Chapter 13 Count Models

Paul E. Johnson

April 16, 2016

1 Glossary

Notation continuing from previous

ξ_{ij}: "xi" total (or combined) individual row-level error term for group i row j,

Longitudinal terminology: "group" is often a single person (j is group of rows) and the rows for each person are differentiated by time (which is i in this book). Economists tend to call these i, and t, whereas Laird-Ware/Bates would call them i j

ζ_j: "zeta" a group level random error, j indexes a grouping variable, $Var(\zeta_j) = \psi$

ϵ_{ij}: "epsilon" $Var(\epsilon_{ij}) = \theta$. Individual row-level error uncorrelated with ζ_j

$Cov(y_{ij}, y_{i'j}|\zeta_j) = 0$. Apart from ζ_j, the observed scores are "conditionally uncorrelated"

2 Count Data

1. 13.2.1 This chapter is about counts, NOT r/n (binomial), which xtmelogit can handle
   a) PJ Note: Proportion with offset in Poisson framework possible
   b) Also proportions (Beta regression)
   c) Compositional data
2. 13.2.2 Count can be seen as aggregated individual data, and the grouping naturally leads to multilevel issues
   a) Caution: ecological fallacy
3. 13.3 Poisson model. λ is incidence rate
   a) Interesting notation here is linked to time interval of "exposure", they use $\mu_i = \lambda t_i$.
   b) Poisson is written up separately in my slides on count data, don't want to re type all that.
   c) p 689: $Pr(y_1|\lambda) \times Pr(y_2\lambda) = Pr(y_1 + y_2)$. The probability of the aggregated counts across 2 times is product of 2 probabilities.
   d) log link function

e) $exp(\beta)$is INCIDENCE RATE Ratio (IRR) due to increase in x.

f) p. 690 If exposure times differ among observation units, offset is used to synchronize that.

g) p 690 Overdispersion more common than underdispersion (I agree).

4. Longitudinal data structure

   a) Note distracting tendency to generate new, unusual symbols. p. 692 introduces ν_{ij} where usually we would have η_{ij} in the MLM literature.

5. ROBUST: need to get better understanding of standard errors and what magic Stata is doing on "vce(cluster id)".

   a) p 694. "The confidnece intervals are based on the sandwich estimator taking the dependence of the repeated counts (given the covariates) into account"). Frustratingly vague amid all of the other very precise elements here.

6. Random intercept Poisson regression

   a) Frailty.

      i. Cases with equivalent predictors always can generate different counts. That's not the special thing here.

      ii. The special thing is that frailty allows 2 cases with same predictors to have different probability distributions, like $Poisson(x+1)$ versus $Poisson(x+10)$. Here, we suppose the differences between cases are simple parametric differences, but it would be just a small conceptual step to suppose they are different probability models (leading to a "mixture model").

   b) ζ_{1j} is a group-level error, $\zeta_{1j}|x_{ij} \sim N(0, \psi_{11})$

   $$\mu_{ij} = exp(X_{ij}\beta + \zeta_{1j}) \qquad (1)$$

   c) Marginal (population averaged) models line up with unit-level results, unlike in logistic regression. This a a point I had not considered before reading this book.

   p. 696 "Interestingly, we can also interpret the coefficients as marginal or population-averaged effects because the relationship between the marginal expectation of the count (given x_{ij} but averaged over ζ_{1j}) and the covariates is

   $$\mu_{ji}^{M} = exp\{(\beta_1 + \psi_{11}/2) + \beta_2 x_{2i} + \ldots\} \qquad (2)$$

   The intercept is the only parameter that is not the same in the marginal and conditional models [in general the marginal intercept is $E\{exp(\beta_1 + \zeta_{ji})\}$, which becomes $exp(\beta_1 + \psi_{11}/2)$ for a normally distributed random intercept".

   NOTE: need go get derivation on that.

   Then the implication p. 696 "Because marginal and conditional effects coincide for random-intercept Poisson regression models, in contrast to random-intercept logistic regression, consistent estimation of regression parameters (apart from the intercept) does not hinge on the correct choice of the random-intercept distribution. What is required is correct specification of the mean structure and lack of correlation between the random intercpet and the covariates."

d) The marginal (population averaged) variance

$$Var(y_{ij}|x_{ij}) = \mu_{ij}^M + (\mu_{ij}^M)^2\{exp(\psi_{11}) - 1\} \tag{3}$$

Note that, like the negative binomial, the variance of the random effect outcome is the poisson variance plus another piece

e) ICC hard to define, no "latent variable" characterization similar to logit that would allow similar ICC.

   i. Option 1. Analytical derivation. Tough! See Stryhn 2006, given covariates, correlation between two counts in same cluster that have all same predictors is still a very complicated formula (p. 697)

   ii. Option 2. Simulation. (p. 697) randomly draw individuals with same predictors, form an IRR.

$$IRR_{median} = exp\{\sqrt{2\psi_{11}}\Phi^{-1}(3/4)\} \tag{4}$$

   See Section 10.9.2, the median incidence-rate ratio

f) Stata estimation example inserts normally distibuted error in a poisson model.

g) **Hypo test** of the variance estimate. Same idea as other models, "null hypothesis is on the border of the parameter space" (p. 698). From commentary, cannot tell for sure if software is supposed to make corrections described or if use is supposed to remember that they would be necessary.

   Takeaway point: If The **LR test** gives a very small p value, then the random effect variance is not 0, but if the p value is larger, where we might conclude "there is no random effect", we should be somewhat cautious.

   Similar commentary p. 703 "...the asymptotic null distribution for this test is $1/2\chi_1^2 + 1/2\chi_2^2$, and we conclude that the random-intercept Poisson model is rejected in favor of the random-coefficient Poisson model."

7. Random Slopes

a) Suppose linear predictor has intercept and slope variance

$$(\beta_1 + \zeta_{1j}) + (\beta_2 + \zeta_{2j})x_{21} + \ldots \tag{5}$$

b) Unstructured correlation between 2 group-level random effects

$$\left[\begin{array}{c} \zeta_{1j} \\ \zeta_{2j} \end{array}\right] \sim MVN\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \left[\begin{array}{cc} \psi_{11} & \psi_{21} \\ \psi_{21} & \psi_{22} \end{array}\right]\right) \tag{6}$$

c) Marginal expected count is

$$E[y_{ji}|x_{ij}] = exp\{\beta_1 + \beta_2 x_{2i} + \dots (\psi_{11} + 2\psi_{21}x_{2i} + \psi_{22}x_{2i}^2)/2\} \tag{7}$$

NOTE: Get derivation on that.

factor out the x_{2i}, reorganize.

d) The variable x_2 is dichotomous, {0,1}. "Hence, the multiplicative effect of the reform on the marginal epected count now no longer equals $exp(\beta_2)$ as in the random-intercept model but $exp(\beta_2 + \psi_{21} + \psi_{22}/2)$. In general, teh marginal expected count is given by the exponential of the sum of the linear predictor and half the variance of the random part of the model." (p. 702).

e) "Interpretation of estimates" p. 705

   i. I don't understand part after "Instead of thinking". Isn't this just plugging in 0 and 1 as two values of x_2?

f) Big point about identification. (TODO Come back, clean this up).

   i. In LMM and Logit models, the individual level variance is not separately estimable

   ii. p. 705 Individual level variance θ...

8. Overdispersion (single level models)

   a) Normally distributed individual noise implies "GLMM", but log Gamma noise implies Negative Binomial

   b) p. 707 Shows how to use xtpoisson to estimate individual-level normal error.

      i. could we get both individual and group level errors? Big question.

   c) Note we get more zeroes out of these models, possibly solving the "Zero Inflation" issue.

   d) Negative Binomial (p. 707)

      i. individual frailty is log Gamma distribution, which implies outcomes are Negative Binomial.

         A. Say it this way. The random effects are determined FIRST.

         B. Then they are nserted into $Pois(X\beta + Z\zeta)$

         C. Hence the observed count is Poisson, conditional on ζ.

         D. Since we don't know ζ (not observed), we have to imagine the range of possible draws from the Poisson process that would happen if ζ takes on many values. Hence, the overall "marginal" distribution of y happens to follow a Neg. Binomial distribution.

      ii. **TODO:** think about writing one clean derivation of log Gamma error to NB distribution!

9. Quasilikelihood:

   a) Either GLMM or NB models are based on a specific function for inserted random noise. Can we avoid that assumption?

   b) Quasi models try to build a regression model without making that detailed assumption. Suppose instead one assumes

      i. The expected value is same as in all models discussed so far, the log link: $E[y_{ji}] = ln(\mu_{ij}) = X_{ij}\beta$

      ii. But we know the variance of y_{ji} is not same as Poisson, because there is some uncertainty. So ASSERT

$$Var(y_{ij}|X_{ji}) = \phi\mu_{ji} \tag{8}$$

      Variance is proportional to mu. We'd like to estimate ϕ.

   c) Wedderburn (197x) proposed this

   d) In my mind, it is tied to generalized estimating equations.

10. Level-1 overdispersion

   a) I've been wondering a long time, what if the data generator has random error at various levels.

   b) We are used to hierarchical error in regressions,

      i. Linear Mixed model $X_j\beta + \zeta_{1j} + \zeta_{2j} + \epsilon_{ij}$

      ii. It seems superficial/silly now to use different letters ζ and ϵ , these are just errors we assume exist

   c) In a count model, say poisson, why can't we introduce similar? Could fit a multilevel model. If the software would work, there's no conceptual problem.

   d) Alternative math model spelled out p. 711. Hausman, Hall and Griliches (1984) with a special Stata function `xtnbreg`. If you read into the details, you see this is a quasilikelihood model with an assertion of the EV and the variance function. The interesting bit is that the variance component has a draw from a Beta distribution embedded in it. Neat! However, after fighting though that, we learn that RHS don't like to interpret these models. p. 712 "We therefore do not recommend using this model." Wasted effort.

   e) Alternative estimation approach is to fit same old random intercept model, but correct the standard errors for individual-level variability. "Robust" standard errors again.

11. Conditional Poisson: The Fixed Effects equivalent.

   a) Could estimate dummy variables, one for each unit.

b) Can suppress those dummies, as in linear models (regression on deviation scores) or in logistic regression (clogit, Chapter 10.14.1, which I did not understand either).

c) `xtpoisson y x1 x2, fe` : MUST omit all predictors that are constant within units. No level 2 predictors.

d) p. 715: Conditional NB regression.

12. Generalized Estimating Equations. Wish I could find my other notes on this.

13. Missing values

    a) p. 716 "From the discussion in section 13.7.1, we would expect the estimated conditional effects using random-intercept Poisson models to be similar to the estimated marginal effects using ordinary Poisson regression or GEE (apart from the intercept). However, it is eviden from table 13.1 that the estimates are quite different." Some guesses follow, including unbalanced data.

    b) p. 717 "It is often claimed that GEE requires data to be missing completely at random (MCAR) for consistency, but missingness can actually depend on the covariates."

    c) Simulation study p 717.

    d) **TODO:** This might be worth some further study. Will do in R since Stata programming still seems horrible to me. Anyway, it addresses problem related to resignation from studies and how mixed models correct for it (p. 717). In ordinary Poisson model, intervention appeared effective, but mixed model says it was not, and allegation is that the missingness structure is fooling the Poisson model.

14. Estimating a proportion (ratio) model. Higher-than-expected rates of lip cancer take on same pattern as any proportion model estimated by offets with Poisson. Section 13.13 - 13.15 should probably be in one section.

    a) It is an example of using an "offset" to estimate a model of a proportion.

$$ln(\frac{r_j}{n_j}) = ln(r_j) - ln(n_j) = X_j\beta \tag{9}$$

$$ln(r_j) - ln(n_j) = X_j\beta \tag{10}$$

$$ln(r_j) = ln(n_j) + X_j\beta \tag{11}$$

    b) on the right hand side, we have a predictor for which we do not estimate a coefficient, $ln(n_j)$. That's an "offset", it is just an intercept we assume exists.

    c) p. 724: shows gllamm estimation of grouped random effect with offset.

    d) Predicted values: some details worked out. I'm not writing out much here.

i. Posterior expectation:

$$\int exp(\hat{\beta}_1 + \zeta_j) \times Posterior(\zeta_j | observed) d\zeta_j \tag{12}$$

What's that? we know ζ_j varies randomly, we have to use its various values to weight $exp(\hat{\beta}_1 + \zeta_j)$.

ii. gllapred will calculate those for us

15. NPML: Nonparametric Maximum Likelihood

a) PJ Note: after reading the Murray Aitkin papers and the GLIM book, the NPML idea seems mis-named. It should be something like "discrete random effects" models.

b) Instead of assuming $\zeta_j \sim N(0,1)$ or such, just suppose there's an unmeasured variable that takes on a few discrete values. Then the challenge is to try to figure out

i. what are the values that happened to produce the data we got, and

ii. can we guess which rows were affected by which of the random effects, and

iii. in light of those random effects, do our other regression estimates change?

c) That sounds horribly complicated, but Aitkin showed these models can be estimated with much less elaborate software than LMM or GLMM.

d) Big conceptual problem is that the estimation process can proceed in steps. "Allow 2 discrete valued random effects" then "Allow 3 values" then "Allow 4 values" and the calculations do not always stabilize and the conclusions we might reach about the models can be affected by this uncertain calculation.

e) See the beautiful graph p. 731!

f)

3 Stata notes

1. p. 691 `xtdescribe`

2. Fitting Counts: without random effects

   a) p. 693 `poisson`

   b) `glm` Same thing, general purpose name

   c) option "vce(cluster id)". WTF

3. Random Intercept: Poisson with random effects

   a) p. 697 xtpoisson

i. **NOTE** CRITICAL observation: xtpoisson can fit a gamma distributed $exp(\zeta)$ term, meaning it is equivalent to negative binomial in that case.

A. KEY: check if same model can have individual-level log-gamma error as well as group level errors. WOW!

B. Hence necessary to include argument "normal" in call (p. 698)

b) p. 699 xtmepoisson `|| id:`

i. does not require declaration that error is normal, that's assumed.

A. Default uses 7 quadrature points

B. Could use Laplace method equivalent of 1 quadrature point

c) gllamm. Interesting design

```
generate cons = 1
eq ri: cons
gllamm numvisit x1 x2, family(poisson) link(log) i(id) eqs
    (ri) eform adapt
```

i. Seems to offer avenue to fit model with both individual level and group level random effects.

ii. p. 701. lrtest. "We must divide the p-value by 2, as discussed in section 2.6.2,..."

4. Random slope

a) xtpoisson cannot do this, it only for intercepts

b) xtmepoisson: `|| id: reform`. Don't forget `covariance(unstructured)`

c) gllamm (p. 704)

i. Code shows how to use estimates from random intercept model to start the random slope estimator.

5. Fixed effects estimator:

a) `xtpoisson y x1 x1, fe`

6. post estimation simulation in gllamm "gllasim" similar to simulate() in R functions.

a) TODO: type in their code, learn from it.