Lecture 14: Extensions to Loglinear Models

MATH3823 Generalised Linear Models

Richard P Mann

MATH3823 Generalised Linear Models

Reading

Course notes: Chapter 7

`www.richardpmann.com/MATH3823`

The Problem: Fixed Marginals

In many studies, some totals are fixed by design:

- Clinical trial: Fixed number per treatment group
- Case-control study: Fixed cases and controls
- Survey: Fixed sample size

When totals are fixed:

- Cell counts are *not* independent Poisson
- The standard Poisson model is theoretically incorrect
- But we can still use Poisson GLM.

Example: Fixed Total

Melanoma study: Total of 400 patients (fixed).

Questions:

- Can we still use Poisson loglinear models?
- What distribution do the counts actually follow?
- How do we interpret the results?

Answer: The counts follow a *multinomial* distribution, but Poisson gives the same parameter estimates.

From Poisson to Multinomial

Proposition

If $Y_1, \ldots, Y_k$ are independent Poisson(λ_i), then conditional on $Y_+ = n$:

$$(Y_1, \ldots, Y_k) \mid Y_+ = n \sim \text{Multinomial}\left(n; \frac{\lambda_1}{\lambda_+}, \ldots, \frac{\lambda_k}{\lambda_+}\right)$$

Key insight: Conditioning independent Poissons on their sum gives a multinomial.

From Poisson to Binomial

Proposition

If $Y_1 \sim \text{Poisson}(\lambda_1)$ and $Y_2 \sim \text{Poisson}(\lambda_2)$ independently, then:

$$Y_1 \mid (Y_1 + Y_2 = n) \sim \text{Binomial}\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$$

Application: In a 2×2 table with fixed row totals, the row distributions are binomial.

Product-Multinomial Distribution

When row totals y_{i+} are fixed:
Each row follows an independent multinomial:

$$(Y_{i1}, \ldots, Y_{ic}) \sim \text{Multinomial}\left(y_{i+}; \pi_{i1}, \ldots, \pi_{ic}\right)$$

This is called a product-multinomial distribution.

The joint distribution is the product of row-wise multinomials.

The Key Theorem

Theorem

Tables with fixed margin sums can be analyzed using a multinomial or product-multinomial model as though they were independent Poisson models, provided terms corresponding to the fixed margins are included in the model.

What this means:

- Fixed total $y_{++} \Rightarrow$ include μ (intercept)
- Fixed row totals $y_{i+} \Rightarrow$ include row main effects α_i
- Fixed column totals $y_{+j} \Rightarrow$ include column main effects β_j

Why Does This Work?

Mathematical reason:
When we condition on fixed margins, the parameters for those margins “absorb” the constraint. The remaining parameters (e.g., interactions) have the same MLEs whether we use:

- Poisson likelihood
- Multinomial/product-multinomial likelihood

Practical implication:
We can use `glm(..., family = poisson)` even when the Poisson assumption is technically violated.

Application: Flu Vaccine Trial

Data: 73 participants, fixed group sizes

Group	Low	Moderate	High	Total
Placebo	25	8	5	38
Vaccine	6	18	11	35
Total	31	26	16	73

Row totals fixed: 38 placebo, 35 vaccine (by design).

Flu Data: R Analysis

```r
# Create data
flu <- data.frame(
  Group = factor(rep(c("Placebo", "Vaccine"),
                     each = 3)),
  Response = factor(rep(c("Low", "Mod", "High"), 2),
                    levels = c("Low", "Mod", "High")),
  Count = c(25, 8, 5, 6, 18, 11)
)

# Independence model (must include Group for fixed rows)
model <- glm(Count ~ Group + Response,
             family = poisson, data = flu)
summary(model)
```

Flu Data: Results

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.6461     0.1796  14.729  < 2e-16 ***
GroupVaccine -0.0816     0.2399  -0.340    0.734
ResponseMod  -0.1757     0.2721  -0.646    0.518
ResponseHigh -0.6614     0.3164  -2.091    0.037 *

Residual deviance: 18.643  on 2  degrees of freedom
```

Test of independence:

- Deviance: 18.64 on 2 df
- p-value < 0.001
- Strong evidence of association between group and response

Model Simplification

Observation: Perhaps "Moderate" and "High" can be combined?

```
# Combine Moderate and High
flu$Response2 <- flu$Response
levels(flu$Response2) <- c("Low", "ModHigh", "ModHigh")

# Refit
model2 <- glm(Count ~ Group + Response2,
              family = poisson, data = flu)
```

Result: Deviance 2.405 on 2 df ($p = 0.30$).
Interpretation: Vaccine affects Low vs. Moderate/High, not Moderate vs. High.

Guidelines for Fixed Margins

Two-way table:

Fixed	Include in model
Total only (y_{++})	Intercept μ
Row totals (y_{i+})	Row effects α_i
Column totals (y_{+j})	Column effects β_j
Both margins	Both α_i and β_j

Three-way table: Similar principle — include main effects and interactions corresponding to fixed margins.

Cautions

1. **Small expected counts:**
   - χ^2 approximation may be poor
   - Rule of thumb: Expected counts ≥ 5
   - Consider exact tests or combining categories
2. **Overdispersion:**
   - If deviance $\gg$ df, consider `quasipoisson`
   - Or investigate causes (missing covariates, clustering)
3. **Sparse tables:**
   - Many zeros cause problems
   - May need specialized methods

Summary

Key points:

- Fixed marginals $\Rightarrow$ counts are not independent Poisson
- Conditioning Poisson on sum gives multinomial
- Product-multinomial when row totals are fixed
- **Key theorem:** Can use Poisson GLM if fixed-margin terms included
- Include row effects if row totals are fixed
- Include column effects if column totals are fixed
- Parameter estimates for interactions are the same.

Next lecture: Course revision and summary.