Today

- HW 2: due March 4, 11.59 pm.
- Generalized Linear Models Chs. 6 and 7 SM 10.2,3
- after mid-term break: random effects, mixed linear and non-linear models, nonparametric regression methods
- In the News: measles

Generalized linear models: theory

- $$f(y_i; \mu_i, \phi_i) = \exp\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i; \phi_i)\}$$
- $E(y_i \mid x_i) = b'(\theta_i) = \mu_i$ defines μ_i as a function of θ_i
- $g(\mu_i) = x_i^T\beta = \eta_i$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_i is the linear predictor
- $\text{Var}(y_i \mid x_i) = \phi_i b''(\theta_i) = \phi_i V(\mu_i)$
- $V(\cdot)$ is the variance function

Examples

- Normal
- Binomial
- Poisson
- Gamma/Exponential
- Inverse Gaussian

family {stats} R Documentation

Family Objects for Models

Description

Family objects provide a convenient way to specify the details of the models used by functions such as glm. See the documentation for glm for the details on how such model fitting takes place.

Usage

```
family(object, ...)

binomial(link = "logit")
gaussian(link = "identity")
Gamma(link = "inverse")
inverse.gaussian(link = "1/mu^2")
poisson(link = "log")
quasi(link = "identity", variance = "constant")
quasibinomial(link = "logit")
quasipoisson(link = "log")
```

Examples

- Normal: $f(y_i;\mu_i,\sigma^2) = \frac{1}{\sqrt(2\pi)\sigma}\exp\{-\frac{1}{2\sigma^2}(y_i-\mu_i^2)\}$

  $= \exp\{\frac{y_i\mu_i - (1/2)\mu_i^2}{\sigma^2} - (1/2)\log\sigma^2 - y_i^2/2\sigma^2 - (1/2)\log\sqrt(2\pi)\}$

  $\phi_i = \sigma^2, \quad \theta_i = \mu_i, \quad b(\mu_i) = \mu_i^2/2\sigma^2$ note $b''(\mu_i) = 1$

- Binomial: $f(r_i;p_i) = \binom{m_i}{r_i} p_i^{r_i}(1-p_i)^{m_i-r_i}; \quad y_i = r_i/m_i$

  $= \exp[m_i y_i \log\{p_i/(1-p_i)\} + m_i\log(1-p_i) + \log\binom{m_i}{m_i y_i}]$

  $\phi_i = 1/m_i, \quad \theta_i = \log\{p_i/(1-p_i)\}, \quad b(p_i) = -\log(1-p_i)$

  Note $p_i = \mu_i = \mathrm{E}(y_i)$

- ELM (p.115) uses $a_i(\phi)$ in place of ϕ_i, later (p.117) $a_i(\phi) = \phi/w_i$; later (p.118) w_i used for weights in IRWLS algorithm;
  SM uses ϕ_i, later (p. 483) $\phi_i = \phi a_i$

Inference

using Blackboard Notes, Feb 11

- $\ell(\beta; y) = \sum\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i)\}$
- $b'(\theta_i) = \mu_i; \quad g(\mu_i) = g(b'(\theta_i)) = \eta_i = x_i^{\mathrm{T}}\beta$
- $\dfrac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \dfrac{\partial \ell_i}{\partial \theta_i}\dfrac{\partial \theta_i}{\partial \beta_j} = \sum \dfrac{y_i - b'(\theta_i)}{\phi_i}\dfrac{\partial \theta_i}{\partial \beta_j}$
- $g'(b(\theta_i))b''(\theta_i)\dfrac{\partial \theta_i}{\partial \beta_j} = x_{ij} = g'(\mu_i)V(\mu_i)$ See Slide 2
- $\dfrac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \dfrac{y_i - \mu_i}{\phi_i g'(\mu_i)V(\mu_i)}x_{ij} = \sum \dfrac{y_i - \mu_i}{a_i g'(\mu_i)V(\mu_i)}x_{ij}$

  when $\phi_i = a_i\phi$
- matrix notation:

$$\frac{\partial \ell(\beta)}{\partial \beta} = X^{\mathrm{T}}u(\beta), \quad X = \frac{\partial \eta}{\partial \beta^{\mathrm{T}}}, \quad u = (u_1, \dots, u_n)$$

Scale parameter ϕ_i

- in most cases, either ϕ_i is known, or $\phi_i = \phi a_i$, where a_i is known

- Normal distribution, $\phi = \sigma^2$

- Binomial distribution $\phi_i = m_i^{-1}$

- Gamma distribution, $\phi = 1/\nu$

- $$\frac{\partial \ell(\beta; y)}{\partial \beta_j} = \sum \frac{y_i - \mu_i}{\phi_i g'(\mu_i) V(\mu_i)} x_{ij} = \sum \frac{y_i - \mu_i}{a_i g'(\mu_i) V(\mu_i)} x_{ij}$$
  when $\phi_i = a_i \phi$

- if $\theta_i = g(\mu_i)$ canonical link, then $g'(\mu_i) = 1/V(\mu_i)$, and
  $$\sum \frac{y_i x_{ij}}{a_i} = \sum \frac{y_i \hat{\mu}_i x_{ij}}{a_i}$$

Solving maximum likelihood equation

- Newton-Raphson: $\ell'(\hat\beta) = 0 \approx \ell'(\beta) + (\hat\beta - \beta)\ell''(\beta)$

  defines iterative scheme

- $\hat\beta^{(t+1)} = \hat\beta^{(t)} - \{\ell''(\hat\beta^{(t)})\}^{-1}\ell'(\hat\beta^{(t)})$

- Fisher scoring: $-\ell''(\beta) \leftarrow \mathrm{E}\{-\ell''(\beta)\} = i(\beta)$

  many books use $I(\beta)$

- $\hat\beta^{(t+1)} = \hat\beta^{(t)} + \{i(\hat\beta^{(t)})\}^{-1}\ell'(\hat\beta^{(t)})$

- applied to matrix version:

  $X^{\mathrm{T}}u(\hat\beta) = 0 \doteq X^{\mathrm{T}}u(\beta) + (\hat\beta - \beta)X^{\mathrm{T}}\dfrac{\partial u(\beta)}{\partial\beta^{\mathrm{T}}}$ slide 5

- change to Fisher scoring: $\hat\beta = \beta + i(\beta)^{-1}X^{\mathrm{T}}u(\beta)$

... maximum likelihood equation

- $\hat{\beta} = \beta + i(\beta)^{-1} X^{\mathrm{T}} u(\beta)$

$$\frac{\partial^2 \ell(\beta; y)}{\partial\beta_j \partial\beta_k} = \sum \frac{-b''(\theta_i)}{\phi_i} \left(\frac{\partial\theta_i}{\partial\beta_j}\right) \left(\frac{\partial\theta_i}{\partial\beta_k}\right) + \sum \frac{y_i - b'(\theta_i)}{\phi_i} \frac{\partial^2\theta_i}{\partial\beta_j\partial\beta_k}$$

- $$\mathrm{E}\left(-\frac{\partial^2 \ell(\beta; y)}{\partial\beta_j \partial\beta_k}\right) = \sum \frac{V(\mu_i)}{\phi_i} \frac{x_{ij}}{g'(\mu_i)V(\mu_i)} \frac{x_{ik}}{g'(\mu_i)V(\mu_i)} = \sum \frac{x_{ij}x_{ik}}{\phi_i\{g'(\mu_i)\}^2 V(\mu_i)}$$

- $$\begin{aligned} \hat{\beta} &= \beta + (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}u(\beta) = (X^{\mathrm{T}}WX)^{-1}\{X^{\mathrm{T}}WX\beta + X^{\mathrm{T}}u(\beta)\} \\ &= (X^{\mathrm{T}}WX)^{-1}\{X^{\mathrm{T}}W(X\beta + W^{-1}u(\beta)\} \\ &= (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}Wz \end{aligned}$$

- does not involve ϕ_i

- iterative ly re-weighted least squares — W, z both depend on β
- derived response $z = X\beta + W^{-1}u$ — linearized version of y

Nonlinear least squares

SM §10.2

- $y_i \sim N(\mu_i, \sigma^2)$, independently, $i = 1, \ldots, n$
- generalized linear model with $\theta_i = \mu_i$
- link function $g(\mu_i) = x_i^{\mathrm{T}}\beta = \eta_i$ is non-canonical link

- can be more natural to think of
  $y_i = \eta_i(\beta) + \epsilon_i, i = 1, \ldots, n, \quad \epsilon_i \sim N(0, \sigma^2)$

- as with glms $\hat{\beta}$ can be computed by iteratively re-weighted LS

- $\hat{\beta} = (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}Wz \quad X = X(\hat{\beta}) = \left.\dfrac{\partial\eta(\beta)}{\partial\beta^{\mathrm{T}}}\right|_{\hat{\beta}}$
- as before $W = W(\hat{\beta}) = \mathrm{diag}(w_i); \quad w_i = \mathrm{E}(-\partial^2\ell_i/\partial\eta_i^2)$
- as before $z = z(\hat{\beta}) = (X\beta + W^{-1}u); \quad u_i(\hat{\beta}) = \partial\ell_i(\eta)/\partial\eta_i$

Calcium data: Example 10.1

10.1 · Introduction

Table 10.1 Calcium uptake (nmoles/mg) of cells suspended in a solution of radioactive calcium, as a function of time suspended (minutes) (Rawlings, 1988, p. 403).

Time (minutes)	Calcium uptake (nmoles/mg)		
0.45	0.34170	−0.00438	0.82531
1.30	1.77967	0.95384	0.64080
2.40	1.75136	1.27497	1.17332
4.00	3.12273	2.60958	2.57429
6.10	3.17881	3.00782	2.67061
8.05	3.05959	3.94321	3.43726
11.15	4.80735	3.35583	2.78309
13.15	5.13825	4.70274	4.25702
15.00	3.60407	4.15029	3.42484

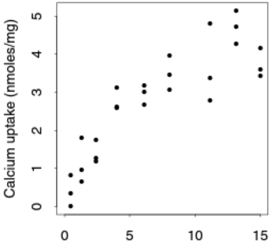


Figure 10.1 Calcium uptake (nmoles/mg) of cells suspended in a solution of radioactive calcium, as a function of time suspended (minutes).

... calcium data

- model
  $E(y_i) = \beta_0\{1 - \exp(-x_i/\beta_1)\}, \quad y_i = E(y_i) + \epsilon_i, \ \epsilon_i \sim N(0, \sigma^2)$
- fitting:

$$\min_{\beta_0,\beta_1} \sum_{j=1}^{n} (y_i - \eta_i)^2$$

- use `nls` or `nlm`; requires starting values
-
```
> library(SMPracticals); data(calcium)
> fit = nls(cal ~ b0*(1-exp(-time/b1)), data = calcium, start = list(b0=5,b1=5))
> summary(fit)
Formula: cal ~ b0 * (1 - exp(-time/b1))

Parameters:
   Estimate Std. Error t value Pr(>|t|)
b0   4.3094     0.3029  14.226 1.73e-13 ***
b1   4.7967     0.9047   5.302 1.71e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5464 on 25 degrees of freedom

Number of iterations to convergence: 3
Achieved convergence tolerance: 9.55e-07
```

... calcium data

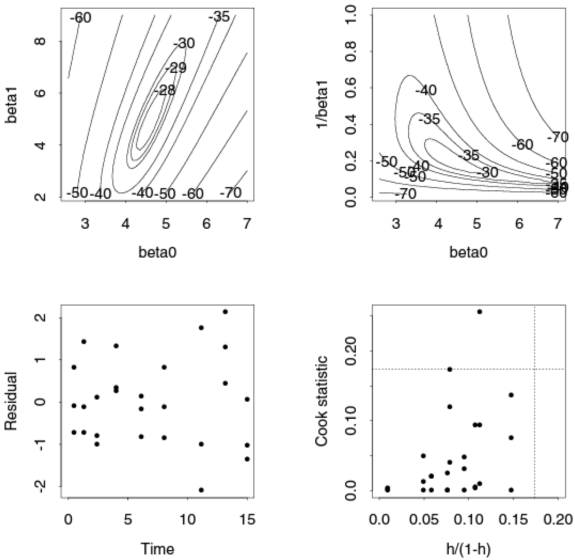


Figure 10.4 Fit of a nonlinear model to the calcium data. Upper left: contours for $\ell_p(\beta_0, \beta_1)$. Upper right: contours for $\ell_p(\beta_0, \gamma_1)$, where $\gamma_1 = 1/\beta_1$. Lower left: standardized residuals plotted against time. Lower right: plot of Cook statistics against $h/(1-h)$, where h is leverage.

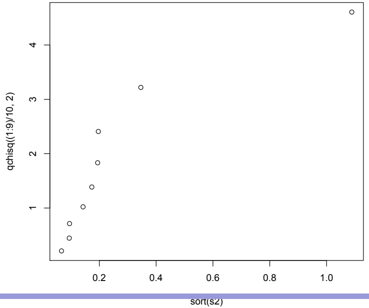
... calcium data

- there are 3 observations at each time point
- can fit a model with a different parameter for each time: $E(y_i) = \eta_i + \epsilon_i$
- the nonlinear model is nested within this; constrains η_i as above
- `anova(lm(cal ~ factor(time), data = calcium))`
- ```
  Analysis of Variance Table

  Response: cal
                Df Sum Sq Mean Sq F value    Pr(>F)
  factor(time)   8 48.437  6.0546  22.720 6.688e-08 ***
  Residuals     18  4.797  0.2665
  ```
- ```
  > deviance(fit) # 7.464514 (mistake in Davison)
  > sum(residuals(fit)^2) # 7.464514
  > (7.464514 - 4.797)/(25 - 18) # 0.3811
  > .3811/.2665
  [1] 1.429919  ## Davison has 1.53
  > pf(1.430,7,18, lower=F)
  [1] 0.2538313
  ```

... calcium data

- checking constant variance assumption
- estimates of σ^2 at each time, each with 2 degrees of freedom
- ```
  > s2 = tapply(calcium$cal, factor(calcium$time), var)
  > s2
  > s2
        0.45        1.3        2.4          4        6.1       8.05
  0.17367258 0.34616902 0.09523507 0.09422579 0.06686923 0.19656739
       11.15      13.15         15
  1.08876166 0.19415027 0.14279290
  > plot(sort(s2),qchisq((1:9)/10,2))
  ```


  ```

Diagnostics

ELM §6.4

- residuals $r_{Pi} = (y_i - \hat{\mu}_i)/\sqrt{V(\hat{\mu}_i)}$ $\quad$ $\mathrm{E}(y_i) = \mu_i, \mathrm{Var}(y_i) = \phi V(\mu_i)$
- $r_{Di} = \mathrm{sign}(y_i - \hat{\mu}_i)\sqrt{d_i}$ $\quad$ $\Sigma r_{Pi}^2 = X^2; \quad \Sigma r_{Di}^2 = \text{Deviance}$
- response residuals: $y_i - \hat{y}_i$ not usually of interest
- working residuals: residuals in last iteration of weighted LS `myglm$residuals`
- instead use
  `residuals(myglm, type = c("deviance", "Pearson"))`
- plot residuals in the usual way: look for non-constant variance, outliers
- plot residuals vs linear predictor; use `qqnorm` or `halfnorm` for outliers

... diagnostics

- linear model $y = X\beta + \epsilon$: $\hat{y} = X\hat{\beta} = X(X^{\mathrm{T}}X)^{-1}X^{T}y$
- hat matrix $H = X(X^{\mathrm{T}}X)^{-1}X^{T}$

- generalized linear model: $\hat{\beta} = (X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}Wz$
- hat matrix $H = W^{1/2}X(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}W^{1/2}$

- leverage of point i = $H_{ii} = h_i$
- `influence(myglm)$hat`

- measures influence of y_i on fitted model
- in the linear model, depend only on X; in glm, depend as well on $\hat{\beta}$

... diagnostics

- case influence: effect of y_i on estimate of β
- `influence(myglm)$coef` $n \times p$ matrix

- Cook's distances:

$$D_i = \frac{2}{p}\{\ell(\hat{\beta}) - \ell(\hat{\beta}_{-i})\}$$

- effect of case i on the 'average' estimation of β
-
$$D_i \approx \frac{h_i}{p(1-h_i)} r_{Pi}^2$$

$h_i = H_{ii}; H = W^{1/2}X(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}W^{1/2}$

- `cooks.distance(myglm)`

see ELM §6.4 for partial residuals, equivalent expresson for D_i

Choosing models

- generalized linear models have two structural components:
- the probability distribution for the response   or just mean and variance
- the regression component: how does the response depend on x

- it is often very helpful to separate these two features
- probability distribution depends on: convenience, standard in the field, consistency with known generating mechanisms, plausible simple starting point, ...

- two or more plausible distributions may lead to same, or different, conclusions
- see e.g. ELM example `wafer` p.137, where log-normal or gamma model give same conclusions
- and `motoring` p.138, where they do not

- example: inverse Gaussian density arises in boundary-crossing problems
- has $V(\mu_i) = \mu_i^3$
- see §7.2 where this is too rapid an increase to fit the data

... models

- fitting generalized linear models uses only $g(\mathrm{E}(y_i)) = x_i^{\mathrm{T}}\beta$ and $\mathrm{var}(y_i) = \phi V(\mu_i)$
- recall score equation:

$$\sum_{i=1}^{n} x_i u_i(\beta) = \sum_{i=1}^{n} x_i \frac{y_i - \mu_i}{g'(\mu_i)\phi_i V(\mu_i)} = 0$$

- $$\mathrm{E}(u_i) = 0, \quad \mathrm{E}\left(-\frac{\partial u_i}{\partial \mu_i}\right) = \mathrm{var}(u_i)$$

- these two properties mimic those of a log-likelihood

$\mathrm{E}(\ell'(\theta)) = 0; \mathrm{var}(\ell'(\theta)) = -\mathrm{E}(\ell''(\theta))$

- suggests that we can use

$$Q(\beta) = \sum Q_i(\beta) = \sum \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi_i V(t)} dt$$

- as a “log-likelihood”, based only on mean and variance
- see §7.4 where this is used for proportion data

Example: poisson regression

SM 10.28

Set 2

Table 2 (Bissell, 1972) gives the numbers of faults in rolls of textile fabric. The distribution of number of faults is of interest, especially in its relation to that expected if faults occur at random at a fixed rate per metre.

Table 2. Numbers of faults in rolls of textile fabric

Roll No.	Roll length (metres)	No. of faults	Roll No.	Roll length (metres)	No. of faults
1	551	6	17	543	8
2	651	4	18	842	9
3	832	17	19	905	23
4	375	9	20	542	9
5	715	14	21	522	6
6	868	8	22	122	1
7	271	5	23	657	9
8	630	7	24	170	4
9	491	7	25	738	9
10	372	7	26	371	14
11	645	6	27	735	17
12	441	8	28	749	10
13	895	28	29	495	7
14	458	4	30	716	3
15	642	10	31	952	9
16	492	4	32	417	2

... Poisson regression

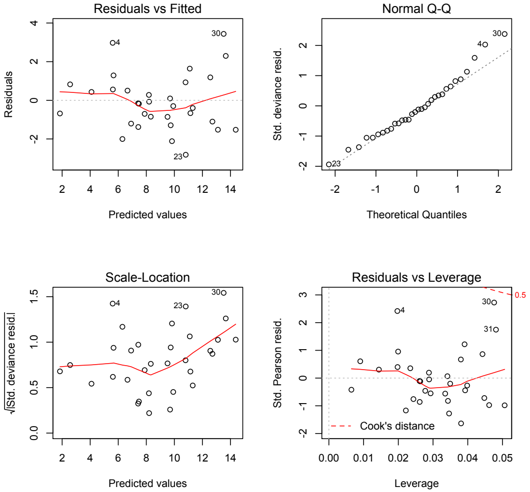
```
> data(cloth)
> cloth[1:5,]
     x  y
1 1.22  1
2 1.70  4
3 2.71  5
4 3.71 14
5 3.72  7
> with(cloth,plot(x,y)) # gives Fig 10.11
> cloth.glm0 = glm(y ~ x - 1, family = poisson(link = identity), data = cloth)
> summary(cloth.glm0)
Coefficients:
  Estimate Std. Error z value Pr(>|z|)
x  1.51024    0.08962   16.85   <2e-16 ***
---

(Dispersion parameter for poisson family taken to be 1)

    Null deviance:    Inf  on 32  degrees of freedom
Residual deviance: 64.537  on 31  degrees of freedom
> cloth.glm1 = glm(y ~ x - 1, family = quasipoisson(link = identity), data = cloth)
> summary(cloth.glm1)
Coefficients:
  Estimate Std. Error t value Pr(>|t|)
x   1.5102     0.1328   11.38 1.35e-12 ***
---

(Dispersion parameter for quasipoisson family taken to be 2.194371)
```

Quasi-Poisson model fit



Measles

Original article

Measles vaccination and antibody res autism spectrum disorders

G Baird,[1] A Pickles,[2] E Simonoff,[3] T Charman,[4] P Sullivan,[5] S (
D Meldrum,[7] M Afzal,[8] B Thomas,[9] L Jin,[9] D Brown[9]

[1] Newcomen Centre, Guy's & St Thomas' NHS Foundation Trust, London, UK; [2] Biostatistics Group, Division of Epidemiology & Health Sciences, University of Manchester, Manchester, UK; [3] Department of Child and Adolescent Psychiatry, Institute of Psychiatry, King's College London, UK; [4] Behavioural and Brain Sciences Unit, UCL Institute of Child Health, London,

ABSTRACT

Objective: To test the hypothesis that measles vaccination was involved in the pathogenesis of autism spectrum disorders (ASD) as evidenced by signs of a persistent measles infection or abnormally persistent immune response shown by circulating measles virus or raised antibody titres in children with ASD who had been vaccinated against measles, mumps and rubella (MMR) compared with controls.

Design: Case–control study, community based.

What is already kn

- Public concern ab mumps, measles a and autism spectr resulted in lower
- Epidemiological st between MMR an

▸ Link


Happy Heart Day!
(Happy Valentine's Day, too!)
Replay
© AGCM, Inc.