The Abdus Salam International Centre for Theoretical Physics

United Nations Educational, Scientific and Cultural Organization
IAEA International Atomic Energy Agency

1863-11

Advanced School and Conference on Statistics and Applied Probability in Life Sciences

24 September - 12 October, 2007

Generalized linear models III

Peter McCullagh

Department of Statistics
University of Chicago
Chicago IL 60637, USA

Generalized linear models III

Peter McCullagh

Department of Statistics
University of Chicago

Trieste, October 2007

Outline

Binary response models

Model:

$Y_1, \ldots, Y_n$ independent Bernoulli variables
$E(Y_i) = \pi(x_i)$ is the mean
Need a model for the response function $\pi(x)$

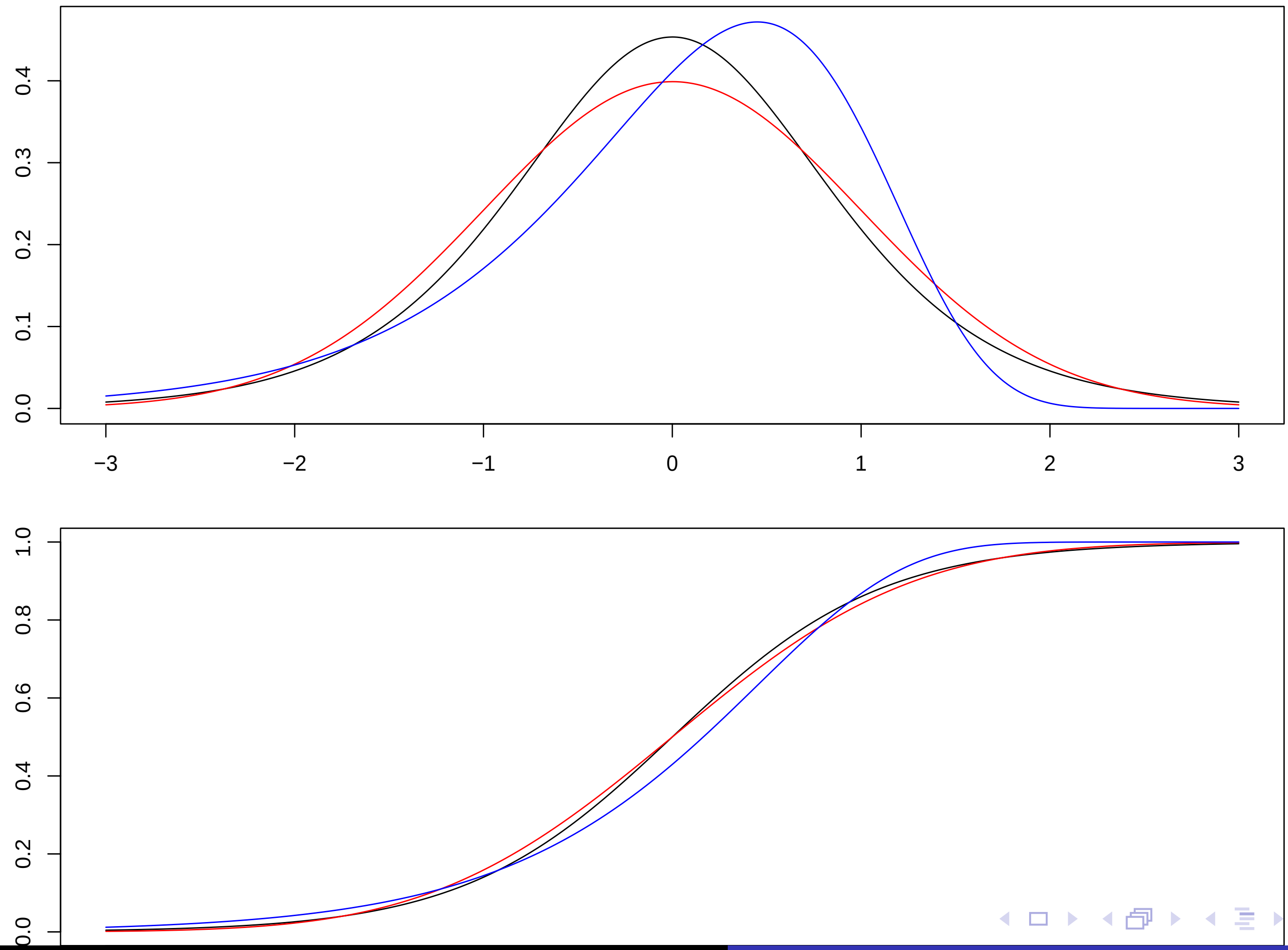
Constraint: $0 \le \pi(x) \le 1$

Standard response functions: $\pi(x) = F(\beta' x)$

F is a cumulative distribution function
$f = F'$ is a density function on $\mathcal{R}$

logit: $F(\eta) = e^\eta/(1 + e^\eta)$; $\eta = \log(\pi/(1-\pi))$
probit: $F(\eta) = \Phi(\eta)$; $\eta = \Phi^{-1}(\pi)$
c-log log: $F(\eta) = 1 - \exp(-e^\eta)$; $\eta = \log(-\log(1-\pi))$

Associated standardized densities



Logistic models

$$\mathrm{odds}(E) = \frac{\mathrm{pr}(E)}{\mathrm{pr}(\bar{E})}, \qquad \mathrm{pr}(E) = \frac{\mathrm{odds}(E)}{1 + \mathrm{odds}(E)}$$

$$\mathrm{logit}(E) = \log \mathrm{odds}(E)$$

$$\mathrm{logit}\,\pi(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

$$\mathrm{odds}(Y = 1; x) = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)$$

Parameter interpretation:

β_p is the change in the log odds... per unit change in x_p when $x_1, \ldots, x_{p-1}$ are held fixed

e^{β_p} is the multiplicative change in the odds per unit change in x_p...

The change in the probability is approximately

$$d\pi(x) = \pi(x)(1 - \pi(x))\beta_p dx_p < \beta_p dx_p/4$$

Example: Titanic survival rates

Data for 1316 passengers and 885 crew
Data listed by individual

Name	class	child/adult	female/male	survived
Andrews, C	1	0	1	1
Andrews, T	1	0	0	0
Rothschild, M	1	0	0	0
Young, M	1	0	1	1
Abelson, S	2	0	0	0
Abbott, E	3	0	0	0
⋮	⋮	⋮	⋮	⋮

Coding: class 0=crew; 1=first, 2, 3
adult=0, child=1; male=0, female=1; died=0, survived=1

Summary table

Summary data for 1316 passengers and 885 crew

	crew	class1	class2	class3	child	female	surv	pct
crew	885	0	0	0	0	23	212	24
class1	0	325	0	0	6	145	203	63
class2	0	0	285	0	24	106	118	41
class3	0	0	0	706	79	196	178	25
child	0	6	24	79	109	45	57	52
female	23	145	106	196	45	470	344	73
surv	212	203	118	178	57	344	711	32

Logistic model

Main effects only:

$$\begin{aligned} \mathrm{odds}(\text{survival}) &= \text{class effect} \times \text{sex effect} \times \text{age effect} \\ \log \mathrm{odds}(\text{survival}) &= \text{class effect} + \text{sex effect} + \text{age effect} \end{aligned}$$

Class treated as a 4-level factor
model formula used: `class+child+female`

Parameter estimates and S.E.s

	`class+child+female`		`class+child+female-1`	
(Intercept)	-1.23	0.08	–	–
crew	–	–	-1.23	0.080
class1	0.86	0.157	-0.38	0.136
class2	-0.16	0.174	-1.39	0.155
class3	-0.92	0.149	-2.15	0.127
child	1.06	0.244	1.06	0.244
female	2.42	0.140	2.42	0.140

Parameter estimates and S.E.s

	class+child+female		class+child+female-1	
(Intercept)	-1.23	0.08	–	–
crew	–	–	-1.23	0.080
class1	0.86	0.157	-0.38	0.136
class2	-0.16	0.174	-1.39	0.155
class3	-0.92	0.149	-2.15	0.127
child	1.06	0.244	1.06	0.244
female	2.42	0.140	2.42	0.140

Fitted survival odds:

class1 adult female: $\exp(-0.38 + 2.42) = 7.69$, prob=0.89
class1 adult male: $\exp(-0.38) = 0.68$, prob=0.41
class2 adult male: $\exp(-1.39) = 0.25$, prob=0.20
class3 adult male: $\exp(-2.15) = 0.12$, prob=0.10
class3 male child: $\exp(-2.15 + 1.06) = 0.34$, prob=0.25

Does the additive model fit?

Children and females given preferential treatment

Does the degree of preference depend on class?

Model formula: `child+class+female+female*class`

Parameter	estimates	S.E.
class0	-1.25	0.082
class1	-0.68	0.158
class2	-1.91	0.219
class3	-1.70	0.124
child	1.05	0.230
female	3.15	0.625
class1:female	1.09	0.820
class2:female	0.64	0.725
class3:female	-1.78	0.652

Fitted survival odds:

class1 adult female: $\exp(-0.68 + 3.15 + 1.09) = 35.16$ pr=0.97

class3 adult female: $\exp(-1.70 + 3.15 - 1.78) = 0.72$ pr=0.41

Formal test of significance

Likelihood ratio or deviance test: Deviances:

Model	Deviance	d.f.	Dev diff
`child+class+female`	2210	2195	
`child+class*female`	2143	2192	67 (3)
`class+child*female`	2192	2194	18 (1)
`class*child+female`	2174	2193	36 (2)
`all 2-factors`	2097	2187	113 (8)

All 2-factor interactions strongly significant.

Some findings:

Preferential treatment given to females;

Degree of preference higher in 1st class than 3rd

Effects of nuisance parameters

Binary matched pairs problem:

Subjects come in matched pairs (blocks of size 2)

Pair effects $\lambda_1, \ldots, \lambda_n$

Common treatment effect Δ

$$\mathrm{pr}(Y_{i1} = 1) = \frac{e^{\lambda_i}}{1 + e^{\lambda_i}}, \quad \mathrm{pr}(Y_{i2} = 1) = \frac{e^{\lambda_i + \Delta}}{1 + e^{\lambda_i + \Delta}}$$

Likelihood

$$\prod_i \frac{e^{\lambda_i y_{i1}}}{1 + e^{\lambda_i}} \frac{e^{\lambda_i y_{i2} + \Delta y_{i2}}}{1 + e^{\lambda_i + \Delta}} = \prod_i \frac{e^{\lambda_i y_{i\cdot}}}{(1 + e^{\lambda_i})(1 + e^{\lambda_i + \Delta})} \times e^{\Delta y_{\cdot 2}}$$

So $(\{y_{i\cdot}\}, y_{\cdot 2})$ is sufficient
and $(\{y_{i\cdot}\}$ is sufficient for fixed Δ.

Binary matched pairs (contd)

Maximum likelihood estimation: Sufficient statistic $(\{y_{i.}\}, y_{.2})$

$$
\begin{aligned}
E(Y_{i.};\lambda,\Delta) &= \frac{e^{\lambda_i}}{1+e^{\lambda_i}} + \frac{e^{\lambda_i+\Delta}}{1+e^{\lambda_i+\Delta}} \\
E(Y_{.2};\lambda,\Delta) &= \sum \frac{e^{\lambda_i+\Delta}}{1+e^{\lambda_i+\Delta}}
\end{aligned}
$$

Maximum likelihood equations for fixed Δ

$$
\hat\lambda_i(\Delta) = \begin{cases} -\infty & y_{i.} = 0 \\ -\Delta/2 & y_{i.} = 1 \\ \infty & y_{i.} = 2 \end{cases}
$$

Profile likelihood is a product over the mixed-response pairs

$$
\mathrm{pr}(0,1;\hat\lambda,\Delta) = \frac{e^{\Delta/2}}{\text{[illegible]}}
$$

Conditional likelihood

$$\begin{aligned}
\mathrm{pr}(Y_{i1}, Y_{i2} = (0,1)) &= \frac{e^{\lambda_i+\Delta}}{(1+e^{\lambda_i})(1+e^{\lambda_i+\Delta})} \\
\mathrm{pr}(Y_{i1}, Y_{i2} = (1,0)) &= \frac{e^{\lambda_i}}{(1+e^{\lambda_i})(1+e^{\lambda_i+\Delta})} \\
\mathrm{pr}(Y_{i1}, Y_{i2} = (0,1) \,|\, \mathrm{mixed}) &= \frac{e^{\lambda_i+\Delta}}{e^{\lambda_i}+e^{\lambda_i+\Delta}} \\
&= e^{\Delta}/(1+e^{\Delta})
\end{aligned}$$

Hence the mixed pairs are conditionally iid

The conditional distribution given $n_{10} + n_{01} = S$

$$n_{01} \sim \mathrm{Bin}(S, e^{\Delta}/(1+e^{\Delta}))$$

Conditional mle is $\hat{\Delta}_c = \log(n_{01}/n_{10}) = \hat{\Delta}/2$