The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

1863-12

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

24 September - 12 October, 2007

Generalized linear models IV

Peter McCullagh

Department of Statistics
University of Chicago
Chicago IL 60637, USA

Generalized linear models IV

Peter McCullagh

Department of Statistics
University of Chicago

Trieste, October 2007

Outline

Log-linear models

Model:

$Y_1, \ldots$ independent Poisson variables
x_i is the covariate value for unit i
$E(Y_i) = \mu(x_i)$ is the mean for unit i
Components independent but not identically distributed
Need a model for the response function $\mu(x)$

Constraint: $\mu(x) \geq 0$

Standard response function: $\mu(x) = \exp(\beta' x)$
i.e. $\log E(Y(u)) = \beta' x(u)$ for unit u
not $\log Y(u) = \beta' x(u) + \epsilon(u)$

Poisson process

What is a Poisson process?

probabilistic model for the occurrence of a series of events

Example: arrival of calls at a telephone exchange

calls arrive at constant rate λ per unit time

Events in non-overlapping intervals are independent

Expected number of events in $(t, t+dt)$ is $\lambda\, dt$

event occurs in dt with probability $\lambda\, dt$

$$\mathrm{pr}(\text{no event in } (0,t)) = \lim_{k\to\infty} (1 - \lambda t/k)^k = e^{-\lambda t}$$

(divide interval into k equal parts of length t/k)

$$\begin{aligned} \mathrm{pr}(n \text{ events at } dt_1, \ldots, dt_n) &= \lambda^n e^{-\lambda t} dt_1 \cdots dt_n \\ \mathrm{pr}(n \text{ events}) &= \lambda^n t^n e^{-\lambda t}/n! \end{aligned}$$

Poisson process (contd)

What is a Poisson process?

probabilistic model for random subset $Y \subset \mathcal{X}$

No of events in A: $Y(A) = \#(Y \cap A)$ for $A \subset \mathcal{X}$

$Y(A), Y(A')$ for disjoint subsets of $\mathcal{X}$ are independent

$\Rightarrow Y$ is a Poisson process

$Y(A) \sim \mathrm{Po}(\Lambda(A))$

$Y(A \cup A') = Y(A) + Y(A') \sim \mathrm{Po}(\Lambda(A \cup A'))$ (disjoint)

Example: $\mathcal{X} = \mathcal{R}$, $\Lambda(A) = \lambda \times$ length of A

$Y((0, t]) \sim \mathrm{Po}(\lambda t)$

Time T_1 to first event:

$\mathrm{pr}(T_1 > t) = \mathrm{Po}(\lambda t)(0) = e^{-\lambda t}$

Density function: $-d/dt\,(e^{-\lambda t}) = \lambda e^{-\lambda t}$

Applications of Poisson processes

Applications of Poisson processes

- Queueing theory (telephone and computer networks,...)
- Insurance: occurrences of car accidents
- Insurance: major disasters (earthquakes, hurricanes, floods)
- Rare events: death by horsekick in the Prussian army
- Rare events: Space shuttle explosions
- Medical/epidemiological: incidence of anencephalus
- Radioactive decay: No of α-particles emitted in 1 sec
- Genetics: Number of crossovers on chromosome 1
- Warehousing: No of orders for product X in 1 week
- Counting: No. of votes for G.B. in S. Dakota

Example: Ship damage incidents

Type	Constr	Oper	Months service	Incidents
A	1960–64	1960–74	127	0
A	1960–64	1975–79	63	0
A	1965–69	1960–74	1095	3
A	1965–69	1975–79	1095	4
A	1970–74	1960–74	1512	6
A	1970–74	1975–79	3353	18
A	1975–79	1960–74	0	0*
A	1975–79	1975–79	2244	11
B	1960–64	1960–74	44882	39
B	1960–64	1975–79	17176	29
B	1965–69	1960–74	28609	58
B	1965–69	1975–79	20370	53
B	1970–74	1960–74	7064	12
B	1970–74	1975–79	13099	44
B	1975–79	1960–74	0	0*
B	1975–79	1975–79	7117	18

Statistical modelling

Response: number of incidents — suggests Poisson model

each ship type with its own rate

each construction period with its own rate

OPEC effect for period of operation

Effects on $E(Y)$ should be multiplicative

Expected value proportional to period at risk

Initial model:

$$
\begin{aligned}
\log(E(Y)) = \beta_0 &+ \log\text{(aggregate months service)} \\
&+ \text{(effect due to ship type)} \\
&+ \text{(effect due to year of construction)} \\
&+ \text{(OPEC effect due to service period).}
\end{aligned}
$$

Remarks: factors and offsets

Stationarity, Independence

Testing for stationarity

Extended non-stationary model:

$$\begin{aligned}\log(E(Y)) = \beta_0 + \beta_1 \log(\text{aggregate months service}) \\ + (\text{effect due to ship type}) \\ + (\text{effect due to year of construction}) \\ + (\text{OPEC effect due to service period}).\end{aligned}$$

Stationarity: $\beta_1 = 1$

$\hat\beta_1 = 0.9, \quad \text{s.e.}(\hat\beta_1) = 0.1$

consistent with stationarity

All subsequent work uses stationary model – why?

Zero values

Zero exposure/risk implies zero count

$t = 0$ implies $y = 0$ (non-informative components can be ignored)

Could have $t > 0$, $\mu > 0$ with $\mathrm{pr}(y = 0) = e^{-\mu} > 0$

likelihood contribution $e^{-\mu}$

Impossible factor combinations: $t = 0$
Possible factor combinations that do not occur: $t = 0$

Model specification in R:

```
y = y[t>0],....
glm(y~stype+cons+period, family=poisson(),
offset=log(t[t>0]))
```

`stype, cons, period` as factors

Over-dispersion

Meaning of over-dispersion: $\mathrm{var}(Y_i) > E(Y_i)$

Sources of over-dispersion: correlations and unmodelled inhomogeneities

Modelling of over-dispersion: $\mathrm{var}(Y_i = \sigma^2 E(Y_i)$

Effect of over-dispersion: $\mathrm{cov}(\hat\beta) \simeq \sigma^2 (X'WX)^{-1}$

Detection of over-dispersion:

$X^2 = \sum (Y_i - \hat\mu_i)^2/\mu_i$ Pearson statistic

$E(X^2) \simeq (n-p)\sigma^2$; approx distribution $\sigma^2\chi^2_{n-p}$

$s^2 = X^2/(n-p)$ estimates σ^2

$s^2 = 1.69$ for ship damage data

Accommodation of over-dispersion:

$\mathrm{cov}(\hat\beta) \simeq s^2 (X'WX)^{-1}$

Conclusions for ship damage data

(1) Examination of interactions: Not very large

(2) Parameter estimates and s.e.s (with over-dispersion factor)

Parameter		Estimate	S.E.	$\exp(\hat\beta)$
Ship type	A	0.00	—	1.00
	B	−0.54	0.23	0.58
	C	−0.69	0.43	0.50
	D	−0.08	0.38	0.92
	E	0.33	0.31	1.39
Year of construction	1960–64	0.00	—	1.00
	1965–69	0.70	0.19	2.01
	1970–74	0.82	0.22	2.27
	1975–79	0.45	0.30	1.57
Service period	1960–74	0.00	—	1.00
	1975–79	0.38	0.15	1.46

Post OPEC rate $= e^{0.38} \times$ rate before 1974

Example with continuous response