A Reliable Constrained Method for Identity Link Poisson Regression

Ian Marschner

Macquarie University, Sydney

Australasian Region of the International Biometrics Society, Taupo, NZ, Dec 2009.

Identity Link Poisson Model

- Useful for count data where the mean depends additively on a collection of covariates

- Useful as an approximation to identity link binomial model for modelling additive probabilities

- Epidemiological applications:
  - Rate difference regression
  - Risk difference regression

- Alternative to the usual multiplicative models:
  - Multiplicative incidence (log link Poisson)
  - Multiplicative risk (log link binomial)
  - Multiplicative odds (logistic link binomial)

Advantages and Disadvantages

- **Modelling advantages**:

    - Rate/risk differences may be preferred to rate/risk/odds ratios

    - Additive model may fit the data better than a multiplicative model

Advantages and Disadvantages

- **Modelling advantages**:

  - Rate/risk differences may be preferred to rate/risk/odds ratios

  - Additive model may fit the data better than a multiplicative model

- **Computational disadvantages**:

  - Positivity constraints may lead to non-convergence of standard computational methods e.g. IRLS

  - Standard errors and confidence intervals may be invalid if MLE is on or near parameter space boundary

  - Even if standard methods do converge for a single data set, they often won't work for thousands of bootstrap replications

Categorical Model

- **Count data**: $\{Y_i; i = 1, \ldots, n\}$ with covariates $\mathbf{x}_i = (x_{i1}, \ldots, x_{iA})$
  - $x_{ij} \in \{1, 2, \ldots, k_j\}$ and $\mathbf{x}_i \in \mathcal{X} = \prod_j \{1, 2, \ldots, k_j\}$

Categorical Model

- **Count data**: $\{Y_i; i = 1, \ldots, n\}$ with covariates $\mathbf{x}_i = (x_{i1}, \ldots, x_{iA})$
  - $x_{ij} \in \{1, 2, \ldots, k_j\}$ and $\mathbf{x}_i \in \mathcal{X} = \prod_j \{1, 2, \ldots, k_j\}$

- **Poisson model**: Y_i independent Poisson$(N_i \lambda_i)$ $\quad i = 1, \ldots, n$
  - Additive means: $\lambda_i = \Lambda(\mathbf{x}_i; \theta) = \alpha_0 + \sum_{j=1}^{A} \alpha_j(x_{ij})$
  - N_i are known positive standardisation constants e.g. expected number of events or number of binomial trials
  - Log-likelihood: $L(\theta) = \sum_{i=1}^{n} Y_i \log[N_i \Lambda(\mathbf{x}_i; \theta)] - N_i \Lambda(\mathbf{x}_i; \theta)$

Categorical Model

- **Count data**: $\{Y_i; i = 1, \ldots, n\}$ with covariates $\mathbf{x}_i = (x_{i1}, \ldots, x_{iA})$

  - $x_{ij} \in \{1, 2, \ldots, k_j\}$ and $\mathbf{x}_i \in \mathcal{X} = \prod_j \{1, 2, \ldots, k_j\}$

- **Poisson model**: Y_i independent $\mathrm{Poisson}(N_i \lambda_i)$ $\qquad i = 1, \ldots, n$

  - Additive means: $\lambda_i = \Lambda(\mathbf{x}_i; \theta) = \alpha_0 + \sum_{j=1}^A \alpha_j(x_{ij})$

  - N_i are known positive standardisation constants e.g. expected number of events or number of binomial trials

  - Log-likelihood: $L(\theta) = \sum_{i=1}^n Y_i \log[N_i \Lambda(\mathbf{x}_i; \theta)] - N_i \Lambda(\mathbf{x}_i; \theta)$

- **Identifiability restriction**: for some $\mathbf{r} = (r_1, \ldots, r_A) \in \mathcal{X}$ define $\alpha_j(r_j) = 0$ for each $j = 1, \ldots, A$

Categorical Model

- **Count data**: $\{Y_i; i = 1, \ldots, n\}$ with covariates $\mathbf{x}_i = (x_{i1}, \ldots, x_{iA})$
  - $x_{ij} \in \{1, 2, \ldots, k_j\}$ and $\mathbf{x}_i \in \mathcal{X} = \prod_j \{1, 2, \ldots, k_j\}$

- **Poisson model**: Y_i independent Poisson$(N_i \lambda_i)$ $i = 1, \ldots, n$
  - Additive means: $\lambda_i = \Lambda(\mathbf{x}_i; \theta) = \alpha_0 + \sum_{j=1}^{A} \alpha_j(x_{ij})$
  - N_i are known positive standardisation constants e.g. expected number of events or number of binomial trials
  - Log-likelihood: $L(\theta) = \sum_{i=1}^{n} Y_i \log[N_i \Lambda(\mathbf{x}_i; \theta)] - N_i \Lambda(\mathbf{x}_i; \theta)$

- **Identifiability restriction**: for some $\mathbf{r} = (r_1, \ldots, r_A) \in \mathcal{X}$ define $\alpha_j(r_j) = 0$ for each $j = 1, \ldots, A$

- **Parameter space**: $\Theta = \{\theta : \Lambda(\mathbf{x}; \theta) \geq 0 \text{ for all } \mathbf{x} \in \mathcal{X}\}$
  - Interpretation: overall means $E(Y_i)$ are constrained to be non-negative but the individual $\alpha_j(x_{ij})$ can be negative

Model with Non-negative Coefficients

- Suppose we did constrain the model: $\alpha_j(x_{ij}) \geq 0$ for all i and j

- This type of identity link Poisson model is well known in specialised applications involving Poisson deconvolution

- The EM algorithm provides a computationally stable and convenient method to fit such models

- This is great if you have a specialised application that calls for non-negative coefficients

- In general, identity link Poisson regression models do not have this constraint so the specialised methods are not applicable

- We discuss how to adapt the specialised methods to the general case

Parameter Space Subset

- Given the chosen identifiability restriction $\mathbf{r}$, define the following subset of the parameter space Θ:

$$\Theta(\mathbf{r}) = \{\theta \in \Theta : \Lambda(\mathbf{r}; \theta) \leq \Lambda(\mathbf{x}; \theta) \text{ for all } \mathbf{x} \in \mathcal{X}\}$$

- $\Theta(\mathbf{r})$ is the subset of the parameter space in which no covariate pattern has a smaller (standardised) Poisson mean than the covariate pattern $\mathbf{r}$

- $\Theta(\mathbf{r})$ is the subset of the parameter space in which all $\alpha_j(x_{ij}) \geq 0$ for $x_{ij} \neq r_j$

- Given the chosen identifiability restriction $\mathbf{r}$, fitting a non-negative coefficient model is the same as fitting the model subject to the constraint $\theta \in \Theta(\mathbf{r})$

- What use is this for the general model with $\theta \in \Theta$?

Fitting the General Model

- For one of the choices of $\mathbf{r} \in \mathcal{X}$ the maximum of $L(\theta)$ over Θ is within $\Theta(\mathbf{r})$ since

$$\Theta = \bigcup_{\mathbf{r} \in \mathcal{X}} \Theta(\mathbf{r})$$

- Cycle through the finite number of possible choices of $\mathbf{r} \in \mathcal{X}$ and maximise $L(\theta)$ over $\Theta(\mathbf{r})$ in each case

- This involves a sequence of (stable and convenient) EM algorithm applications to yield a sequence of contrained estimates $\hat{\theta}(\mathbf{r})$

- The MLE $\hat{\theta}$ is then the $\hat{\theta}(\mathbf{r})$ that yields the highest value of L. That is, $\hat{\theta} = \hat{\theta}(\mathbf{r}^*)$ where:

$$\mathbf{r}^* = \{\mathbf{r} \in \mathcal{X} : L(\hat{\theta}(\mathbf{r})) \geq L(\hat{\theta}(\mathbf{s})) \text{ for all } \mathbf{s} \in \mathcal{X}\}$$

- If you find a stationary point then the process can terminate but often the MLE will not be a stationary point

EM Algorithm

- Since the Poisson means are additive, Y_i can be viewed as the sum of unobserved Poisson variates:

$$Y_i = \sum_{j=0}^{A} Y_i^{(j)} \qquad \text{where} \qquad E(Y_i^{(j)}) = N_i \alpha_j(x_{ij})$$

- EM algorithm: $\{Y_i\}$ = observed data and $\{Y_i^{(j)}\}$ = complete data

- Only works for the non-negative coefficient model since the complete data means must be non-negative

- Yields a stable multiplicative algorithm which increases L within $\Theta(\mathbf{r})$ as long as $\hat{\theta}^{(0)} \in \Theta(\mathbf{r})$

- At convergence: $\hat{\theta}^{(\infty)} = \hat{\theta}(\mathbf{r})$, the maximum of $L(\theta)$ over $\Theta(\mathbf{r})$

EM Algorithm

- At iteration c:

$$\hat{Y}_{i(c+1)}^{(j)} = E(Y_i^{(j)}|Y_i;\hat{\theta}^{(c)}) = \hat{\alpha}_j^{(c)}(l)Y_i\Big/\Lambda(\mathbf{x}_i;\hat{\theta}^{(c)})$$

$$
\begin{aligned}
\hat{\alpha}_j^{(c+1)}(l) &= \sum_{i\in I_{jl}} \hat{Y}_{i(c+1)}^{(l)} \Big/ \sum_{i\in I_{jl}} N_i \\
&= \hat{\alpha}_j^{(c)}(l) \left(\sum_{i\in I_{jl}} N_i \right)^{-1} \sum_{i\in I_{jl}} \left(Y_i \Big/ \Lambda(\mathbf{x}_i;\hat{\theta}^{(c)}) \right)
\end{aligned}
$$

where $I_{jl} = \{i : x_{ij} = l\}$ identifies the observations that have covariate j equal to l

Extension to Continuous Covariates

- **Isotonic regression**:
    - Covariate with $C + 1$ distinct observed values $w_0 < \cdots < w_C$
    - Additive contribution is $f(w_i)$ for unspecified non-decreasing f
    - Add C dummy covariates to the model corresponding to the C non-negative increments $\gamma_i = f(w_i) - f(w_{i-1})$
    - EM algorithm maximises subject to the isotonicity constraint $\gamma_i \geq 0$
    - Works for ordinal categorical or continuous covariates

Extension to Continuous Covariates

- **Isotonic regression**:
  - Covariate with $C + 1$ distinct observed values $w_0 < \cdots < w_C$
  - Additive contribution is $f(w_i)$ for unspecified non-decreasing f
  - Add C dummy covariates to the model corresponing to the C non-negative increments $\gamma_i = f(w_i) - f(w_{i-1})$
  - EM algorithm maximises subject to the isotonicity constraint $\gamma_i \geq 0$
  - Works for ordinal categorical or continuous covariates

- **Linear regression**:
  - Shift each covariate by subtracting the minimum (or maximum)
  - Use the EM algorithm to maximise subject to a non-negative (or non-positive) gradient
  - Cycle through all the combinations and find the highest likelihood

Epidemiological data on respiratory cancer mortality in 8047 Montana smelter workers

Table 2. Observed and Expected Numbers of Deaths in 40 States Defined by Birthplace and Cumulative Years Working in Heavy and Moderate Arsenic Areas

Years Moderate Arsenic	Years Heavy Arsenic									
	0		<1		1–4		5+		Total	
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
U.S. Born										
0	28	20.86	2	1.21	3	.56	6	.60	39	23.23
<1	7	4.91	2	.76	1	.20	2	.29	12	6.16
1–4	8	3.10	4	.33	1	.10	1	.11	15	3.64
5–14	4	1.58	0	.12	0	.08	0	.01	4	1.79
15+	4	1.14	1	.11	0	.05	0	.03	5	1.33
Totals	51	31.59	9	2.53	5	.99	10	1.04	75	36.15ᵃ
Foreign Born										
0	33	7.34	1	.39	0	.11	2	.28	36	8.12
<1	2	1.31	0	.10	0	.02	0	.05	2	1.48
1–4	4	.91	0	.01	0	.07	0	.04	4	1.03
5–14	6	1.05	0	.02	0	.13	0	.04	6	1.24
15+	16	1.60	3	.20	0	.10	0	.01	19	1.91
Total	61	12.21	4	.72	0	.43	2	.42	67	13.78ᵃ

```
> glm.fit(Design*Expected,Observed,family=poisson(link="identity"))
Error:  no valid set of coefficients has been found:  please supply starting values
In addition:  Warning message:  In log(ifelse(y == 0, 1, y/mu)) :  NaNs produced
```

Example 1: Respiratory Cancer Mortality

Illustration of calculations for constrained identity link model with 2 covariates:

| r | Constant | Birthplace | | Years heavy arsenic | | | | Log-likelihood |
		Foreign	US	0	<1	1–4	5+	
$(1,1)$	2.56	–	0.00	–	1.44	0.96	4.98	345.02
$(1,2)$	2.66	–	0.00	0.00	–	0.86	4.88	343.99
$(1,3)$	2.59	–	0.00	0.00	1.41	–	4.95	344.80
$(1,4)$	2.72	–	0.00	0.00	1.28	0.80	–	340.63
(2,1)	**1.65**	**3.05**	–	–	**1.77**	**1.87**	**5.47**	**360.61**
$(2,2)$	1.79	2.99	–	0.00	–	1.73	5.34	358.73
$(2,3)$	1.74	2.97	–	0.00	1.69	–	5.40	359.54
$(2,4)$	1.87	2.92	–	0.00	1.57	1.65	–	354.35

Example 1: Respiratory Cancer Mortality

Constrained identity link analysis with all covariates and isotonic or unrestricted arsenic exposure effects

| Covariate	Unrestricted analysis		Isotonic analysis	
	Estimate	95% CI	Estimate	95% CI
Constant	1.41	$0.91 - 1.90$	1.37	$0.89 - 1.78$
Foreign born (vs. U.S.)	2.48	$1.26 - 3.70$	2.48	$1.29 - 3.67$
Years moderate arsenic (vs. 0)				
<1 year	-0.17	$-1.18 - 0.85$	0.00	$0.00 - 1.03$
1–4 years	1.60	$-0.17 - 3.37$	1.35	$0.00 - 2.63$
5–14 years	0.86	$-1.16 - 2.88$	1.35	$0.31 - 3.50$
15+ years	4.03	$1.09 - 6.96$	4.06	$1.59 - 7.18$
Years heavy arsenic (vs. 0)				
<1 year	1.17	$-0.80 - 3.13$	1.21	$0.00 - 2.96$
1–4 years	1.93	$-1.47 - 5.32$	1.95	$0.00 - 5.26$
5+ years	5.68	$1.11 - 10.25$	5.70	$1.90 - 10.59$
Identity link deviance	26.05		26.46	
Log link deviance	30.36		31.95	

Example 1: Respiratory Cancer Mortality

Results of 5000 bootstrap replications of the isotonic identity link model
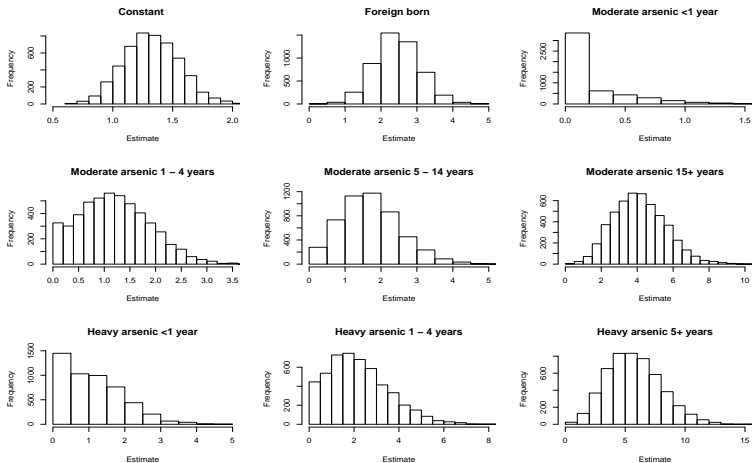
Example 2: Crab Population Counts Agresti (1996)

- **Outcome**: Number of partners for $n = 173$ horseshoe crabs
- **Model**: Identity link Poisson model for the mean number of partners versus crab size (cm), adjusted for colour and spine condition
- **Analysis**: IRLS fails to converge; identity link better than log link

Linear and Isotonic Identity Link Models

Bootstrap 95% Confidence Intervals
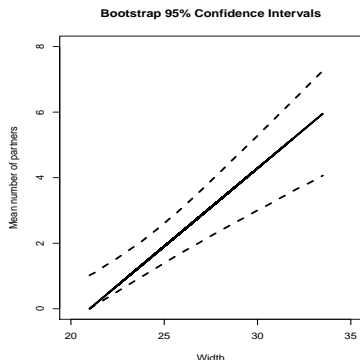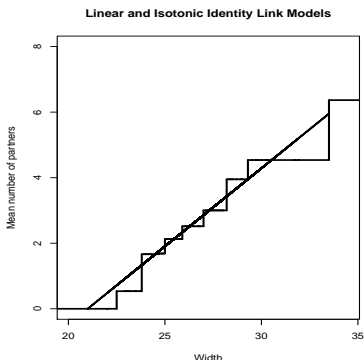
Summary

- **Reliability**:
  - Computationally stable method for fitting constrained identity link Poisson models when standard methods fail
  - Stationary or non-stationary maximum
  - Boundary or interior of parameter space

- **Flexibility**:
  - Categorical, ordinal or continuous covariates
  - Unspecified isotonic relationships
  - Smoothing step could easily be added for non-parameteric smooth relationships

- **Validity**:
  - Computationally reliable bootstrap analysis yields confidence intervals that satisfy parameter constraints
  - Even when standard methods converge for the main analysis, they are often too unreliable for a resampling analysis