Summingup2018

1 Univariate exponential family (MLL-8)

$f(y_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi} w_i + c(y_i, w_i, \phi)\right)$

parameter of interest; Not θ_i; dispersion parameter; not of interest (variance)

$E(Y_i) = b'(\theta_i)$

$Var(Y_i) = b''(\theta_i) \cdot \frac{\phi}{w_i}$

Normal, Bernoulli, Binomial (n given), Poisson, Gamma, ⋮

canonical link $\theta_i = \eta_i$ } nice, things happen

- concave log-likelihood
- $H(\beta) = F(\beta)$ (on exp)

Normal regression : MR
Binary regression : MB
Poisson regression, Gamma } : MM

2 MLR (MLR)

$Y = X\beta + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I)$ (response, design matrix, parameters of interest)

$E(Y) = X\beta$, $Cov(Y) = \sigma^2 I$

$Y \sim N(\mu, \sigma^2 I)$

or

$Y_i \sim N(\mu_i, \sigma^2)$

$Y_1, \dots, Y_n$ independent (univariate exp fam)

add correlated responses

3 MLR (prepared for GLM)

1) $Y_i \sim N$, $\mu_i = E(Y_i)$, $Var(Y_i) = \sigma^2$
2) $\eta_i = x_i^T\beta$
3) $\eta_i = \mu_i$

4 GLM Model

1) $Y_i \sim f$ univ. exp fam, $E(Y_i) = \mu_i$, $Var(Y_i) = \sigma_i^2$
2) $\eta_i = x_i^T\beta$
3) $\eta_i = g(\mu_i)$ (link) and $\mu_i = h(\eta_i)$ (response)

canonical link: $E(Y_i) = b'(\theta_i)$, $\eta_i = g(\mu_i)$, $\mu_i = h(\eta_i)$

$\theta_i \longleftrightarrow \mu_i \longleftrightarrow \eta_i \longleftrightarrow \beta$

$E(Y_i)$

7 Requirements GLM model

+ Y_i's independent
+ X full rank ($(X^TX)^{-1}$ possible)
+ one-to-one and twice differentiable link function

5 GLM likelihood inference. AIM ML $\hat\beta$ (maximum likelihood)

$L(\beta) = \prod_{i=1}^n L_i(\beta)$ and $l(\beta) = \sum_{i=1}^n \ln(L_i(\beta)) = \sum_{i=1}^n l_i(\beta)$

Work with $l_i(\beta)$ and chain rule for differentiation:

$s(\beta) = \frac{\partial l}{\partial \beta} = \sum_{i=1}^n \frac{(y_i - \mu_i)\, x_i\, h'(\eta_i)}{Var(Y_i)} = X^T D \Sigma^{-1}(y - \mu)$ ($p \times 1$)

$D = diag(h'(\eta_i))$ ($n \times n$)

$\Sigma = diag(Var(Y_i))$ ($n \times n$)

$s(\beta)$ = score function

$E(s(\beta)) = 0$ (also $E(s_i(\beta)) = 0$)

solve $s(\hat\beta) = 0$ to get $\hat\beta$ MLE

$F(\beta)$ = expected Fisher info. matrix ($p \times p$)

$F(\beta) = Cov(s(\beta)) = E\left(-\frac{\partial^2 l}{\partial\beta\partial\beta^T}\right) = X^T W X$ ($p \times n$, $n \times n$, $n \times p$)

$W = diag\left(\frac{h'(\eta_i)^2}{Var(Y_i)}\right)$ ($n \times n$)

6 Fisher-scoring to find $\hat\beta$ numerically

$\hat\beta^{(t+1)} = \hat\beta^{(t)} + F^{-1}(\hat\beta^{(t)}) \cdot s(\hat\beta^{(t)})$

For GLM this can be written as an IRWLS (iteratively reweighted least squares)

$\hat\beta^{(t+1)} = (X^T \underbrace{W(\hat\beta^{(t+1)})}_{\text{working weights}} X)^{-1} X^T W(\hat\beta^{(t+1)}) \tilde y^{(t)}$

working response vector: $x_i^T\hat\beta^{(t)} + \frac{y_i - h(x_i^T\hat\beta^{(t+1)})}{h'(x_i^T\hat\beta^{(t+1)})}$

Asymptotic distribution of $\hat\beta$: $\hat\beta \approx N(\beta, F^{-1}(\hat\beta))$ — Two important uses of $F^{-1}(\hat\beta)$

Requirements: $n \to \infty$, $\lambda_{min} F_n \to \infty$ } too advanced for us!

for all $\delta > 0$ $\max_\beta \|U_n(\beta) - I\| \to 0$ $U_n(\beta) = F_n^{-\frac{1}{2}} F_n(\beta) F_n^{-\frac{1}{2}}$ (normed info matrix)

15 Binary regression

$f(y_i) = \binom{1}{y_i} \pi_i^{y_i}(1-\pi_i)^{1-y_i}$

1) $Y_i \sim bin(1, \pi_i)$, $\mu_i = E(Y_i) = \pi_i$ ← univariate exponential (show!) (family)
2) $\eta_i = x_i^T\beta$ [x_i is $p \times 1$ column vector]
3) $\eta_i = g(\mu_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right)$ } logit link (= canonical)

$\pi_i = h(\eta_i) = \frac{\exp(\eta_i)}{1+\exp(\eta_i)}$

Interpretation of β with logit link

If x_{ij} increase from x_{ij} to $x_{ij}+1$ then $\frac{\pi_i}{1-\pi_i}$ (odds) change with a multiplicative factor e^{β_j}

$\frac{P(Y_i = 1 \mid x_{i1}+1)}{P(Y_i = 0 \mid x_{i1}+1)} : \frac{P(Y_i = 1 \mid x_{i1})}{P(Y_i = 0 \mid x_{i1})} = e^{\beta_1}$

Extension to

$Y_i \sim bin(n_i, \pi_i)$ for grouped data →

still use logit link $\eta_i = logit(\pi_i)$

Same covariate pattern for all n_i, $i = 1, \dots, G$ (groups)

$n_i = 1$: for only continuous covariates

Likelihood: $L(\beta) = \prod_{i=1}^G \pi_i^{y_i}(1-\pi_i)^{n_i - y_i}\binom{n_i}{y_i}$

Score function: $s(\beta) = \sum_{i=1}^G x_i(y_i - n_i\pi_i)$

Expected Fisher info: $F(\beta) = \sum_{i=1}^G x_i x_i^T n_i \pi_i(1-\pi_i)$

Deviance: $D = 2\sum_{i=1}^G\left[y_i \ln\left(\frac{y_i}{n_i\hat\pi_i}\right) + (n_i - y_i)\ln\left(\frac{n_i - y_i}{n_i - n_i\hat\pi_i}\right)\right]$

saturated model has $\tilde\pi_i = \frac{y_i}{n_i}$ (inserted in D above)

16 Poisson regression

1) $Y_i \sim Poisson(\mu_i)$. $f(y_i) = \frac{\mu_i^{y_i}}{y_i!}e^{-\mu_i}$, $y_i = 0, 1, 2, \dots$, $\mu_i > 0$

$E(Y_i) = Var(Y_i) = \mu_i$

2) $\eta_i = x_i^T\beta$
3) $\eta_i = \log(\mu_i)$ ← canonical link (= log link)

$\mu_i = \exp(\eta_i)$

Interpretation: $\mu_i = \exp(\eta_i) = \exp(x_i^T\beta)$

$= e^{\beta_0} \cdot e^{\beta_1 x_{i1}} \dots e^{\beta_k x_{ik}} = e^{\beta_0}(e^{\beta_1})^{x_{i1}} \dots (e^{\beta_k})^{x_{ik}}$

If $\beta_1 > 0$: then $\exp(\beta_1) > 1$ so μ_i increases if x_{i1} increases

$\beta_1 < 0$: < 1 μ_i decreases

$\beta_1 = 0$: $= 1$ no change in μ_i if x_{i1} change

Likelihood: $L(\beta) = \prod_{i=1}^n \frac{\mu_i^{y_i}}{y_i!}e^{-\mu_i}$

Log-likelihood: $l(\beta) = \sum_{i=1}^n[y_i \ln\mu_i - \mu_i - \ln(y_i!)] = \sum_{i=1}^n y_i x_i^T\beta - \sum_{i=1}^n \exp(x_i^T\beta) + C$

Score function: $s(\beta) = \sum_{i=1}^n(y_i - \mu_i)x_i$

Expected Fisher info: $F(\beta) = \sum_{i=1}^n x_i x_i^T \mu_i$

Deviance: $D = 2\sum_{i=1}^n\left[y_i \ln\left(\frac{y_i}{\hat y_i}\right) - (y_i - \hat y_i)\right]$ ($\hat y_i = \exp(x_i^T\hat\beta)$)

$\sum y_i = \sum \hat y_i$ for log link when intercept present

Rate models and offset:

Want model for $\frac{Y_i}{t_i}$ instead of Y_i, so $E(\frac{Y_i}{t_i}) = \frac{\mu_i}{t_i}$

Let $\log(\mu_i/t_i) = x_i^T\beta \Leftrightarrow x_i^T\beta = \underbrace{\log(t_i)}_{\text{offset}} + \log(\mu_i)$

Proceed as usual, only add offset = $\log(t_i)$ in R

17 Gamma-regression

1) $Y_i \sim Gamma(\mu_i, \nu)$. $f(y_i) = \frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu_i}\right)^\nu y_i^{\nu-1}\exp\left(-\frac{\nu}{\mu_i}y_i\right)$, $y_i > 0$
2) $\eta_i = x_i^T\beta$
3) $\eta_i = g(\mu_i) = -\frac{1}{\mu_i}$ is canonical link, but

$\eta_i = g(\mu_i) = \log(\mu_i)$ is more popular ← "same" interpretation as for Poisson with log-link (similar)

The gamma has ν as nuisance parameter.

Categorical regression

— TO BE WRITTEN —
NEW 2018

8 Further GLM inference [CI, hyp., deviance]

CI for β_j: $[\hat\beta_j \pm z_{\frac{\alpha}{2}}\hat{SD}(\hat\beta_j)]$ where $\hat{SD}(\hat\beta_j) = \sqrt{F^{-1}(\hat\beta)_{[j,j]}}$

Hyp: $H_0: C\beta = d$ (r × p, r × 1) vs $H_1: C\beta \neq d$ (r hypotheses)

$H_0: \beta_j = 0$ vs $H_1: \beta_j \neq 0$ ($C = [0 \dots 0\ 1\ 0 \dots 0]$)

NB: $\left(\frac{\hat\beta_j - 0}{\hat{SD}(\hat\beta_j)}\right)^2 \sim \chi^2_1$ is also Wald test (summary print-out glm in R)

Wald test: $Z^2 = (C\hat\beta - d)^T[CF^{-1}(\hat\beta)C^T]^{-1}(C\hat\beta - d) \approx \chi^2_r$

P-value: $P(\chi^2_r > z^2)$ (observed value)

Note that the LRT can be calculated from the deviances of two models

9 Deviance (test) for GLM

Likelihood-ratio test (LRT): B nested in A

A: larger model ["H_0 true"] → $\hat\beta_A$ with $L(\hat\beta_A)$

B: smaller model ["H_0 true"] → $\hat\beta_B$ with $L(\hat\beta_B)$

$LRT = -2[\ln L(\hat\beta_B) - \ln L(\hat\beta_A)] \approx \chi^2_{\#A - \#B}$ (par in A − par in B = r)

P-value: $P(\chi^2_r > lrt)$ (obs. value of LRT)

Requirement $LRT \approx \chi^2_r$: $n \to \infty$

A: saturated model = same number of param. as number of covariate patterns n ← G patterns (perfect fit to the data)

B: candidate model = our model

Deviance $D = -2[\ln L(\hat\beta_B) - \ln L(\hat\beta_A)] \approx \chi^2_{G-p}$ (observed = d; $\hat\beta_A$ a function of the y's in each cov pattern)

If we find that $P(\chi^2_{G-p} > d) \leq \alpha$ (e.g. 0.05) then our model is very far from the saturated model — and that is a bad thing. Then we "reject: H_0 that our model is good". NB!!

NB: for Poisson and binomial (grouped) GLM we need the number of obs for each covariate pattern to be large to trust the χ^2-distribution assumption.

10 AIC for GLM

: can be used to choose between models — both nested and not (but Y needs to be the same in both models)

$AIC = -2l(\hat\beta, \hat\phi) + 2p^*$

$p^* = p + \#$par in ϕ (1 for Bin, Pois, + gamma)

can also compare link functions

same!

11 LMM

$Y_i = X_i\beta + U_i\gamma_i + \varepsilon_i$ where $\varepsilon_i \sim N(0, \sigma^2 I)$ (design matrices; Conditional model $Y_i \mid \gamma_i \sim N(X_i\beta + U_i\gamma_i, \sigma^2 I)$)

($n_i \times p$, $p \times 1$, $n_i \times q$, $q \times 1$, $n_i \times 1$; $\gamma_i \sim N(0, Q)$ ($q \times 1$), $q \times q$)

$N = \sum_{i=1}^m n_i$

and ε_i and γ_i independent for $i = 1, \dots, m$

Marginal model:

$Y_i = X_i\beta + \varepsilon_i^*$ where $\varepsilon_i^* = U_i\gamma_i + \varepsilon_i \sim N(0, \underbrace{U_iQU_i^T + \sigma^2 I}_{V_i})$

so $Y_i \sim N(X_i\beta, V_i)$ ($n_i \times 1$, $n_i \times n_i$)

Likelihood: $L(\beta, Q, \sigma^2) = \prod_{i=1}^m f(y_i \mid \beta, \theta)$ (θ: Q, σ^2's par)

ML: $\hat\beta = [\sum_{i=1}^m (X_i^T\hat V_i^{-1}X_i)]^{-1}\sum_{i=1}^m(X_i^T\hat V_i^{-1}Y_i)$

$\hat\beta \sim N(\beta, (\sum_{i=1}^m(X_i^TV_i^{-1}X_i))^{-1})$ and ≈ when $\hat V_i$ replace V_i (asymp.)

REML: $\hat V_i$: found from maximizing $A^TY \sim N(0, A^TCov(Y)A)$ (matrix A could be a $(N-p) \times N$ "version" of $(I-H)$) where and gives less biased estimator than using MLL

Prediction of random effects γ (BLUP) and ε

$\hat\gamma_i = \hat QU_i^T\hat V_i^{-1}(Y_i - X_i\hat\beta)$
$\hat\varepsilon_i = Y_i - X_i\hat\beta - U_i\hat\gamma_i$
} can be plotted to check model assumptions

12 LMM: random intercept model

$q = 0$

$Y_{ij} = \underbrace{X_{ij}^T\beta}_{\beta_0 + \beta_1 x_{ij} \text{ if } p=2} + \gamma_{0i} + \varepsilon_{ij}$

$X_i = \begin{bmatrix} X_{i1}^T \\ \vdots \\ X_{in_i}^T \end{bmatrix}$ ($n_i \times p$; X_{ij} ($p \times 1$ vector), $1 \times p$ vector)

γ_{0i} i.i.d $N(0, \tau_0^2)$

For the random intercept model $V_i = \sigma^2 I + \tau_0^2 11^T$ and

$Cov(Y_{ij}, Y_{il}) = \frac{\tau_0^2}{\tau_0^2 + \sigma^2}$ called intraclass correlation (ICC)

Plots of x_{ij} vs $\beta_0 + \gamma_{0i} + \beta_1 x_{ij}$ (many parallel lines)

LMM: random intercept and slope (one x_{ij}) ($p = 2$), $q = 1$

$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_{0i} + \gamma_{1i}x_{ij} + \varepsilon_{ij}$

$\begin{bmatrix}\gamma_{0i} \\ \gamma_{1i}\end{bmatrix} \sim N\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2\end{pmatrix}\right)$

$Cov(Y_{ij}, Y_{il})$ not so easy to interpret since dependent on x_{ij} and x_{il}.

13 Model selection for LMM

Fixed effects can be tested with Wald or LRT from ML estimation

Wald: $H_0: C\beta = d$ (r × p) vs $H_1: C\beta \neq d$

$(C\hat\beta - d)^T[C(\sum_{i=1}^m X_i^T\hat V_i^{-1}X_i)^{-1}C^T]^{-1}(C\hat\beta - d) \approx \chi^2_r$

LRT: $-2[\ln L(\hat\beta_B) - \ln L(\hat\beta_A)] \approx \chi^2_{\#\text{par } A - \#\text{par } B}$

A: model under H_1
B: model under H_0

For testing of random effects: trouble because on boundary of param space and normality of $\hat\theta$ problematic

→ instead: use AIC to compare models

need to either choose ML or REML

$AIC_{ML} = -2\underbrace{l(\hat\beta, \hat\theta)}_{ML} + 2p^*$ (both β's and # par in θ) $AIC_{REML} = -2\underbrace{l(\hat\beta, \hat\theta)}_{REML} + 2p^*$

so $p + 1 + 1$ for random intercept
$p + 1 + 2$ or 3 for random intercept and slope ($Q = \begin{bmatrix}\tau_0^2 & 0 \\ 0 & \tau_1^2\end{bmatrix}$ or $\begin{bmatrix}\tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2\end{bmatrix}$)

14 GLMM

: correlated non-normal responses Y_{ij} $i = 1, \dots, m$, $j = 1, \dots, n_i$

1) Distributional assumptions

$Y_{ij} \mid \gamma_i$ has conditional distribution $f(y_{ij} \mid \gamma_i)$ [binomial, Poisson, ...]

2) Structural assumptions:

$\mu_{ij} = E(Y_{ij} \mid \gamma_i)$ linked to $\eta_{ij} = X_{ij}^T\beta + U_{ij}^T\gamma_i$ ($p \times 1$, $p \times 1$, $q \times (q+1)$, $(q+1) \times 1$)

$\eta_{ij} = g(\mu_{ij})$

$\mu_{ij} = h(\eta_{ij})$

$X_i = \begin{bmatrix}X_{i1}^T \\ X_{i2}^T \\ \vdots \\ X_{in_i}^T\end{bmatrix}$ $U_i = \begin{bmatrix}U_{i1}^T \\ \vdots \\ U_{in_i}^T\end{bmatrix}$

3) Distributional assumption for random effects

γ_i i.i.d $N(0, Q)$ ($q \times 1$, $q \times (q+1)$)

FIG? Remember simulation results to show correction between conditional and marginal mean

Marginal model: $f(y_{ij}) = \int_{\gamma_i} f(y_{ij} \mid \gamma_i) f(\gamma_i)\, d\gamma_i$

not on closed form!

($f(y_{ij} \mid \gamma_i)$: bin, Pois, Gamma; $f(\gamma_i)$: always (m)normal)

Which makes maximum likelihood estimation numerically challenging. Several solutions exists.

We have not looked at CI and testing since asymptotic distr. of parameters challenging for small samples. Use asymptotic results on $\hat\beta \approx N(\beta, F(\hat\beta)^{-1})$ for fixed effects.