The Generalized Linear Model 15 Oct 01

McCullagh + Nelder (1989)
Fahrmeir and Tutz (1994)

Data $(y_i, \underset{\sim}{x}_i)$ $i = 1, \ldots, n$

regression like
but y values don't cover $(-\infty, \infty)$

Moving beyond

$$Y_i = \underset{\sim}{x}_i' \underset{\sim}{\beta} + \epsilon_i$$

eg. $Y = 0, 1$

Assumptions.

1. If some of the entries of $\underset{\sim}{x}$ are r.v.'s, then given the $\underset{\sim}{x}_i$ the Y_i are independent

2. Again given the $\underset{\sim}{x}_i$, Y_i belongs to a simple exponential family, $E\{Y_i | \underset{\sim}{x}_i\} = \mu_i$, possibly containing a common scale parameter.

15 Oct. 01

3. The <u>linear predictor</u> $\eta_i = \underset{\sim}{x}_i' \underset{\sim}{\beta}$

$$E\{Y_i | \underset{\sim}{x}_i\} = \mu_i = h(\eta_i) = h(\underset{\sim}{x}_i' \underset{\sim}{\beta})$$

h: the <u>inverse link</u>, 1-1

$\eta_i = g(\mu_i)$ g: the <u>link</u>

<u>Exponential family</u>

pdf/pmf

$$f_Y(y_i | \theta_i, \varphi) = \exp\left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\varphi)} + c(y_i, \varphi) \right\}$$

θ_i : <u>natural parameter</u>

φ: scale or dispersion parameter, φ, σ^2

$a(.)$, $b(.)$, $c(.)$ correspond to the type of family

15 Oct 01

Important cases: normal, binomial (binary), Poisson, gamma, inverse Gaussian

The mean and variance may be expressed by

$$E(Y) = \mu = \frac{db(\theta)}{d\theta}$$

$$Var(Y) = a(\varphi)\, v(\mu) \quad \text{with} \quad v(\mu) = \frac{d^2 b(\theta)}{d\theta^2} \qquad (*)$$

Perhaps there is a weight and

$$Var(Y) = \varphi\, v(\mu)/w$$

15 Oct 01

Proof of (*).

$$\ell(\theta, \varphi | y) = \log f_Y(y | \theta, \varphi)$$

$$\int f_Y \, dy = 1$$

$$E\left(\frac{\partial \ell}{\partial \theta}\right) = \int \frac{1}{f_Y} \frac{\partial f_Y}{\partial \theta} f_Y \, dy = 0$$

$$\int \left(\frac{\partial}{\partial \theta} \log f_Y\right) f_Y \, dy = 0$$

$$E\left(\frac{\partial^2 \ell}{\partial \theta \partial \theta'}\right) + E\left(\frac{\partial \ell}{\partial \theta}\right)\left(\frac{\partial \ell}{\partial \theta}\right)' = 0$$

For exponential family

$$\ell(\theta | y) = \frac{[y\theta - b(\theta)]}{a(\varphi)} + c(y, \varphi)$$

$$\frac{\partial \ell}{\partial \theta} = [y - b'(\theta)] / a(\varphi)$$

$$\frac{\partial^2 \ell}{\partial \theta^2} = -b''(\theta) / a(\varphi)$$

$$E\left(\frac{\partial \ell}{\partial \theta}\right) = [E(Y) - b'(\theta)] / a(\varphi) = 0$$

15 Oct 01

$$\mathrm{var}\left(\frac{\partial l}{\partial \theta}\right) = E\left(\frac{\partial l}{\partial \theta}\right)^2 = \frac{1}{a(\varphi)^2}\,\mathrm{var}\,Y$$

$$= -E\left(\frac{\partial^2 l}{\partial \theta^2}\right)$$

$$= \frac{b''(\theta)}{a(\varphi)}$$

16 Oct 01

Other glm models.

Poisson $f(y|\theta) = \frac{\lambda^y e^{-\lambda}}{y!}$ $y = 0, 1, 2, \ldots$

$$\log f(y|\theta) = y \log \lambda - \lambda - \log y!$$

cp. $$\frac{y\theta - b(\theta)}{a(\varphi)} + c(y, \varphi)$$

$$\theta = \log \lambda, \quad \lambda = e^{\theta} > 0$$

$$b(\theta) = \lambda = e^{\theta}$$

$$\frac{db'}{d\theta} = e^{\theta} = \lambda$$

$$\frac{d^2 b}{d\theta^2} = e^{\theta} = \lambda.$$

Particular cases of glm

Sample Family ① Take HW

5 16 Oct 01

Models for binary and binomial responses

1. Binary / Bernoulli $Y = {0 \atop 1}$ $\underset{\sim}{x}$ covariate

$E\{Y \mid \underset{\sim}{X}\} = \text{Prob}\{Y = 1 \mid \underset{\sim}{X}\} = \pi$ $\pi(\underset{\sim}{x})$

$\text{var}\{Y \mid \underset{\sim}{X}\} = \pi(1-\pi)$

2. Binomial $Y : B(m, \pi)$

frequency Y/m

$E\{Y/m \mid \underset{\sim}{X}\} = \pi$ $\pi(\underset{\sim}{x})$

$\text{var}\{Y/m \mid \underset{\sim}{X}\} = \pi(1-\pi)/m$ $w = m$

m: weight

16 Oct 01

$\underline{B(1, \pi)}$.

$\theta(\mu) = \log \pi/(1-\pi)$

$b(\theta) = \log(1+e^{\theta})$

$\phi = 1$

$E(Y) = b'(\theta) = \dfrac{e^{\theta}}{1+e^{\theta}}$

$w(\mu) = \pi(1-\pi) = b''(\theta)$

$f_Y(y|\pi) = \pi^{y}(1-\pi)^{1-y} \qquad y = 0, 1$

$\log f_Y(y|\pi) = y \log \pi + (1-y)\log(1-\pi)$

$= y \log \pi/1-\pi + \log(1-\pi)$

cp. $\dfrac{[y\theta - b(\theta)]}{a(\varphi)} + c(y, \varphi)$

$\theta = \log \pi/1-\pi \qquad \pi = e^{\theta}/(1+e^{\theta})$

$b(\theta) = \log(1-\pi)$

16 Oct 01

Linear predictor. $\eta = \underset{\sim}{x}^T \underset{\sim}{\beta}$

Need $0 \le \pi \le 1$

Write $\pi = F(\eta)$ F: ~~strictly~~ monotone $0 \le F \le 1$ cdf

Probit model.

$$\pi = \Phi(\eta) = \Phi(\underset{\sim}{x}^T \underset{\sim}{\beta})$$

$$h(.) \equiv \Phi(.)$$
$$g(.) \equiv \Phi^{-1}(.)$$

Logit model.

$$\pi = h(\eta) = \exp\{\eta\} / [1 + \exp\{\eta\}]$$

$$\eta = g(\pi) = \log \frac{\pi}{1-\pi}$$

logistic distribution function

16 Oct 06

Complementary log-log model

$$h(x) = 1 - \exp\{-\exp\{x\}\}$$

$$g(\pi) = \log(-\log(1-\pi))$$

extreme minimal value distribution

20 Feb. 02

Example 2.6.1 The "O-Ring Data"

Space shuttle

Sometimes an O-ring failed

Data for 23 pre-Challenger launches

Challenger blew up on take-off

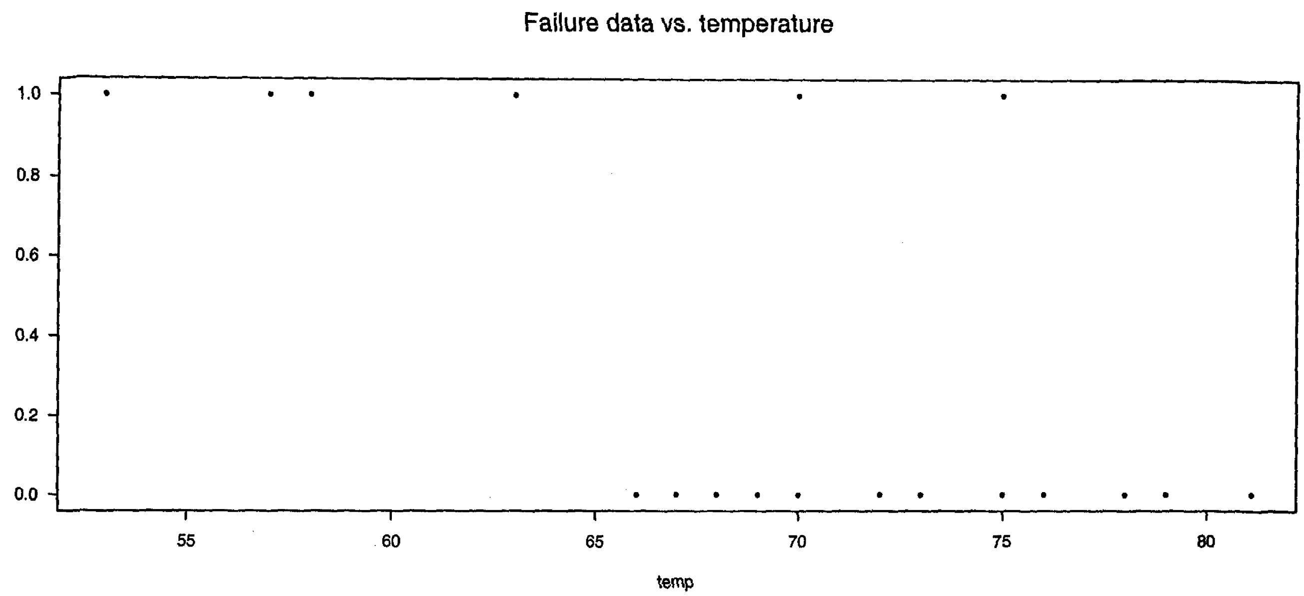
Richard Feynman thought temperature might be part of the cause

It was 31°F when Challenger crashed.

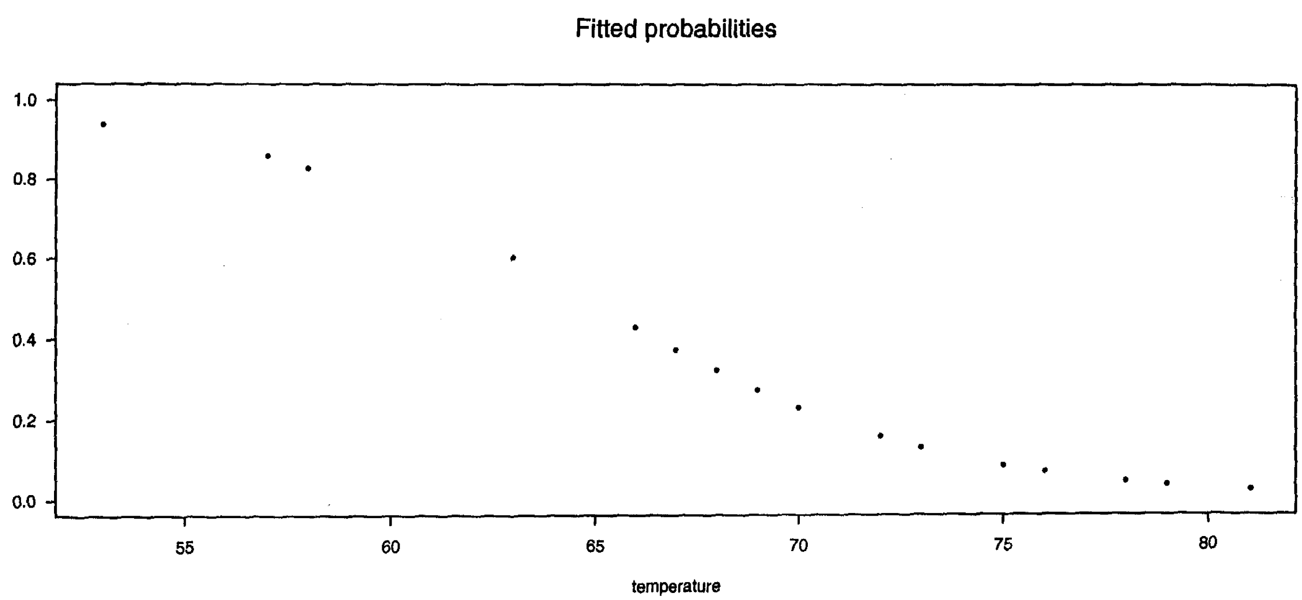
The predicted prob of O-ring failure at 31°F is .9996

- big extrapolation!

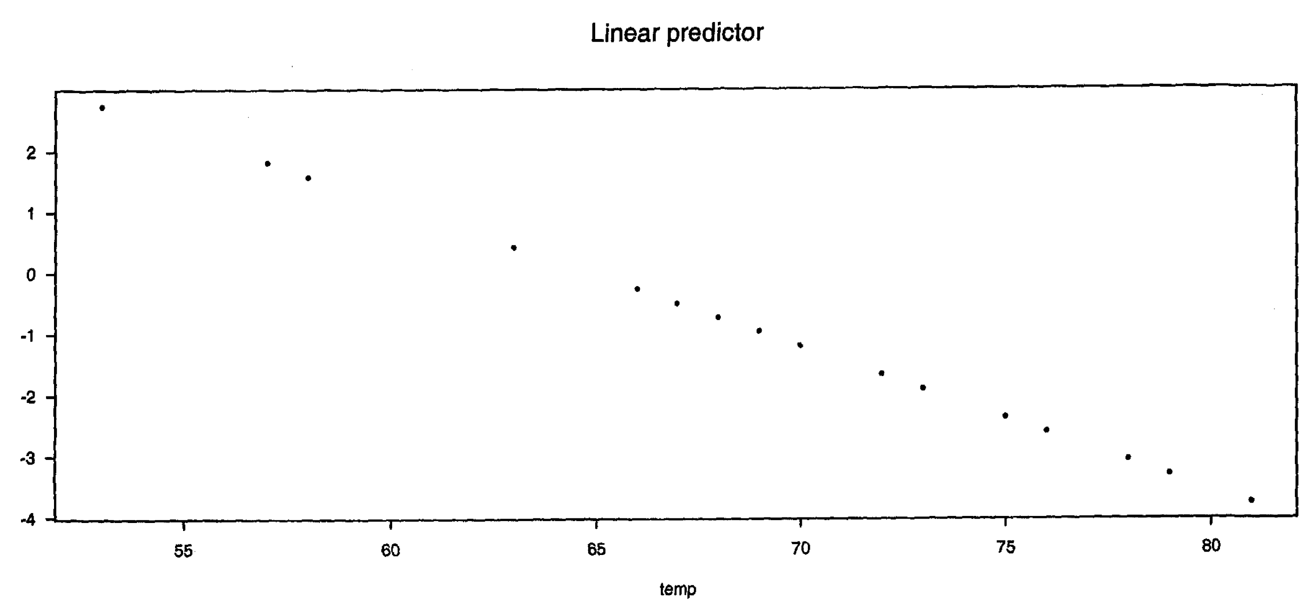
```
   fail temp
 1    1   53
 2    1   57
 3    1   58
 4    1   63
 5    0   66
 6    0   67
 7    0   67
 8    0   67
 9    0   68
10    0   69
11    0   70
12    0   70
13    1   70
14    1   70
15    0   72
16    0   73
17    0   75
18    1   75
19    0   76
20    0   76
21    0   78
22    0   79
23    0   81

Call: glm(formula = fail ~ temp, family = "binomial")
Deviance Residuals:
       Min        1Q     Median        3Q      Max
 -1.061112 -0.76128 -0.3782822 0.4523828 2.217506

Coefficients:
                  Value Std. Error   t value
(Intercept)  15.0422911  7.3366528  2.050293
       temp  -0.2321537  0.1076141 -2.157279

(Dispersion Parameter for Binomial family taken to be 1 )

    Null Deviance: 28.26715 on 22 degrees of freedom

Residual Deviance: 20.31519 on 21 degrees of freedom

Number of Fisher Scoring Iterations: 4
Points out of bounds X= 75 Y= 3.252558
[1] "95% CI for risk probability for Challenger at 31 degrees F"
         1         1         1
 0.3614313 0.9996087 0.9999999
```



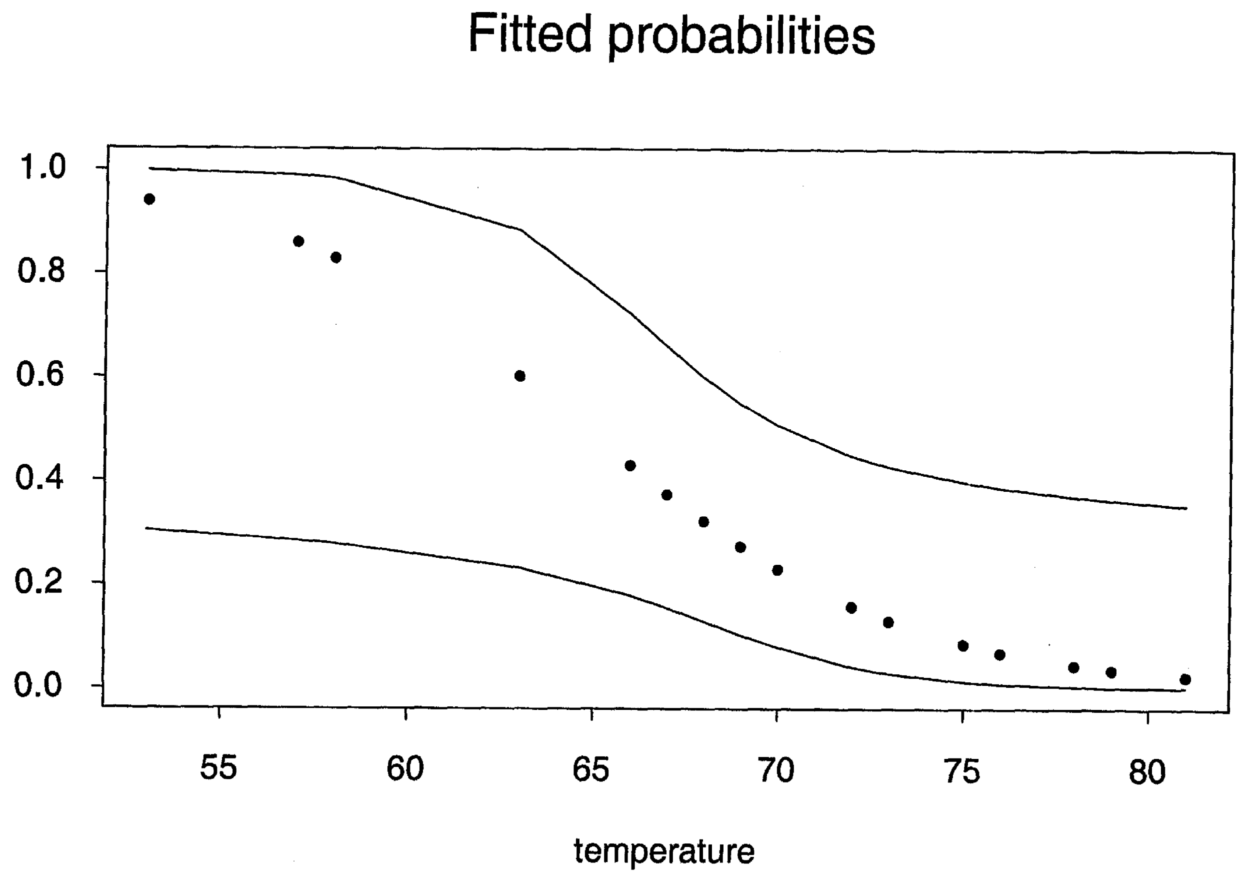
Failure data vs. temperature
1.0
0.8
0.6
0.4
0.2
0.0
55
60
65
70
75
80
temp

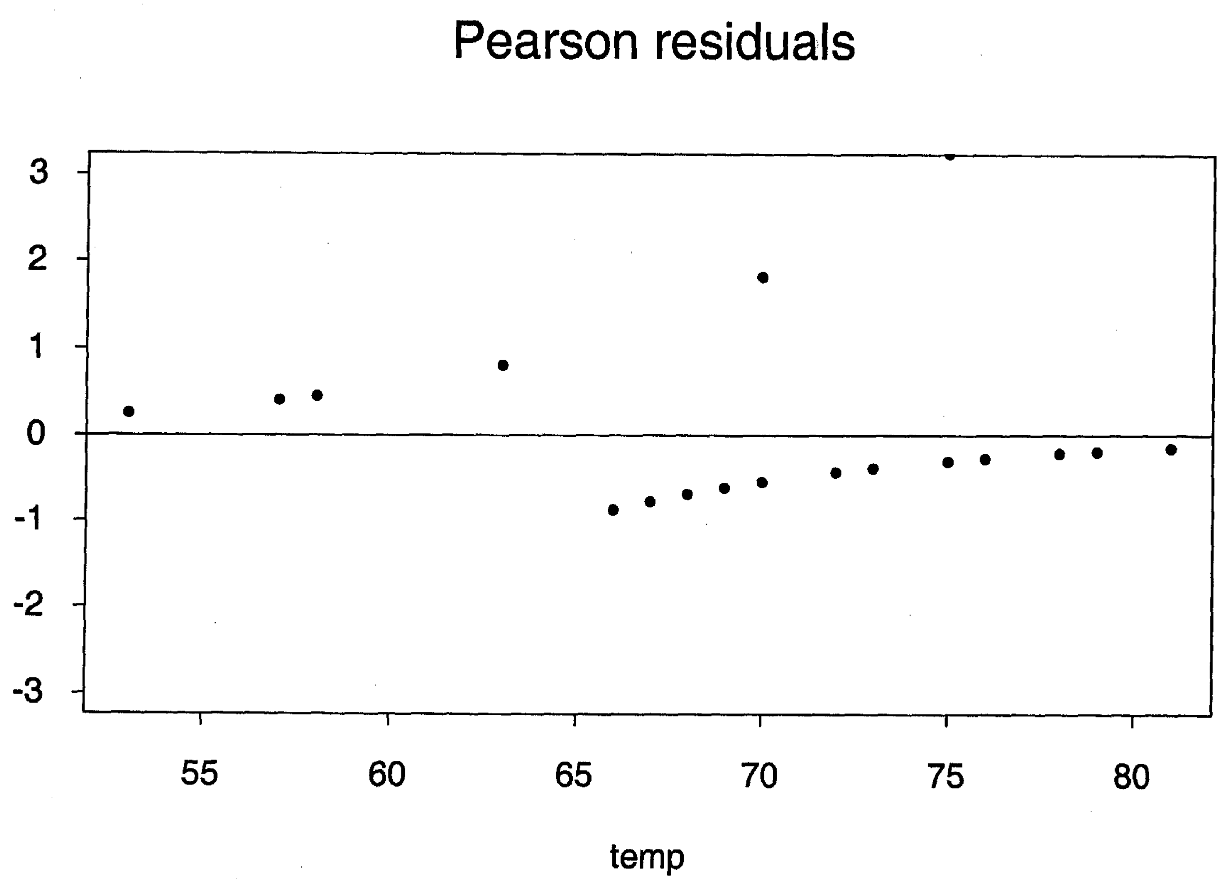


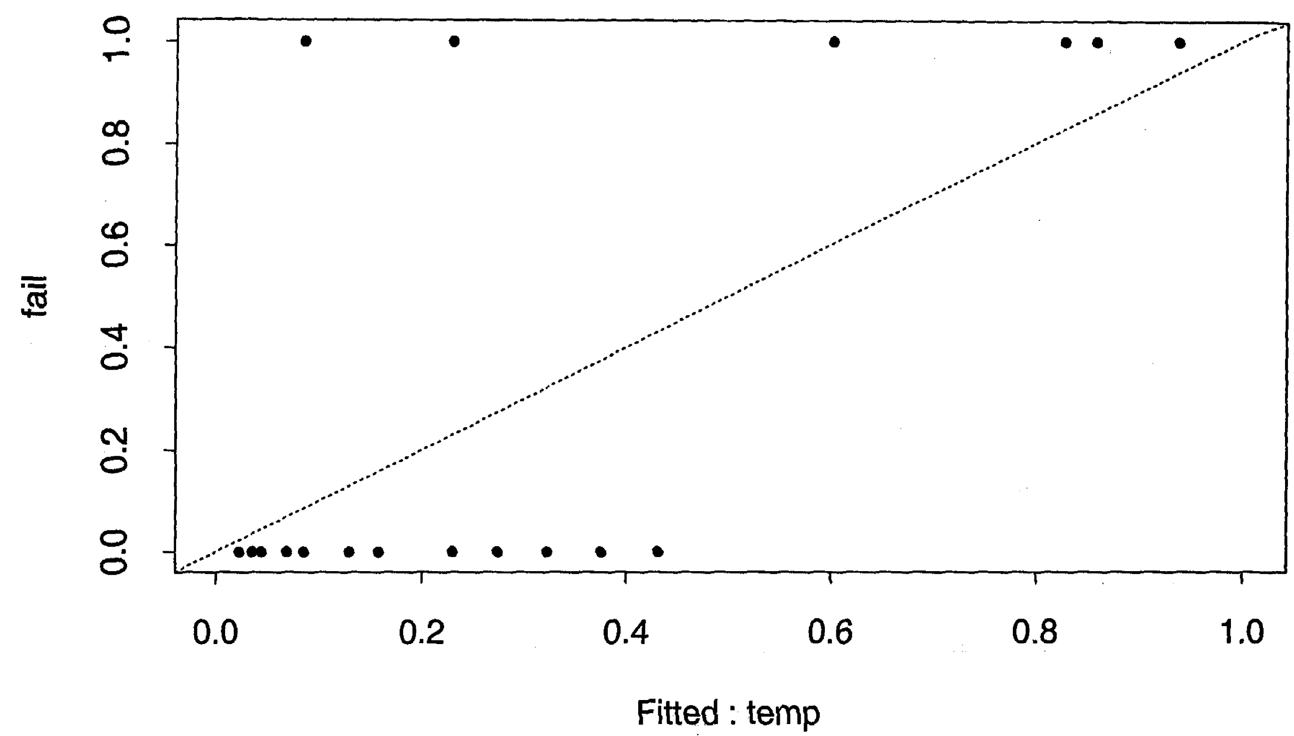
Fitted probabilities
1.0
0.8
0.6
0.4
0.2
0.0
55
60
65
70
75
80
temperature

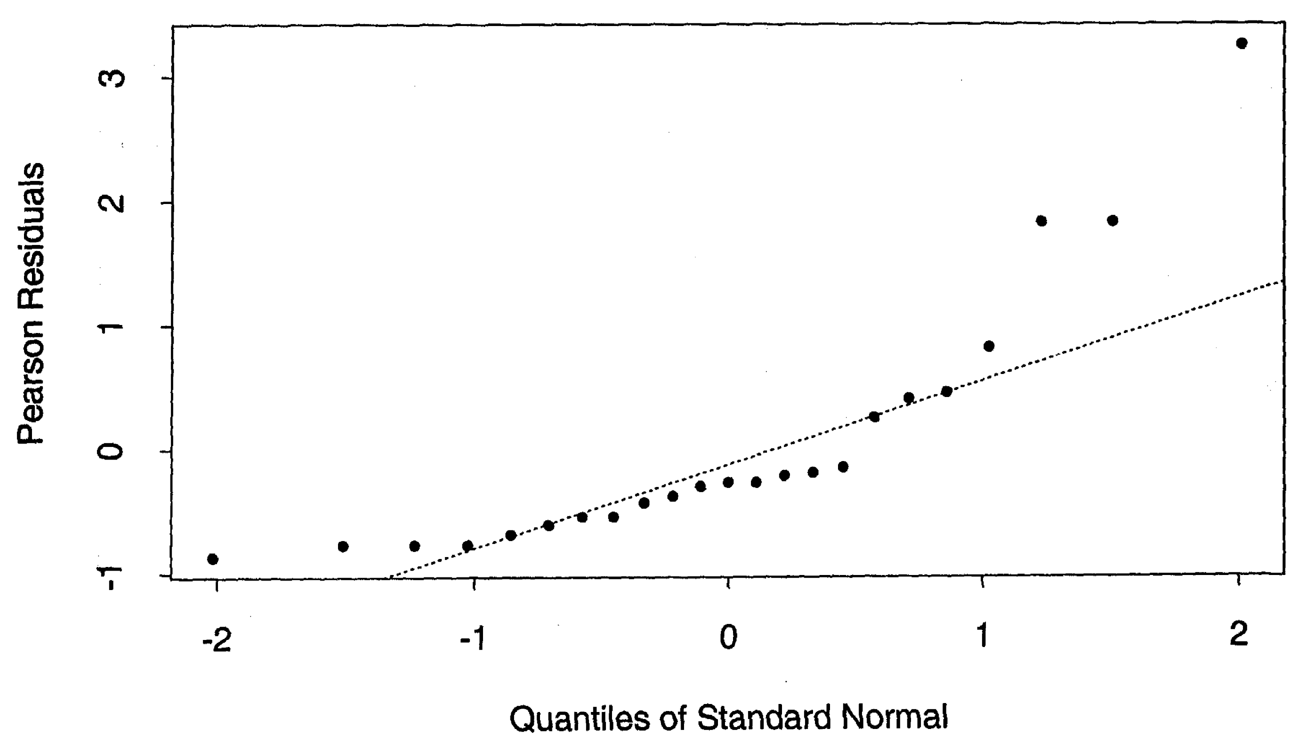


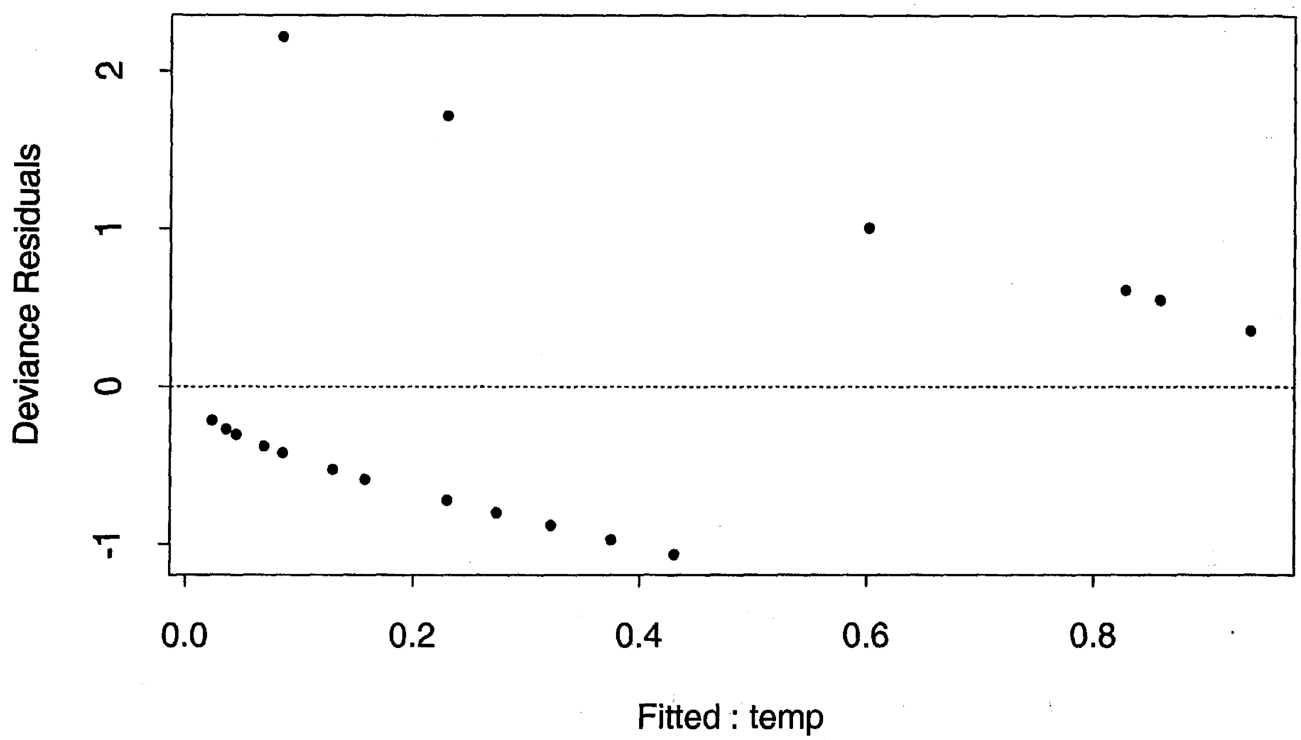
Linear predictor
2
1
0
-1
-2
-3
-4
55
60
65
70
75
80
temp


Fitted probabilities
1.0
0.8
0.6
0.4
0.2
0.0
55
60
65
70
75
80
temperature

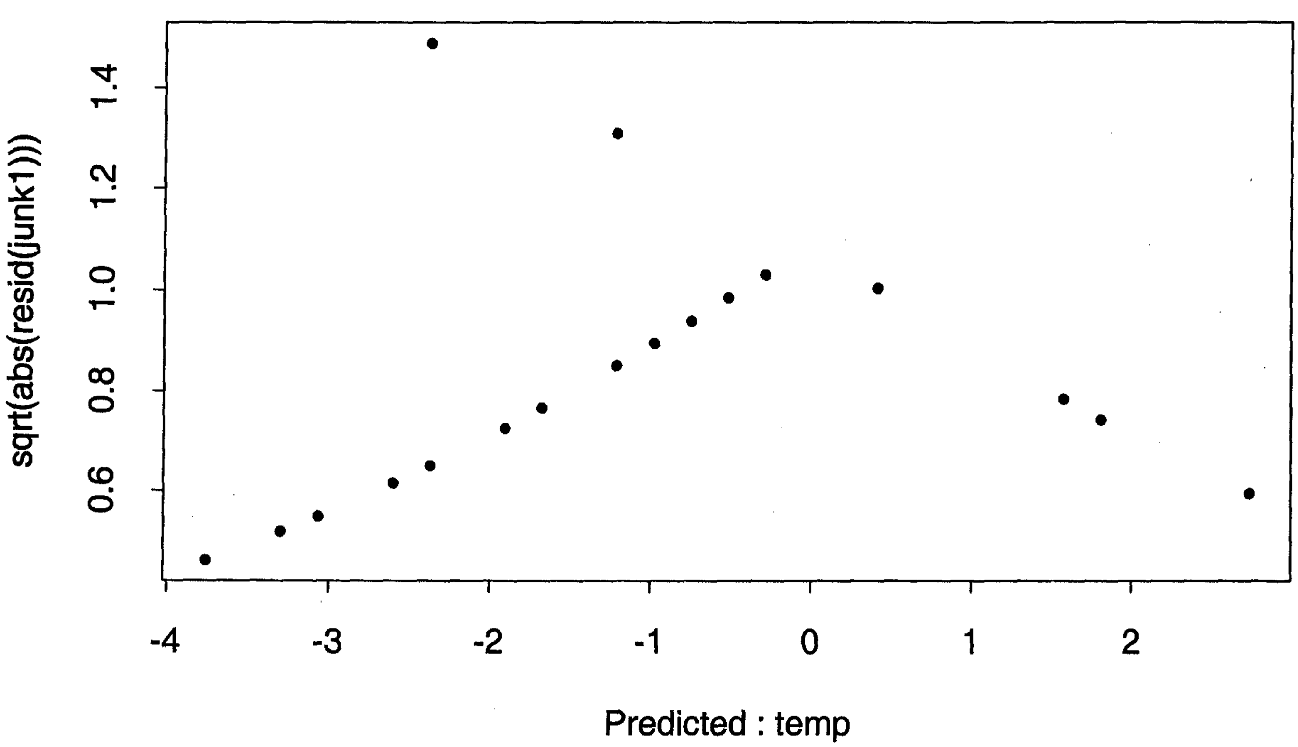

Pearson residuals
3
2
1
0
-1
-2
-3
55
60
65
70
75
80
temp


fail
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
Fitted : temp


Pearson Residuals
-1
0
1
2
3
-2
-1
0
1
2
Quantiles of Standard Normal



Deviance Residuals
-1
0
1
2
0.0
0.2
0.4
0.6
0.8
Fitted : temp



sqrt(abs(resid(junk1)))
0.6
0.8
1.0
1.2
1.4
-4
-3
-2
-1
0
1
2
Predicted : temp

16 Oct 01

Binary models as threshold models of latent linear models,

latent variate

$$\zeta = \underset{\sim}{x}^{\tau} \underset{\sim}{\gamma} + \sigma \epsilon$$

where ϵ distributed as $F(.)$

Set $Y = 1 \quad \zeta \le \theta$
$\quad\;\; 0 \quad \zeta > \theta$

$$\text{Prob}\{Y=1 \mid \underset{\sim}{X}\} = \text{Prob}\{\underset{\sim}{x}^{\tau} \underset{\sim}{\gamma} + \sigma\epsilon \le \theta\}$$

$$= \text{Prob}\left\{\epsilon \le \frac{\theta - \underset{\sim}{x}^{\tau} \underset{\sim}{\gamma}}{\sigma}\right\}$$

$$= F\left(\frac{\theta - \underset{\sim}{x}^{\tau} \underset{\sim}{\gamma}}{\sigma}\right)$$

If $x_1 \equiv 1$, absorb θ into γ
write $-\underset{\sim}{\gamma}/\sigma = \underset{\sim}{\beta}$

$$\text{Prob}\{Y=1 \mid \underset{\sim}{X}\} = F(\underset{\sim}{x}^{\tau} \underset{\sim}{\beta})$$

Conceptual Biological Model

Action potentials of neuron $M = \{\sigma_j\}$ contribute to the *membrane potential*, $U(t)$, in neuron N at its trigger zone.

$$U(t) = \sum_j a(t - \sigma_j)$$

$a(.)$: <u>*summation function*</u>

$\theta(.)$: *threshold function*

time since last fired

N fires when $U(t)$ crosses $\gamma + \dot{\theta}(t)$. The threshold is reset to a high level on N's firing.

Description
Association
Regression
- **Likelihood**

For computations, discretize time to $t=0,\pm1,\pm2,\ldots$ and M, N to 0–1 processes.

$$\sum_j a(t-\sigma_j) \approx \sum_u a_{t-u} M_u$$

$$\theta_t = \sum_{v=1}^{\gamma_t} b_v N_{t-v}$$

Suppose there is noise with c.d.f $P(.)$ superposed on the *threshold*, θ_t.

Consider the conditional probability

$$P_t = Prob\{N_t = 1 \mid the\ past\} = P(\psi_t)$$

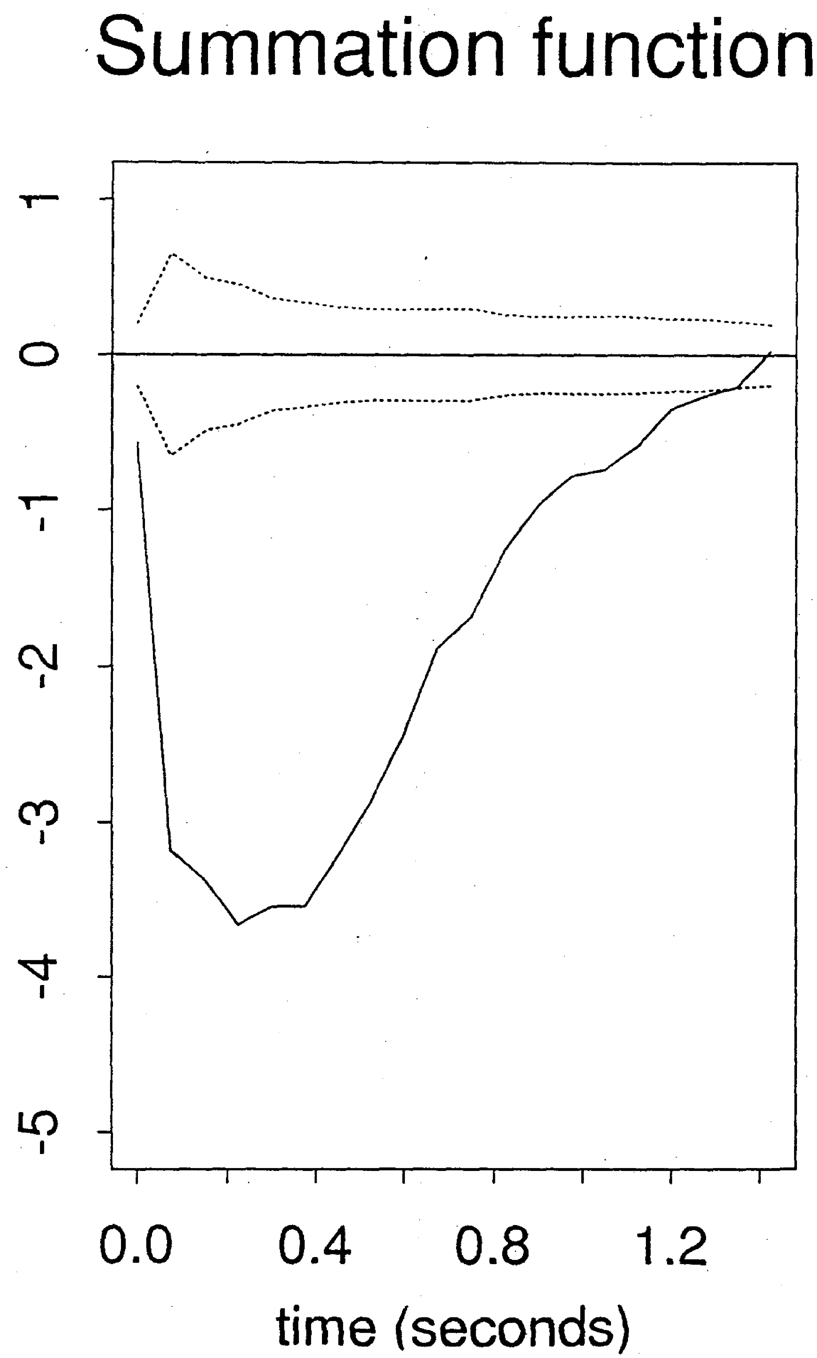
where

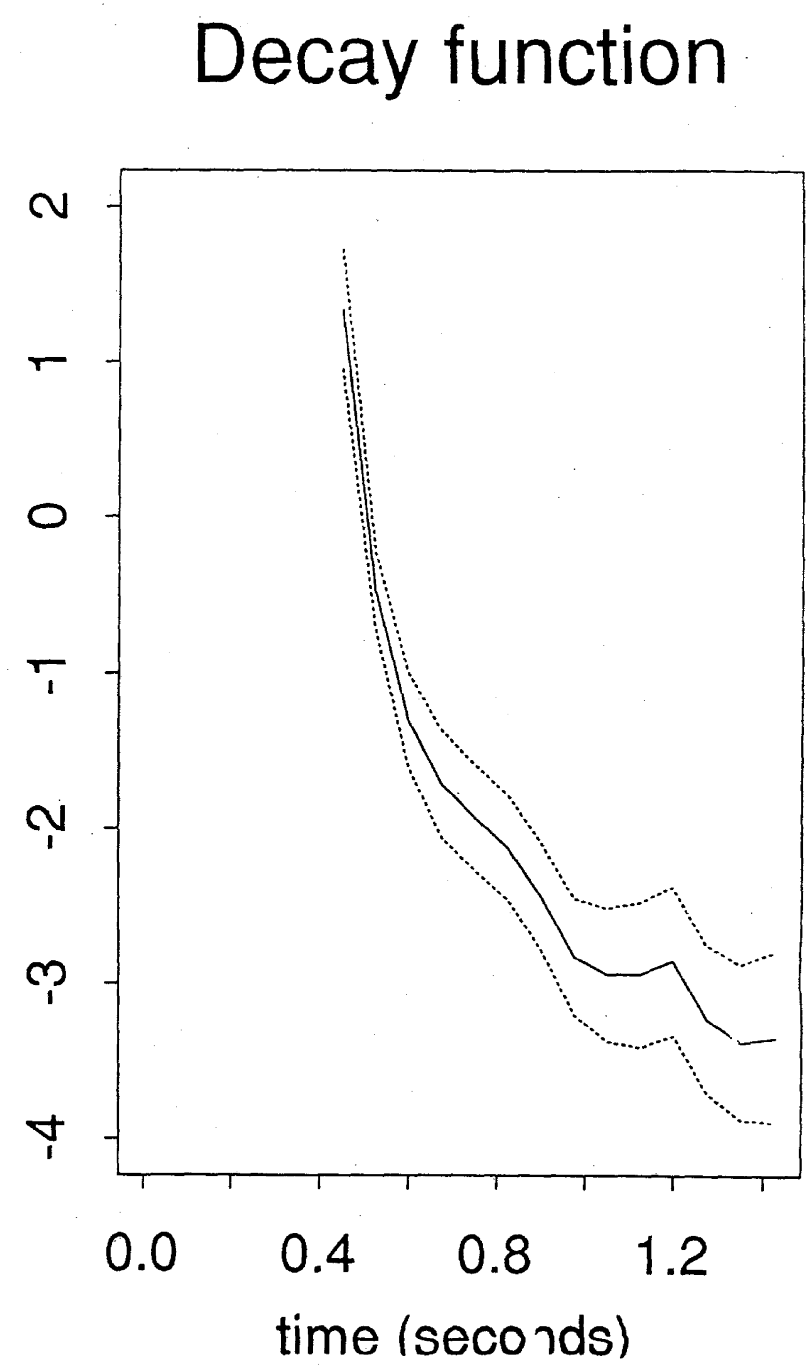
$$\psi_t = \sum a_u M_{t-u} - \theta_t$$

Maximize the loglikelihood

$$\sum [N_t \log P_t + (1 - N_t) \log (1 - P_t)]$$

Initially take $P(.) = \Phi(.)$, the normal cumulative.


Summation function
1
0
-1
-2
-3
-4
-5
0.0
0.4
0.8
1.2
time (seconds)


Decay function
2
1
0
-1
-2
-3
-4
0.0
0.4
0.8
1.2
time (seconds)

16 Oct 01

Overdispersion

$$\text{var}\{Y/m \mid \underset{\sim}{X}\} = \phi \frac{\pi(1-\pi)}{m}$$

overdispersion $\phi > 1$

16 Oct 01

Count data

Log-linear Poisson

$$\log \mu = \eta = \underset{\sim}{x}^{T} \underset{\sim}{\beta}$$

$$\underset{\sim}{\mu} = \exp\{\eta\}$$

16 Oct 01

Poisson $P(\lambda)$

$\theta(\mu) = \log \lambda$

$b(\theta) = \exp\{\theta\}$

$\phi = 1$

$E\{Y\} = b'(\theta) = \lambda = e^{\theta}$

$v(\mu) = \lambda$

16 Oct 01

Maximum likelihood estimation.

log-likelihood

$$l_i(\theta_i) = \frac{y_i\theta_i - b(\theta_i)}{\phi}$$ up to an additive constant

$$= \frac{y_i\theta(\mu_i) - b(\theta(\mu_i))}{\phi}$$ ϕ : known

eg. binary $y_i \log \pi_i + (1-y_i)\log(1-\pi_i)$

Poisson $y_i \log \lambda_i - \lambda_i$

Mean structure $\mu_i = h(\underset{\sim}{x}_i^{T}\beta)$

$$l_i(\beta) = l_i(h(\underset{\sim}{x}_i^{T}\beta))$$

Sample log likelihood

$$l(\beta) = \sum_i l_i(\beta)$$

$$\frac{\partial l}{\partial \beta} = \sum_i s_i(\beta) = s(\beta)$$ score function

$$\frac{\partial l_i}{\partial \beta_i} = \frac{1}{\phi}\left[y_i\, \theta'(\mu_i)\frac{\partial \mu_i}{\partial \beta} - b'(\theta(\mu_i))\theta'(\mu_i)\frac{\partial \mu_i}{\partial \beta}\right] = \frac{1}{\phi} \ldots$$

16 Oct 01

$$s_i(\beta) = x_i D_i(\beta) \sigma_i^{-2} [y_i - \mu_i(\beta)]$$

$$\mu_i(\beta) = h(x_i^T \beta)$$

$$\sigma_i^2(\beta) = v(h(x_i^T \beta))\phi$$

$$D_i(\beta) = \frac{\partial h(x_i^T \beta)}{\partial \eta}$$

Expected Fisher information matrix

$$F(\beta) = \operatorname{var} s(\beta)$$ use for scoring

$$= \sum_i F_i(\beta)$$

where $F_i(\beta) = x_i x_i^T w_i(\beta)$

$$w_i(\beta) = D_i^2(\beta) \sigma_i^{-2}(\beta)$$

For natural link functions, i.e. θ_i

$$s(\beta) = \frac{1}{\phi} \sum_i x_i^T [y_i - \mu_i(\beta)]$$ $D_i(\beta) = \phi \sigma_i^2$

$$F(\beta) = \frac{1}{\phi} \sum_i v(\mu_i(\beta)) x_i x_i^T$$

16 Oct 01

In matrix notation

$$\underset{\sim}{y} = \begin{bmatrix} y_1 \\ \vdots \\ \vdots \end{bmatrix} \qquad \underset{\sim}{\mu}(\underset{\sim}{\beta}) = \begin{bmatrix} \mu_i(\underset{\sim}{\beta}) \\ \vdots \\ \vdots \end{bmatrix}$$

$$\underset{\sim}{\Sigma}(\underset{\sim}{\beta}) = \begin{bmatrix} \sigma_1^2(\underset{\sim}{\beta}) & 0 & \\ 0 & \ddots & \\ & & \ddots \end{bmatrix}$$

$$\underset{\sim}{D}(\underset{\sim}{\beta}) = \begin{bmatrix} D_1(\underset{\sim}{\beta}) & 0 & \\ 0 & \ddots & \\ & & \ddots \end{bmatrix}$$

$$\underset{\sim}{W}(\underset{\sim}{\beta}) = \begin{bmatrix} w_1(\underset{\sim}{\beta}) & 0 & \\ 0 & \ddots & \\ & & \ddots \end{bmatrix}$$

$$s(\underset{\sim}{\beta}) = \underset{\sim}{X}^T D(\underset{\sim}{\beta}) \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\beta}) [\underset{\sim}{y} - \underset{\sim}{\mu}(\underset{\sim}{\beta})]$$

$$F(\underset{\sim}{\beta}) = \underset{\sim}{X}^T \underset{\sim}{W}(\beta) \underset{\sim}{X}$$

Compute $\hat{\beta}$ as solution of

$$\underset{\sim}{s}(\hat{\beta}) = \underset{\sim}{0}$$