1. 概率论

§1 特殊分布的性质

1. χ^2: $X_1,\cdots,X_n \overset{i.i.d.}{\sim} N(0,1)$, $\sum_{i=1}^n X_i^2 \sim \chi^2(n)$, $\mu=n$, $\sigma^2=2n$
2. t: $X\sim N(0,1)$, $Y\sim\chi^2(n)$ 独, $Z=\frac{X}{\sqrt{Y/n}}\sim t(n)$, $\mu=0$, $\sigma^2=\frac{n}{n-2}$
3. F: ① $X\sim\chi^2(n_1)$, $Y\sim\chi^2(n_2)$ 独, $Z=\frac{X/n_1}{Y/n_2}\sim F(n_1,n_2)$
   ② $F_{1-\alpha}(n_1,n_2)=\frac{1}{F_\alpha(n_2,n_1)}$

§2 重要定理

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

1. Cauchy-Schwarz inequality: $[E(XY)]^2\le E(X^2)E(Y^2)$
2. Law of large number: $Z=\frac{1}{n}\sum_{i=1}^n X_i \to_p \mu$
3. Central limit theorem: $\frac{\bar X-\mu}{\sigma/\sqrt n}\to_d N(0,1)$ $(n>30)$
4. The Delta method: $Z=f(Y)\approx f(\mu)+(Y-\mu)f'(\mu)$
   $E(Z)\approx f(\mu)$, $var(Z)\approx\sigma_Y^2[f'(\mu)]^2$
   假设 $E(Y_i)=\mu_i$, $var(Y_i)\propto V(\mu_i)$ 求 $Z=f(Y)$
   $\Rightarrow f(\mu)\propto\int\frac{1}{\sqrt{V(\mu)}}d\mu$

2. GLM 基础

§1 GLM 基础

1. GLM的常见组成部分: ① Random component: 用于
   · 模拟 non-Gaussian data · 允许 Y_i 的 mean-variance dependence
   ② Systematic component: 实现 predictors 和 response 间的复杂关系
2. GLM的选取: ① $Y_i\in\{0,1\}$: Ber/Bin ② $Y_i\in N^+$: Poisson
   ③ $Y_i\in R^+$: Gamma/Exp/log-normal

§2 GLM 的 random component

1. Exponential dispersion model (EDM) ③ $\phi>0$ dispersion parameter
   $$Y\sim P(y;\theta,\phi)=a(y,\phi)\exp\left\{\frac{y\theta-\kappa(\theta)}{\phi}\right\}$$
   ① θ canonical parameter ② $\kappa(\theta)$ cumulant function
2. EDM的例子
   ① Normal distribution $Y\sim N(\mu,\sigma^2)$ $(\mu;\sigma^2)$
   $f(y;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(y-\mu)^2}{2\sigma^2}\}=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{y^2}{2\sigma^2}\}\cdot\exp\{\frac{\mu y-\frac12\mu^2}{\sigma^2}\}$
   其中, $\theta=\mu$, $\phi=\sigma^2$, $\kappa(\theta)=\frac12\mu^2=\frac12\theta^2$, $a(y,\phi)=\frac{1}{\sqrt{2\pi\phi}}\exp\{-\frac{y^2}{2\phi}\}$
   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$)$ $(\gamma\Gamma(1+\frac1\alpha),\ \gamma^2[\Gamma(1+\frac2\alpha)-(\Gamma(1+\frac1\alpha))^2])$
   $f(y;\gamma)=\frac1\gamma\exp(-\frac y\gamma)=\exp\{-\frac y\gamma-\log\gamma\}$
   其中, $\theta=-\frac1\gamma$, $\phi=1$, $\kappa(\theta)=\log\gamma=-\log(-\theta)$, $a(y,\phi)=1$
   ⑥ Gamma distribution $Y\sim Gamma(\alpha,\beta)$ $f(y;\alpha,\beta)=\frac{1}{\Gamma(\alpha)\beta^\alpha}y^{\alpha-1}e^{-\frac y\beta}$
   $=\frac{1}{\Gamma(\alpha)}y^{\alpha-1}\cdot\exp\{\alpha\log\alpha\}\exp\left\{\frac{-\frac{1}{\alpha\beta}y-\log\alpha\beta}{\frac1\alpha}\right\}$ $(\frac\alpha\beta, \frac{\alpha}{\beta^2})$
   其中, $\theta=-\frac{1}{\alpha\beta}$, $\phi=\frac1\alpha$, $\kappa(\theta)=\log\alpha\beta=-\log(-\theta)$
   $a(y,\phi)=\frac{1}{\Gamma(\alpha)}y^{\alpha-1}\cdot\exp\{\alpha\log\alpha\}=\frac{1}{\Gamma(\frac1\phi)}y^{\frac1\phi-1}\cdot\exp\{\frac1\phi\log\frac1\phi\}$
   ⑦ Geometric distribution $Y\sim Geo(p)$ $(\frac1p, \frac{1-p}{p^2})$
   $f(y;p)=p(1-p)^{y-1}=\exp\{\log(1-p)\cdot y-(\log(1-p)-\log p)\}$
   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)
   $f(y;\mu)=\frac1\mu\exp(-\frac y\mu)=\exp(-\frac y\mu-(-\log\frac1\mu))$
   其中, $\theta=-\frac1\mu$, $\phi=1$, $\kappa(\theta)=-\log(-\theta)$, $a(y,\phi)=1$
   ⑨ Inverse Gaussian $Y\sim IG(\mu,\lambda)$ $(\mu, \frac{\mu^3}{\lambda})$
   $f(y;\mu,\lambda)=\sqrt{\frac{\lambda}{2\pi y^3}}\exp\{-\frac{\lambda(y-\mu)^2}{2\mu^2y}\}=\sqrt{\frac{\lambda}{2\pi y^3}}\exp\{-\frac{\lambda}{2y}\}\cdot\exp\left\{\frac{-\frac{1}{2\mu^2}y+\frac1\mu}{\frac1\lambda}\right\}$
   其中, $\theta=-\frac{1}{2\mu^2}$ $(\mu=(-2\theta)^{-\frac12})$, $\phi=\frac1\lambda$, $\kappa(\theta)=-\frac1\mu=-\sqrt{-2\theta}$
   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. EDM 的 MGF, CGF, cumulants, mean, 和 variance
   ① MGF: $M(t)=E[e^{tY}]=\int_S a(y,\phi)\exp\{\frac{y(\theta+t\phi)-\kappa(\theta+t\phi)+\kappa(\theta+t\phi)-\kappa(\theta)}{\phi}\}dy$
   $=\exp\{\frac{\kappa(\theta+t\phi)-\kappa(\theta)}{\phi}\}\cdot\int_S a(y,\phi)\exp\{\frac{y(\theta+t\phi)-\kappa(\theta+t\phi)}{\phi}\}dy=\exp\{\frac{\kappa(\theta+t\phi)-\kappa(\theta)}{\phi}\}$
   ② CGF: $K(t)=\log M(t)=\frac{\kappa(\theta+t\phi)-\kappa(\theta)}{\phi}$
   ③ rth cumulant: $\kappa_r=\frac{d^rK(t)}{dt^r}\Big|_{t=0}=\phi^{r-1}\kappa^{(r)}(\theta)$
   ④ mean: $E[Y]=\mu=\kappa'(\theta)$
   ⑤ variance: $var[Y]=\phi\kappa''(\theta)=\phi V(\mu)$
4. Variance function: $V(\mu)=\frac{d\mu}{d\theta}=\kappa''(\theta)$
   $\kappa''(\theta)=\frac{d\kappa'(\theta)}{d\theta}=\frac{d\mu}{d\theta}>0\Rightarrow\mu$ 关于 θ 单调增 $\Rightarrow\mu$ 和 θ one-to-one mapping
   注: ① 若 $\left(\frac{d\mu}{d\theta}\right)^{-1}=V^{-1}(\mu)\Rightarrow$ 积分求出 θ 关于 μ 的表达式 $\Rightarrow$ 求出 $\kappa'(\theta)=\mu$
   ② EDM 可以和 $V(\mu)$ 相互确定
   eg. variance function $V(\mu)=\mu^2$ 对应了 EDM distribution 中的 gamma distribution
   ① $V(\mu)=\mu^2\Rightarrow\frac{d\mu}{d\theta}=\mu^{-2}\Rightarrow\theta=\int\mu^{-2}d\mu=-\mu^{-1}+C_0\Rightarrow\mu=\frac{1}{C_0-\theta}$
   ② $\kappa'(\theta)=\mu=\frac{1}{C_0-\theta}\Rightarrow\kappa(\theta)=\int\frac{1}{C_0-\theta}d\theta=-\log(C_0-\theta)+C_1=\log(\mu)+C_1$
   ③ 对比 Gamma(α,β) 的 pdf $\frac{1}{\Gamma(\alpha)\beta^\alpha}y^{\alpha-1}e^{-\frac y\beta}=\exp\{\frac{y(-\mu^{-1})-\log(\mu)}{1/\alpha}\}\cdot a$
5. The unit deviation: $d(y,\mu)=2\{t(y,y)-t(y,\mu)\}$
   其中 $t(y,\mu)=y\theta-\kappa(\theta)$
6. dispersion model form (DMF) $f(y;\mu,\phi)=b(y,\phi)\exp\{-\frac{1}{2\phi}d(y,\mu)\}$
   其中 $b(y,\phi)=a(y,\phi)\cdot\exp\{t(y,y)/\phi\}$
   eg. 求出 $Bin(n,p)$ 的 dispersion model form (n已知)
   $t(y,\mu)=y\log\frac{p}{1-p}+n\log(1-p)=y\log\frac{\mu}{n-\mu}+n\log(1-\frac\mu n)$ (利用 $p=\frac\mu n$)
   $p\in(0,1)\Rightarrow\mu=np\in(0,n)$
   ① 若 $y\ne0$ 且 $y\ne n$, $t(y,y)=y\log\frac{y}{n-y}+n\log(1-\frac yn)$
   ② 若 $y=0$, $t(y,y)=\lim_{\varepsilon\to0}t(\varepsilon,\varepsilon)=\lim_{\varepsilon\to0}\varepsilon\log\frac{\varepsilon}{n-\varepsilon}+n\log(1-\frac\varepsilon n)$
   $=\lim_{\varepsilon\to0}\varepsilon\log\varepsilon-\varepsilon\log(n-\varepsilon)+n\log(1-\frac\varepsilon n)=0$
   ③ 若 $y=n$, $t(y,y)=\lim_{\varepsilon\to0}t(n+\varepsilon,n+\varepsilon)=\lim_{\varepsilon\to0}(n+\varepsilon)\log(n+\varepsilon)-n\log n-\varepsilon\log(-\varepsilon)=0$
   $d(y,\mu)=2\{t(y,y)-t(y,\mu)\}=2\{y\log\frac y\mu\cdot\frac{n-\mu}{n-y}+n\log(\frac{n-y}{n-\mu})\}$
   $b(y,\phi)=a(y,\phi)\exp\{\frac{t(y,y)}{\phi}\}=\binom{n}{y}\exp\{y\log\frac{y}{n-y}+n\log(1-\frac yn)\}$
7. Saddlepoint approximation $f(y;\mu,\phi)=b(y,\phi)\exp\{-\frac{1}{2\phi}d(y,\mu)\}$
   $\Rightarrow\tilde P(y;\mu,\phi)=\frac{1}{\sqrt{2\pi\phi V(y)}}\exp\{-\frac{1}{2\phi}d(y,\mu)\}$
   即用 $\frac{1}{\sqrt{2\pi\phi V(y)}}$ 来近似 $b(y,\phi)=a(y,\phi)\exp\{\frac{t(y,y)}{\phi}\}$
   注: ① $V(\mu)=\frac{d\mu}{d\theta}=\kappa''(\theta)$ 是一个关于 μ 的函数. $V(y)=V(\mu)|_{\mu=y}$
   ② $V(y)=var(\mu)|_{\mu=y}$: 先将 $var(y)$ 为 μ 表示, 再把 μ 替换为 y.
   eg. 对 $Poi(\mu)$ 进行 saddlepoint approximation
   $\kappa(\theta)=\mu=e^\theta\Rightarrow V(y)=V(\mu)|_{\mu=y}=\kappa''(\theta)|_{\mu=y}=y$
   $\Rightarrow\frac{1}{\sqrt{2\pi\phi V(y)}}=\frac{1}{\sqrt{2\pi y}}\Rightarrow\tilde P(y;\mu)=\frac{1}{\sqrt{2\pi y}}\exp\{-y\log\frac y\mu+(y-\mu)\}$
   eg. $Poi(\mu)$ 的 saddlepoint approximation 不为 exact:
   $b(y,\phi)=\frac{1}{y!}\exp\{y\log y-y\}=\frac{1}{y!}\cdot y^y\cdot e^{-y}$, $\frac{1}{\sqrt{2\pi\phi V(y)}}=\frac{1}{\sqrt{2\pi y}}$
   $b(y,\phi)\approx\frac{1}{\sqrt{2\pi\phi V(y)}}\Rightarrow y!\approx y^y\cdot e^{-y}\cdot\sqrt{2\pi y}\Rightarrow y$ not too small
8. EDM 的 saddlepoint approximation 的 accuracy (或 adequacy)
   ① normal 和 inverse Gaussian distributions 的 saddlepoint approximation 为 exact
   ② 对于 Poisson 和 Binomial distributions:
   · Poisson distribution: y 不太小, i.e. $y>3$
   · Binomial distribution: y 和 $m-y$ 不太小, i.e. $y\ge3$ 且 $m-y\ge3$
   ③ 否则需要 small ϕ 与 not small y? (Smyth and Verbyla theorem)
   定义 $\tau=\frac{\phi V(y)}{(y-\text{boundary})^2}$, 若 $\tau\le\frac13$, 则 $\tilde P\approx P$ $(V(y)=V(\mu)|_{\mu=y})$
   eg. ① Poisson distribution: $\tau=\frac{1\cdot y}{(y-0)^2}=\frac1y\le\frac13\Rightarrow y\ge3$
   ② Binomial distribution: $var(y)=mp(1-p)=\mu(1-\frac\mu m)\Rightarrow\phi V(y)=y(1-\frac ym)$
   $\begin{cases}\tau_0=\frac{y(1-\frac ym)}{y^2}\le\frac13\\ \tau_1=\frac{y(1-\frac ym)}{(y-m)^2}\le\frac13\end{cases}\Rightarrow\begin{cases}y\ge\frac{1}{\frac1m+\frac13}\approx3\\ m-y\ge\frac{1}{\frac1m+\frac13}\approx3\end{cases}\Rightarrow\begin{cases}y\ge3\\ m-y\ge3\end{cases}$
   ③ Gamma distribution: $\phi=\frac1\alpha$, $V(\mu)=\mu^2$
   $\tau=\frac{\phi\cdot y^2}{y^2}\le\frac13\Rightarrow\phi\le\frac13$
9. unit deviance 的分布: $\frac{d(y,\mu)}{\phi}\sim\chi_1^2$ 若 spa accurate
10. Deviance function & total deviance:
   ① deviance function / deviance / total deviance: $D(y,\mu)=\sum_{i=1}^n d(y_i,\mu_i)$
   ② scaled deviance function: $D^*(y,\mu)=\frac{D(y,\mu)}{\phi}$
11. scaled deviance 的分布: $\frac{D(y,\mu)}{\phi}=D^*(y,\mu)\sim\chi^2_n$
   注: ① $\frac{D(y,\hat\mu)}{\phi}\sim\chi^2_{n-p'}$ (approximately) $\Rightarrow\hat\phi=\frac{D(y,\hat\mu)}{n-p'}$
   ② $l(\mu;y)=\sum_{i=1}^n\log b(y_i,\phi)-\frac{D(y,\mu)}{2\phi}$

Distribution	Domain	$\mu=E[Y\mid x]$	$v(\mu)$	$\theta(\mu)$	$b(\theta)$	ϕ
Binomial $B(n,p)$	$0,1,\ldots,n$	np	$\mu-\frac{\mu^2}{n}$	$\log\frac{p}{1-p}$	$n\log(1+e^\theta)$	1
Poisson $P(\mu)$	$0,1,\ldots,\infty$	μ	μ	$\log(\mu)$	e^θ	1
Neg. Binom. $NB(\mu,\alpha)$	$0,1,\ldots,\infty$	μ	$\mu+\alpha\mu^2$	$\log(\frac{\alpha\mu}{1+\alpha\mu})$	$-\frac1\alpha\log(1-\alpha e^\theta)$	1
Gaussian/Normal $N(\mu,\sigma^2)$	$(-\infty,\infty)$	μ	1	μ	$\frac12\theta^2$	σ^2
Gamma $N(\mu,\nu)$	$(0,\infty)$	μ	μ^2	$-\frac1\mu$	$-\log(-\theta)$	$\frac1\nu$
Inv. Gauss. $IG(\mu,\sigma^2)$	$(0,\infty)$	μ	μ^3	$-\frac{1}{2\mu^2}$	$-\sqrt{-2\theta}$	σ^2
Tweedie $p\ge1$	depends on p	μ	μ^p	$\frac{\mu^{1-p}}{1-p}$	$\frac{\alpha-1}{\alpha}\left(\frac{\theta}{\alpha-1}\right)^\alpha$	ϕ

§3 GLM 的 systematic component

1. systematic component: $g(\mu_i)=\eta_i=\beta_0+\sum_{j=1}^p\beta_jx_{ij}$
2. Canonical link function: $\eta_i=g(\mu_i)=\theta_i$ $\Rightarrow g(\cdot)=\kappa'^{-1}(\cdot)$
   注: 求 canonical link function 的方法: ① 写成 EDM 形式
   ② 找出 θ, κ 关于 parameters 的表达式, 将 κ 化为 $\kappa(\theta)$
   ③ 求出 $\mu=\kappa'(\theta)$ ④ 求出反函数 $\theta=g(\mu)=\kappa'^{-1}(\mu)$
   eg. 求 Binomial distribution 的 canonical link function
   $\mu=\kappa'(\theta)=\frac{ne^\theta}{1+e^\theta}\Rightarrow e^\theta=\frac{\mu}{n-\mu}\Rightarrow\theta=\log\frac{\mu}{n-\mu}=g(\mu)$
3. offset: GLMs 形式变为 $\eta_i=o_i+\beta_0+\sum_{j=1}^p\beta_jx_{ij}$
   offset o_i 被视作一个值事先已知的"structural predictor". 其 coefficient 值为1.

3. 参数求解

§1 关于 β 的 Likelihood function 与 score equation

1. likelihood function: $L(\beta)=\prod_{i=1}^n a(y_i,\phi)\exp\{\frac{y_i\theta_i-\kappa(\theta_i)}{\phi}\}$
   $\Rightarrow l(\beta)=\sum_{i=1}^n l_i(\beta)=\sum_{i=1}^n\{\frac{y_i\theta_i-\kappa(\theta_i)}{\phi}\}+\sum_{i=1}^n\log a(y_i;\phi)$
2. score equation (general case):
   ① 将 $\frac{\partial l(\beta)}{\partial\beta}$ 用 $\frac{\partial l_i}{\partial\beta}$ 表示: $\frac{\partial l(\beta)}{\partial\beta}=\sum_{i=1}^n\frac{\partial l_i}{\partial\beta}$
   ② $\begin{cases}l_i=\{\frac{y_i\theta_i-\kappa(\theta_i)}{\phi}\}+\log a(y_i;\phi)\\ \mu_i=\kappa'(\theta_i)\\ \eta_i=g(\mu_i)\ \text{(general link)}\\ \eta_i=x_i^T\beta\end{cases}\Rightarrow\begin{cases}\frac{\partial l_i}{\partial\theta_i}=\frac{y_i-\kappa'(\theta_i)}{\phi}=\frac{y_i-\mu_i}{\phi}\\ \frac{\partial\mu_i}{\partial\theta_i}=\kappa''(\theta_i)=\frac{Var(Y_i)}{\phi}\ (=V(\mu_i))\\ \frac{\partial\eta_i}{\partial\mu_i}=g'(\mu_i)\\ \frac{\partial\eta_i}{\partial\beta}=x_i\end{cases}$
   $\Rightarrow\frac{\partial l_i}{\partial\beta}=\frac{\partial l_i}{\partial\theta_i}\cdot\frac{\partial\theta_i}{\partial\mu_i}\cdot\frac{\partial\mu_i}{\partial\eta_i}\cdot\frac{\partial\eta_i}{\partial\beta}=\frac{y_i-\mu_i}{\phi}\cdot\frac{\phi}{Var(Y_i)}\cdot\frac{1}{g'(\mu_i)}\cdot x_i=\frac{(y_i-\mu_i)\cdot x_i}{Var(Y_i)\,g'(\mu_i)}$
   ③ score equations: $U(\beta)=\sum_{i=1}^n\frac{(y_i-\mu_i)}{Var(Y_i)\,g'(\mu_i)}x_i=0$ $(\in R^{p+1})$
   ④ 矩阵形式: $U(\beta)=[x_1,\cdots,x_n]\begin{bmatrix}\frac{1}{Var(Y_1)g'(\mu_1)}&&\\&\ddots&\\&&\frac{1}{Var(Y_n)g'(\mu_n)}\end{bmatrix}\begin{bmatrix}y_1-\mu_1\\ \vdots\\ y_n-\mu_n\end{bmatrix}=X^T\Delta(y-\mu)$
   其中 $X^T=[x_1,\cdots,x_n]$, $\Delta=diag\{\frac{1}{Var(Y_i)g'(\mu_i)}\}$
3. score equation (canonical link function)
   ① $\frac{\partial l_i}{\partial\beta}=\frac{\partial l_i}{\partial\theta_i}\cdot\frac{\partial\theta_i}{\partial\mu_i}\cdot\frac{\partial\mu_i}{\partial\eta_i}(=\frac{\partial\theta_i}{\partial\eta_i}=1)\cdot\frac{\partial\eta_i}{\partial\beta}=\frac{\partial l_i}{\partial\theta_i}\cdot\frac{\partial\eta_i}{\partial\beta}=\frac{y_i-\mu_i}{\phi}x_i$
   ② $U(\beta)=\frac1\phi\sum_{i=1}^n(y_i-\mu_i)x_i=0\Rightarrow X^T(y-\mu)=0$

§2 Regression coefficient β 的估计

1. Newton-Raphson algorithm Hessian matrix, $H(\beta^{(t)})=\frac{\partial U(\beta)^T}{\partial\beta}\Big|_{\beta=\beta^{(t)}}=\frac{\partial^2l(\beta)}{\partial\beta\partial\beta^T}$
   ① 基本思路: $0=U(\hat\beta)=U(\beta^{(t)})+H(\beta^{(t)})\cdot(\hat\beta-\beta^{(t)})+o(\|\hat\beta-\beta\|)$
   $\Rightarrow\hat\beta\approx\beta^{(t)}-H^{-1}(\beta^{(t)})U(\beta^{(t)})$
   ② Newton-Raphson algorithm 的算法流程
   (1) 令迭代轮数 $t=0$, 设置 initial value $\beta^{(0)}$ (可以设置为0)
   求出 $\mu^{(0)}=g^{-1}(X\beta^{(0)})$, $U(\beta^{(0)})=X^T\Delta^{(0)}(y-\mu^{(0)})$ 与 $H(\beta^{(0)})$
   (2) Update: $\beta^{(t+1)}=\beta^{(t)}-H^{-1}(\beta^{(t)})U(\beta^{(t)})$
   $\left.\begin{array}{l}U(\beta^{(t+1)})=X^T\Delta^{(t+1)}(y-\mu^{(t+1)})\\ H(\beta^{(t+1)})=\frac{\partial U(\beta)}{\partial\beta}\Big|_{\beta=\beta^{(t+1)}}\end{array}\right\}$ 其中 $\mu^{(t+1)}=g^{-1}(X\beta^{(t+1)})$
   令迭代轮数 $t=t+1$
   (3) 重复直到 $\|\beta^{(t+1)}-\beta^{(t)}\|<\varepsilon$ 或 $\|U(\beta^{(t+1)})(\cdot\phi)\|<\varepsilon$ (如 10^{-3})
   ③ $H^{-1}(\beta)$ 的形式 & 为什么上述步骤可忽视未知的 dispersion ϕ?
   $H(\beta)=\sum_{i=1}^n\frac{\partial^2l_i(\beta)}{\partial\beta\partial\beta^T}=\sum_{i=1}^n\frac{\partial}{\partial\beta^T}(\frac{\partial l}{\partial\beta})=\sum_{i=1}^n[\frac{\partial}{\partial\theta_i}(\frac{\partial l}{\partial\beta})]\cdot\frac{\partial\theta_i}{\partial\mu_i}\cdot\frac{\partial\mu_i}{\partial\eta_i}\cdot\frac{\partial\eta_i}{\partial\beta^T}$
   $=\sum_{i=1}^n[\frac{\partial}{\partial\theta_i}(\frac{\partial l}{\partial\beta})]\cdot\frac{1}{\kappa''(\theta_i)}\cdot\frac{1}{g'(\mu_i)}\cdot x_i^T$ $(\kappa''(\theta_i)=V(\mu_i)$ 与 ϕ 无关$)$
   $\frac{\partial}{\partial\theta_i}(\frac{\partial l}{\partial\beta})=\frac{\partial}{\partial\theta_i}(\frac{y_i-\mu_i}{\phi}\cdot\frac{1}{\kappa''(\theta_i)}\cdot\frac{1}{g'(\mu_i)}\cdot x_i)$ $(\mu=\kappa'(\theta)$ 也含有 $\theta)$
   $=\left\{\frac{-\kappa''(\theta_i)}{\phi}\cdot\frac{1}{\kappa''(\theta_i)}\cdot\frac{1}{g'(\mu_i)}+\frac{y_i-\mu_i}{\phi}\cdot\frac{\kappa'''(\theta_i)}{[\kappa''(\theta_i)]^2}\cdot\frac{1}{g'(\mu_i)}+\frac{y_i-\mu_i}{\phi}\cdot\frac{1}{\kappa''(\theta_i)}\cdot\frac{-g''(\mu_i)\cdot\kappa''(\theta_i)}{[g'(\mu_i)]^2}\right\}x_i$
   $=-\left\{1+\frac{[y_i-\kappa'(\theta_i)]\cdot\kappa'''(\theta_i)}{[\kappa''(\theta_i)]^2}+\frac{[y_i-\kappa'(\theta_i)]g''(\mu_i)}{g'(\mu_i)}\right\}\cdot\frac{1}{\phi g'(\mu_i)}\cdot x_i=-\delta(y_i;\mu_i)\cdot\frac{1}{\phi g'(\mu_i)}\cdot x_i$
   $\Rightarrow-H(\beta)=\sum_{i=1}^n\frac{\delta(y_i;\mu_i)}{\phi\kappa''(\theta_i)[g'(\mu_i)]^2}\cdot x_ix_i^T\Rightarrow-H^{-1}(\beta^{(t)})U(\beta^{(t)})$ 中不含有 ϕ
2. Fisher scoring algorithm
   ① $E[-H(\beta)]$ 的形式:
   (1) 第一种算法: $E[y_i-\kappa'(\theta_i)]=0\Rightarrow E[\delta(y_i;\mu_i)]=1$
   $\Rightarrow E[-H(\beta)]=\sum_{i=1}^n\frac{1}{\phi\kappa''(\theta_i)[g'(\mu_i)]^2}\cdot x_ix_i^T=\phi^{-1}X^TWX$
   (2) 第二种算法: $E[-H(\beta)]=E[U(\beta)U(\beta)^T]=Var[U(\beta)]=Var[X^T\Delta(Y-\mu)]$
   $=X^T\Delta\,Var[Y]\Delta X=X^T diag\{\frac{1}{Var(Y_i)[g'(\mu_i)]^2}\}X=\phi^{-1}X^TWX$
   (3) 与 Fisher information 的关系: $I_n(\beta)=E[-H(\beta)]=\phi^{-1}X^TWX$
   其中 $W=diag\{\frac{1}{V(\mu_i)[g'(\mu_i)]^2}\}$, $X=[x_1,\cdots,x_n]^T\in R^{n\times p'}$
   注: 若使用 canonical link, 则 $W=diag\{V(\mu_i)\}$ $(g(\cdot)=(\kappa')^{-1}(\cdot), V(\cdot)=(\kappa')'(\cdot))$
   ② Fisher scoring algorithm 的算法流程
   (1) 令迭代轮数 $t=0$, 设置 initial value $\beta^{(0)}$ (可以设置为0)
   求出 $\mu^{(0)}=g^{-1}(X\beta^{(0)})$, $U(\beta^{(0)})=X^T\Delta^{(0)}(y-\mu^{(0)})$ 与 $I_n(\beta^{(0)})$
   (2) Update: $\beta^{(t+1)}=\beta^{(t)}\boxed{+I_n^{-1}(\beta^{(t)})}U(\beta^{(t)})$ 记得取 inverse & 符号改变!!!
   $\left.\begin{array}{l}U(\beta^{(t+1)})=X^T\Delta^{(t+1)}(y-\mu^{(t+1)})\\ \boxed{I_n(\beta^{(t+1)})=\phi^{-1}X^TW^{(t+1)}X}\end{array}\right\}$ 其中 $\mu^{(t+1)}=g^{-1}(X\beta^{(t+1)})$
   令迭代轮数 $t=t+1$
   (3) 重复直到 $\|\beta^{(t+1)}-\beta^{(t)}\|<\varepsilon$ 或 $\|U(\beta^{(t+1)})(\cdot\phi)\|<\varepsilon$ (如 10^{-3})
3. Iteratively reweighted least squares (IRLS)
   ① 基本思路: $\beta^{(t+1)}=\beta^{(t)}+I_n^{-1}(\beta^{(t)})\cdot U(\beta^{(t)})$
   $=\beta^{(t)}+\phi[X^TW^{(t)}X]^{-1}\cdot X^T\Delta^{(t)}(y-\mu^{(t)})$
   $=\beta^{(t)}+\phi[X^TW^{(t)}X]^{-1}\cdot X^TW^{(t)}\phi^{-1}\cdot diag\{g'(\mu_i^{(t)})\}(y-\mu^{(t)})$
   $=\beta^{(t)}+[X^TW^{(t)}X]^{-1}\cdot X^TW^{(t)}\cdot diag\{g'(\mu_i^{(t)})\}(y-\mu^{(t)})$
   $=[X^TW^{(t)}X]^{-1}\cdot X^TW^{(t)}\{X\beta^{(t)}+diag\{g'(\mu_i^{(t)})\}(y-\mu^{(t)})\}$
   $=[X^TW^{(t)}X]^{-1}\cdot X^TW^{(t)}z^{(t)}$
   其中 $z^{(t)}=X\beta^{(t)}+diag\{g'(\mu_i^{(t)})\}(y-\mu^{(t)})=X\beta^{(t)}+(W^{(t)})^{-1}\phi\Delta^{(t)}(y-\mu^{(t)})$
   $z^{(t)}$ 的第 i 个元素为 $z_i^{(t)}=\eta_i^{(t)}+g'(\mu_i^{(t)})(y_i-\mu_i^{(t)})$
   z_i 被称为 working response, W_i 被称为 working weight
   ② Iteratively reweighted least square 的算法流程
   (1) 令迭代轮数 $t=0$, 设置 initial value $\beta^{(0)}$ (或直接设置 $\mu_i^{(0)}=y_i$)
   求出 $\mu^{(0)}=g^{-1}(X\beta^{(0)})$, $W^{(0)}=diag\{\frac{1}{V(\mu_i^{(0)})[g'(\mu_i^{(0)})]^2}\}$ 与 $z^{(0)}$
   (2) Update: $\beta^{(t+1)}=[X^TW^{(t)}X]^{-1}\cdot X^TW^{(t)}z^{(t)}$
   $\left.\begin{array}{l}W^{(t+1)}=diag\{\frac{1}{V(\mu_i^{(t+1)})[g'(\mu_i^{(t+1)})]^2}\}\\ z^{(t+1)}=g(\mu^{(t+1)})+diag\{g'(\mu_i^{(t+1)})\}(y-\mu^{(t+1)})\end{array}\right\}$ 其中 $\mu^{(t+1)}=g^{-1}(X\beta^{(t+1)})$
   令迭代轮数 $t=t+1$
   (3) 重复直到 $\|\beta^{(t+1)}-\beta^{(t)}\|<\varepsilon$ 或 $\|U(\beta^{(t+1)})(\cdot\phi)\|<\varepsilon$ (如 10^{-3})

§3 Dispersion ϕ 的估计

1. Modified profile log-likelihood (MPL) estimator

$\hat\beta(\phi)=\arg\max_{\beta} l(\beta,\phi;y) \Rightarrow l(\hat\beta(\phi),\phi,y)=\max_{\beta} l(\beta,\phi;y)$

$\Rightarrow l^0(\phi)=l(\hat\beta(\phi),\phi,y)+\frac{p'}{2}\log\phi$ (防止ϕ过小) $\Rightarrow \hat\phi^0=\arg\max\, l^0(\phi)$

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

若 spa accurate, residual deviance $D(y,\hat\mu)$ 大致服从 $\frac{D(y,\hat\mu)}{\phi}\sim\chi^2_{n-p'}$

$\Rightarrow E[\frac{D(y,\hat\mu)}{\phi}]=n-p' \Rightarrow \tilde\phi$ 为 ϕ 的 unbiased estimator

e.g. 求 Gamma GLM (shape $=\alpha$, scale $=\beta$) 的 ϕ 的 estimator

$f(y;\alpha,\beta)=\frac{1}{\Gamma(\alpha)}y^{\alpha-1}\cdot\exp\{\alpha\log\alpha\}\exp\{\frac{-\frac{1}{\mu}y-\log\mu}{\frac{1}{\alpha}}\}$, $E[Y]=\kappa'(\theta)=-\frac{1}{\theta}=\alpha\beta$

$\Rightarrow t(y,\mu)=-\frac{1}{\mu}y-\log\mu,\ t(y,y)=-1-\log y \Rightarrow d(y,\mu)=2(-1+\frac{y}{\mu}-\log\frac{y}{\mu})$

$\Rightarrow D(y,\hat\mu)=\sum_{i=1}^n d(y_i,\hat\mu_i)=2\sum_{i=1}^n(-1+\frac{y_i}{\hat\mu_i}-\log\frac{y_i}{\hat\mu_i}) \Rightarrow \tilde\phi=\frac{D(y,\hat\mu)}{n-p'}$

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

① 可以将 Pearson statistics 视作 working RSS

$X^2=\sum_{i=1}^n w_i(z_i-\hat\eta_i)^2=\sum_{i=1}^n\frac{(z_i-\hat\eta_i)^2}{V(\hat\mu_i)[g'(\hat\mu_i)]^2}=\sum_{i=1}^n\frac{[g'(\hat\mu_i)]^2(y_i-\hat\mu_i)^2}{V(\hat\mu_i)[g'(\hat\mu_i)]^2}=\sum_{i=1}^n\frac{(y_i-\hat\mu_i)^2}{V(\hat\mu_i)}$

② $\bar\phi$ 为 approximately unbiased: $Var[Y_i]=\phi\cdot V(\mu_i)\Rightarrow E[\frac{(y_i-\mu_i)^2}{V(\mu_i)}]=\phi$

$\Rightarrow E[\frac{1}{n-p'}\sum_{i=1}^n\frac{(y_i-\hat\mu_i)^2}{V(\hat\mu_i)}]=\phi$

③ $\frac{X^2}{\phi}=\sum_{i=1}^n\frac{(y_i-\hat\mu_i)^2}{\phi V(\hat\mu_i)}=\sum_{i=1}^n\frac{(y_i-\hat\mu_i)^2}{\widehat{var}(y_i)}\sim\chi^2_{n-p'}$ (CLT holds for individual observation)

Gamma: $\phi\le\frac13$, Poisson: $\min\{y_i\}\ge5$, Binomial: $\min\{y_i\}\ge5$ & $\min\{n_i-y_i\}\ge5$

e.g. 考虑 Gaussian GLM fitted for independent samples $Y_1,\ldots,Y_n$ using logarithmic link function. linear predictor 为 $\eta=X\beta$, $\beta\in R^{p'}$, $X\in R^{n\times p'}$. 求 (a) score function $U(\beta)$ 和 Fisher information $I_{jk}(\beta)$ (b) working response z_i, $i=1,\ldots,n$ (c) ϕ 的 MLE, Mean-deviance estimator, Pearson estimator

(a) logarithmic link $\Rightarrow g(\mu_i)=\log\mu_i=\eta_i \Rightarrow \frac{\partial l_i}{\partial\beta_j}=\frac{(y_i-\mu_i)\cdot\mu_i}{Var(Y_i)}\cdot x_{ij}$

Gaussian GLM $\Rightarrow \mu_i=\kappa'(\theta_i)=\theta_i,\ V(\mu_i)=\kappa''(\theta_i)=1 \Rightarrow U(\beta_j)=\sum_{i=1}^n\frac{(Y_i-e^{\eta_i})\cdot e^{\eta_i}}{\phi}\cdot x_{ij}$

$I(\beta)=\phi^{-1}X^TWX=\phi^{-1}X^T diag\{\frac{1}{V(\mu_i)[g'(\mu_i)]^2}\}X=\phi^{-1}X^T diag\{\mu_i^2\}X$

$I_{jk}(\beta)=\frac{1}{\phi}\sum_{i=1}^n\mu_i^2\cdot x_{ij}\cdot x_{ik}=\frac{1}{\phi}\sum_{i=1}^n e^{2\eta_i}\cdot x_{ij}\cdot x_{ik}$

(b) $g(\mu_i)=\log\mu_i=\eta_i \Rightarrow g'(\mu_i)=\frac{1}{\mu_i}=e^{-\eta_i} \Rightarrow z_i=\eta_i+g'(\mu_i)(y_i-\mu_i)=\eta_i-1+e^{-\eta_i}y_i$

(c) $\frac{\partial l}{\partial\phi}=\frac{\partial}{\partial\phi}\sum\{-\frac12\log(2\pi\phi)-\frac{(y_i-\mu_i)^2}{2\phi}\}=\sum\{-\frac{1}{2\phi}+\frac{(y_i-\mu_i)^2}{2\phi^2}\}\Rightarrow\hat\phi_{MLE}=\frac{\sum_{i=1}^n(Y_i-\hat\mu_i)^2}{n}$

$t(y,\mu)=\mu y-\frac12\mu^2,\ t(y,y)=\frac12y^2 \Rightarrow d(y,\mu)=2\{t(y,y)-t(y,\mu)\}=(y-\mu)^2$

$\Rightarrow D(y,\hat\mu)=\sum_{i=1}^n d(y_i,\hat\mu_i)=\sum_{i=1}^n(y_i-\hat\mu_i)^2 \Rightarrow \tilde\phi_{MDE}=\frac{D(y,\hat\mu)}{n-p'}=\sum_{i=1}^n\frac{(Y_i-\hat\mu_i)^2}{n-p'}$

$V(\mu)=1 \Rightarrow \bar\phi_{Pearson}=\frac{X^2}{n-p'}=\frac{1}{n-p'}\sum_{i=1}^n\frac{(y_i-\hat\mu_i)^2}{V(\hat\mu_i)}=\sum_{i=1}^n\frac{(Y_i-\hat\mu_i)^2}{n-p'}$

4. 假设检验

§1 STA 2004 中的假设检验

1. Fisher information: $I(\theta)=E[(\frac{\partial l(\theta)}{\partial\theta})^2]=-E[\frac{\partial^2 l(\theta)}{\partial\theta^2}]$

$\sqrt{I(\theta)}(\hat\theta-\theta)\xrightarrow{d}N(0,1) \Rightarrow \hat\theta_{MLE}\xrightarrow{d}N(\theta,\frac{1}{I(\theta)})$

2. Likelihood ratio test: $\hat T=2[l(\hat\theta)-l(\theta_0)]\xrightarrow{d}\chi^2_1 \Rightarrow R:=\{\hat T>\chi^2_{\alpha,1}\}$

3. Wald test: $\hat I(\theta)=-\frac{d^2l(\theta)}{d\theta^2} \Rightarrow \hat W=\hat I(\hat\theta)(\hat\theta-\theta_0)^2\xrightarrow{d}\chi^2_1$ (或用 $I(\hat\theta)$)

4. Score test: $\hat T=\frac{U^2(\theta_0)}{I(\theta_0)}\xrightarrow{d}\chi^2_1 \Rightarrow R:=\{\hat T>\chi^2_{\alpha,1}\}$

§2 β 的 MLE 的 asymptotic distribution

1. $\hat\beta$: $\sqrt n(\hat\beta-\beta)\xrightarrow{d}N(0,nI^{-1}(\beta))$ ($I^{-1}(\beta)=\phi(X^TWX)^{-1}$)

2. $\hat\beta_j$: $\sqrt n(\hat\beta_j-\beta_j)\xrightarrow{d}N(0,n\phi v_j^2)$ 其中 $v_j=\sqrt{[(X^TWX)^{-1}]_{jj}}$

3. $h(\hat\beta)$: $\sqrt n(h(\hat\beta)-h(\beta))\xrightarrow{d}N(0,n[\frac{\partial h}{\partial\beta}]^TI^{-1}(\beta)[\frac{\partial h}{\partial\beta}])$ (要求 $h(\cdot)$ 连续)

4. $c^T\hat\beta$: $\sqrt n(c^T\hat\beta-c^T\beta)\xrightarrow{d}N(0,nc^TI^{-1}(\beta)c)$

5. $\hat\eta_i=x_i^T\hat\beta$: $\sqrt n(\hat\eta_i-\eta_i)\xrightarrow{d}N(0,nx_i^TI^{-1}(\beta)x_i)$

§3 Wald test ($I^{-1}(\beta^0)=\phi(X^TWX)^{-1}|_{\beta=\beta^0}$, $W=diag\{\frac{1}{V(\mu_i)[g'(\mu_i)]^2}\}$)

1. $H_0:\beta_j=\beta_j^0$ v.s. $H_1:\beta_j\ne\beta_j^0$ (ϕ已知)

$T=\frac{\hat\beta_j-\beta_j^0}{\sqrt\phi\cdot\hat v_j}=\frac{\hat\beta_j-\beta_j^0}{\sqrt\phi\cdot\sqrt{[(X^T\hat WX)^{-1}]_{jj}}}\xrightarrow{d}N(0,1)$ ($\hat W=diag\{\frac{1}{V(\hat\mu_i)[g'(\hat\mu_i)]^2}\}$)

$\Rightarrow$ (1−α) CI: $\hat\beta_j\pm z_{\alpha/2}\cdot\sqrt\phi\cdot\sqrt{[(X^T\hat WX)^{-1}]_{jj}}$, $\hat p=\inf\{\alpha\in[0,1]:|T|>z_{\alpha/2}\}=1-\Phi(|T|)+\Phi(-|T|)$

2. $H_0:\beta_j=\beta_j^0$ v.s. $H_1:\beta_j\ne\beta_j^0$ (ϕ未知)

$T=\frac{\hat\beta_j-\beta_j^0}{\sqrt{\bar\phi}\cdot\hat v_j}=\frac{\hat\beta_j-\beta_j^0}{\sqrt{\bar\phi}\cdot\sqrt{[(X^T\hat WX)^{-1}]_{jj}}}\xrightarrow{d}t_{n-p'}$ ($\bar\phi=\frac{X^2}{n-p'}=\frac{1}{n-p'}\sum_{i=1}^n\frac{(y_i-\hat\mu_i)^2}{V(\hat\mu_i)}$)

$\Rightarrow$ (1−α) CI: $\hat\beta_j\pm t_{\alpha/2}(n-p')\cdot\sqrt{\bar\phi}\cdot\sqrt{[(X^T\hat WX)^{-1}]_{jj}}$, $\hat p=\inf\{\cdots|T|>t_{\alpha/2}(n-p')\}=1-\Phi(|T|)+\Phi(-|T|)$

3. $H_0:\beta=\beta^0$ v.s. $H_1:\beta\ne\beta^0$ (ϕ已知)

$T=(\hat\beta-\beta^0)^T\widehat{Cov}(\hat\beta)^{-1}(\hat\beta-\beta^0)=\frac{(\hat\beta-\beta^0)^T(X^T\hat WX)(\hat\beta-\beta^0)}{\phi}\sim\chi^2_{p'}$

$\Rightarrow \hat p=\inf\{\alpha\in[0,1]:T>\chi^2_\alpha(p')\}=1-\Phi(T)$

4. $H_0:\beta=\beta^0$ v.s. $H_1:\beta\ne\beta^0$ (ϕ未知)

$T=\frac{(\hat\beta-\beta^0)^T(X^T\hat WX)(\hat\beta-\beta^0)/p'}{\bar\phi}\xrightarrow{d}F(p',n-p')$

$\Rightarrow \hat p=\inf\{\alpha\in[0,1]:T>F_\alpha(p',n-p')\}=1-\Phi(T)$

5. $H_0:c^T\beta=c^T\beta^0$ v.s. $H_1:c^T\beta\ne c^T\beta^0$ (ϕ已知)

$T=\frac{c^T\hat\beta-c^T\beta^0}{\sqrt\phi\cdot\sqrt{c^T(X^T\hat WX)^{-1}c}}\sim N(0,1) \Rightarrow c^T\hat\beta\pm z_{\alpha/2}\cdot\sqrt\phi\cdot\sqrt{c^T(X^T\hat WX)^{-1}c}$

6. $H_0:c^T\beta=c^T\beta^0$ v.s. $H_1:c^T\beta\ne c^T\beta^0$ (ϕ未知)

$T=\frac{c^T\hat\beta-c^T\beta^0}{\sqrt{\bar\phi}\cdot\sqrt{c^T(X^T\hat WX)^{-1}c}}\sim t_{n-p'} \Rightarrow c^T\hat\beta\pm t_{\alpha/2}(n-p')\cdot\sqrt{\bar\phi}\cdot\sqrt{c^T(X^T\hat WX)^{-1}c}$

7. $H_0:x_i^T\beta=x_i^T\beta^0$ v.s. $H_1:x_i^T\beta\ne x_i^T\beta^0$ (ϕ已知)

$T=\frac{\hat\eta_i-\eta_i}{\sqrt\phi\cdot\sqrt{x_i^T(X^T\hat WX)^{-1}x_i}}\sim N(0,1) \Rightarrow x_i^T\hat\beta\pm z_{\alpha/2}\cdot\sqrt\phi\cdot\sqrt{x_i^T(X^T\hat WX)^{-1}x_i}$

8. $H_0:x_i^T\beta=x_i^T\beta^0$ v.s. $H_1:x_i^T\beta\ne x_i^T\beta^0$ (ϕ未知)

$T=\frac{\hat\eta_i-\eta_i}{\sqrt{\bar\phi}\cdot\sqrt{x_i^T(X^T\hat WX)^{-1}x_i}}\sim t_{n-p'} \Rightarrow x_i^T\hat\beta\pm t_{\alpha/2}(n-p')\cdot\sqrt{\bar\phi}\cdot\sqrt{x_i^T(X^T\hat WX)^{-1}x_i}$

注: $\mu_i=g^{-1}(x_i^T\beta)$ 的 CI 可以通过考虑 transformation $g^{-1}(\cdot)$ 得到

9. Wald test 的缺点: Not invariant to reparameterization

§4 Score test

1. $H_0:\beta=\beta^0\in R^{p'}$ v.s. $H_1:\beta\ne\beta^0$: $T_S=U(\hat\beta^0)^TI^{-1}(\hat\beta^0)U(\hat\beta^0)\sim\chi^2_{p'}$

注: ① $H_0:\beta\in B_0$ v.s. $H_1:\beta\in B_1$: 令 d 为 number of unique restrictions on β (若 $H_0:\beta=\beta^0$, 则 $d=p'$), $\hat\beta^0=\arg\max_{\beta\in B_0}l(\beta) \Rightarrow T_S=U(\hat\beta^0)^TI^{-1}(\hat\beta^0)U(\hat\beta^0)\sim\chi^2_d$

② 若ϕ未知, 则可以用 Pearson estimator $\bar\phi$ 代替 ϕ, 采用 F-test. (分子上除以 d.f.)

2. $H_0:\beta_1=\beta_1^0\in R^{d_1}$ v.s. $H_1:\beta_1\ne\beta_1^0$ (令 $\beta=\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix}$, 其中 $\beta_1\in R^{d_1}$, $\beta_2\in R^{p'-d_1}$)

$T_S=U_1(\hat\beta^0)^TI^{11}(\hat\beta^0)U_1(\hat\beta^0)\sim\chi^2_{d_1}$

其中, $\hat\beta^0=\begin{bmatrix}\beta_1^0\\ \hat\beta_2(\beta_1^0)\end{bmatrix}$ 为 H_0下的MLE ($\hat\beta_2(\beta_1^0)=\arg\max_{\beta_1=\beta_1^0}l(\beta_2)$)

$U(\hat\beta^0)=\begin{bmatrix}U_1(\hat\beta^0)\\ U_2(\hat\beta^0)\end{bmatrix}$ ($U_1(\hat\beta^0)\in R^{d_1}$) $I^{-1}(\beta)=\begin{bmatrix}I^{11}&I^{12}\\ I^{21}&I^{22}\end{bmatrix}=\begin{bmatrix}var(U_1)&cov(U_1,U_2)\\ cov(U_2,U_1)&var(U_2)\end{bmatrix}$

§5 Likelihood ratio test (LRT)

1. $H_0:\beta=\beta^0\in R^{p'}$ v.s. $H_1:\beta\ne\beta^0$: $T_{LR}=2\log\frac{L(\hat\beta)}{L(\hat\beta^0)}\sim\chi^2_{p'}$

其中, $\hat\beta=\arg\max_{\beta\in H_0\cup H_1}l(\beta)=\arg\max_{\beta\in R^{p'}}l(\beta)$ ($\{\beta:\beta=\beta^0\}\cup\{\beta:\beta\ne\beta^0\}=R^{p'}$)

注: ① $H_0:\beta\in B_0$ v.s. $H_1:\beta\in B_1$: $\hat\beta^0=\arg\max_{\beta\in B_0}l(\beta)$, $\hat\beta=\arg\max_{\beta\in B_0\cup B_1}l(\beta)$

$\Rightarrow T_{LR}=2\log\frac{L(\hat\beta)}{L(\hat\beta^0)}\sim\chi^2_d$

② 若ϕ未知, 则可以用 Pearson estimator $\bar\phi$ 代替 ϕ, 采用 F-test. (分子上除以 d.f.)

2. $H_0:\beta_1=\beta_1^0\in R^{d_1}$ v.s. $H_1:\beta_1\ne\beta_1^0$ (令 $\beta=\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix}$, 其中 $\beta_1\in R^{d_1}$, $\beta_2\in R^{p'-d_1}$)

$T_{LR}=2\log\frac{L(\hat\beta)}{L(\hat\beta^0)}\sim\chi^2_{d_1}$

其中, $\hat\beta=\arg\max_{\beta\in H_0\cup H_1}l(\beta)=\arg\max_{\beta\in R^{p'}}l(\beta)$ $\hat\beta^0=\begin{bmatrix}\beta_1^0\\ \hat\beta_2(\beta_1^0)\end{bmatrix}$ 为 H_0下的MLE

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

$l=-\frac{1}{2\phi}D(y,\mu)+\sum_i\log b(y_i,\phi) \Rightarrow$

① (ϕ已知) $T_{LR}=2(l(B)-l(A))=\frac{D(y,\hat\mu_A)-D(y,\hat\mu_B)}{\phi}\sim\chi^2(p_B-p_A)$

② (ϕ未知) $\tilde F=\frac{[D(y,\hat\mu_A)-D(y,\hat\mu_B)]/(p_B-p_A)}{\tilde\phi}\sim F(p_B-p_A,n-p_B')$

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)

或 $\bar F=\frac{[D(y,\hat\mu_A)-D(y,\hat\mu_B)]/(p_B-p_A)}{\bar\phi}\sim F(p_B-p_A,n-p_B')$

其中 $\bar\phi=\frac{X_B^2}{n-p_B'}=\frac{1}{n-p_B'}\sum_{i=1}^n\frac{(y_i-\hat\mu_{Bi})^2}{V(\hat\mu_{Bi})}$, $\frac{X_B^2}{\phi}\sim\chi^2(n-p_B')$ (若 CLT holds for ⋯)

③ 多个 Models 的比较 ($M_0\subseteq M_1\subseteq\cdots M_k\subseteq\cdots M_N$) $df_k=n-p_k'$

Models	Deviance D	d.f.	ΔD	Δd.f.
M_0	D_0	df_0	—	—
M_1	D_1	df_1	D_0-D_1	df_0-df_1
M_k	D_k	df_k	$D_{k-1}-D_k$	$df_{k-1}-df_k$
M_N	D_N	df_N	$D_{N-1}-D_N$	$df_{N-1}-df_N$

对于 M_{k-1} v.s. M_k:

$\tilde F=\frac{(D_{k-1}-D_k)/(df_{k-1}-df_k)}{\tilde\phi}\sim F(\underbrace{df_{k-1}-df_k}_{=p_k-p_{k-1}},n-p_N')$, $\tilde\phi\longrightarrow=\frac{D_N}{n-p_N'}$

4. goodness-of-fit test

① 关于 saturated model: $\tilde\mu_i=y_i \Rightarrow$ d.f. $=0$, $D(y,\tilde\mu)=0$

$\Rightarrow T_{LR}=\frac{D(y,\hat\mu)}{\phi}$ ($\hat\mu$ 为当前 model 的 fitted values)

② Deviance gof test: $D(y,\hat\mu)\sim\chi^2_{n-p'}$ ($P<0.05$ 或 $D(y,\hat\mu)\gg n-p'$, 则不足)

③ Pearson gof test: $X^2=\sum_{i=1}^n\frac{(y_i-\hat\mu_i)^2}{V(\hat\mu_i)}\sim\chi^2_{n-p'}$

e.g. $P(y;\mu)=\frac{1}{\mu}\exp(-\frac{y}{\mu})$, 对于 $H_0:\mu=1$, $H_1:\mu>1$, 求 Wald & LR.

$L(\mu|y)=\frac{1}{\mu^n}\exp(-\frac{\sum_{i=1}^n y_i}{\mu}) \Rightarrow l(\mu|y)=-n\log\mu-\frac{\sum_{i=1}^n y_i}{\mu} \Rightarrow u(\mu|y)=-\frac{n}{\mu}+\frac{\sum_{i=1}^n y_i}{\mu^2}$

$\Rightarrow \hat\mu=\frac{\sum_{i=1}^n y_i}{n}$, $\hat I(\hat\mu)=-\frac{\partial^2 l}{\partial\mu^2}|_{\mu=\hat\mu}=\frac{n^3}{(\sum_{i=1}^n y_i)^2}$

① $\hat T_W=\sqrt{\hat I(\hat\mu)}(\hat\mu-\mu_0)=\sqrt n(1-\frac{n}{\sum_{i=1}^n y_i})\sim N(0,1)$

② Since $L(\mu)$ increases when $\mu<\hat\mu$, decreases when $\mu>\hat\mu$

$\hat T_{LR}=2\log\frac{\sup_{\mu\in[1,\infty)}L(\mu)}{L(\mu_0)}=\begin{cases}0, & \text{when } \hat\mu<1\\ 2\log\frac{L(\hat\mu)}{L(1)}, & \text{when } \hat\mu\ge1\end{cases}=\cdots\cdots$

5. 模型诊断

§1 GLM 的 Leverage

1. Leverage: $H=W^{\frac12}X(X^TWX)^{-1}X^TW^{\frac12}$ ($W=diag\{\frac{1}{V(\mu_i)[g'(\mu_i)]^2}\}$)

H 的对角线元素 $h_i=H_{ii}$ 被定义为 leverage of z_i ($z=\eta+diag\{g'(\mu_i)\}(y-\mu)$)

2. 关于 Leverage: ① $\hat\beta=(X^TWX)^{-1}X^TWz \Rightarrow W^{\frac12}X\hat\beta=W^{\frac12}X(X^TWX)^{-1}X^TW^{\frac12}(W^{\frac12}z) \Rightarrow W^{\frac12}\hat\eta=H(W^{\frac12}z) \Rightarrow \hat\eta_i/\sqrt{w_i}=\sum_j H_{ij}\cdot z_j/\sqrt{w_j}$

其中 $w_i=V(\mu_i)[g'(\mu_i)]^2$ (W 对角线元素的倒数)

② $h_i=H_{ii}$ 表示计算 $\hat\eta_i$ 时 z_i 占的 weight

§2 GLM 的 residuals

1. Pearson residuals: $r_{Pi}=\frac{y_i-\hat\mu_i}{\sqrt{V(\hat\mu_i)}}$ ~ Normal roughly (若 CLT holds for ⋯)

2. deviance residuals: $r_{Di}=sign(y_i-\hat\mu_i)\sqrt{d(y_i,\hat\mu_i)}$ (若 spa accurate)

3. quantile residuals (continuous): $r_{Qi}=\Phi^{-1}(F(y_i;\hat\mu_i,\hat\phi))$ (可用 $\bar\phi$ 若未知)

quantile residuals (discrete): $a=\lim_{\varepsilon\to0^-}F(y_i+\varepsilon,\hat\mu_i,\hat\phi)$, $b=F(y_i;\hat\mu_i,\hat\phi)$

$u\sim U(a,b)$, $r_{Qi}=\Phi^{-1}(u)$

4. Pearson residuals 和 deviance residuals 的 variance

$X^2=\sum_{i=1}^n w_i(z_i-\hat\eta_i)^2=\sum_{i=1}^n\frac{(y_i-\hat\mu_i)^2}{V(\hat\mu_i)} \Rightarrow r_{Pi}=\frac{y_i-\hat\mu_i}{\sqrt{V(\hat\mu_i)}}=w_i^{\frac12}(z_i-\hat\eta_i)$, $r_P=W^{\frac12}(z-\hat\eta)$

$\Rightarrow var(r_P)=var(W^{\frac12}(z-\hat\eta))=var((I-H)W^{\frac12}z)=(I-H)W^{\frac12}var(z)W^{\frac12}(I-H)$

$var(z)=diag\{g'(\mu_i)\}var(y)diag\{g'(\mu_i)\}=\phi W^{-1}$ ($W=diag\{\frac{1}{V(\mu_i)[g'(\mu_i)]^2}\}$)

$\Rightarrow var(r_P)=\phi(I-H)$, $var(r_{Pi})=\phi(1-h_i)$ (r_D 一样)

5. standardized residuals

① $r_{Pi}'=\frac{r_{Pi}}{\sqrt{\phi(1-h_i)}}=\frac{y_i-\hat\mu_i}{\sqrt{\phi V(\hat\mu_i)(1-h_i)}}$ ② $r_{Di}'=\frac{r_{Di}}{\sqrt{\phi(1-h_i)}}=\frac{sign(y_i-\hat\mu_i)\cdot\sqrt{d(y_i,\hat\mu_i)}}{\sqrt{\phi(1-h_i)}}$

③ $r_{Qi}'=\frac{r_{Qi}}{\sqrt{(1-h_i)}}$ (没有 ϕ)

6. 三种 residuals 的选取 ③ 对于 discrete EDMs, 建议使用 r_Q

① 若 $Y\sim N(\mu,\sigma^2)$, 则均 Normal, 若 $Y\sim IG$, 则 r_D 和 r_Q Normal

② 若 CLT 和 saddlepoint approximation 条件不满足, 则使用 r_Q

§3 Model assumptions 的检验

1. Check random component: Q-Q plots for quantile residual r_Q

2. Check systematic component: overall structure:

standardized residual plots r_D'/r_Q' v.s. $\hat\mu$ 与 r_D'/r_Q' v.s. x_j

3. Check systematic component: link function: plot working responses z_i v.s. $\hat\eta_i$ ($=\hat\eta_i+e_i$)

4. Check systematic component: inclusion form of x_j: partial residual $u_{\cdot j}$ v.s. $x_{\cdot j}$

5. Check systematic component: missing variables: F 检验 $u_{\cdot j}=e_i+\hat\beta_jx_{\cdot j}$

§4 Outlier 与 influential points

(quantile 若 discrete data)

1. Outlier 的识别 ① 法一: 作出 standardized (deviance) residuals 的 Q-Q plot

② 法二: 求出 approximate studentized residuals: $|r_i''|>3 \Rightarrow$ outlier

$r_i''=sign(y_i-\hat\mu_i)\sqrt{\frac{1}{\phi}(r_{D,i}^2+\frac{h_i}{1-h_i}r_{P,i}^2)}$, $\bar\phi(i)=\frac{D(y,\hat\mu)-r_{D,i}^2/(1-h_i)}{n-p'-1}$ (若ϕ未知)

2. Influential points 的识别 (若 ϕ 未知, 则估计为 Pearson estimator $\bar\phi$)

① approximated Cook's distance: $D_i\approx(\frac{r_{P,i}}{1-h_i})^2\frac{h_i}{\phi p'}=\frac{(r_{P,i}')^2}{p'}\frac{h_i}{1-h_i}>F(p',n-p')$ 的 50-th percentile

② $|DFFITS_i|=|\frac{\hat\mu_i-\hat\mu_{i(i)}}{\bar\phi_{(i)}}|=|r_{D,i}''\frac{h_i}{1-h_i}|\ge1$ (小的 dataset) 或 $\ge2\sqrt{p'/n}$

③ $|DFBETAS_i|=|\frac{\hat\beta_j-\hat\beta_{j(i)}}{se(\hat\beta_{j(i)})}|>1$ ⑤ leverage $h_i>3p'/n \Rightarrow$ high-leverage

④ $CR_i=\frac{1}{1-h_i}\{\frac{n-p}{n-p'+(r_{D,i}'')^2}\}^p>3p'/(n-p')$

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

④ $X=\begin{bmatrix}1&x_{11}&\cdots&x_{1p}\\ 1&x_{21}&\cdots&x_{2p}\\ \vdots&\vdots&\ddots&\vdots\\ 1&x_{n1}&\cdots&x_{np}\end{bmatrix}\in R^{n\times p'}$

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

④ $X=\begin{bmatrix}1&x_{11}&\cdots&x_{1p}\\ 1&x_{21}&\cdots&x_{2p}\\ \vdots&\vdots&\ddots&\vdots\\ 1&x_{C1}&\cdots&x_{Cp}\end{bmatrix}\in R^{C\times p'}$

3. Link function 的选取: 常见的 link function (此处为 η 和 π 的关系) 包括:

① Logit link: $\eta=\log\frac{\pi}{1-\pi}$ (等价于 $B(\pi_i)$ 和 $Bin(n_c,\pi_c)$ 的 canonical link)

② Probit link: $\eta=\Phi^{-1}(\pi)$, 其中 Φ 为 $N(0,1)$ 的 CDF $\Rightarrow \pi=\Phi(\eta)$

③ Complementary log-log link: $\eta=\log(-\log(1-\pi)) \Rightarrow \pi=1-\exp\{-\exp\{\eta\}\}$

§2 Binomial GLM 的 estimation 和 inference

1. Binomial GLM 的 MLE

① ungrouped: $L(\beta)=\prod_{c=1}^C\pi_c^{y_c}(1-\pi_c)^{n_c-y_c} \Rightarrow l(\beta)=\sum_{c=1}^C\{y_c\log\frac{\pi_c}{1-\pi_c}+n_c\log(1-\pi_c)\}$

② grouped: $L(\beta)=\prod_{c=1}^C\binom{n_c}{y_c}\pi_c^{y_c}(1-\pi_c)^{n_c-y_c} \Rightarrow l(\beta)=\sum_{c=1}^C\{y_c\log\frac{\pi_c}{1-\pi_c}+n_c\log(1-\pi_c)\}+\sum_{c=1}^C\log\binom{n_c}{y_c}$

2. Binomial GLM 的 MLE 的 variance

① General cases: $Var(\hat\beta)=(X^TWX)^{-1}$, $W=diag\{\frac{1}{n_c\pi_c(1-\pi_c)[g'(n_c\pi_c)]^2}\}$

② Canonical link (grouped): $g(\mu_c)=\log\frac{\mu_c}{n_c-\mu_c} \Rightarrow g'(\mu_c)=\frac{n_c}{\mu_c(n_c-\mu_c)} \Rightarrow W=diag\{n_c\pi_c(1-\pi_c)\}$

3. Binomial GLM 的 inference

① Single β_j 的 CI: $\hat\beta_j\pm z_{\alpha/2}\hat{s.e.}(\hat\beta_j)$

② Wald test: $H_0:\beta_j=0$, $Z_W=\frac{\hat\beta}{\hat{s.e.}(\hat\beta)}\sim N(0,1)$ 或 $(Z_W)^2\sim\chi^2_1$

③ Predicted probability at x ($\hat\pi=\frac{e^{\hat\eta}}{1+e^{\hat\eta}}$, with $\hat\eta=x^T\hat\beta$) 的 CI

$Var(\hat\eta)=x^TVar(\hat\beta)x \Rightarrow \eta$ 的 CI $(\eta_L,\eta_U)=\hat\eta\pm z_{\alpha/2}\hat{s.e.}(\hat\eta)$

$\Rightarrow \pi$ 的 CI 为 $(\pi_L,\pi_U)=(\frac{e^{\eta_L}}{1+e^{\eta_L}},\frac{e^{\eta_U}}{1+e^{\eta_U}})$

§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 ($x=0,1,2$ (A, B, C) 定义 $x_1=1_{x=1}$, $x_2=1_{x=2}$)

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

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

3. 情况三: one continuous covariate $\log\frac{\pi(x)}{1-\pi(x)}=\beta_0+\beta_1x$

$\begin{cases}\text{odds for }x=\frac{\pi(x)}{1-\pi(x)}=\exp(\beta_0+\beta_1x)\\ \text{odds for }x+1=\frac{\pi(x+1)}{1-\pi(x+1)}=\exp(\beta_0+\beta_1x+\beta_1)\end{cases} \Rightarrow OR_{x+1\text{ vs. }x}=\frac{\exp(\beta_0+\beta_1x+\beta_1)}{\exp(\beta_0+\beta_1x)}=e^{\beta_1}$

$\Rightarrow$ Interpretation: 1 unit of increase in x leads to e^{β_1} times increases in odds

4. 情况四: one binary covariate 和 one continuous covariate (with interaction)

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

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

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

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

$\Rightarrow l(\hat\mu)=\sum_{i=1}^n t(y_i,y_i)-\frac12D(y,\hat\mu)$

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

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

2. grouped data model 的 deviance

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

$\Rightarrow l^G(\hat\mu^G)=\sum_{c=1}^C t(y_c,y_c)-\frac12D^G(y,\hat\mu^G)+\sum_{c=1}^C\log\binom{n}{y_c}$

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

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

3. 两种 model 的 deviance 不同

$l^G(\hat\mu^G)=l(\hat\mu)+\sum_{c=1}^C\log\binom{n}{y_c} \Rightarrow D(y,\hat\mu)-D^G(y,\hat\mu^G)=2\{\sum_{i=1}^n t(y_i,y_i)-\sum_{c=1}^C t(y_c,y_c)\}$

$\begin{cases}t(y_i,y_i)=y_i\log\frac{y_i}{1-y_i}+\log(1-y_i)=\log y_i^{y_i}(1-y_i)^{1-y_i}=0, & \forall y_i=0\text{ 或 }1\\ t(y_c,y_c)=y_c\log\frac{y_c/n_c}{1-y_c/n_c}+n_c\log(1-y_c/n_c)\ne0\end{cases}$

$\Rightarrow D(y,\hat\mu)\ne D^G(y,\hat\mu^G)$

4. ungrouped data model 没有 goodness of fit test

对于 ungrouped data model, 有

$D(y,\hat\mu)=-2\sum_i\{y_i\,logit(\hat\mu_i)+\log(1-\hat\mu_i)\}$

$=-2\{n_1\sum_{i,y_i=1}logit(\hat\mu_i)+\sum_i\log(1-\hat\mu_i)\}$

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

§5 overdispersion

1. Overdispersion 及其后果: variation in data $>n\pi(1-\pi)$

2. Overdispersion 的成因: ① π_i not constant ② n cases 正相关

3. Overdispersion 的识别: goodness of fit test fail + 包含所有变量 + 没有 outlier 和 influential points

4. Overdispersion 的处理: quasi-binomial: $Var(y)=\phi V(y)$