1. GLM 基础

$A=\begin{bmatrix} a & b \\ c & d\end{bmatrix} \Rightarrow A^{-1}=\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a\end{bmatrix}$

§1 GLM 的 random component

1. Exponential dispersion model (EDM)

$$Y\sim P(y;\theta,\phi)=a(y,\phi)\exp\left\{\frac{y\theta-\kappa(\theta)}{\phi}\right\}$$

① θ canonical parameter ② $\kappa(\theta)$ cumulant function ③ $\phi>0$ dispersion parameter

2. EDM 的例子

① Normal distribution $Y\sim N(\mu,\sigma^2)$ $(\mu;\sigma^2)$

$f(y;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(y-\mu)^2}{2\sigma^2}\}=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{y^2}{2\sigma^2}\}\cdot\exp\{\frac{\mu y-\frac{1}{2}\mu^2}{\sigma^2}\}$

其中，$\theta=\mu$, $\phi=\sigma^2$, $\kappa(\theta)=\frac{1}{2}\mu^2=\frac{1}{2}\theta^2$, $a(y,\phi)=\frac{1}{\sqrt{2\pi\phi}}\exp\{-\frac{y^2}{2\phi}\}$

② Bernoulli distribution $Y\sim Ber(p)$ $(p, p(1-p))$

$f(y;p)=p^y(1-p)^{1-y}=\exp\{y\log\frac{p}{1-p}+\log(1-p)\}$

其中，$\theta=\log\frac{p}{1-p}$ $(p=\frac{e^\theta}{1+e^\theta})$, $\phi=1$, $\kappa(\theta)=-\log(1-p)=\log(1+e^\theta)$, $a(y,\phi)=1$

③ Binomial distribution $Y\sim Bin(n,p)$, n 已知 $(np; np(1-p))$

$f(y;p)=\binom{n}{y}\cdot p^y(1-p)^{n-y}=\binom{n}{y}\cdot\exp\{y\log\frac{p}{1-p}+n\log(1-p)\}$

其中，$\theta=\log\frac{p}{1-p}$ $(p=\frac{e^\theta}{1+e^\theta})$, $\phi=1$, $\kappa(\theta)=-n\log(1-p)=n\log(1+e^\theta)$, $a(y,\phi)=\binom{n}{y}$

④ Poisson distribution $Y\sim Poi(\mu)$ $(\mu;\mu)$

$f(y;\mu)=\frac{e^{-\mu}\mu^y}{y!}=\frac{1}{y!}\exp\{y\log\mu-\mu\}$

其中，$\theta=\log\mu$ $(\mu=e^\theta)$ $\phi=1$, $\kappa(\theta)=\mu=e^\theta$, $a(y,\phi)=\frac{1}{y!}$

⑤ Weibull distribution $Y\sim Weibull(\alpha,\gamma)$ $(\alpha=1$ 时$)$ $(\gamma\Gamma[1+\frac{1}{\alpha}],\ \gamma^2[\Gamma(1+\frac{2}{\alpha})-(\Gamma(1+\frac{1}{\alpha}))^2])$

$f(y;\gamma)=\frac{1}{\gamma}\exp(-\frac{y}{\gamma})=\exp\{-\frac{y}{\gamma}-\log\gamma\}$

其中，$\theta=-\frac{1}{\gamma}$, $\phi=1$, $\kappa(\theta)=\log\gamma=-\log(-\theta)$, $a(y,\phi)=1$

⑥ Gamma distribution $Y\sim Gamma(\alpha,\beta)$ $f(y;\alpha,\beta)=\frac{1}{\Gamma(\alpha)\beta^\alpha}y^{\alpha-1}e^{-\frac{y}{\beta}}$

$=\frac{1}{\Gamma(\alpha)}y^{\alpha-1}\cdot\exp\{\alpha\log\alpha\}\exp\left\{\frac{-\frac{1}{\alpha\beta}y-\log\alpha\beta}{\frac{1}{\alpha}}\right\}$ $(\frac{\alpha}{\beta}, \frac{\alpha}{\beta^2})$

其中，$\theta=-\frac{1}{\alpha\beta}$, $\phi=\frac{1}{\alpha}$, $\kappa(\theta)=\log\alpha\beta=-\log(-\theta)$

$a(y,\phi)=\frac{1}{\Gamma(\alpha)}y^{\alpha-1}\cdot\exp\{\alpha\log\alpha\}=\frac{1}{\Gamma(\frac{1}{\phi})}y^{\frac{1}{\phi}-1}\cdot\exp\{\frac{1}{\phi}\log\frac{1}{\phi}\}$

⑦ Geometric distribution $Y\sim Geo(p)$ $(\frac{1}{p},\frac{1-p}{p^2})$

$f(y;p)=p(1-p)^{y-1}=\exp\{\log(1-p)\cdot y-(\log(1-p)-\log p)\}$

其中，$\theta=\log(1-p)$ $(p=1-e^\theta)$, $\phi=1$, $\kappa(\theta)=\theta-\log(1-e^\theta)$, $a(y,\phi)=1$

⑧ Exponential distribution $Y\sim Exp(\mu)$ (μ,μ^2)

$f(y;\mu)=\frac{1}{\mu}\exp(-\frac{y}{\mu})=\exp(-\frac{y}{\mu}-(-\log\frac{1}{\mu}))$

其中，$\theta=-\frac{1}{\mu}$, $\phi=1$, $\kappa(\theta)=-\log(-\theta)$, $a(y,\phi)=1$

⑨ Inverse Gaussian $Y\sim IG(\mu,\lambda)$ $(\mu,\frac{\mu^3}{\lambda})$

$f(y;\mu,\lambda)=\sqrt{\frac{\lambda}{2\pi y^3}}\exp\{-\frac{\lambda(y-\mu)^2}{2\mu^2y}\}=\sqrt{\frac{\lambda}{2\pi y^3}}\exp\{-\frac{\lambda}{2y}\}\cdot\exp\left\{\frac{-\frac{1}{2\mu^2}y+\frac{1}{\mu}}{\frac{1}{\lambda}}\right\}$

其中，$\theta=-\frac{1}{2\mu^2}$ $(\mu=(-2\theta)^{-\frac{1}{2}})$, $\phi=\frac{1}{\lambda}$, $\kappa(\theta)=-\frac{1}{\mu}=-\sqrt{-2\theta}$

⑩ Negative Binomial $Y\sim NB(r,p)$ $(r$ 已知$)$ $(\frac{r(1-p)}{p},\frac{r(1-p)}{p^2})$

$f(y;r,p)=\binom{y+r-1}{y}\cdot(1-p)^y\cdot p^r=\binom{y+r-1}{y}\cdot\exp\{y\log(1-p)+r\log p\}$

其中，$\theta=\log(1-p)$ $(p=1-e^\theta)$, $\phi=1$, $\kappa(\theta)=-r\log p=-r\log(1-e^\theta)$, $a(y,\phi)=\binom{y+r-1}{y}$

3. deviation: 令 $t(y,\mu)=y\theta-\kappa(\theta)$，若 spa holds：

① unit deviance: $d(y,\mu)=2\{t(y,y)-t(y,\mu)\}$ $\Rightarrow$ $\frac{d(y,\mu)}{\phi}\sim\chi^2_1$

② deviance (function) / total deviance: $D(y,\mu)=\sum_{i=1}^n d(y_i,\mu_i)$

③ scaled deviance function: $D^*(y,\mu)=\frac{D(y,\mu)}{\phi}\sim\chi^2_n$

注: $\ell(\mu;y)=\sum_{i=1}^n\log b(y_i,\phi)-\frac{D(y,\mu)}{2\phi}$

4. dispersion model form (DMF) $f(y;\mu,\phi)=b(y,\phi)\exp\{-\frac{1}{2\phi}d(y,\mu)\}$

其中 $b(y,\phi)=a(y,\phi)\cdot\exp\{t(y,y)/\phi\}$

$\tilde{P}(y;\mu,\phi)=\frac{1}{\sqrt{2\pi\phi V(y)}}\exp\{-\frac{1}{2\phi}d(y,\mu)\}$ 用 $\frac{1}{\sqrt{2\pi\phi V(y)}}$ 近似 $b(y,\phi)$

- Poisson: $y\geq 3$ • Binomial: $y\geq 3$ 且 $m-y\geq 3$ • Gamma: $\phi\leq\frac{1}{3}$

§2 Regression coefficient β 的估计

1. Score: $U(\beta)=\sum_{i=1}^n\frac{(y_i-\mu_i)}{Var(Y_i)g'(\mu_i)}x_i=X^T\Delta(y-\mu)$ $\Delta=diag\{\frac{1}{Var(Y_i)g'(\mu_i)}\}$

2. Fisher Info: $I_n(\beta)=E[-H(\beta)]=\phi^{-1}X^TWX$ $W=diag\{\frac{1}{V(\mu_i)[g'(\mu_i)]^2}\}$

§3 Dispersion ϕ 的估计

1. Mean deviance estimator: $\tilde{\phi}=\frac{D(y,\hat{\mu})}{n-p'}\sim\chi^2_{n-p'}$

2. Pearson estimator: $X^2=\sum_{i=1}^n\frac{(y_i-\hat{\mu}_i)^2}{V(\hat{\mu}_i)}$, $\bar{\phi}=\frac{X^2}{n-p'}=\frac{1}{n-p'}\sum_{i=1}^n\frac{(y_i-\hat{\mu}_i)^2}{V(\hat{\mu}_i)}$

§4 Likelihood ratio test (LRT)

1. LRT 的一个应用: Compare nested models (H_0: $\beta_{P_A+1}=\cdots=\beta_{P_B}=0$)

① (ϕ 已知) $T_{LR}=2(\ell(B)-\ell(A))=\frac{D(y,\hat{\mu}_A)-D(y,\hat{\mu}_B)}{\phi}\sim\chi^2(P_B-P_A)$

② (ϕ 未知) $F=\frac{[D(y,\hat{\mu}_A)-D(y,\hat{\mu}_B)]/(P_B-P_A)}{\tilde{\phi}}\sim F(P_B-P_A, n-P_B')$

其中 $\tilde{\phi}=\frac{D(y,\hat{\mu}_B)}{n-P_B'}$, $\frac{D(y,\hat{\mu}_B)}{\phi}\sim\chi^2(n-P_B')$ (若 saddlepoint approximation 为 accurate)

2. goodness-of-fit test

① Deviance gof test: $D(y,\hat{\mu})\sim\chi^2_{n-p'}$ ② Pearson test: $X^2=\sum_{i=1}^n\frac{(y_i-\hat{\mu}_i)^2}{V(\hat{\mu}_i)}\sim\chi^2_{n-p'}$

2. Binomial GLM

§1 Binomial GLM 的模型设定

1. ungrouped data model:

① $Y_i\overset{iid}{\sim}B(\pi_i)\Rightarrow Y_i=\begin{cases}0 & w.p.\ 1-\pi_i\\ 1 & w.p.\ \pi_i\end{cases}$ ② $g(\pi_i)=\beta_0+\sum_{j=1}^p\beta_jx_{ij}$

③ $f(y_i;\pi_i)=\exp\{y_i\cdot\log\frac{\pi_i}{1-\pi_i}+\log(1-\pi_i)\}$

2. grouped data model

① $Y_1,\cdots,Y_C\overset{iid}{\sim}Bin(n_c,\pi_c)$ ② $g(\pi_c)=\beta_0+\sum_{j=1}^p\beta_jx_{cj}$

③ $f(y_c;\pi_c)=\binom{n_c}{y_c}\exp\{y_c\cdot\log\frac{\pi_c}{1-\pi_c}+n_c\log(1-\pi_c)\}$

§3 Binomial GLM 的 interpretation

1. 情况一：one binary covariate $\log\frac{\pi(x)}{1-\pi(x)}=\beta_0+\beta_1x$

$\begin{cases}\text{odds for group } A=\frac{\pi(0)}{1-\pi(0)}=\exp(\beta_0)\\ \text{odds for group } B=\frac{\pi(1)}{1-\pi(1)}=\exp(\beta_0+\beta_1)\end{cases}\Rightarrow OR_{x=1\text{ vs. }x=0}=\frac{\exp(\beta_0+\beta_1)}{\exp(\beta_0)}=e^{\beta_1}$

$\Rightarrow$ Interpretation: group B (x=1) 的 odds 是 group A (x=0) 的 odds 的 e^{β_1} 倍

2. 情况二：one categorical covariate with multiply levels ($X=0,1,2$ 定义 $x_1=1_{x=1}$, $x_2=1_{x=2}$) A, B, C

$\log\frac{\pi(x_1,x_2)}{1-\pi(x_1,x_2)}=\beta_0+\beta_1x_1+\beta_2x_2$

$\Rightarrow\begin{cases}OR_{x=1\text{ vs. }x=0}=\frac{\exp(\beta_0+\beta_1)}{\exp(\beta_0)}=e^{\beta_1} & \text{group}(x=1)\text{ 的 odds 是 group}(x=0)\text{ 的 odds 的 }e^{\beta_1}\text{ 倍}\\ OR_{x=2\text{ vs. }x=0}=\frac{\exp(\beta_0+\beta_2)}{\exp(\beta_0)}=e^{\beta_2} & \text{group}(x=2)\text{ 的 odds 是 group}(x=0)\text{ 的 odds 的 }e^{\beta_2}\text{ 倍}\\ OR_{x=1\text{ vs. }x=2}=\frac{\exp(\beta_0+\beta_1)}{\exp(\beta_0+\beta_2)}=e^{\beta_1-\beta_2} & \text{group}(x=1)\text{ 的 odds 是 group}(x=2)\text{ 的 odds 的 }e^{\beta_1-\beta_2}\text{ 倍}\end{cases}$

3. 情况三：one binary covariate 和 one continuous covariate (with interaction)

$x=0,1$, y continuous with interaction $\Rightarrow\log\frac{\pi(x,y)}{1-\pi(x,y)}=\beta_0+\beta_1x+\beta_2y+\beta_3xy$

$y=y_0$: $OR_{x=1,y=y_0\text{ vs. }x=0,y=y_0}=\frac{\exp(\beta_0+\beta_1+\beta_2y_0+\beta_3y_0)}{\exp(\beta_0+\beta_2y_0)}=e^{\beta_1+\beta_3y_0}$

$y=y_0+1$: $OR_{x=1,y=y_0+1\text{ vs. }x=0,y=y_0+1}=\frac{\exp(\beta_0+\beta_1+\beta_2y_0+\beta_2+\beta_3y_0+\beta_3)}{\exp(\beta_0+\beta_2y_0+\beta_2)}=e^{\beta_1+\beta_3y_0+\beta_3}$

$\Rightarrow$ 1 unit of increase in y leads to e^{β_3} times increases in odds ratio between x=0 and x=1

§4 Binomial GLM 的 deviance

1. ungrouped data model 的 deviance

对于 $B(\pi)$, $a(y,\pi)=1$, $\phi=1\Rightarrow b(y,\phi)=\exp\{t(y,y)\}$

$\Rightarrow\ell(\hat{\mu})=\sum_{i=1}^n t(y_i,y_i)-\frac{1}{2}D(y,\hat{\mu})$ $\Rightarrow D(y,\hat{\mu})=2\{\ell(\tilde{\mu})-\ell(\hat{\mu})\}$

对于 saturated model: $D(y,\tilde{\mu})=0\Rightarrow\ell(\tilde{\mu})=\sum_{i=1}^n t(y_i,y_i)$

2. grouped data model 的 deviance

对于 $Bin(n,\pi)$, $a(y,\pi)=\binom{n}{y}$, $\phi=1\Rightarrow b(y,\phi)=\binom{n}{y}\exp\{t(y,y)\}$

$\Rightarrow\ell^G(\hat{\mu}^G)=\sum_{c=1}^C t(y_c,y_c)-\frac{1}{2}D^G(y,\hat{\mu}^G)+\sum_{c=1}^C\log\binom{n_c}{y_c}$

对于 saturated model: $D^G(y,\tilde{\mu}^G)=0\Rightarrow\ell^G(\tilde{\mu}^G)=\sum_{c=1}^C t(y_c,y_c)+\sum_{c=1}^C\log\binom{n_c}{y_c}$

$\Rightarrow D^G(y,\hat{\mu}^G)=2\{\ell^G(\tilde{\mu}^G)-\ell^G(\hat{\mu}^G)\}$

3. 两种 model 的 deviance 不同

$\ell^G(\hat{\mu}^G)=\ell(\hat{\mu})+\sum_{c=1}^C\log\binom{n_c}{y_c}\Rightarrow D(y,\hat{\mu})-D^G(y,\hat{\mu}^G)=2[\sum t(y_i,y_i)-\sum t(y_c,y_c)]$

$\begin{cases}t(y_i,y_i)=y_i\log\frac{y_i}{1-y_i}+\log(1-y_i)=\log y_i^{y_i}(1-y_i)^{1-y_i}=0,\ \forall y_i=0\text{ 或 }1\\ t(y_c,y_c)=y_c\log\frac{y_c/n_c}{1-y_c/n_c}+n_c\log(1-y_c/n_c)\neq 0\end{cases}\Rightarrow D(y,\hat{\mu})\neq D^G(y,\hat{\mu}^G)$

4. ungrouped data model 没有 goodness of fit test

对于 ungrouped data model，有

$D(y,\hat{\mu})=-2\sum\{y_i\,\text{logit}(\hat{\mu}_i)+\log(1-\hat{\mu}_i)\}$

$=-2\{\sum_{i:y_i=1}\text{logit}(\hat{\mu}_i)+\sum\log(1-\hat{\mu}_i)\}$

$X^2=\sum_{i=1}^n\frac{(y_i-\hat{\mu}_i)^2}{V(\hat{\mu}_i)}=\sum_{i:y_i=0}\frac{\hat{\mu}_i^2}{V(\hat{\mu}_i)}+\sum_{i:y_i=1}\frac{(1-\hat{\mu}_i)^2}{V(\hat{\mu}_i)}$

3. Logit model for multinomial response

§1 Multinomial distribution

1. Multinomial distribution: $Y|X=(Y_1,Y_2,\cdots,Y_J)\sim Multinomial(m,\pi_1(X),\cdots,\pi_J(X))$

① $Z\in\{1,\cdots,J\}$ ② $\pi_j(X)=P(Z=j|X)$: $\sum_{j=1}^J\pi_j(X)=1$ ③ $Y_j=\sum_{i=1}^m 1\{Z_i=j\}$

2. EDM 形式:

$P(Y,\pi)=\frac{m!}{\prod_{j=1}^J(y_j!)}\prod_{j=1}^J\pi_j^{y_j}=\frac{m!}{\prod_{j=1}^J(y_j!)}\cdot\exp\{\sum_{j=1}^{J-1}y_j\log\frac{\pi_j}{\pi_J}+m\log\pi_J\}$ $\tilde{Y}^T\theta-\kappa(\theta)$

其中 $\tilde{Y}=(Y_1,\cdots,Y_{J-1})^T\in R^{J-1}$, $\theta=(\log\frac{\pi_1}{\pi_J},\log\frac{\pi_2}{\pi_J},\cdots,\log\frac{\pi_{J-1}}{\pi_J})^T\in R^{J-1}$

$\Rightarrow e^{\theta_j}=\frac{\pi_j}{\pi_J}\Rightarrow\sum_{j=1}^{J-1}e^{\theta_j}=\frac{1-\pi_J}{\pi_J}\Rightarrow\pi_J=\frac{1}{1+\sum_{j=1}^{J-1}e^{\theta_j}}$ & $\pi_j=\frac{e^{\theta_j}}{1+\sum_{j=1}^{J-1}e^{\theta_j}}$, $j=1,\cdots,J-1$

$\Rightarrow\kappa(\theta)=m\log(1+\sum_{j=1}^{J-1}e^{\theta_j})\Rightarrow E[Y_j]=m\cdot\frac{e^{\theta_j}}{1+\sum_{j=1}^{J-1}e^{\theta_j}}=m\pi_j$

§2 Nominal responses: Baseline logit model

1. Baseline logit model 的形式: $J-1$ 个 regression $(J-1)\times(p+1)$ 个 coefficients

$\begin{cases}Y_i|X_i\sim Multinomial(m_i,\pi_1(X_i),\pi_2(X_i),\cdots,\pi_J(X_i))\\ \log\frac{\pi_j(X_i)}{\pi_J(X_i)}=\alpha_j+\beta_j^TX_i,\ j=1,\cdots,J-1\end{cases}$ $(\beta_j, X_i\in R^p)$

① 任意一对 response categories 间的 (generalized) logits:

$\log\frac{\pi_{j_1}(X)}{\pi_{j_2}(X)}=\log\frac{\pi_{j_1}(X)}{\pi_J(X)}-\log\frac{\pi_{j_2}(X)}{\pi_J(X)}=(\alpha_{j_1}+\beta_{j_1}^TX)-(\alpha_{j_2}+\beta_{j_2}^TX)$

② Response probability: $\pi_j(X)=\frac{\exp(\alpha_j+\beta_j^TX)}{1+\sum_{h=1}^{J-1}\exp(\alpha_h+\beta_h^TX)}$, $j=1,\cdots,J$ 令 $\alpha_J=0$, $\beta_J=0$

2. Baseline logit model 的求解和 inference

① Likelihood: $\ell_i(\alpha,\beta)=\sum_{j=1}^{J-1}y_{ij}(\alpha_j+\beta_j^Tx_i)-m_i\log(1+\sum_{j=1}^{J-1}\exp(\alpha_j+\beta_j^Tx_i))+\log(\frac{m_i!}{\prod_{j=1}^J(y_{ij}!)})$

$\ell(\alpha,\beta)=\sum_i\{\sum_{j=1}^{J-1}y_{ij}(\alpha_j+\beta_j^Tx_i)-m_i\log(1+\sum_{j=1}^{J-1}\exp(\alpha_j+\beta_j^Tx_i))\}+C$

② Goodness of fit:

1° Chi-squared statistic: $X^2=\sum_i\sum_j\frac{(y_{ij}-\hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}\xrightarrow{d}\chi^2_{N-p}$ 若 CLT

2° Deviance: $D=2\{\ell(\tilde{\mu})-\ell(\hat{\mu})\}\xrightarrow{d}\chi^2_{N-p}$ $\ell(\tilde{\mu})$ 为 saturated model

3. Baseline logit model 的 interpretation β_j interpret 为 OR: X=1 vs. X=0

① $\log(\frac{\pi_j(X)}{\pi_J(X)})=\alpha_j+\beta_jX$ (binary) $\Rightarrow OR_j=\frac{\pi_{j1}}{\pi_{J1}}/\frac{\pi_{j0}}{\pi_{J0}}=e^{\beta_j}$

OR between exposure's presence and absense for response j relative to reference category J

注: $\log(\frac{\pi_j(X)}{\pi_J(X)})=\alpha_j+\beta_{j1}x_1+\cdots+\beta_{jp}x_p\Rightarrow OR_{jk}=\frac{\pi_j|X_k=1}{\pi_J|X_k=1}/\frac{\pi_j|X_k=0}{\pi_J|X_k=0}=e^{\beta_{jk}}$

② 若 $\beta_j=0$, 则 $OR_j=1\Rightarrow$ 表示相应的 exposure factor 没有 effect

③ OR_j 的 95% CI 为: $\exp\{\hat{\beta}_j\pm 1.96\,s.e.(\hat{\beta}_j)\}$

§3 Ordinal response: cumulative logit

1. cumulative logits: cumulative prob: $Pr(Y\leq j|X)=\pi_1(X)+\cdots+\pi_j(X)$

$\Rightarrow\text{logit}[Pr(Y\leq j|X)]=\log\frac{Pr(Y\leq j|X)}{Pr(Y>j|X)}=\log\frac{\sum_{i=1}^j\pi_i(X)}{\sum_{i=j+1}^J\pi_i(X)}$ $j=1,2,\cdots,J-1$

2. Proportional odds model: 只有一组 β，α_j 关于 j 递增

$\text{logit}[Pr(Y\leq j|X)]=\alpha_j+\beta^TX$ $j=1,2,\cdots,J-1$

注: nonproportional odds model: $\text{logit}[Pr(Y\leq j|X)]=\alpha_j+\beta_j^TX$

3. cumulative odds ratio (Proportional odds model 的 intepretation)

$\log\frac{Pr(Y\leq j|X_1)/Pr(Y>j|X_1)}{Pr(Y\leq j|X_2)/Pr(Y>j|X_2)}=\text{logit}[Pr(Y\leq j|X_1)]-\text{logit}[Pr(Y\leq j|X_2)]=\beta^T(X_1-X_2)$

Interpretation: $X=X_1$ 时 response $\leq j$ 的 odds 是 $X=X_2$ 时 response $\leq j$ 的 odds 的 $e^{\beta^T(X_1-X_2)}$ 倍

4. Proportional odds model 的 inference

检验 proportional odds assumption: H_0: proportional vs. H_1: nonproportional

$T_{LR}=2\{\ell_1-\ell_0\}\sim\chi^2_{df_0-df_1}$ (under H_0)

4. Poisson GLM

§1 Poisson GLM

1. Poisson GLM $\begin{cases}y\sim Poisson(\mu)\\ \log\mu=\beta_0+\beta_1x_1+\cdots+\beta_px_p\end{cases}\Longleftrightarrow\mu=e^{\beta_0}(e^{\beta_1})^{x_1}\cdots(e^{\beta_p})^{x_p}$

2. estimate 和 test ① test coefficient: $Var(\hat{\beta})=(X^T\hat{W}X)^{-1}$, $W_{ii}=\mu_i$

② test model sufficiency: $D(y,\hat{\mu})=\sum_i d(y_i,\hat{\mu}_i)=2\sum_i[y_i\log(\frac{y_i}{\hat{\mu}_i})-y_i+\hat{\mu}_i]$

若含 intercept $\Rightarrow\sum_i(y_i-\hat{\mu}_i)=0\Rightarrow D(y,\hat{\mu})=2\sum_i y_i\log(\frac{y_i}{\hat{\mu}_i})\sim\chi^2_{n-p'}$

§2 Poisson linear model: two-way contingency table

1. Notations: $\pi_{ij}=P(A=i,B=j)$, $\mu_{ij}=m\pi_{ij}=E[Y_{ij}]$

2. variables 间独立: $\pi_{ij}=\pi_{i+}\pi_{+j}\Rightarrow\mu_{ij}=m\pi_{i+}\pi_{+j}$

$\Rightarrow\log\mu_{ij}=\log m+\log\pi_{i+}+\log\pi_{+j}=\beta_0+\beta_i^A+\beta_j^B$ $(\beta_1^A=\beta_1^B=0)$

$\Rightarrow$ 共有 $1+(I-1)+(J-1)=I+J-1$ 个 parameters

$L(\mu)=\sum_i\sum_j y_{ij}\log(\mu_{ij})-\sum_i\sum_j\mu_{ij}$

$\Rightarrow\begin{cases}\sum_{i,j}(y_{ij}-\mu_{ij})=0\\ \sum_i(y_{ij}-\mu_{ij})=0, & j=1,\cdots,J\\ \sum_j(y_{ij}-\mu_{ij})=0, & i=1,\cdots,I\end{cases}\Rightarrow\begin{cases}\hat{\mu}_{ij}=\frac{y_{i+}y_{+j}}{m}\\ \hat{\pi}_{i+}=y_{i+}/m\\ \hat{\pi}_{+j}=y_{+j}/m\end{cases}$

3. variables 间有 interaction: $\log\mu_{ij}=\beta_0+\beta_i^A+\beta_j^B+\gamma_{ij}^{AB}$ $(\beta_1^A=\beta_1^B=\gamma_{1j}^{AB}=\gamma_{i1}^{AB}=0)$

$\Rightarrow$ 共有 $1+(I-1)+(J-1)+(I-1)(J-1)=IJ$ 个 parameters (saturated: $\hat{\mu}_{ij}=y_{ij}$)

此时 interaction 可以 interprete 成 odds ratio:

$\log\frac{\pi_{i_1j_1}}{\pi_{i_1j_2}}/\frac{\pi_{i_2j_1}}{\pi_{i_2j_2}}=\log\frac{\mu_{i_1j_1}}{\mu_{i_1j_2}}/\frac{\mu_{i_2j_1}}{\mu_{i_2j_2}}=\gamma_{i_1j_1}^{AB}+\gamma_{i_2j_2}^{AB}-\gamma_{i_1j_2}^{AB}-\gamma_{i_2j_1}^{AB}$

§3 Poisson linear model: three-way contingency table

1. mutual independence (A, B, C) $\pi_{ijk}=\pi_{i++}\pi_{+j+}\pi_{++k}$

$\Rightarrow\begin{cases}\log\mu_{ijk}=\beta_0+\beta_i^A+\beta_j^B+\beta_k^C\\ \beta_1^A=\beta_1^B=\beta_1^C=0\end{cases}\Rightarrow df=IJK-(1+(I-1)+(J-1)+(K-1))=IJK-I-J-K+2$

2. joint independence (A, BC) $\pi_{ijk}=\pi_{i++}\pi_{+jk}$

$\Rightarrow\begin{cases}\log\mu_{ijk}=\beta_0+\beta_i^A+\beta_j^B+\beta_k^C+\gamma_{jk}^{BC}\\ \beta_1^A=\beta_1^B=\beta_1^C=0,\ \gamma_{j1}^{BC}=\gamma_{1k}^{BC}=0\end{cases}\Rightarrow df=IJK-I-J-K+2-(J-1)(K-1)$

3. conditional independence (AC, BC) $\pi_{ijk}=\pi_{i+k}\pi_{+jk}/\pi_{++k}$

$\Rightarrow\begin{cases}\log\mu_{ijk}=\beta_0+\beta_i^A+\beta_j^B+\beta_k^C+\gamma_{ik}^{AC}+\gamma_{jk}^{BC}\\ \beta_1^A=\beta_1^B=\beta_1^C=0,\ \gamma_{i1}^{AC}=\gamma_{1k}^{AC}=\gamma_{j1}^{BC}=\gamma_{1k}^{BC}=0\end{cases}\Rightarrow df=IJK-I-J-K+2-(J-1)(K-1)-(I-1)(K-1)$

4. homogeneous association (AB, AC, BC)

$\Rightarrow\begin{cases}\log\mu_{ijk}=\beta_0+\beta_i^A+\beta_j^B+\beta_k^C+\gamma_{ik}^{AC}+\gamma_{jk}^{BC}+\gamma_{ij}^{AB}\\ \beta_1^A=\beta_1^B=\beta_1^C=0,\ \gamma_{i1}^{AC}=\gamma_{1k}^{AC}=\gamma_{j1}^{BC}=\gamma_{1k}^{BC}=\gamma_{1j}^{AB}=\gamma_{i1}^{AB}=0\end{cases}\Rightarrow df=(I-1)(J-1)(K-1)$

无论第三个 variable 取哪个 level，任意两个 variables 间的 conditional odds ratios 不变

$\log(\frac{\pi_{i_1|j_1k}}{\pi_{i_2|j_1k}}/\frac{\pi_{i_1|j_2k}}{\pi_{i_2|j_2k}})=\log(\frac{\pi_{i_1j_1k}}{\pi_{i_2j_1k}})-\log(\frac{\pi_{i_1j_2k}}{\pi_{i_2j_2k}})=\log(\frac{\mu_{i_1j_1k}}{\mu_{i_2j_1k}})-\log(\frac{\mu_{i_1j_2k}}{\mu_{i_2j_2k}})$

$=[(\beta_{i_1}^A-\beta_{i_2}^A)+(\gamma_{i_1j_1}^{AB}-\gamma_{i_2j_1}^{AB})+(\gamma_{i_1k}^{AC}-\gamma_{i_2k}^{AC})]-[(\beta_{i_1}^A-\beta_{i_2}^A)+(\gamma_{i_1j_2}^{AB}-\gamma_{i_2j_2}^{AB})+(\gamma_{i_1k}^{AC}-\gamma_{i_2k}^{AC})]=\gamma_{i_1j_1}^{AB}-\gamma_{i_2j_1}^{AB}-\gamma_{i_1j_2}^{AB}+\gamma_{i_2j_2}^{AB}$ (与 k 取值无关)

5. saturated model (ABC)

$\log\mu_{ijk}=\beta_0+\beta_i^A+\beta_j^B+\beta_k^C+\gamma_{ik}^{AC}+\gamma_{jk}^{BC}+\gamma_{ij}^{AB}+\gamma_{ijk}^{ABC}\Rightarrow df=0$

After model fitting,

$$\log\mu=5.18+1.16*x_2^A-0.25*x_3^A-0.254*x_2^B-0.5*x_3^B$$
$$+0.38*x_2^A*x_2^B+0.24*x_2^A*x_3^B+0.86*x_3^A*x_2^B$$
$$+1.1*x_3^A*x_3^B$$

where $\forall i\geq 2, x_i^A=\begin{cases}1, A=i\\ 0, otherwise,\end{cases}$ $\forall j\geq 2, x_j^B=\begin{cases}1, B=j\\ 0, otherwise.\end{cases}$

- $i=1, j=3, k=1$

$\log\frac{\hat{P}(B=3|A=1)}{\hat{P}(B=1|A=1)}=\log\frac{\hat{\mu}_{13}}{\hat{\mu}_{11}}=\hat{\beta}_3^B+\underbrace{\hat{\gamma}_{13}^{AB}-(\hat{\beta}_1^B+\hat{\gamma}_{11}^{AB})}_{=0}=\hat{\beta}_3^B=-0.5$

- $i'=3, j=3, k=1$

$\log\frac{\hat{P}(B=3|A=3)}{\hat{P}(B=1|A=3)}=\log\frac{\hat{\mu}_{33}}{\hat{\mu}_{31}}=\hat{\beta}_3^B+\hat{\gamma}_{33}^{AB}-\underbrace{(\hat{\beta}_1^B+\hat{\gamma}_{31}^{AB})}_{=0}$
$=\hat{\beta}_3^B+\hat{\gamma}_{33}^{AB}$
$=0.6$

- The OR of Liberal to Fundamental between people with education $\geq$ *Bachelor* and people with education < *HS* is $\exp(\underbrace{1.1}_{=\hat{\gamma}_{33}^{AB}})=3$, which is the same as the observed OR:

$\frac{y_{11}*y_{33}}{y_{31}*y_{13}}=\frac{178*252}{138*108}=3$

§4 Loglinear 和 logit model 间的转化

接下来考虑 $Y\in\{1,2\}$，即 $I\times 2\times K$ contingency table

1. Loglinear model $\Rightarrow$ logit model (实际上更常使用 $\text{logit}(P(Y=2|X=i,Z=k))$)

$\text{logit}(P(Y=1|X=i,Z=k))=\log\frac{P(Y=1|X=i,Z=k)}{P(Y=2|X=i,Z=k)}=\log\frac{P(Y=1,X=i,Z=k)}{P(Y=2,X=i,Z=k)}=\log(\mu_{i1k})-\log(\mu_{i2k})$

e.g. 以 homogeneous model $\log\mu_{ijk}=\beta_0+\beta_i^X+\beta_j^Y+\beta_k^Z+\gamma_{ik}^{XZ}+\gamma_{jk}^{YZ}+\gamma_{ij}^{XY}$ 为例:

$\text{logit}(P(Y=1|X=i,Z=k))$ $(df=(I-1)(J-1)(K-1)=(I-1)(K-1))$

$=\log(\mu_{i1k})-\log(\mu_{i2k})$

$=(\beta_0+\beta_i^X+\beta_1^Y+\beta_k^Z+\gamma_{ik}^{XZ}+\gamma_{1k}^{YZ}+\gamma_{i1}^{XY})-(\beta_0+\beta_i^X+\beta_2^Y+\beta_k^Z+\gamma_{ik}^{XZ}+\gamma_{2k}^{YZ}+\gamma_{i2}^{XY})$

$=\underbrace{(\beta_1^Y-\beta_2^Y)}_{:=\alpha}+\underbrace{(\gamma_{1k}^{YZ}-\gamma_{2k}^{YZ})}_{:=\gamma_k^Z}+\underbrace{(\gamma_{i1}^{XY}-\gamma_{i2}^{XY})}_{:=\gamma_i^X}$

$=\alpha+\gamma_k^Z+\gamma_i^X$ $(df=I(J-1)K-1-(I-1)-(K-1)=IK-I-K+1)$

e.g. conditional independence model: 两者不等价

2. logit model $\Rightarrow$ Loglinear model

① 利用实际意义转化

e.g. $\text{logit}(P(Y=1|X=i,Z=k))=\alpha+\gamma_i^X$ $(df=I(J-1)K-1-(I-1)=IK-I)$ 表示 $Y\perp Z|X$

$\Longleftrightarrow\log\mu_{ijk}=\beta_0+\beta_i^X+\beta_j^Y+\beta_k^Z+\gamma_{ij}^{XY}+\gamma_{ik}^{XZ}$ $(df=2(I-1)(K-1)+(J-1)(K-1)=IK-I)$

其中 $\alpha=\beta_1^Y-\beta_2^Y$, $\gamma_i^X=\gamma_{i1}^{XY}-\gamma_{i2}^{XY}$

e.g. $\text{logit}(P(Y=1|X=i,Z=k))=\alpha+\gamma_i^X+\gamma_i^Z$ 表示 effect of X on Y $\perp$ effect of X on Z

$\Longleftrightarrow\log\mu_{ijk}=\beta_0+\beta_i^X+\beta_j^Y+\beta_k^Z+\gamma_{ij}^{XY}+\gamma_{ik}^{XZ}+\gamma_{jk}^{YZ}$

② 利用法则: (1) 加上 β_0

(2) 加上 response 和所有出现在 logit model 中的 explanatory variable 之间的 interaction

(3) 加上所有 explanatory variable (无论是否出现在 logit model 中) 之间的 interaction

§5 Overdispersion

$$Y_i|\lambda_i\sim Poisson(\lambda_i)$$
$$\lambda_i\sim Gamma(\mu_i,\phi)$$

It is straightforward to show, under the hierarchical model, that

$$E[Y_i]=\mu_i,\quad var[Y_i]=\mu_i+\phi\mu_i^2$$

where the variance contains an overdisperion term $\phi\mu_i^2$. The larger ϕ, the greater the overdispersion. The PMF of this NB is:

$$P(y_i;\mu_i,\alpha)=\frac{\Gamma(y_i+\alpha)}{\Gamma(y_i+1)\Gamma(\alpha)}(\frac{\mu_i}{\mu_i+\alpha})^{y_i}(1-\frac{\mu_i}{\mu_i+\alpha})^\alpha.$$

where $\alpha=1/\phi$ and and $\Gamma()$ is the gamma function, so that $var[Y_i]=\mu_i+\mu_i^2/\alpha$.

5. Gamma GLM

§1 Gamma GLM

1. Gamma 分布的分布函数 $P(y;\alpha,\beta)=\frac{1}{\Gamma(\alpha)\beta^\alpha}y^{\alpha-1}e^{-\frac{y}{\beta}}\Rightarrow\begin{cases}E[y]=\alpha\beta\\ Var(y)=\alpha\beta^2\end{cases}$

$\Longleftrightarrow P(y;\mu,\phi)=\frac{1}{y\Gamma(\frac{1}{\phi})}\cdot(\frac{y}{\phi\mu})^{1/\phi}\exp\{-\frac{y}{\phi\mu}\}$ $(\alpha=1/\phi, \beta=\mu\phi)$

$=(\frac{y}{\phi})^{1/\phi}\frac{1}{\Gamma(1/\phi)}\cdot\frac{1}{y}\cdot\exp\{\frac{y(-\frac{1}{\mu})-\log\mu}{\phi}\}$ canonical link 为 $g(\mu)=-\frac{1}{\mu}$

2. Gamma 分布的性质

① coefficient of variation (CV) 为: $CV=\frac{Var(Y)}{E[Y]^2}$

② Gamma 分布的 CV 不取决于 mean: $\frac{Var(Y)}{E[Y]^2}=\frac{\alpha\beta^2}{(\alpha\beta)^2}=\frac{1}{\alpha}=\phi$

$\Rightarrow\log(var(y))=2\log\mu+\log\phi$

3. Gamma 分布的 deviance

① unit deviance 为: $d(y,\mu)=2\{-\log(\frac{y}{\mu})+\frac{y-\mu}{\mu}\}$

② scaled residual deviance 为: $\tilde{D}(y,\hat{\mu})=\frac{D(y,\hat{\mu})}{\phi}=\sum_i d(y_i,\hat{\mu}_i)/\phi\sim\chi^2_{n-p'}$

4. Gamma 分布的 dispersion estimate

① mean deviance estimator: $\tilde{\phi}=\frac{D(y,\hat{\mu})}{n-p'}$ ② Pearson: $\bar{\phi}=\frac{1}{n-p'}\cdot\sum_{i=1}^n\frac{(y_i-\hat{\mu}_i)^2}{\hat{\mu}_i^2}$

§2 Inverse Gaussian GLM

1. Inverse Gaussian 分布的分布函数 $P(y;\mu,\phi)=(2\pi y^3\phi)^{-\frac{1}{2}}\exp\{-\frac{1}{2\phi}\frac{(y-\mu)^2}{y\mu^2}\}$

$=(2\pi y^3\phi)^{-\frac{1}{2}}\exp\{-\frac{1}{2\phi y}\}\cdot\exp\{\frac{y(-\frac{1}{2\mu^2})+\frac{1}{\mu}}{\phi}\}$ canonical link 为 $g(\mu)=-\frac{1}{2\mu^2}$

2. Inverse Gaussian 分布的性质 ① $Var(Y)=\phi\mu^3$ ② $V(\mu)=\mu^3$

3. Inverse Gaussian 分布的 deviance

① unit deviance 为: $d(y,\mu)=\frac{(y-\mu)^2}{y\mu^2}$

② scaled residual deviance 为: $\tilde{D}(y,\hat{\mu})=\frac{D(y,\hat{\mu})}{\phi}=\sum_i d(y_i,\hat{\mu}_i)/\phi\sim\chi^2_{n-p'}$ (exactly)

4. Inverse Gaussian 分布的 dispersion estimate

① mean deviance estimator: $\tilde{\phi}=\frac{D(y,\hat{\mu})}{n-p'}$ ② Pearson: $\bar{\phi}=\frac{1}{n-p'}\cdot\sum_{i=1}^n\frac{(y_i-\hat{\mu}_i)^2}{\hat{\mu}_i^3}$

5. Testing nested model: $F=\frac{[D(y,\hat{\mu}_{M_0})-D(y,\hat{\mu}_{M_1})]/(df_0-df_1)}{\tilde{\phi}}\sim F(df_0-df_1, df_1)$

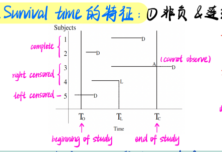
Under H_0: M_0 为 sufficient 其中 $\tilde{\phi}$ 为 M_1 的 estimate of dispersion

6. Survival Analysis

§1 关于 survival analysis

1. Survival time：从一个 starting point 到某个 event（被称为"failure"）发生的时间 Y

2. Survival time 的特征：① 非负 & 连续 ② Right skewed ③ Censoring



T_i: observed time
Y_i: true survive time
$\delta_i: 1\{Y_i \le C_i\} = \begin{cases} 1 & \text{uncensored} \\ 0 & \text{censored} \end{cases}$
C_i: censoring time
（仅考虑 right censored）
此处 $T_1 = Y_1$, $T_2 = Y_2$, $T_3 = Y_3$, $T_4 = C_4$, $T_5 = Y_5$
$C_3 = C_4 = C_5 = T_2$
$C_4 = T_2$

§2 Survival function 与 hazard function 令随机变量 Y 表示 survival time，$f(y)$ 表示其 density function

1. Survival function $S(y) = Pr(Y \ge y) = \int_y^{+\infty} f(t)\,dt$

2. Hazard function：$h(y) = \lim_{\delta y \to 0} \frac{Pr(y \le Y \le y+\delta y \mid Y > y)}{\delta y}$ Hazard function 可被改写为：

① $h(y) = \lim_{\delta y \to 0} \frac{Pr(y \le Y \le y+\delta y)}{\delta y} \cdot \frac{1}{Pr(Y>y)} = \frac{f(y)}{S(y)}$

② $h(y) = -\frac{d}{dy}\{\log[S(y)]\}$ （$S'(y) = \frac{d}{dy}\int_y^{+\infty} f(t)\,dt = -f(y)$）

3. cumulative harzard function $H(y) = \int_0^y h(t)\,dt \iff h(y) = \frac{d}{dy}H(y)$

因此有 $H(y) = -\log[S(y)] \iff S(y) = \exp(-H(y))$

§3 Exponential distribution, Weibull distribution, 与 proportional hazards model

1. Exponential distribution ① PDF 与 CDF：$f(y;\lambda) = \lambda e^{-\lambda y} \Rightarrow F(y;\lambda) = 1 - e^{-\lambda y}$

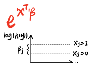
② Survival 与 hazard function：$S(y;\lambda) = e^{-\lambda y}$（表明 median survival 为 $y(50) = \frac{1}{\lambda}\log 2$）

$\Rightarrow h(y;\lambda) = \frac{f(y;\lambda)}{S(y;\lambda)} = \lambda \quad \Rightarrow H(y;\lambda) = \lambda y$

③ 检验 exponential model 的合理性：plot $\hat{H}(y)$ v.s. y（$\hat{H}(y) = -\log[\hat{S}(y)]$ 应 linear）

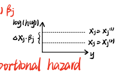
④ 模型构建：$\log(\lambda) = \beta_0 + \sum_{j=1}^p \beta_j x_j \Rightarrow h(y;\lambda) = \lambda = e^{\beta_0} \cdot e^{x^T\beta}$

⑤ Interpretation：(1) Binary explanatory variables：$\frac{h(y|x_j=1)}{h(y|x_j=0)} = e^{\beta_j}$



为此时 $x_j = 0$（exposure absence）v.s. $x_j = 1$（exposure presence）的 hazard ratio

(2) Continous explanatory variables：$\frac{h(y|x_j=x_j^{(1)})}{h(y|x_j=x_j^{(2)})} = e^{(x_j^{(1)} - x_j^{(2)})\beta_j}$



为此时 $x_j = x_j^{(1)}$ v.s. $x_j = x_j^{(2)}$ 的 hazard ratio

注：由于 $\log(h(y)) \propto x_j$（fix 其他 x_i, $i \ne j$），因此也被称为 proportional hazard

2. Weibull distribution ① PDF：$f(y;\alpha,\gamma) = \frac{\alpha}{\gamma}\left(\frac{y}{\gamma}\right)^{\alpha-1}\exp\{-(\frac{y}{\gamma})^\alpha\}$

令 $\phi = \gamma^{-\alpha}$，则 $f(y;\alpha,\gamma) = \alpha\phi y^{\alpha-1}\exp\{-\phi y^\alpha\}$，$y \ge 0, \alpha > 0, \phi > 0$

② Survival 与 hazard function：$S(y;\alpha,\gamma) = \int_y^{+\infty} \alpha\phi u^{\alpha-1}\exp\{-\phi u^\alpha\}\,du = \exp\{-\phi y^\alpha\}$

$\Rightarrow h(y;\alpha,\gamma) = \alpha\phi y^{\alpha-1} \Rightarrow H(y;\alpha,\gamma) = \phi y^\alpha$ （$y(50) = \phi^{-\frac{1}{\alpha}}(\log 2)^{\frac{1}{\alpha}}$）

（表明若 $\alpha \ne 1$，则 $h(y;\alpha,\gamma)$ 会随着 y 的增加而增加/减少 $\Rightarrow$ accelerated failure time model）

③ 检验 Weibull model 的合理性：plot $\log(\hat{H}(y))$ v.s. $\log(y)$（应 linear）

④ 模型构建：$\log(\phi) = \beta_0 + \sum_{j=1}^p \beta_j x_j \Rightarrow h(y;\alpha,\gamma) = \alpha\phi y^{\alpha-1} = \underbrace{\alpha y^{\alpha-1}e^{\beta_0}}_{=:h_0(y)} \cdot e^{x^T\beta}$

⑤ Interpretation：同 exponential

3. Proportional hazards model：$h(y) = h_0(y)e^{x^T\beta}$，$h_0(y)$ 为 baseline hazard

$\Rightarrow H(t) = \int_0^y h(t)\,dt = H_0(y)e^{x^T\beta}$（$H_0(y) := \int_0^y h_0(y)\,dt$）

$\Rightarrow \log(H(t)) = \log(H_0(y)) + x^T\beta$

注：① 若仅有一个 covariate 且 X 为 binary，则 $\log(H_P(y)) = \log(H_A(y)) + \beta$

② 若要检验 Proportional hazards model 的合理性，可以

(1) plot $\log(\hat{H}(t)) = \log(-\log(\hat{S}(y))) = \log(H_0(y)) + x^T\beta$，关于每个 x_j 为 linear

(2) plot $\begin{cases} \log(-\log(\hat{S}_A(y))) & \text{v.s.} \quad \log(y) \\ \log(-\log(\hat{S}_B(y))) & \text{v.s.} \quad \log(y) \end{cases}$ 两者平行

§4 Empirical survival function

1. Empirical survival function：$\tilde{S}(t) = \frac{\text{subjects with survival time} \ge t \text{ 的总数}}{\text{subjects 的总数}}$

2. Kaplan-Meier estimate（product limit estimate）$t_{(1)} \le t_{(2)} \le \cdots \le t_{(J)}$

① 找出所有 observed survival time（将 censored time 从 $t_1, \cdots, t_n$ 中剔除后剩下的），并升序排列：

② 找出每个 time $t_{(j)}$ 时 failures/events 的发生数（要剔除 censored times）：d_j, $j = 1, \ldots, J$

③ 找出每个 time $t_{(j)}$ 及以后还存活的 subjects 数（不需要剔除 censored times）：n_j, $j = 1, \ldots, J$

④ 求出 the estimated probability of survival past $t_{(j)}$：$P(\text{passing } t_{(j)}) = \frac{n_j - d_j}{n_j}$

⑤ Kaplan-Meier estimate of the survival function at time t 为：$\hat{S}(t) = \prod_{j: t_{(j)} \le t} \left(\frac{n_j - d_j}{n_j}\right)$

注：① $\hat{S}(t)$ 为一个左连续的函数 ② control group 一般没有 censored times，

$n_j - d_j = n_{j+1}$，$\hat{S}(t)$ 可直接估计为：$\frac{\text{subjects with survival time} > t \text{ 的总数}}{\text{subjects 的总数}}$

eg. ① Control group：

$\{t_{(1)}, \cdots, t_{(J)}\} = \{1, 2, 3, 4, 5, 8, 11, 12, 15, 17, 22, 23\}$

For $\forall t \in [0, 1)$，$\hat{S}(t) = 1$

$\forall t \in [1, 2)$，$\hat{S}(t) = \frac{21-2}{21} = \frac{19}{21}$

$\forall t \in [2, 3)$，$\hat{S}(t) = \frac{19}{21} \cdot \frac{19-2}{19} = \frac{17}{21} \cdots$

Controls

1	1	2	2	3	4	4	5	5	8	8
8	8	11	11	12	12	15	17	22	23	

Treatment

6	6	6	6*	7	9*	10	10*	11*	13	16
17*	19*	20*	22	23	25*	32*	32*	34*	35*	

* indicates censoring

② Treatment group：$\{t_{(1)}, \cdots, t_{(J)}\} = \{6, 7, 10, 13, 16, 22, 23\}$

For $\forall t \in [0, 6)$，$\hat{S}(t) = 1$ $\cdots$ $\forall t \in [23, +\infty)$，$\hat{S}(t) = \frac{18}{21} \cdot \frac{16}{17} \cdot \frac{14}{15} \cdot \frac{11}{12} \cdot \frac{10}{11} \cdot \frac{6}{7} \cdot \frac{6-1}{6}$

$\forall t \in [6, 7)$，$\hat{S}(t) = \frac{21-3}{21} = \frac{18}{21}$

$\forall t \in [7, 10)$，$\hat{S}(t) = \frac{18}{21} \cdot \frac{17-1}{17} = \frac{18}{21} \cdot \frac{16}{17}$

§5 Estimation, inference, 和 checking $\delta_j = 1\{\text{survival time 为 uncensored}\} = 1\{Y_j \le C_j\}$

令：Y_j 为 true time to failure，C_j 为 time to censoring，$T_j = \min\{Y_j, C_j\}$ 为 recorded data，

1. Likelihood 令 $Y_j \overset{i.i.d.}{\sim} f_Y(y_j)$，$C_j \overset{i.i.d.}{\sim} f_C(C_j)$，则

① uncensored 的 single likelihood：$P(T \le t, \delta = 1) = P(Y \le t, Y \le C) = \int_0^t f_Y(y) \cdot S_C(y)\,dy$

$\Rightarrow f(T = t, \delta = 1) = f_Y(t) \cdot S_C(t)$

② censored 的 single likelihood：$P(T \le t, \delta = 0) = P(C \le t, Y > C) = \int_0^t f_C(c) \cdot S_Y(c)\,dc$

$\Rightarrow f(T = t, \delta = 0) = f_C(t) \cdot S_Y(t)$

$L = \prod_{i=1}^n [f(T_i = t_i, \delta_i = 1)]^{1\{\delta_i = 1\}} \cdot [f(T_i = t_i, \delta_i = 0)]^{1\{\delta_i = 0\}} \propto \prod_{i=1}^n [f_Y(t_i)^{\delta_i} \cdot S_Y(t_i)^{1-\delta_i}]$

$\Rightarrow l \propto \sum_{i=1}^n \{\delta_i \log \frac{f_Y(t_i)}{S_Y(t_i)} + \log S_Y(t_i)\} \Rightarrow l_Y = \sum_{i=1}^n \{\delta_i \log h_Y(t_i) + \log S_Y(t_i)\}$

2. Exponential distribution 的 estimation

$\begin{cases} l_Y(\lambda) = \sum_{i=1}^n \{\delta_i \log \lambda_i - \lambda_i t_i\} = \underline{\sum_{i=1}^n \{\delta_i \log \lambda_i t_i - \lambda_i t_i\}} - \sum_{i=1}^n \delta_i \log t_i \\ \log \lambda_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \end{cases}$ fit Poisson regression on censoring indicator $D_i = \delta_i$

$\iff \begin{cases} D_i \overset{ind.}{\sim} Poi(\mu_i) \\ \log \mu_i = \underline{\log(t_i)} + \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \end{cases}$

3. Weibull distribution 的 estimation

$\begin{cases} l_Y(\lambda) = \sum_{i=1}^n \{\delta_i \log(\alpha\phi_i t_i^{\alpha-1}) - \phi_i t_i^\alpha\} = \underline{\sum_{i=1}^n \{\delta_i \log(\phi_i t_i^\alpha) - \phi_i t_i^\alpha\}} + \sum_{i=1}^n \delta_i \log \alpha - \sum_{i=1}^n \delta_i \log t_i \\ \log \lambda_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \end{cases}$ fit Poisson regression on censoring indicator $D_i = \delta_i$

$\iff \begin{cases} D_i \overset{ind.}{\sim} Poi(\mu_i) \\ \log \mu_i = \underline{\alpha \log(t_i)} + \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \end{cases}$

注：fit 思路：① Initialize $\alpha^{(0)}$ ($=1$) ② Fix α，通过 fit Poisson GLM 来得到 β

③ Fix β，求出 α ④ 重复 ② ③ 步直到收敛

$l_Y = \sum_{i=1}^n \{\delta_i \log(\phi_i t_i^\alpha) - \phi_i t_i^\alpha\} + \sum_{i=1}^n \delta_i \log \alpha - \sum_{i=1}^n \delta_i \log t_i$

$\Rightarrow \frac{\partial l_Y}{\partial \alpha} = \sum_{i=1}^n \{\delta_i \log t_i - \phi_i t_i^\alpha \log t_i\} + \sum_{i=1}^n \frac{\delta_i}{\alpha} = \frac{m}{\alpha} + \sum_{i=1}^n [(\delta_i - \mu_i)\log t_i] = 0$

$\Rightarrow \hat{\alpha} = \frac{m}{\sum_{i=1}^n [(\hat{\mu}_i - \delta_i)\log t_i]}$ 其中 m 为非零的 δ 数，即 uncensored observations 数

4. Inference 对 nested model 的 test

M_1: p parameters, $\hat{l}_1$ v.s. M_0: $q < p$ parameters, $\hat{l}_0$

$\Rightarrow D = 2(\hat{l}_1 - \hat{l}_0) \sim \chi^2_{p-q}$ Under H_0: M_0 is sufficient

7. GEE 和 GLMM

§1 Generalized estimating equation (GEE)

1. GEE 的模型假设

① $\begin{cases} \mu_{ij} := E[Y_{ij}] \\ g(\mu_{ij}) = x_{ij}^T\beta \end{cases}$ ② $Var(Y_{ij}) = \phi V(\mu_{ij})$ ③ $Corr(Y_i) = R_i(\alpha)$ 为 working correlation matrix

$\rho_{jk}^{(i)}(\alpha) = corr(Y_{ij}, Y_{ik})$ 为关于 unknown α 的 known function

2. $Corr(Y_i)$ 与 $Var(Y_i)$ 间的关系 $A_i = diag(V(\mu_{i1}), \cdots, V(\mu_{in_i}))$

$corr(Y_{ij}, Y_{ik}) = \frac{cov(Y_{ij}, Y_{ik})}{\sigma_{ij} \cdot \sigma_{ik}} = \rho_{jk}^{(i)}(\alpha) \Rightarrow Var(Y_i) = \phi A_i^{1/2} \cdot R_i(\alpha) \cdot A_i^{1/2}$

3. Working correlation 的种类：① Independent：$\rho_{jk}^{(i)} = 0$ ② Unstructured：$\rho_{jk}^{(i)} = \alpha_{jk}$

③ Exchangeable：$\rho_{jk}^{(i)} = \alpha$ ④ m-dependent：$\rho_{jk}^{(i)} = \alpha_{|j-k|}$ for $|j-k| \le m$ and 0 otherwise

⑤ Autoregressive of order 1：$Corr(Y_{ij}, Y_{ik}) = \alpha^{|j-k|}$ ⑥ Markov：$Corr(Y_{ij}, Y_{ik}) = \alpha^{|j-k|}$

4. GEE 的 estimation：$U(\beta) = \sum_{i=1}^M D_i^T V_i^{-1}(Y_i - \mu_i) = 0$，$D_i = \frac{\partial \mu_i}{\partial \beta^T} \in R^{n_i \times p}$

§2 Generalized linear mixture model (GLMM) $\iff D^T V^{-1}(Y - \mu) = 0$

1. GLMM 的模型假设 ① $Y_{ij}|b_i \overset{ind.}{\sim} EDM(b_i)$ ② $b_i \overset{ind.}{\sim} MVN(0, D)$

③ $\begin{cases} \mu_{ij} := E[Y_{ij}|b_i] \\ g(\mu_{ij}) = x_{ij}^T\beta + z_{ij}^T b_i \end{cases}$ x_{ij}：cluster i 在 t_{ij} 时刻的 fixed effect；z_{ij}：cluster i 在 t_{ij} 时刻的 random effect（z_{ij} 是 x_{ij} 的 subset）

2. linear mixture model (LMM)

① data normally distributed：$Y_{ij} = x_{ij}^T\beta + z_{ij}^T b_i + \varepsilon_{ij}$ $\varepsilon_i \overset{ind}{\sim} MVN(0, \Sigma_i)$

$\Rightarrow g(\cdot)$ 为 identical link：$E[Y_{ij}|b_i] = x_{ij}^T\beta + z_{ij}^T b_i$

② $b_i \overset{ind.}{\sim} MVN(0, D)$ 且 $b_i \perp \varepsilon_i$

3. GLMM 的性质：$cov(Y_{ij}, Y_{ik}) = E[Y_{ij}Y_{ik}] - E[Y_{ij}]E[Y_{ik}]$

$= E_{b_i}\{E[Y_{ij}|b_i] \cdot E[Y_{ik}|b_i]\} - E_{b_i}\{E[Y_{ij}|b_i]\}E_{b_i}\{E[Y_{ik}|b_i]\}$

$= E_{b_i}\{g^{-1}(x_{ij}^T\beta + z_{ij}^T b_i) \cdot g^{-1}(x_{ik}^T\beta + z_{ik}^T b_i)\} - E_{b_i}\{g^{-1}(x_{ij}^T\beta + z_{ij}^T b_i)\} \cdot E_{b_i}\{g^{-1}(x_{ik}^T\beta + z_{ik}^T b_i)\}$

4. LMM 的例子：Clustered data

$Y_{ij} = \mu + a_i + \varepsilon_{ij}$，$a_i \overset{i.i.d.}{\sim} N(0, \sigma_a^2)$，$\varepsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_e^2)$，$a_i \perp \varepsilon_{ij}$

① $E[Y_{ij}] = \mu$ ② $Var[Y_{ij}] = \sigma_a^2 + \sigma_e^2$ ③ $cov(Y_{ij}, Y_{ik}) = \sigma_a^2$

5. LMM 的例子：longitudinal data

$Y_{ij} = \beta_0 + a_i + (\beta_1 + b_i)t_{ij} + \varepsilon_{ij}$，$a_i \overset{i.i.d.}{\sim} N(0, \sigma_a^2)$，$b_i \overset{i.i.d.}{\sim} N(0, \sigma_b^2)$，$\varepsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_e^2)$

① $E[Y_{ij}] = \beta_0 + \beta_1 t_{ij}$ ② $Var[Y_{ij}] = var(a_i) + t_{ij}^2 var(b_j) + var(e_{ij}) = \sigma_a^2 + t_{ij}^2\sigma_b^2 + \sigma_e^2$

③ $cov(Y_{ij}, Y_{ik}) = cov(a_i + b_i t_{ij} + \varepsilon_{ij}, a_i + b_i t_{ik} + \varepsilon_{ik}) = \sigma_a^2 + t_{ij}t_{ik}\sigma_b^2$

6. GLMM 的 estimation：以 Bernoulli 为例

① general form：$P(Y_i) = \int_{b_i} [\prod_{j=1}^{n_i} P(Y_{ij}|b_i)] \cdot f(b_i|D)\,db_i$

$\Rightarrow L(\beta, D) = \prod_{i=1}^M \{\int_{b_i} [\prod_{j=1}^{n_i} P(Y_{ij}|b_i)] \cdot f(b_i|D)\,db_i\}$

② Bernoulli distribution：$P(Y_{ij}|b_i) = (e^{x_{ij}^T\beta + b_i})^{y_{ij}} \left(\frac{1}{1 + e^{x_{ij}^T\beta + b_i}}\right)$

§3 Conditional effect 和 marginal effect 的比较

1. Conditional effect 和 marginal effect 的比较

GLMM 中的 β：$E[Y_{ij}|b_i] = g^{-1}(x_{ij}^T\beta + z_{ij}^T b_i)$

$\Rightarrow E[Y_{ij}] = E[E[Y_{ij}|b_i]] = \int g^{-1}(x_{ij}^T\beta + z_{ij}^T b_i) \cdot f(b_i|D)\,db_i$

GEE 中的 β^*：$E[Y_{ij}] = g^{-1}(x_{ij}^T\beta)$ $\Rightarrow \beta = \beta^*$

① 若 $g(\cdot)$ 为 identity link：$E[Y_{ij}] = \int (x_{ij}^T\beta + z_{ij}^T b_i) \cdot f(b_i|D)\,db_i = x_{ij}^T\beta$

② 若 $g(\cdot)$ 为 non-linear，此时

GEE 的 population-averaged effects < GLMM 的 cluster-specific effects

8. Bayesian Analysis

Definition (Semi-conjugate prior)

Suppose $\theta = (\theta_1, \theta_2)^T$, the semi-conjugate prior is a product of priors,

$$\pi(\theta) = \pi(\theta_1)\pi(\theta_2)$$

each of which, conditioned on the other parameters, is conjugate to $p(y|\theta)$.

The Semi-conjugate prior for (β, σ^2) in normal linear regression is

$$\beta \sim MVN(\beta_0, \Sigma_0)$$

$$\sigma^2 \sim Inverse-gamma\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right).$$

Theorem (Semi-conjugate prior)

For normal linear regression, with previously defined semi-conjugate prior, it can be shown that

$$\beta|y, X, \sigma^2 \sim MVN(\beta_n, V_{\beta_n})$$

$$\sigma^2|y, X, \beta \sim Inverse-Gamma\left(\frac{\nu_0 + n}{2}, \frac{\nu_0\sigma_0^2 + SSR(\beta)}{2}\right)$$

where $\beta_n = (\Sigma_0^{-1} + \frac{X^TX}{\sigma^2})^{-1}(\Sigma_0^{-1}\beta_0 + \frac{X^Ty}{\sigma^2})$, $V_{\beta_n} = (\Sigma_0^{-1} + \frac{X^TX}{\sigma^2})^{-1}$, and $SSR(\beta) = (y - X\beta)^T(y - X\beta)$.

Proof. By plugging in $\beta_0 = 0$ and $\Sigma_0^{-1} = \frac{1}{g\sigma^2}X^TX$, we have

$\beta|y, X, \sigma^2 \sim MVN\left(\left(\frac{1}{g\sigma^2}X^TX + \frac{1}{\sigma^2}X^TX\right)^{-1}\frac{1}{\sigma^2}X^Ty, \left(\frac{1}{g\sigma^2}X^TX + \frac{1}{\sigma^2}X^TX\right)^{-1}\right)$

$\Rightarrow \beta|y, X, \sigma^2 \sim MVN\left(\frac{g}{g+1}(X^TX)^{-1}X^Ty, \frac{g\sigma^2}{g+1}(X^TX)^{-1}\right)$

Similarly, we have

$\tilde{\beta}|y, X, \sigma^2 \sim MVN\left(\frac{g}{g+1}(X^TX)^{-1}X^Ty, \frac{g\sigma^2}{g+1}(X^TX)^{-1}\right)$

$\Rightarrow H\tilde{\beta}|y, X, \sigma^2 \sim MVN\left(\frac{g}{g+1}H^{-1}(X^TX)^{-1}X^Ty, \frac{g\sigma^2}{g+1}H^{-1}(X^TX)^{-1}(H^T)^{-1}\right)$

$\Rightarrow H\tilde{\beta}|y, X, \sigma^2 \sim MVN\left(\frac{g}{g+1}(X^TX)^{-1}X^Ty, \frac{g\sigma^2}{g+1}(X^TX)^{-1}\right)$

Thus, the posterior distribution of β is the same as that of $H\tilde{\beta}$ under the g-prior.

$p(y|X, \sigma^2) = \int f(y|X, \beta, \sigma^2)p(\beta|X, \sigma^2)\,d\beta$

$= \int (2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)\right]$

$\cdot |2\pi g\sigma^2(X^TX)^{-1}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2g\sigma^2}\beta^TX^TX\beta\right]d\beta$

$= \int (2\pi\sigma^2)^{-n/2} \cdot |2\pi g\sigma^2(X^TX)^{-1}|^{-\frac{1}{2}}$

$\cdot \exp\left[-\frac{1}{2\sigma^2}\left(\frac{g+1}{g}\beta^TX^TX\beta - 2y^TX\beta + y^Ty\right)\right]d\beta$

$= \int (2\pi\sigma^2)^{-n/2} \cdot |2\pi g\sigma^2(X^TX)^{-1}|^{-\frac{1}{2}}$

$\cdot \exp\left[-\frac{1}{2}(\beta - m)^TV^{-1}(\beta - m) + \frac{1}{2}m^TV^{-1}m - \frac{1}{2\sigma^2}y^Ty\right]d\beta$

$= (2\pi\sigma^2)^{-n/2} \cdot (g+1)^{-\frac{p}{2}}\exp\left[\frac{1}{2}m^TV^{-1}m - \frac{1}{2\sigma^2}y^Ty\right]$

$\cdot \int |2\pi V|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\beta - m)^TV^{-1}(\beta - m)\right]d\beta$

$= (2\pi\sigma^2)^{-n/2} \cdot (g+1)^{-\frac{p}{2}}\exp\left[\frac{1}{2}m^TV^{-1}m - \frac{1}{2\sigma^2}y^Ty\right]$

$= (2\pi\sigma^2)^{-n/2} \cdot (g+1)^{-\frac{p}{2}}\exp\left[-\frac{SSR_g}{2\sigma^2}\right]$

where $V = \frac{g}{g+1}\sigma^2(X^TX)^{-1}$, $m = \frac{g}{g+1}(X^TX)^{-1}X^Ty$ and

$SSR_g = y^Ty - m^TV^{-1}m = y^T(I - \frac{g}{g+1}X(X^TX)^{-1}X^T)y$.

Note that the term SSR_g decreases to $SSR = y^T(I - X(X^TX)^{-1}X^T)y$ as $g \to \infty$.

Then, combining the prior term with $p(y|X, \sigma^2)$, we have

$p(\sigma^2|y, X) \propto (\sigma^2)^{-n/2}(g+1)^{-p/2}\exp(-\frac{SSR_g}{2\sigma^2}) \cdot (\sigma^2)^{-\nu_0/2-1}\exp(-\frac{\nu_0\sigma_0^2}{2\sigma^2})$

$= (\sigma^2)^{-(\nu_0+n)/2-1}\exp(-\frac{\nu_0\sigma_0^2 + SSR_g}{2\sigma^2})$

Thus,

$\sigma^2|y, X \sim \text{inverse-gamma}(\frac{\nu_0 + n}{2}, \frac{\nu_0\sigma_0^2 + SSR_g}{2})$

Definition (Unit information prior)

If a prior distribution contains the same amount of information as that would be contained in only a single observation, then the prior distribution is called a unit information prior.

The unit informative prior for (β, σ^2) in LM is

$$\beta|\sigma^2 \sim MVN(\hat{\beta}_{OLS}, n\sigma^2(X^TX)^{-1})$$

$$\sigma^2 \sim Inverse-Gamma(\frac{1}{2}, \frac{\hat{\sigma}^2_{OLS}}{2})$$

where $\hat{\beta}_{OLS} = (X^TX)^{-1}X^Ty$, an estimate of β_0 based on the marginal mean of y, and $\hat{\sigma}^2_{OLS} = \frac{SSR(\hat{\beta}_{OLS})}{n-p}$.

The conjugate prior of (β, σ^2) is

$$\beta|\sigma^2 \sim MVN(0, g\sigma^2(X^TX)^{-1})$$

$$\sigma^2 \sim Inverse-Gamma(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2})$$

Above prior for β is a g-prior.

Definition (g-prior)

In normal LM, the g-prior for β is a MVN with mean as a hyperparameter β_0 and covariance matrix proportional to $\sigma^2(X^TX)^{-1}$, i.e.,

$$\beta|\sigma^2 \sim MVN(\beta_0, g\sigma^2(X^TX)^{-1})$$

Theorem

Posterior of β based on g-prior is invariant to changes in the scale of regressors.

$$\beta|y, X, \sigma^2 \sim MVN(\frac{g}{g+1}(X^TX)^{-1}X^Ty, \frac{g\sigma^2}{g+1}(X^TX)^{-1})$$

$$\sigma^2|y, X \sim IG(\frac{\nu_0 + n}{2}, \frac{\nu_0\sigma_0^2 + y^T[I_n - \frac{g}{g+1}X(X^TX)^{-1}X]y}{2})$$

Gibbs sampling For $(s+1)$ iteration (given $z^{(s)}, \beta^{(s)}, \sigma^{2(s)}$)

① For each j，$z_j^{(s+1)} \sim P(z_j|Y, X, z_{-j}^{(s)})$

$P(z_j = 1|Y, X, z_{-j}^{(s)}) = \frac{P(Y|X, z_j = 1, z_{-j})}{P(Y|X, z_j = 1, z_{-j}) + P(Y|X, z_j = 0, z_{-j})}$

$P(Y|X, z) = \int_{\sigma^2}\int_\beta P(Y|X, z, \beta, \sigma^2) \cdot \pi(\beta|X, z, \sigma^2) \cdot \pi(\sigma^2)\,d\beta\,d\sigma^2$

② $\sigma^{2(s+1)}|Y, X, z \sim IG(\frac{\nu_0 + n}{2}, \frac{\nu_0\sigma_0^2 + SSR_g(z)}{2})$

$P(\sigma^2|Y, X, z) = \frac{P(Y|X, z, \sigma^2) \cdot \pi(\sigma^2)}{\int_{\sigma^2} P(Y|X, z, \sigma^2) \cdot \pi(\sigma^2)\,d\sigma^2}$

$P(y|X, z, \sigma^2) = \int_\beta P(Y|X, z, \beta, \sigma^2) \cdot \pi(\beta|X, z, \sigma^2)\,d\beta$

③ update β：$\beta^{(s+1)}|X, z^{(s+1)}, \sigma^{2(s+1)}, y \sim MVN(\frac{g}{g+1}(X_z^TX_z)^{-1}X_z^Ty, \frac{g\sigma^2}{g+1}(X_z^TX_z)^{-1})$

若 $P(z_j = 1|X, z) = \frac{1}{S}\sum_{s=1}^S 1\{z_j^{(s)} = 1\} > P_0$，then x_j will be included，$\hat{\beta} = \frac{1}{S}\sum_{s=1}^S \beta^{(s)}$

· Metropolis algorithm.

Given $\{\theta^{(1)}, \cdots, \theta^{(s)}\}$.

① Sample $\theta^* \sim \mathcal{Z}(\theta|\theta^{(s)})$，$\mathcal{Z}(\theta|\theta^{(s)})$ is a symmetric proposal dist.

② Calculate $\gamma = \min\left\{\frac{P(\theta^*|y)}{P(\theta^{(s)}|y)}, 1\right\}$

③ If $\gamma = 1$，$\theta^{(s+1)} = \theta^*$。

If $\gamma < 1$，sample $u \sim U(0, 1)$，then $\theta^{(s+1)} = \begin{cases} \theta^*, & \text{if } u \le \gamma \\ \theta^{(s)}, & \text{if } u > \gamma \end{cases}$

MH-algorithm.

① $\theta_1^* \sim \mathcal{Z}_1(\theta_1|\theta_1^{(s)}, \theta_2^{(s)})$，may not be symmetric

② $\gamma_1 = \min\left\{\frac{P(\theta_1^*, \theta_2^{(s)}|y)}{P(\theta_1^{(s)}, \theta_2^{(s)}|y)} \cdot \frac{\mathcal{Z}_1(\theta_1^{(s)}|\theta_1^*, \theta_2^{(s)})}{\mathcal{Z}_1(\theta_1^*|\theta_1^{(s)}, \theta_2^{(s)})}, 1\right\}$

若 $\gamma_1 = 1$，$\theta_1^{(s+1)} = \theta_1^*$，$\theta_2^{(s+1)} = \theta_2^{(s)}$

若 $\gamma_1 <$，$\theta_1^{(s+1)} = \begin{cases} \theta_1^*, & \text{if } \gamma_1 \ge u \\ \theta_1^{(s)}, & \text{if } \gamma_1 < u \end{cases}$ $u \sim U(0, 1)$

③ $\theta_2^* \sim \mathcal{Z}_2(\theta_2|\theta_1^{(s+1)}, \theta_2^{(s)})$，may not be symmetric

④ $\gamma_2 = \min\left\{\frac{P(\theta_1^{(s+1)}, \theta_2^*|y)}{P(\theta_1^{(s+1)}, \theta_2^{(s)}|y)} \cdot \frac{\mathcal{Z}_2(\theta_2^{(s)}|\theta_1^{(s+1)}, \theta_2^*)}{\mathcal{Z}_2(\theta_2^*|\theta_1^{(s+1)}, \theta_2^{(s)})}, 1\right\}$

若 $\gamma_2 = 1$，…

若 $\gamma_2 < 1$，$\theta_2^{(s+1)} = \begin{cases} \theta_2^*, & \text{if } \gamma_2 \ge u \\ \theta_2^{(s)}, & \text{if } \gamma_2 < u \end{cases}$

(Question4): (10 points) Consider the loglinear model (WX, XY, YZ) established for a four-way contingency table $I \times J \times K \times L$. Explain why W and Z are independent given X alone or given both X and Y. When W and Y are conditionally independent?

Then we find out the condition for W and Y's conditionally independence

$\begin{cases} \log \mu_{ij++} = \beta_0 + \beta_i^W + \beta_j^X + \gamma_{ij}^{WX} + \log(\sum_{l,k} e^{\beta_k^Y + \gamma_{jk}^{XY} + \gamma_{kl}^{YZ} + \beta_l^Z}) \\ \log \mu_{+jk+} = \beta_0 + \beta_j^X + \beta_k^Y + \gamma_{jk}^{XY} + \log(\sum_i e^{\beta_i^W + \gamma_{ij}^{WX}}) + \log(\sum_l e^{\beta_l^Z + \gamma_{kl}^{YZ}}) \\ \log \mu_{+j++} = \beta_0 + \beta_j^X + \log(\sum_i e^{\beta_i^W + \gamma_{ij}^{WX}}) + \log(\sum_{l,k} e^{\beta_k^Y + \gamma_{jk}^{XY} + \gamma_{kl}^{YZ} + \beta_l^Z}) \end{cases}$

$\Rightarrow \log \frac{\mu_{ij++} \cdot \mu_{+jk+}}{\mu_{+j++}} = \log \mu_{ij++} + \log \mu_{+jk+} - \log \mu_{+j++}$

$= [\beta_0 + \beta_i^W + \beta_j^X + \gamma_{ij}^{WX} + \log(\sum_{l,k} e^{\beta_k^Y + \gamma_{jk}^{XY} + \gamma_{kl}^{YZ} + \beta_l^Z})]$

$+ [\beta_0 + \beta_j^X + \beta_k^Y + \gamma_{jk}^{XY} + \log(\sum_i e^{\beta_i^W + \gamma_{ij}^{WX}}) + \log(\sum_l e^{\beta_l^Z + \gamma_{kl}^{YZ}})]$

$- [\beta_0 + \beta_j^X + \log(\sum_i e^{\beta_i^W + \gamma_{ij}^{WX}}) + \log(\sum_{l,k} e^{\beta_k^Y + \gamma_{jk}^{XY} + \gamma_{kl}^{YZ} + \beta_l^Z})]$

$= \beta_0 + \beta_i^W + \beta_j^X + \beta_k^Y + \gamma_{ij}^{WX} + \gamma_{jk}^{XY} + \log(\sum_l e^{\beta_l^Z + \gamma_{kl}^{YZ}}) = \log \mu_{ijk+}$

$\Rightarrow \frac{\mu_{ij++} \cdot \mu_{+jk+}}{\mu_{+j++}} = \mu_{ijk+} \Rightarrow \frac{\pi_{ij++} \cdot \pi_{+jk+}}{\pi_{+j++}} = \pi_{ijk+} \Rightarrow \frac{\pi_{ij++}}{\pi_{+j++}} \cdot \frac{\pi_{+jk+}}{\pi_{+j++}} = \frac{\pi_{ijk+}}{\pi_{+j++}}$

$\Rightarrow \pi_{i++|j} \cdot \pi_{+k+|j} = \pi_{ik+|j} \Rightarrow$ W and Y are independent given X alone

---

(a) Find the odds of survival beyond time y for the exponential, Weibull and log-logistic distributions (Hint: consider $f(y)$ in **Q1** as the density of log-logistic distribution).

(b) In **proportional odds models**, explanatory variables $\{x_1, \ldots, x_p\}$ have multiplicative effect on the odds of survival beyond time y

$$O_i = \eta_i O_0.$$

① exponential：$S(y;\lambda) = \int_y^{+\infty} \lambda e^{-t\lambda}\,dt = -e^{-t\lambda}\Big|_y^{+\infty} = e^{-\lambda y} \Rightarrow O(y;\lambda) = \frac{e^{-\lambda y}}{1 - e^{-\lambda y}} = \frac{1}{e^{\lambda y} - 1}$

② Weibull：$S(y;\alpha,\gamma) = \int_y^{+\infty} \alpha\phi u^{\alpha-1}\exp\{-\phi u^\alpha\}\,du = -\exp\{-\phi u^\alpha\}\Big|_y^\infty = \exp\{-\phi y^\alpha\}$ (define $\phi = \gamma^{-\alpha}$)

$\Rightarrow O(y;\alpha,\gamma) = \frac{\exp\{-\phi y^\alpha\}}{1 - \exp\{-\phi y^\alpha\}} = \frac{1}{\exp\{\phi y^\alpha\} - 1} = \frac{1}{\exp\{\gamma^{-\alpha}y^\alpha\} - 1}$

③ log-logistic：$S(y;\theta,\lambda) = \int_y^{+\infty} f(t)\,dt = \int_y^{+\infty} \frac{e^\theta \lambda t^{\lambda-1}}{(1 + e^\theta t^\lambda)^2}\,dt = \frac{1}{1 + e^\theta t^\lambda}\Big|_\infty^y = \frac{1}{1 + e^\theta y^\lambda}$

$\Rightarrow O(y;\theta,\lambda) = \frac{(1 + e^\theta y^\lambda)^{-1}}{1 - (1 + e^\theta y^\lambda)^{-1}} = \frac{1}{e^\theta y^\lambda}$

$\theta = \beta_0 + \sum_{j=1}^p \beta_j x_j \Rightarrow O(y;\theta,\lambda) = \frac{1}{e^\theta y^\lambda} = \frac{1}{\exp\{\beta_0 + \sum_{j=1}^p \beta_j x_j\}y^\lambda}$

$\Rightarrow$ Define $\eta_i = \frac{1}{\exp\{\sum_{j=1}^p \beta_j x_{ij}\}} \Rightarrow$ the form of $O_i = \eta_i O_0 = \eta_i \cdot \frac{1}{\exp\{\beta_0\}y^\lambda}$

(Question3): (10 points) For a binary response, consider the random effects model

$$logit[P(Y_{it} = 1|b_i)] = \alpha + \beta_t + b_i, \quad t = 1, 2, ..., T$$

where $b_i \overset{ind.}{\sim} N(0, \sigma^2)$, and the marginal model

$$logit[P(Y_{it} = 1)] = \alpha + \beta_t^*, \quad t = 1, 2, ..., T$$

For identifiability, $\beta_T = \beta_T^* = 0$. Explain why all $\beta_t = 0$ implies that all $\beta_t^* = 0$? Is the converse true?

① $\beta_t = 0 \Rightarrow logit[P(Y_{it} = 1|b_i)] = \alpha + b_i \Rightarrow P(Y_{it} = 1|b_i) = \frac{\exp(\alpha + b_i)}{1 - \exp(\alpha + b_i)}$

$\Rightarrow P(Y_{it} = 1) = \int_{-\infty}^{+\infty} \frac{\exp(\alpha + b_i)}{1 - \exp(\alpha + b_i)} \cdot \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{b_i^2}{2\sigma^2}\}\,db_i$ does not depend on t

$\Rightarrow$ The value of $logit[P(Y_{it} = 1)]$, $t = 1, \cdots, T$ does not depend on t

$\Rightarrow$ all $\beta_t^* = 0$ in the model $logit[P(Y_{it} = 1)] = \alpha + \beta_t^*$, $t = 1, \cdots, T$

② The converse is not true.

Suppose $\beta_t^* = 0$, then we have the constraints on marginal distributions of b_i and Y_{it}：

$\begin{cases} b_i \sim N(0, \sigma^2) \\ P(Y_{it} = 1) = \frac{\exp(\alpha)}{1 - \exp(\alpha)} \end{cases}$ the joint distribution $P(Y_{it}, b_i)$ is not unique. assume that different t corresponds to different $P(Y_{it}, b_i)$

$\Rightarrow logit[P(Y_{it} = 1|b_i)]$ depends on $t \Rightarrow \beta_t$ not necessarily to be 0 for t

---

(Question1): (15 points) Suppose we observe (y_i, x_i) for $i = 1, \ldots, n$, where y_i is response and x_i represents the value of an explanatory variable collected for observation i. Suppose,

$$y_i|x_i, \beta_1 \overset{independent}{\sim} N(\beta_0 + \beta_1 x_i, \sigma_e^2)$$

and the prior of β_1 is

$$\beta_1 \sim N(\nu_0, \sigma_0^2)$$

Derive the posterior of β_1, i.e., $P(\beta_1|y, X)$.

$P(y|X, \beta_1) = \prod_{i=1}^n P(y_i|X, \beta_1) = (2\pi\sigma_e^2)^{-\frac{n}{2}} \cdot \exp\{-\frac{1}{2\sigma_e^2}\sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2\}$

$= (2\pi\sigma_e^2)^{-\frac{n}{2}} \cdot \exp\{-\frac{1}{2\sigma_e^2}[\sum_{i=1}^n (y_i - \beta_0)^2 + \sum_{i=1}^n \beta_1^2 x_i^2 - 2\sum_{i=1}^n (y_i - \beta_0)\beta_1 x_i]\}$

$P(\beta_1|y, X) \propto P(y|X, \beta_1) \cdot P(\beta_1) \propto \exp\{-\frac{1}{2\sigma_e^2}[\sum_{i=1}^n \beta_1^2 x_i^2 - 2\sum_{i=1}^n (y_i - \beta_0)\beta_1 x_i] - \frac{1}{2\sigma_0^2}(\beta_1^2 - 2\nu_0\beta_1)\}$

$\propto \exp\{-\frac{1}{2\sigma_0^2\sigma_e^2}[\beta_1^2(\sigma_0^2\sum_{i=1}^n x_i^2 + \sigma_e^2) - 2\beta_1(\sigma_0^2\sum_{i=1}^n (y_i - \beta_0)x_i + \sigma_e^2\nu_0)]\}$

$\propto \exp\{-\frac{\sigma_0^2\sum_{i=1}^n x_i^2 + \sigma_e^2}{2\sigma_0^2\sigma_e^2}[\beta_1 - \frac{\sigma_0^2\sum_{i=1}^n (y_i - \beta_0)x_i + \sigma_e^2\nu_0}{\sigma_0^2\sum_{i=1}^n x_i^2 + \sigma_e^2}]^2\}$

$\Rightarrow$ the kenel of Normal with mean $\frac{\sigma_0^2\sum_{i=1}^n (y_i - \beta_0)x_i + \sigma_e^2\nu_0}{\sigma_0^2\sum_{i=1}^n x_i^2 + \sigma_e^2}$ and variance $\frac{\sigma_0^2\sigma_e^2}{\sigma_0^2\sum_{i=1}^n x_i^2 + \sigma_e^2}$

(Question3): (50 points) (Fit a Bayesian liner regression model) In the provided table, y is the response and x is explanatory variable. Assume

$$y|X, \beta \sim N(X\beta, \sigma^2 C_\rho)$$

$$\beta \sim N(0, \Sigma_0)$$

with $\Sigma_0 = \begin{bmatrix} 5, & 0 \\ 0, & 10 \end{bmatrix}$.

$$\sigma^2 \sim IG(1/2, 1/2)$$

and

$$\rho \sim U(0, 1)$$

Based on above model specification, derive

(a) (8 points) $P(\beta|y, X, \sigma^2, \rho)$.

(b) (8 points) $P(\sigma^2|y, X, \beta, \rho)$.

(b) (4 points) What about $P(\rho|y, X, \beta, \sigma^2)$? Does it have closed form?

(a) $P(y|X, \beta, \sigma^2, \rho) = (2\pi)^{-\frac{n}{2}} \cdot |\sigma^2 C_\rho|^{-\frac{1}{2}} \cdot \exp\{-\frac{1}{2}(y - X\beta)^T(\sigma^2 C_\rho)^{-1}(y - X\beta)\}$

$P(\beta|y, X, \sigma^2, \rho) \propto P(\beta) \cdot P(y|X, \beta, \sigma^2, \rho)$

$\propto \exp\{-\frac{1}{2}\beta^T\Sigma_0^{-1}\beta\}\exp\{-\frac{1}{2}(y - X\beta)^T(\sigma^2 C_\rho)^{-1}(y - X\beta)\}$

$\propto \exp\{-\frac{1}{2}\beta^T\Sigma_0^{-1}\beta - \frac{1}{2\sigma^2}(-2\beta^TX^TC_\rho^{-1}y + \beta^TX^TC_\rho^{-1}X\beta)\}$

$\propto \exp\{-\frac{1}{2}\beta^T(\Sigma_0^{-1} + \frac{1}{\sigma^2}X^TC_\rho^{-1}X)\beta + \beta^T\frac{1}{\sigma^2}X^TC_\rho^{-1}y\}$

$\Rightarrow \beta|y, X, \sigma^2, \rho \sim MVN((\sigma^2\Sigma_0^{-1} + X^TC_\rho^{-1}X)^{-1}X^TC_\rho^{-1}y, (\Sigma_0^{-1} + \frac{1}{\sigma^2}X^TC_\rho^{-1}X)^{-1})$

（$P(\theta) \propto \exp\{-\frac{1}{2}\theta^TA\theta + \theta^Tb\} \Rightarrow \theta \sim MVN(A^{-1}b, A^{-1})$）

(b) $P(\sigma^2|y, X, \beta, \rho) \propto P(\sigma^2) \cdot P(y|X, \beta, \sigma^2, \rho)$

$\propto (\sigma^2)^{-\frac{3}{2}} \cdot \exp\{-\frac{1}{2}\sigma^{-2}\} \cdot (\sigma^2)^{-\frac{n}{2}} \cdot \exp\{-\frac{1}{2}(y - X\beta)^T(\sigma^2 C_\rho)^{-1}(y - X\beta)\}$

$\propto (\sigma^2)^{-\frac{n+1}{2} - 1} \cdot \exp\{-\frac{1}{2}(y - X\beta)^TC_\rho^{-1}(y - X\beta) \cdot \sigma^{-2} - \frac{1}{2}\sigma^{-2}\}$

$\propto (\sigma^2)^{-\frac{n+1}{2} - 1} \cdot \exp\{-\sigma^{-2}[\frac{1}{2}(y - X\beta)^TC_\rho^{-1}(y - X\beta) + \frac{1}{2}]\}$

$\Rightarrow \sigma^2|y, X, \beta, \rho \sim IG(\frac{n+1}{2}, \frac{1}{2}(y - X\beta)^TC_\rho^{-1}(y - X\beta) + \frac{1}{2})$

(c) $P(\rho|y, X, \beta, \sigma^2) \propto P(\rho) \cdot P(y|X, \beta, \sigma^2, \rho)$

$= (2\pi\sigma^2)^{-\frac{n}{2}} |C_\rho|^{-\frac{1}{2}} \cdot \exp\{-\frac{1}{2}(y - X\beta)^T(\sigma^2 C_\rho)^{-1}(y - X\beta)\}$

$\Rightarrow$ Obviously, $P(\rho|y, X, \beta, \sigma^2)$ doesn't have a closed form.

$$p(z_i) = \mathcal{N}(z_i|\mu_0, \Sigma_0)$$

$$p(x_i|z_i, \theta) = \mathcal{N}(Wz_i + \mu, \Psi)$$

To do this, we compute the posterior over the latent factors:

$$p(z_i|x_i, \theta) = \mathcal{N}(z_i|m_i, \Sigma_i)$$

$$\Sigma_i \triangleq (\Sigma_0^{-1} + W^T\Psi^{-1}W)^{-1}$$

$$m_i \triangleq \Sigma_i(W^T\Psi^{-1}(x_i - \mu) + \Sigma_0^{-1}\mu_0)$$

Remark: *Notice that Σ_i is independent of i, so we can denoted it by Σ.*

Inverse Gamma：$\frac{\beta^\alpha}{\Gamma(\alpha)}x^{-\alpha-1}\exp(-\frac{\beta}{x})$