

①

14 April 08

## Components of the glm

Data  $(y, x)$

Distribution (p.d.f., p.m.f) of  $y|x$

$$f(y; \theta, \varphi) = \exp \left\{ \frac{y\theta - b(\theta)}{\varphi} \right\} + c(y; \varphi)$$

$$E(Y) = b'(\theta) = \mu$$

linear predictor  $\eta = x^T \beta$

link function,  $g$ .  $\eta = g(\mu)$

inverse link,  $h$ .  $\mu = h(\eta)$

$$E(Y) = h(x^T \beta)$$

$$\text{var}(Y) = \varphi b''(\theta)$$

$$= \varphi V(\mu)$$

variance function  $V(\mu)$

dispersion parameter,  $\varphi$

(2)

Poisson  $f(y; \mu) = \exp\{y \log \mu - \mu - \log y!\}$   
 $y = 0, 1, 2, \dots$   $\mu > 0$

$$\Theta = \log \mu$$

$$b(\Theta) = e^{\Theta}$$

$$\varphi = 1$$

$$c(y; \varphi) = -\log y!$$

$$E(y) = \mu$$

$$V(\mu) = \mu$$

Canonical link  $\eta = \Theta = b'^{-1}(\mu)$

$$b'(\Theta) = e^{\Theta} = \mu$$

$$\eta = \log \mu$$

link

$$\mu = e^{\eta}$$

③

Binomial

$$f(r; \pi) = \binom{m}{r} \pi^r (1-\pi)^{m-r}$$

$$0 < \pi < 1 \quad r = 0, \dots, m$$

$$\exp \left[ m \left\{ \frac{r}{m} \log \frac{\pi}{1-\pi} + \log(1-\pi) \right\} + \log \binom{m}{r} \right]$$

$$* \quad y = \frac{r}{m}$$

$$q = 1/m$$

$$\theta = \log \left( \frac{\pi}{1-\pi} \right) \quad b(\theta) = \log(1+e^\theta)$$

$$c(y; q) = \log \binom{m}{r}$$

$$E(y) = \mu = \frac{e^\theta}{1+e^\theta} = \pi$$

$$V(\mu) = \mu(1-\mu) = \pi(1-\pi)$$

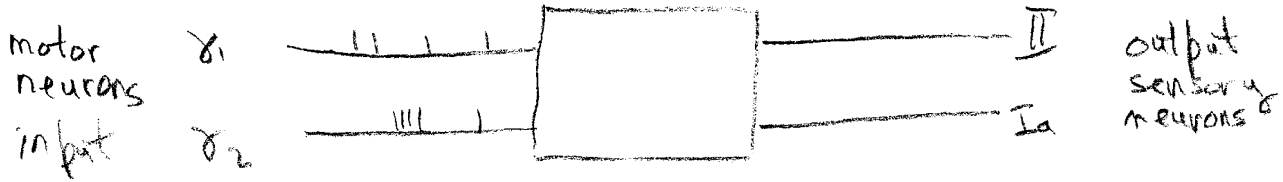
$$b'(\theta) = \frac{e^\theta}{1+e^\theta} = \mu = \pi$$

$$\eta = \log \frac{\mu}{1-\mu} = \text{logit } \pi \quad \underline{\text{link}}$$

$$\pi = \frac{e^\eta}{1+e^\eta}$$

(4)

Muscle spindle.



Output  $y_t = 0, 1 \quad t = 1, 2, \dots$

Input  $x_t = 0, 1 \quad t = 1, 2, \dots$

$$\eta_t = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots$$

Membrane potential

Neuron fires when  $\eta_t$  exceeds threshold  $\theta = \theta + \epsilon_t$   
Reset

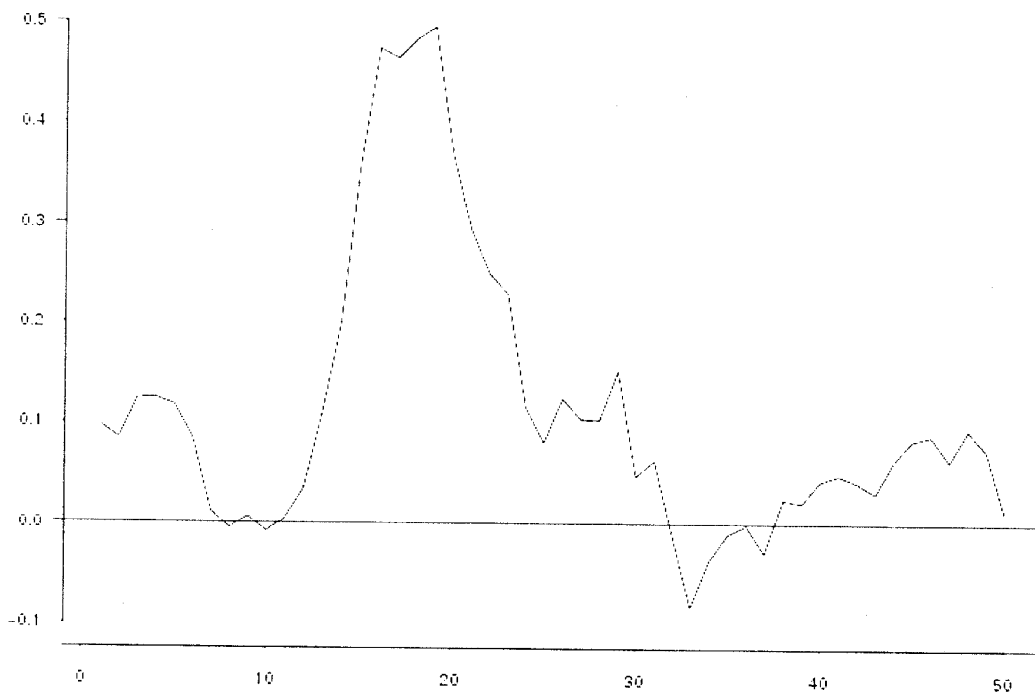
$\epsilon_t$  n.u. cdf  $F$

$$\begin{aligned} \text{Prob}\{y_t = 1/x\} &= P\{\eta_t > \theta\} \\ &= P\{\eta_t > \theta + \epsilon_t\} \\ &= P\{\epsilon_t < \eta_t - \theta\} \\ &= P(\eta_t - \theta) \\ &= E(y_t) \end{aligned}$$

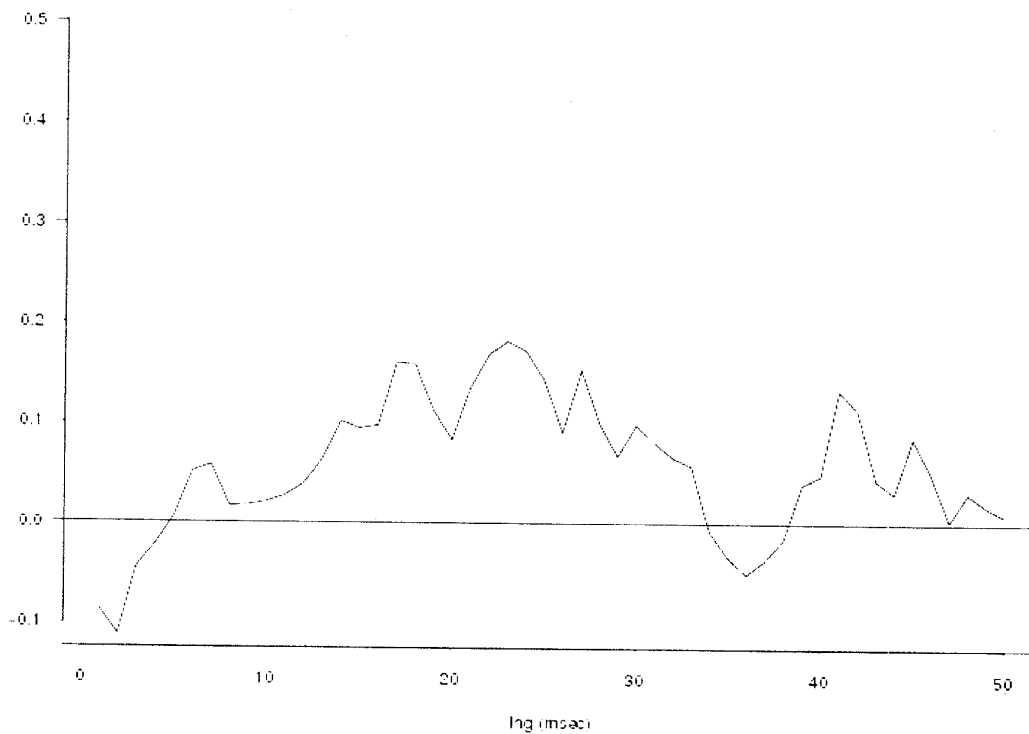
Link: probit, logit  $e^\eta / (1 + e^\eta)$

5

II from g1

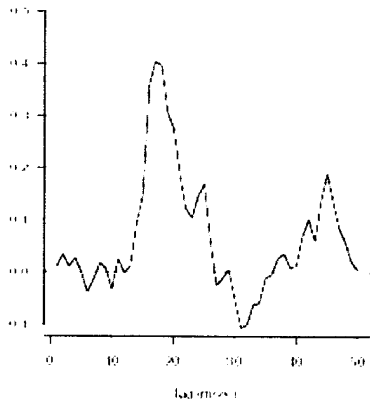


II from g2

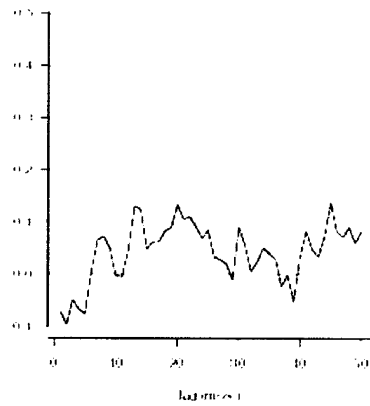


6

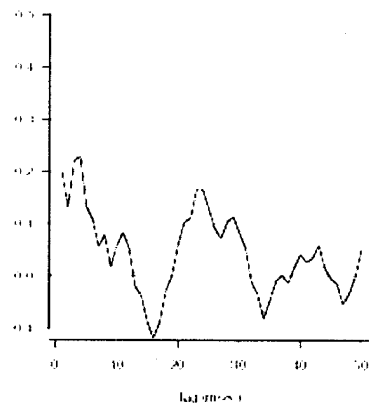
II from Ia. g1. g2 - function g1



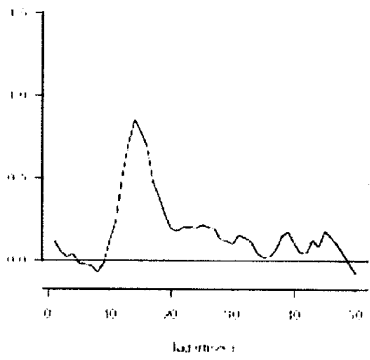
II from Ia. g1. g2 - function g2



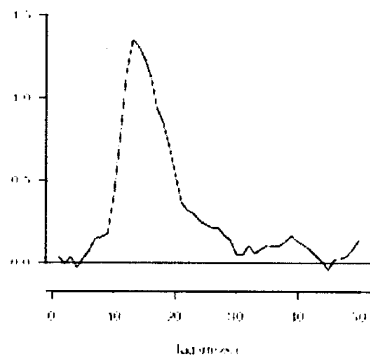
II from Ia. g1. g2 - function Ia



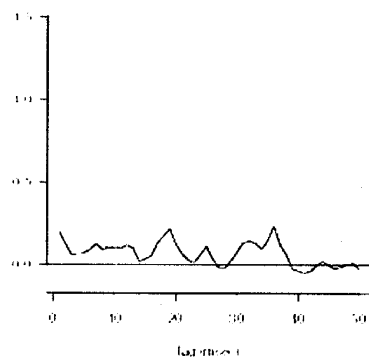
Ia from II. g1. g2 - function g1



Ia from II. g1. g2 - function g2



Ia from II. g1. g2 - function II



(7)

Other links are available in glm()

identity, <sup>inverse</sup> reciprocal,  $\Phi^{-1}$ ,

The usual estimate, given  $\varphi$ , is the mle

The large sample distribution is

$$N(\beta, (X^T W X)^{-1}) \quad W \text{ is below}$$

Might have  $\varphi_j = \varphi a_j$   $a_j$  known, weight

Estimate  $\varphi$  by

$$\frac{1}{n-p} \sum_{j=1}^n (y_j - \hat{\mu}_j) / a_j \sqrt{V(\hat{\mu}_j)}$$

Pearson's statistic

$$P = \frac{1}{\varphi} \sum_{j=1}^n (y_j - \hat{\mu}_j)^2 / V(\hat{\mu}_j) / a_j$$

approx by  $\chi^2_{n-p}$  but

⑧

## Fitting algorithm

Iteratively reweighted least squares

$$\eta = g(\mu) \quad \text{link}$$

$$g(y) \approx g(\mu) + g'(\mu)(y - \mu)$$

$$Eg(y) \approx g(\mu) = x^T \beta$$

$$\text{var } g(y) = [g'(\mu)]^2 \text{var}(y)$$

Consider  $z \approx x^T \beta + g'(\mu)(y - \mu)$  (can evaluate

mean  $x^T \beta$  variance  $[g'(\mu)]^2 \varphi V(\mu)$ )

Estimate  $\Delta \beta$



### 10.3 Generalized linear models

$$y = X\beta + \epsilon$$

$-\infty < y_j < \infty$ , continuous valued, constant variance

Normal density  $-\infty < y < \infty$

Poisson  $0, 1, 2, \dots$

Binomial  $\frac{0}{m}, \frac{1}{m}, \frac{2}{m}, \dots, \frac{m}{m}$

### Jacamar data

	<i>Aphrissa boisduvalli</i> N/S/E	<i>Phoebis argente</i> N/S/E	<i>Dryas iulia</i> N/S/E	<i>Pierella luna</i> N/S/E	<i>Consul fabius</i> N/S/E	<i>Siproeta stelenes</i> † N/S/E
Unpainted	0/0/14	6/1/0	1/0/2	4/1/5	0/0/0	0/0/1
Brown	7/1/2	2/1/0	1/0/1	2/2/4	0/0/3	0/0/1
Yellow	7/2/1	4/0/2	5/0/1	2/0/5	0/0/1	0/0/3
Blue	6/0/0	0/0/0	0/0/1	4/0/3	0/0/1	0/1/1
Green	3/0/1	1/1/0	5/0/0	6/0/2	0/0/1	0/0/3
Red	4/0/0	0/0/0	6/0/0	4/0/2	0/0/1	3/0/1
Orange	4/2/0	6/0/0	4/1/1	7/0/1	0/0/2	1/1/1
Black	4/0/0	0/0/0	1/0/1	4/2/2	7/1/0	0/1/0

Table 10.2 Response of a rufous-tailed jacamar to individuals of seven species of palatable butterflies with artificially coloured wing undersides. (N=not sampled, S = sampled and rejected, E = eaten)

† includes *Phlaethria dido* also.

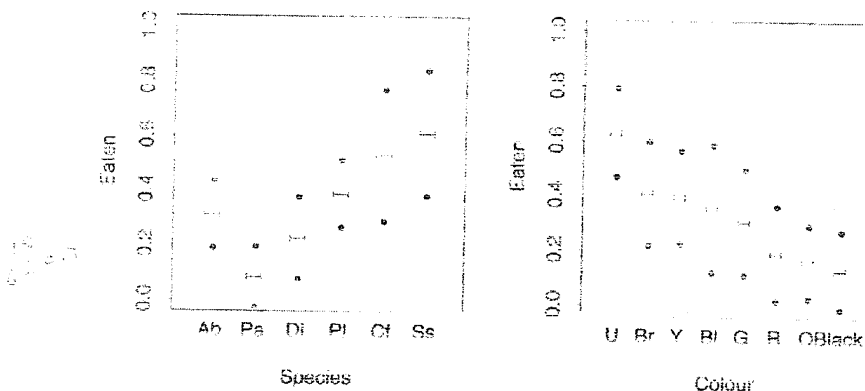


Figure 10.2 Proportion of butterflies eaten (Eaten) for different species and wing colour.

9

## Jacamar data

$r_{cs}$  : number of butterflies of species  $s$   
painted the  $c$ th color and eaten

$$r_{cs} \sim \text{Bin}(\pi_{cs}, m_{cs}) \quad \begin{array}{l} s=5, \dots, 6 \\ c=5, \dots, 8 \end{array} \quad \text{but 4 empty cells}$$

$$y_{cs} = r_{cs} / m_{cs}$$

NB

$$\eta_{cs} = \alpha_c + \gamma_s$$

logit link  $\eta = \log \pi / (1 - \pi)$

factors color and species

models constant  
species  
color  
species \* color

How to assess/compare

---

(10)

## Likelihood inference

log likelihood

$$l(\beta, \varphi) = \sum_{j=1}^n \log f(y_j; \eta_j(\beta), \varphi)$$

$\varphi$ : known

$\hat{\beta}$   $p \times 1$

$$\hat{\eta}_j = \eta_j(\hat{\beta})$$

max like  $l(\hat{\beta})$

Now maximize wrt  $\eta$  ( $n$  parameters)

Scaled deviance

$$D = 2 \sum_{j=1}^n \{ \log f(y_j; \tilde{\eta}_j) - \log f(y_j; \hat{\eta}_j) \}$$

$\sim \chi^2_{n-p}$

11

### ANODEV

species first

Terms	df	Deviance	$\Delta$ df	$\Delta$ dev
1	43	134.24		
1 + species	38	114.59	5	19.65
1 + color	36	108.46	2	6.13
1 + species + color	31	67.28	5	41.18

Compare resid dev 67.28 to  $\chi^2_{31}$

Poor fit

Diagnosis finds and deletes two outliers