{ij}^T B}) \exp(\cancel{-x_{ij}^T B}) \mu_{ij} + \exp(\cancel{2x_{ij}^T B}) \exp(\cancel{-2x_{ij}^T B}) \cdot \gamma^2 \mu_{ij}^2 \\ &= \mu_{ij} + \gamma^2 \mu_{ij}^2 \\ &= \mu_{ij}(1 + \gamma^2 \mu_{ij}) \quad \checkmark \end{aligned}$$

(ii) $Cov(y_{ij}, y_{ik} \mid x_{ij}, x_{ik}) =$

$$= E[y_{ij} \cdot y_{ik} \mid x_{ij}, x_{ik}] - E[y_{ij} \mid x_{ij}]\, E[y_{ij} \mid x_{ik}]$$

<u>Note:</u> $Cov(X, Y) = E[Cov(X, Y \mid Z)] - Cov(E(Y \mid Z), E(X \mid Z))$

<u>Proof:</u>

$$Cov(X, Y \mid Z) = E[XY \mid Z] - E[Y \mid Z]\, E[X \mid Z]$$

$$\begin{aligned} \Rightarrow E[Cov(X, Y \mid Z)] &= E[E[XY \mid Z]] - E[E(Y \mid Z) E[X \mid Z]] \\ &= E[XY] - E[E(Y \mid Z) E(X \mid Z)] \end{aligned}$$

$$\begin{aligned} & Cov(E(Y \mid Z), E(X \mid Z)) \\ &= E[E(Y \mid Z) E(X \mid Z)] - E[E(Y \mid Z)]\, E(E(X \mid Z)) \\ &= E[E(Y \mid Z) E(X \mid Z)] - E(Y) E(X) \end{aligned}$$

$$\begin{aligned} \Rightarrow & E[Cov(X, Y \mid Z)] - Cov[E(Y \mid Z); E(X \mid Z)] \\ &= E(X, Y) - E(X) E(Y) \\ &= Cov(X, Y) \quad \checkmark \end{aligned}$$

$\longrightarrow$

$$\Rightarrow \text{Cov}(y_{ij}, y_{ik}) = E[\text{Cov}(y_{ij}, y_{ik} \mid v_i)] - \text{Cov}(E(y_{ij} \mid v_i), E(y_{ik} \mid v_i))$$

$$= E[\,—\,] - \text{Cov}(\lambda_{ij}, \lambda_{ik})$$

?

$$= 0 - \text{Cov}(\exp(x_{ij}^T B) z_i, \exp(x_{ik}^T \beta) z_i)$$

$$= \exp(x_{ij}^T B + x_{ik}^T B) \, \text{Var}(z_i)$$

$$\left[= \exp(x_{ij}^T B) \exp(x_{ik}^T B) \, \gamma^2 \underbrace{\mu_{ij}^2 \exp(-2x_{ij}^T B)}_{j \text{ or } k\, ?} \right]$$

$$= \exp(x_{ij}^T B + x_{ik}^T B) \, \text{Var}(z_i)$$

$$\frac{\sqrt{\text{Var}(z_i)}}{E(z_i)} = \gamma$$

$$\Rightarrow \text{Var}(z_i) = \gamma^2 E(z_i)^2$$

$$= \exp(x_{ij}^T B + x_{ik}^T B) \, \gamma^2 E(z_i)^2$$

$$= E(\exp(x_{ij}^T B) z_i) \, E(\exp(x_{ik}^T B) z_i) \, \gamma^2$$

$$= \mu_{ij} \mu_{ik} \gamma^2$$

(b) $z_i \sim \text{Gamma}(\alpha, 1/\alpha) \quad \alpha > 0$

(i) Calculate μ_{ij} & γ^2

$$\mu_{ij} = \exp(x_{ij}^T B)\, E[z_i]$$
$$= \exp(x_{ij}^T B)\left(\alpha\left(1/\alpha\right)\right)$$
$$= \exp(x_{ij}^T B)$$

$$\gamma^2 = \left(\frac{\sqrt{Var(z_i)}}{E[z_i]}\right)^2 \qquad E(z_i) = 1$$
$$= \left(\sqrt{\alpha(1/\alpha)^2}\right)^2$$
$$= \left(\sqrt{\alpha^{-1}}\right)^2$$
$$= \alpha^{-1}$$

(ii) Write down the likelihood for (B, α) & show that it can be expressed in closed form using the Gamma function.

$$f(y_{ij}, z_i) = h(y_{ij} | z_i)\, g(z_i)$$
$$= \frac{\lambda_{ij}^{y_{ij}} e^{-\lambda_{ij}}}{y_{ij}!} \cdot \frac{\alpha^\alpha}{\Gamma(\alpha)} z_i^{\alpha-1} \exp(-z_i \alpha)$$
$$= \frac{\exp(y_{ij} x_{ij}^T B)\, z_i^{y_{ij}} \exp(-\exp(x_{ij}^T B) z_i)}{y_{ij}!} \cdot \frac{\alpha^\alpha}{\Gamma(\alpha)} z_i^{\alpha-1} \exp(-z_i \alpha)$$
$$= \frac{\alpha^\alpha}{\Gamma(\alpha)} z_i^{y_{ij}+\alpha-1} \exp(-z_i(\exp(x_i^T B) + \alpha))\, \frac{\exp(y_{ij} x_{ij}^T B)}{y_{ij}!}$$

$$f(y_{ij}) = \int_{z_i} f(y_{ij}, z_i)\, dz_i$$

$\longrightarrow$

$$= \frac{\alpha^\alpha}{\Gamma(\alpha)} \left(\frac{\Gamma(\alpha+y_{ij})}{(\exp(x_{ij}^T B)+\alpha)^{\alpha+y_{ij}}} \right) \cdot \left(\frac{\exp(y_{ij}\, x_{ij}^T B)}{y_{ij}!} \right) \cdot$$

$$\cdot \underbrace{\int_0^\infty \frac{(\exp(x_{ij}^T B)+\alpha)^{y_{ij}+\alpha}}{\Gamma(y_{ij}+\alpha)} z_i^{y_{ij}+\alpha-1} \exp(-z_i(\exp(x_{ij}^T B)+\alpha))\, dz_i}_{=1 \;\; (\text{pdf of Gamma}(y_{ij}+\alpha,\, (\exp(x_{ij}^T B)+\alpha)^{-1}))}$$

$$= \frac{\Gamma(\alpha+y_{ij})}{\Gamma(\alpha)} \frac{\alpha^\alpha}{[\exp(x_{ij}^T B)+\alpha]^{\alpha+y_{ij}}} \left(\frac{\exp(y_{ij}\, x_{ij}^T B)}{y_{ij}!} \right)$$

$$f(y_{i1}, y_{i2}) = \int_{z_i} f(y_{i1}, y_{i2}, z_i)\, dz_i$$

$$= \int_{z_i} f(y_{i1}, y_{i2} \mid z_i)\, g(z_i)\, dz$$

$$f(y_{i1}, y_{i2} \mid z_i)\, g(z_i) = \underbrace{f(y_{i1} \mid z_i)\, f(y_{i2} \mid z_i)\, g(z_i)}_{\text{conditional independence. ?}}$$

Assume for now that y_{ij}, y_{ik} have conditional independence

$$f(y_{i1}, y_{i2} \mid z_i)\, g(z_i)$$

$$= \frac{\lambda_{i1}^{y_{i1}} e^{-\lambda_{i1}}}{y_{i1}!} \cdot \frac{\lambda_{i2}^{y_{i2}} e^{-\lambda_{i2}}}{y_{i2}!} \cdot \frac{\alpha^\alpha}{\Gamma(\alpha)} z_i^{\alpha-1} \exp(-z_i \alpha)$$

$$= \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\exp(y_{i1} x_{i1}^T B)\, z_i^{y_{i1}}}{(y_{i1}!\, y_{i2}!)} \left(\begin{array}{l} \exp(y_{i2} x_{i2}^T B)\, z_i^{y_{i2}}\, z_i^{\alpha-1} \cdot \\ \cdot \exp(-z_i(\exp(x_{i1}^T B)+\exp(x_{i2}^T B)))\exp(-z_i\alpha) \end{array} \right)$$

$$= \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\exp((y_{i1} x_{i1}^T + y_{i2} x_{i2}^T) B)}{(y_{i1}!\, y_{i2}!)} \cdot$$

$$\cdot\, z_i^{(y_{i1}+y_{i2}+\alpha-1)} \exp(-z_i(\exp(x_{i1}^T B)+\exp(x_{i2}^T B)+\alpha))$$

$\longrightarrow$

$$\int_0^\infty (\text{previous})\, dz_i$$

$$= \frac{\alpha^\alpha}{\Gamma(\alpha)} \frac{\exp(y_{i1} x_{i1}^T B + y_{i2} x_{i2}^T B)}{y_{i1}!\, y_{i2}!} \cdot \frac{\Gamma(y_{i1}+y_{i2}+\alpha)}{(\exp(x_{i1}^T B)+\exp(x_{i2}^T B)+\alpha)^{y_{i1}+y_{i2}+\alpha}} \cdot$$

$$\cdot \int_0^\infty \frac{(\exp(x_{i1}^T B)+\exp(x_{i2}^T B)+\alpha)^{y_{i1}+y_{i2}+\alpha}}{\Gamma(y_{i1}+y_{i2}+\alpha)} \cdot z^{y_{i1}+y_{i2}+\alpha-1} \cdot$$

$$\cdot \exp(-z_i(\exp(x_{i1}^T B)+\exp(x_{i2}^T B)+\alpha))\, dz$$

$$= \frac{\Gamma(y_{i1}+y_{i2}+\alpha)}{\Gamma(\alpha)} \frac{\alpha^\alpha}{(\exp(x_{i1}^T B)+\exp(x_{i2}^T B)+\alpha)^{y_{i1}+y_{i2}+\alpha}} \cdot$$

$$\cdot \frac{\exp(y_{i1} x_{i1}^T B + y_{i2} x_{i2}^T B)}{y_{i1}!\, y_{i2}!} \quad (1)$$

$$= f(y_{i1}, y_{i2})$$

$$L(B, \alpha) = \prod_{i=1}^{m} f(y_{i1}, y_{i2})$$

$$= \prod_{i=1}^{m} \left[\frac{\alpha^\alpha\, \Gamma(y_{i1}+y_{i2}+\alpha) \exp(y_{i1} x_{i1}^T B + y_{i2} x_{i2}^T B)}{\Gamma(\alpha)\, (\exp(x_{i1}^T B)+\exp(x_{i2}^T B)+\alpha)^{y_{i1}+y_{i2}+\alpha}\; y_{i1}!\, y_{i2}!} \right]$$

(c) Suggest an algorithm to calculate the MLE of $\theta = (B, \alpha)$
$[\hat{\theta}_M = (\hat{B}_m, \hat{\alpha}_M)]$

- Derive the asymptotic dist of $\hat{B}_m$
  - give explicit form of the cov. of $\hat{B}_m$

Algorithm: Newton Raphson

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + [I(\theta)]^{-1}\Big|_{\theta = \hat{\theta}^{(k)}} \left[\partial/\partial\theta \, \ell(\theta)\right]\Big|_{\theta = \theta^{(k)}}$$

$\hat{\theta}^{(k)}$ = value of $\hat{\theta}$ at current iteration

evaluate $I(\theta)$ & $\frac{\partial}{\partial\theta} \ell(\theta)$ at this value

Continue cycle until $\|\theta^{(k+1)} - \theta^{(k)}\|^2 < 10^{-6}$
(or some other small value of ε)

By MLE theory, we know

$$\sqrt{n}(\hat{\theta}_M - \theta) \xrightarrow{d} N(0, I(\hat{\theta}_M)^{-1})$$

$$I(\hat{\theta}_M) = \begin{bmatrix} \partial^2/\partial B^2 \, \ell(\theta) & \partial^2/\partial B \partial\alpha \, \ell(\theta) \\ \partial^2/\partial B \partial\alpha \, \ell(\theta) & \partial^2/\partial\alpha^2 \, \ell(\theta) \end{bmatrix} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

$$\ell(\theta) = \alpha \log \alpha + \log\Gamma(y_{i1} + y_{i2} + \alpha) - \log\Gamma(\alpha) + {} - (y_{i1} + y_{i2} + \alpha)\log(\exp(x_{i1}^T B) + \exp(x_{i2}^T B) + \alpha) - \log y_{i1}! - \log y_{i2}!$$

$$[I(\hat{\theta}_M)]^{-1} = \begin{bmatrix} (a - b^2 c^{-1})^{-1} & f \\ g & h \end{bmatrix}$$

$$\sqrt{n}(\hat{B}_M - B) \xrightarrow{d} N(0, (a - b^2 c^{-1})^{-1})$$

↳ upper left corner of $I(\hat{\theta}_M)^{-1}$

If $b = 0 \Rightarrow a^{-1}$

(d) $\hat{\beta}_E$ = soln to set of estimating eqns for β

$$\sum_{i=1}^{m} \frac{\partial \mu_i^T}{\partial \beta}(y_i - \mu_i) = \underset{\sim}{0} \rightarrow p \times 1 \text{ vectors of } 0$$

$\mu_i = [\mu_{i1}, \mu_{i2}]^T$, $y_i = [y_{i1}, y_{i2}]^T$

Derive the asymptotic dist of $\hat{\beta}_E$

Taylor series expansion:

Let $S_n(\beta) = \sum_{i=1}^{n} \frac{\partial \mu_i^T}{\partial \beta}(y_i - \mu_i)$

evaluated at $\hat{\beta}_E$, $S_n(\beta) = \underset{\sim}{0}$

Perform a Taylor series expansion of $S_n(\hat{\beta}_E)$ about the true value of $\beta = \beta_0$

(9.4) $$0 = S_n(\hat{\beta}_E) = S_n(\beta_0) + \frac{\partial}{\partial \beta} S_n(\beta_0)(\hat{\beta}_E - \beta_0) +$$

$$+ \frac{1}{2}(\hat{\beta}_E - \beta_0)^T \frac{\partial^2}{\partial \beta^2} S_n(\beta^*)(\hat{\beta}_E - \beta_0)$$

β^* = some value between β_0 & β_E
$= \hat{\beta}_E(t) + \beta_0(1-t)$

By an assumption, $\frac{\partial^2}{\partial \beta^2} S_n(\beta)$ is bounded

$$\left| \partial^2/\partial\beta^2 \, S_n(\beta) \right| \leq f(u)$$

$$\Rightarrow \partial^2/\partial\beta^2 (S_n(\beta)) = O_p(1)$$

(*) $\hat{\beta}_E$ consistent for β_0

- reasoning?

Option 1:

$$\hat{B}_E \xrightarrow{p} B_0 \Rightarrow \hat{B}_E - B_0 \xrightarrow{p} 0 \Rightarrow \hat{B}_E - B_0 = op(1)$$

Option 2: see later

By rules of Op + op,

$$op(1) \cdot Op(1) \cdot op(1) = op(1)$$

$$\Rightarrow \frac{1}{2}(\hat{B}_E - B_0)^T \frac{\partial^2}{\partial B^2} S_n(B^*)(\hat{B}_E - B_0) = op(1)$$

$$0 = \frac{1}{\sqrt{n}} S_n(B_0) + \left(\frac{1}{n}\frac{\partial}{\partial B} S_n(B_0)\right)\sqrt{n}(\hat{B}_E - B_0) + \overbrace{\frac{1}{\sqrt{n}} op(1)}^{op(1)}$$

$$\Rightarrow \sqrt{n}(\hat{B}_E - B_0) = \left(-\frac{1}{\sqrt{n}} S_n(B_0) + op(1)\right)\left(\frac{1}{n}\frac{\partial}{\partial B} S_n(B_0)\right)^{-1}$$

$$S_n(B_0) = \sum_{i=1}^{m} \frac{\partial \mu_i^T}{\partial B}(y_i - \mu_i)\Big|_{B=B_0}$$

$$\frac{\partial}{\partial B} S_n(B_0) = \sum_{i=1}^{m}\left[\left(\frac{\partial^2 \mu_i^T}{\partial B^2}(y_i - \mu_i) + \frac{\partial \mu_i^T}{\partial B}\left(-\frac{\partial \mu_i}{\partial B}\right)\right)\right]\Big|_{B=B_0}$$

$$\frac{1}{n}\frac{\partial}{\partial B} S_n(B_0) \xrightarrow{p} E[\partial/\partial B\, S_n(B_0)] \text{ by WLLN}$$

$$\xrightarrow{p} E[\partial^2 \mu_i^T/\partial B^2 \overbrace{(y_i - \mu_i)}^{0}] - E[(\partial \mu_i^T/\partial B)(\partial \mu_i/\partial B)]$$

$$= 0 - E[(\partial \mu_i/\partial B)^{\otimes 2}]$$

By the CLT,

$$\sqrt{n}\left(\frac{1}{n} S_n(B) - E(S(B))\right) \xrightarrow{d} N(0, Cov(S(B)))$$

$$E(S(B)) = E[\partial \mu_i^T/\partial B (y_i - \mu_i)] = 0$$

$$Cov(S(B)) = \frac{\partial \mu_i^T}{\partial B} Cov(y_i) \frac{\partial \mu_i}{\partial B}$$

$$\circledast \quad Cov(y_i) = \begin{bmatrix} Var(y_{i1} | x_{i1}) & Cov(y_{i1}, y_{i2} | x_{i1}, x_{i2}) \\ Cov(y_{i1}, y_{i2} | x_{i1}, x_{i2}) & Var(y_{i2} | x_{i2}) \end{bmatrix}$$

- replace w/ values found in (a)

$$\Rightarrow \sqrt{n}\left(\frac{1}{n} S_n(B) - 0\right) \xrightarrow{d} N\left(0, \partial \mu_i^T/\partial B \, Cov(y_i)\, \partial \mu_i/\partial B\right)$$

$$\Rightarrow \frac{1}{\sqrt{n}} S_n(B_0) \xrightarrow{d} N\left(0, \partial \mu_i/\partial B \, Cov(y_i)\, \partial \mu_i/\partial B\right)\Big|_{B=B_0}$$

$$\Rightarrow \frac{1}{\sqrt{n}} S_n(B_0) + op(1) \xrightarrow{d} \text{same as above by Slutsky's thm}$$

Consequently,

$$\sqrt{n}(\hat{B}_E - B_0) \equiv \left(\frac{1}{\sqrt{n}} S_n(B_0) + op(1)\right)\left(-\frac{1}{n}\frac{\partial}{\partial B} S_n(B_0)\right)^{-1}$$

$$\xrightarrow{d} N\left(0, \lim_{n\to\infty}\left(\frac{1}{n}\frac{\partial}{\partial B} S_n(B_0)\right)^{-1}\left(\frac{\partial \mu_i^T}{\partial B} Cov(y_i) \frac{\partial \mu_i}{\partial B}\right)\left(\frac{1}{n}\frac{\partial}{\partial B} S_n(B_0)\right)^{-1}\right)$$

By WLLN, $\frac{1}{n}\frac{\partial}{\partial B} S_n(B_0) = -E\left[\left(\partial \mu_i^T/\partial B\right)^{\otimes 2}\right] = -\left(\partial \mu_i^T/\partial B\right)^{\otimes 2}$

$$\Rightarrow \sqrt{n}(\hat{B}_E + B_0) \xrightarrow{d} N\left(0, \left[\left(\frac{\partial \mu_i^T}{\partial B}\right)^{\otimes 2}\right]^{-1}\left(\frac{\partial \mu_i^T}{\partial B} Cov(y_i) \frac{\partial \mu_i}{\partial B}\right)\left[\left(\frac{\partial \mu_i^T}{\partial B}\right)^{\otimes 2}\right]^{-1}\right) \checkmark$$

Option 2:

By Taylor Series expansion,

$$0 = S_n(\hat{B}_E) = S_n(B_0) + \frac{\partial}{\partial B} S_n(B_0)(\hat{B}_E - B_0) + \underbrace{O_p(\|\hat{B}_E - B_0\|^2)}_{op(\|\hat{B}_E - B_0\|^1) \text{ or } O_p(\|\hat{B}_E - B_0\|^2)}$$

Since $R_n = \|\hat{B}_E - B_0\| \xrightarrow{P} 0$ as $n \to \infty$ (consistent)

$\Rightarrow op(\|R_n\|^1) = \|R_n\|^1 op(1) = op(1)$

and $O_p(\|R_n\|^2) = \|R_n\|^2 O_p(1) = op(1) O_p(1) = op(1)$

Theory Exam Section II 2013

2). $Y = X\beta + Z\gamma + \varepsilon$

Y $n\times 1$, X $n\times p$ rank p (full rank)

Z $n\times q$ rank q (full rank)

β $p\times 1$, γ $q\times 1$

$\varepsilon \sim N(0,R)$ indep.

$\gamma \sim N(0,D)$

R, D positive Definite matrices

$N_n(a,b)$ = n-variate normal RV w/ mean vector a & cov matrix b.

(a) For known R & D,

$Y|X,\gamma \sim N(X\beta + Z\gamma, R)$

– Derive the marginal dist of $Y|X$

$Y|X$ will also be normal

$E[Y|X] = E[E(Y|X,\gamma)]$

$= E[X\beta + Z\gamma]$

Since $\gamma \sim N(0,D) \Rightarrow E[Z\gamma] = 0$

$= X\beta + 0$

$= X\beta$

$Cov(Y|X) = E[Cov(Y|X,\gamma)] + Cov(E(Y|X,\gamma))$

$= E[R] + Cov[X\beta + Z\gamma]$

$X\beta$ constant $\Rightarrow$ ignore

$= R + Z\,Cov(\gamma)\,Z'$

$= R + ZDZ'$

$Y|X \sim N(X\beta, R + ZDZ')$

(b) R + D known, γ unknown.

(i) Show predictor of γ, $\hat{\gamma} = DZ'V^{-1}(Y - X\hat{B})$,
satisfies the conditional likelihood eqns for (B, γ)
where $\hat{B}$ = MLE of B + $V = R + ZDZ'$

$$f(y; B, \gamma) = \frac{1}{\sqrt{2\pi}} \frac{1}{|R|^{1/2}} \exp\left(-\frac{1}{2}(Y - XB - Z\gamma)' R^{-1} (Y - XB - Z\gamma)\right)$$

$$= \frac{1}{\sqrt{2\pi}\,|R|^{1/2}} \exp\left(-\frac{1}{2}\left[(Y-XB)'R^{-1}(Y-XB) - 2(Y-XB)'R^{-1}(Z\gamma) + \gamma' Z' R^{-1} Z\gamma\right]\right)$$

sufficient statistic for $\gamma = (Y - XB)'R^{-1}Z$

$$Y = XB + Z\gamma + \varepsilon$$

$$Y | X, \gamma \sim N(XB + Z\gamma, R)$$

R positive def $\Rightarrow R = QQ'$ by spectral decomp
$(R = P\Lambda P' = P\Lambda^{1/2}\Lambda^{1/2}P' = QQ', \quad Q = P\Lambda^{1/2})$

$$Q^{-1}Y = Q^{-1}XB + Q^{-1}Z\gamma + Q^{-1}\varepsilon$$
$$\Rightarrow \quad Y^* = X^*B + Z^*\gamma + \varepsilon^*$$

$\varepsilon^* \sim N(0, Q^{-1}R(Q^{-1})')$
$Q^{-1}R(Q^{-1})' = Q^{-1}QQ'(Q^{-1})' = I$
$\varepsilon^* \sim N(0, I)$

$Y^* \sim N(X^*B + Z^*\gamma, I)$
by same reasoning.

$\hat{\gamma}$ & $\hat{\beta}$ satisfy the following normal eqns:

$$\begin{bmatrix} X^{*\prime} \\ Z^{*\prime} \end{bmatrix} \begin{bmatrix} X^* & Z^* \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} X^{*\prime} \\ Z^{*\prime} \end{bmatrix} Y^*$$

$$\Rightarrow \begin{bmatrix} X^{*\prime}X^* & X^{*\prime}Z^* \\ Z^{*\prime}X^* & Z^{*\prime}Z^* \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} X^{*\prime} \\ Z^{*\prime} \end{bmatrix} Y^*$$

$$\Rightarrow \begin{cases} X^{*\prime}X^*\hat{\beta} + X^{*\prime}Z^*\hat{\gamma} = X^{*\prime}Y^* \\ Z^{*\prime}X^*\hat{\beta} + Z^{*\prime}Z^*\hat{\gamma} = Z^{*\prime}Y^* \end{cases}$$

$\Rightarrow$?

(ii) Derive the exact dist of $\hat{\gamma} = DZ'V^{-1}(Y - X\hat{B})$

$$\hat{\gamma} = DZ'V^{-1}Y - DZ'V^{-1}\underbrace{X(X'X)^{-1}X'}_{M}Y$$

Function of normal rvs $\Rightarrow$ will be normal

$$E[Y] = E[E[Y|X]] = E[XB] = XB \quad \text{(constants)}$$
$$Cov[Y] = E[Cov(Y|X)] + Cov[E(Y|X)]$$
$$= E[R + ZDZ'] + Cov(XB)$$
$$= R + ZDZ' = V$$

$$E[\hat{\gamma}] = DZ'V^{-1}(XB) - DZ'V^{-1}M(XB)$$
M = orthog proj matrix onto $C(X)$
$\Rightarrow MX = X$
$$= DZ'V^{-1}(XB - XB)$$
$$= 0$$

$$Cov(\hat{\gamma}) = Cov(DZ'V^{-1}(I-M)Y)$$
$$= DZ'V^{-1}(I-M)\underbrace{Cov(Y)}_{V}(I-M)V^{-1}ZD$$
$$= DZ'V^{-1}(I-M)V(I-M)ZD$$

$$\hat{\gamma} \sim N(0, \underbrace{DZ'V^{-1}(I-M)V(I-M)ZD}_{?})$$

ⓒ Suppose $R = \sigma^2 I$

B, σ^2, D unknown

Devise a detailed EM algorithm for jointly estimating (B, σ^2, D).

Full joint likelihood:

$$P(Y, \gamma \mid X) = P(Y \mid X)\, P(\gamma)$$

$$\propto \frac{1}{\sigma} \exp\left(\frac{-1}{2\sigma^2} (Y - XB)'(Y - XB)\right) \exp\left(-\frac{1}{2} \gamma' D^{-1} \gamma\right) \cdot \frac{1}{|D|^{1/2}}$$

$$\ell(B, \sigma^2, D) \propto -\frac{1}{2} \log \sigma^2 - \frac{1}{2} \log |D| +$$

$$-\frac{1}{2\sigma^2} (Y - XB)'(Y - XB) - \frac{1}{2} \gamma' D^{-1} \gamma$$

$$\frac{\partial}{\partial \sigma^2} \ell(\sigma^2, B, D) = -\frac{1}{2}\left(\frac{1}{\sigma^2}\right) + \frac{1}{2\sigma^4} (Y - XB)'(Y - XB)$$

$$E\left[\partial / \partial \sigma^2\, \ell(\sigma^2, B, D) \mid \text{past iterations}\right]$$

(d) $A = I - M$ $\quad M =$ orthog proj. operator onto $C(X)$

$W = B'Y$

$A = BB' = I - M$

$B'B = I$

Consider estimation of unknown parameters using the marginal dist of $Y|X$. in (a)

$Y|X \sim N(X\beta,\ R + ZDZ')$

(i) $\hat{\beta}$ = MLE of β when (D, R) are fixed.

Show $\mathrm{Cov}(W, \hat{\beta}) = 0$

Let $V = R + ZDZ'$ (fixed)

- positive definite if R + D positive definite

$$L(Y, \beta) \propto \frac{1}{|V|^{1/2}} \exp\left(-\frac{1}{2}(Y - X\beta)'V^{-1}(Y - X\beta)\right)$$

$V = QQ'$ by spectral decomposition

Note: $V = P\Lambda P' = P\Lambda^{1/2}\Lambda^{1/2}P'$

$= QQ'$ where $Q = P\Lambda^{1/2}$

Transform Variables as follows:

$Y = X\beta + \gamma$

$\gamma \sim N(0, V)$

$\Rightarrow Q^{-1}Y = Q^{-1}X\beta + Q^{-1}\gamma$

$\Rightarrow Y^* = X^*\beta + \gamma^*$

$Y^* \sim N(X^*\beta, \mathrm{Cov}(\gamma^*)) \equiv N(X^*\beta,\ Q^{-1}V(Q^{-1})^T)$

$Q^{-1}V(Q^{-1})^T = Q^{-1}QQ'(Q^{-1})' = I$

$\Rightarrow Y^* \sim N(X^*\beta, I)$ $\rightarrow$

$Y^* \sim N(0, I)$ likewise

$$L(Y^*, B) \propto \exp\left(-\tfrac{1}{2}(Y^* - X^*B)'(Y^* - X^*B)\right)$$

$$\ell(Y^*, B) \propto -\frac{1}{2}(Y^* - X^*B)'(Y^* - X^*B)$$

$$= \frac{-1}{2}\left(Y^{*\prime}Y^* - 2Y^{*\prime}X^*B + B'X^{*\prime}X^*B\right)$$

$$\frac{\partial}{\partial B}\ell(Y^*, B) = -\frac{1}{2}\left(-2Y^{*\prime}X^* + B'(X^{*\prime}X^* + (X^{*\prime}X^*)')\right)$$

$$= Y^{*\prime}X^* - B'(X^{*\prime}X^*) \overset{\text{set}}{=} 0$$

$$\Rightarrow \hat{B} = \underbrace{(X^{*\prime}X^*)^{-1}}_{X^* \text{ full rank } (X \text{ full rank})}X^{*\prime}Y^*$$

$$\hat{B} = ((Q^{-1}X)'(Q^{-1}X))^{-1}(Q^{-1}X)'(Q^{-1}Y)$$
$$= (X'(Q^{-1})'Q^{-1}X)^{-1}X'(Q^{-1})'Q^{-1}Y$$

$$V = QQ'$$
$$\Rightarrow V^{-1} = (QQ')^{-1}$$
$$= ((Q')^{-1}(Q)^{-1})$$
$$= (Q^{-1})'(Q^{-1})$$

$$\hat{B} = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

$$\text{Cov}(W, \hat{B}) = \text{Cov}(B'Y, (X'V^{-1}X)^{-1}X'V^{-1}Y)$$
$$= B'\text{Cov}(Y)[(X'V^{-1}X)^{-1}X'V^{-1}]'$$
$$= B'\underbrace{VV^{-1}}_{I}X(X'V^{-1}X)^{-1}$$

(note: V positive def $\Rightarrow V^{-1}$ positive def $\Rightarrow V$ & V^{-1} symmetric)

$$= B'X(X'V^{-1}X)^{-1}$$

$\longrightarrow$

$A = BB'$ & $B'B = I$

$B'X(X'V^{-1}X)^{-1}$

$= \underbrace{(B'B)}_{I} B'X(X'V^{-1}X)^{-1}$

$= B'\underbrace{(BB')}_{A} X(X'V^{-1}X)^{-1}$

$= B'(I-M)X(X'V^{-1}X)^{-1}$

Since M = orthog proj. matrix onto $C(X)$

$\Rightarrow$ M projects along $C(X)^{\perp}$

$\Rightarrow$ $I-M$ orthog proj matrix onto $C(X)^{\perp}$ & along $C(X)$

$\Rightarrow (I-M)X = 0$

$\Rightarrow B'(I-M)X(X'V^{-1}X)^{-1}$

$= B'(0)(X'V^{-1}X)^{-1}$

$= 0$ ✓

(ii) Derive the density of w.

$Y \sim N(XB, V)$ $\quad V = R + ZDZ'$

~~X~~ (For notation purposes, represent B as q)

$B'Y \sim N(B'XB, B'VB)$

3). $Y_1, \ldots, Y_N$ indep RVs

$Y_i = Bx_i + \varepsilon_i$ $\quad\quad \varepsilon_i \sim N(0, \sigma^2)$

$x_1, \ldots, x_n$ known positive constants

B, σ^2 unknown scalar parameters.

$R_k = I(Y_k \text{ is selected})$

↳ random sample of Y_i

$R_1, \ldots, R_N$ mutually indep & indep of (Y_i, x_i)

$P(R_k = 1) = \pi_k$ for some known positive constant $\pi_k \in (0,1)$

(a) Write the likelihood function of the observed data

$Y_i \sim N(x_i B, \sigma^2)$ if $R_i = 1$

Note: Random sample = sample w/ replacement

Likelihood: (obs. only)

$$\prod_{k=1}^{n} \left(f(y_k)\pi_k\right)^{I(R_k=1)} \quad n = \text{sample size}$$

$$= \prod_{k=1}^{n} \left(\pi_k \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2} (y_k - x_k B)^2 \right) \right)^{R_i}$$

Note: Since we are only concerned w/ the observed data, $I(R_k = 1) = 1$ for all Y_k in this likelihood.

Assuming obs.

$$P(Y_k, R_k = 1) = P(Y_k \mid R_k = 1)\, P(R_k = 1) = \left(f(y_k)\pi_k\right)^{R_i}$$

(b) Compute the MLE of β & σ^2 ($\hat{\beta}$ & $\hat{\sigma}^2$)

- if $n=0 \Rightarrow \hat{\beta}$ & $\hat{\sigma}^2 = 0$

$$l_n(\beta,\sigma^2) = \sum_{k=1}^{n}\left[\log \pi_k - \frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y_i - x_i\beta)^2\right](R_i) \quad (*)$$

$$\frac{\partial}{\partial\beta} l_n(\beta,\sigma^2) = -\frac{1}{\cancel{2\sigma^2}}\sum_{k=1}^{n}\cancel{(2)}(\underbrace{y_i}_{i=k} - x_i\beta)(\cancel{-}x_i) \overset{\text{set}}{=} 0$$

$$\Rightarrow \sum_{k=1}^{n} y_k x_k - \sum_{k=1}^{n} x_k^2\beta = 0$$

$$\Rightarrow \hat{\beta} = \frac{\sum_{k=1}^{n} R_i y_k x_k}{\sum_{k=1}^{n} R_i x_k^2} \longrightarrow \frac{\sum_{k=1}^{n} R_k y_k x_k}{\sum_{k=1}^{n} R_i x_k^2}$$

$$\frac{\partial}{\partial\sigma^2} l_n(\beta,\sigma^2) = -\frac{n}{\cancel{2}}\left(\frac{1}{\sigma^2}\right) + \frac{1}{\cancel{2}\sigma^4}\sum_{k=1}^{n}(\overset{i=k}{y_i} - x_i\beta)^2 \overset{\text{set}}{=} 0$$

$$\Rightarrow -n\sigma^2 + \sum_{k=1}^{n}(y_i - x_i\beta)^2 = 0$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n} R_i(y_k - x_k\hat{\beta})^2$$

by invariance property of MLEs.

(c) Derive the mean + variance of $\hat{B}$

$$E(\hat{B}) = \frac{E[\sum_{k=1}^{n} y_k x_k]}{\sum_{k=1}^{n} x_k^2}$$

$$= \frac{\sum_{k=1}^{n} x_k E(y_k)}{\sum_{k=1}^{n} x_k^2}$$

$$= \frac{B \sum_{k=1}^{n} x_k (x_k)}{\sum_{k=1}^{n} x_k^2}$$

$$= B(1)$$

$$= \boxed{B}$$ ~ true value.

$$Var(\hat{B}) = \left(\frac{1}{\sum_{k=1}^{n} x_k^2}\right)^2 Var(\sum_{k=1}^{n} y_k x_k)$$

$$= \left(\frac{1}{\sum_{k=1}^{n} x_k^2}\right)^2 \sum_{k=1}^{n} x_k^2 Var(y_k)$$

$$= \left(\frac{1}{\sum_{k=1}^{n} x_k^2}\right)^2 (\sigma^2) \sum_{k=1}^{n} x_k^2$$

$$= \boxed{\frac{\sigma^2}{\sum_{k=1}^{n} x_k^2}}$$

(d) Dist of $\hat{B}$

Since $\hat{B}$ is a linear combo of normal RVs ($y_k \sim N(x_k B, \sigma^2)$)

$\Rightarrow$ $\hat{B}$ itself is also normal

$$\hat{B} \sim N(E(\hat{B}), Var(\hat{B}))$$

$$\Rightarrow \hat{B} \sim N(B, \sigma^2 / \sum_{k=1}^{n} x_k^2)$$

↑ true value

(e) Calculate the confidence interval of level $(1-\alpha)$ for B based on the conditional dist of $\hat{B} \mid R_1, \ldots, R_N$

$\hat{B} \mid R_1, \ldots, R_n = \hat{B} \mid$ observed data

$= \hat{B}$ calculated in past sections.

$$\hat{B} \pm 1.96 \sqrt{\hat{Var}(\hat{B})}$$

$$= \hat{B} \pm 1.96 \sqrt{\frac{\hat{\sigma}^2}{\sum_{k=1}^{n} x_k^2}}$$

$$= \hat{B} \pm 1.96 \sqrt{\frac{1}{n} \frac{\sum_{k=1}^{n} (y_k - x_k B)^2}{\sum_{k=1}^{n} x_k^2}}$$

(f) $\tilde{B} = \left[\sum_{i=1}^{N} \frac{R_i}{\pi_i} y_i \right] \left[\sum_{i=1}^{N} x_i \right]^{-1}$

Show $\tilde{B}$ unbiased

Derive the variance of $\tilde{B}$

(i) Show $\tilde{B}$ unbiased

$$E[\tilde{B}] = \frac{\sum_{i=1}^{N} \frac{1}{\pi_i} E[R_i y_i]}{n \bar{x}}$$

$$E[R_i y_i] = E[E[R_i y_i \mid R_i]]$$

$$= E[R_i E[y_i \mid R_i]]$$

$= \ldots$ ↑ y_i, R_i indep $\Rightarrow E[y_i \mid R_i] = E[y_i]$

OR

R_i, y_i indep

$\Rightarrow E[R_i y_i] = E[R_i] E[y_i]$

$= \pi_i (x_i B)$

$\longrightarrow$

$$E[\tilde{B}] = \frac{1}{n\bar{x}} \sum_{i=1}^{N} \frac{1}{\cancel{\pi_i}} (\cancel{\pi_i} x_i B)$$

$$= \left(\frac{\cancel{\Sigma x_i}}{\cancel{\Sigma x_i}}\right) B$$

$$= B \quad \checkmark$$

(ii) Var of $\tilde{B}$

$$Var(\hat{B}) = \frac{1}{(n\bar{x})^2} \sum_{i=1}^{N} \frac{1}{\pi_i^2} Var(R_i y_i)$$

$$Var(R_i y_i) = E[R_i^2 y_i^2] - (E[R_i y_i])^2$$

$$= E[R_i^2] E[y_i^2] - (E(R_i))^2 (E(y_i))^2$$

$$= (Var(R_i) + (E(R_i))^2)(Var(y_i) + (E(y_i))^2) - \pi_i^2 (x_i B)^2$$

$$= (\underbrace{\pi_i(1-\pi_i) + \pi_i^2}_{\pi_i - \cancel{\pi_i^2} + \cancel{\pi_i^2}})(\sigma^2 + (x_i B)^2) - \pi_i^2 (x_i B)^2$$

$$= \pi_i \sigma^2 + \pi_i (x_i B)^2 - \pi_i^2 (x_i B)^2$$

$$= \pi_i \sigma^2 + (x_i B)^2 \pi_i (1-\pi_i)$$

(g) Find the optimal π_i to minimize $var(\hat{\beta})$ under the condition that sample size = n (fixed) & $\sum_{i=1}^{N} \pi_i = n$

· Lagrange multiplier method:

$$\frac{\partial}{\partial \pi_i}\left(Var(\hat{\beta}) + \sum_{i=1}^{n} \pi_i (\lambda)\right) =$$

$$= \sigma^2 + (x_i\beta)^2 - 2\pi_i(x_i\beta)^2 + \lambda = 0$$

$$\Rightarrow \quad 2\pi_i(x_i\beta)^2 = \lambda + (x_i\beta)^2 + \sigma^2$$

$$\Rightarrow \quad \pi_i = \frac{\lambda + (x_i\beta)^2 + \sigma^2}{2(x_i\beta)^2}$$

$$\sum_{i=1}^{n} \pi_i = n = \sum_{i=1}^{n}\left(\frac{\lambda + (x_i\beta)^2 + \sigma^2}{2(x_i\beta)^2}\right)$$

$$\Rightarrow \sum_{i=1}^{n} \frac{\lambda}{2(x_i\beta)^2} = n - \sum_{i=1}^{n} \frac{(x_i\beta)^2 + \sigma^2}{2(x_i\beta)^2}$$

$$\Rightarrow \quad \lambda = \left(n - \sum_{i=1}^{n} \frac{(x_i\beta)^2 + \sigma^2}{2(x_i\beta)^2}\right) \frac{1}{\sum_{i=1}^{n} 2(x_i\beta)^2}$$

$$\pi_i = \frac{\lambda + (x_i\beta)^2 + \sigma^2}{2(x_i\beta)^2} \Bigg|_{\lambda \text{ result above}}$$

(h) For any given function $g(\cdot)$ & π_i, show

$$\tilde{B}(g) \equiv \frac{\sum_{i=1}^N g(x_i) + \sum_{i=1}^N \frac{R_i}{\pi_i}(y_i - g(x_i))}{\sum_{i=1}^n x_i}$$

is unbiased for B & calculate its variance

(i) unbiased

$$E(\tilde{B}(g)) = \frac{\sum_{i=1}^N g(x_i) + \sum_{i=1}^N \frac{1}{\pi_i}(E[R_i y_i] - g(x_i)E(R_i))}{\sum_{i=1}^N x_i}$$

$$= \frac{\sum_{i=1}^N g(x_i)}{\sum_{i=1}^N x_i} + \frac{\sum_{i=1}^N \cancel{\frac{1}{\pi_i}} \cancel{E(R_i)}^{1} E(y_i)}{\sum_{i=1}^N x_i} - \frac{\sum_{i=1}^N g(x_i) \cancel{\frac{1}{\pi_i}} \cancel{E(R_i)}^{1}}{\sum_{i=1}^N x_i}$$

$$= \frac{\sum_{i=1}^N g(x_i)}{\sum_{i=1}^N x_i} + \frac{\sum_{i=1}^N x_i B}{\sum_{i=1}^N x_i} - \frac{\sum_{i=1}^N g(x_i)}{\sum_{i=1}^N x_i}$$

$$= \cancel{\frac{\sum_{i=1}^N g(x_i)}{\sum_{i=1}^N x_i}} + B - \cancel{\frac{\sum_{i=1}^N g(x_i)}{\sum_{i=1}^N x_i}}$$

$$= B \checkmark$$

(ii) $Var(\tilde{B}(g))$

$$Var(\tilde{B}(g)) = \frac{1}{\left(\sum_{i=1}^N x_i\right)^2} \sum_{i=1}^N \frac{1}{\pi_i^2} Var(R_i y_i - R_i g(x_i))$$

$Var(R_i y_i - R_i g(x_i))$

$= Var(R_i y_i) - 2\,Cov(R_i y_i, R_i g(x_i)) + Var(R_i g(x_i))$

$= \pi_i \sigma^2 + (x_i B)^2 \pi_i (1-\pi_i) + (g(x_i))^2(\pi_i(1-\pi_i)) +$

$\overset{*}{\rightarrow} -2\left(E[R_i^2 y_i g(x_i)] - E(R_i y_i)E(R_i g(x_i))\right)$

$*$ $E[R_i^2 y_i g(x_i)] = g(x_i)E[R_i^2]E[y_i]$ (y_i, R_i indep)

$= g(x_i)\pi_i(x_i B)$

$\longrightarrow$

$\mathrm{Var}(\hat{B}(g)) =$

$= \pi_i \sigma^2 + (x_i B)^2 \pi_i (1-\pi_i) + (g(x_i))^2 (\pi_i (1-\pi_i)) +$
$- 2 (g(x_i) \pi_i (x_i B) - \pi_i (x_i B) g(x_i) \pi_i)$

$= \pi_i \sigma^2 + \pi_i (1-\pi_i) ((x_i B)^2 + (g(x_i))^2) - 2 g(x_i)(x_i B) \pi_i (1-\pi_i)$

$= \pi_i \sigma^2 + \pi_i (1-\pi_i) (x_i B - g(x_i))^2$