A Flexible Regression Model for Count Data

Kimberly F. Sellers

Associate Professor of Statistics
Department of Mathematics and Statistics
Georgetown University

Research Mathematical Statistician
Center for Statistical Research & Methodology
U.S. Census Bureau

Adapted from muppet.wikia.com

Issue with Poisson Regression

- Poisson assumption

$$\Rightarrow \quad \text{Var}(Y_i) = \text{E}(Y_i) \quad \Rightarrow \quad GOF = \frac{\text{Var}(Y_i)}{E(Y_i)} = 1$$

- Count data oftentimes do not conform to assumption
  - Over-dispersion: $GOF > 1$
  - Under-dispersion: $GOF < 1$

Alternative I: Negative Binomial Regression

- Benefit:
  - Regression tool available in many statistical software packages
- Drawbacks:
  - Does not allow for under-dispersion
  - Requires fixing r in order to express log-likelihood in form of a generalized linear model (GLM)

Alternative II: Restricted Generalized Poisson Regression

(Famoye 1993)

- Has the form

$$P(Y_i = y_i \mid \mu_i, \alpha) = \left(\frac{\mu_i}{1+\alpha\mu_i}\right)^{y_i} \frac{(1+\alpha y_i)^{y_i-1}}{y_i!} \exp\left(\frac{-\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}\right), \quad y_i = 0, 1, 2, \ldots$$

- link function, $\log \mu_i = \boldsymbol{\beta}' \boldsymbol{X}_i$
- α restricted to $1 + \alpha\mu_i > 0$, and $1 + \alpha y_i > 0$
  - $\alpha = 0$: Poisson regression
  - $\alpha > 0$: over-dispersion
  - $-\frac{2}{\mu_i} < \alpha < 0$: under-dispersion

Alternative II: Restricted Generalized Poisson Regression (cont.)

- Benefits:
  - handles over- or under-dispersion
- Drawbacks:
  - limited ability to handle under-dispersion
  - belongs to exponential family only for constant α

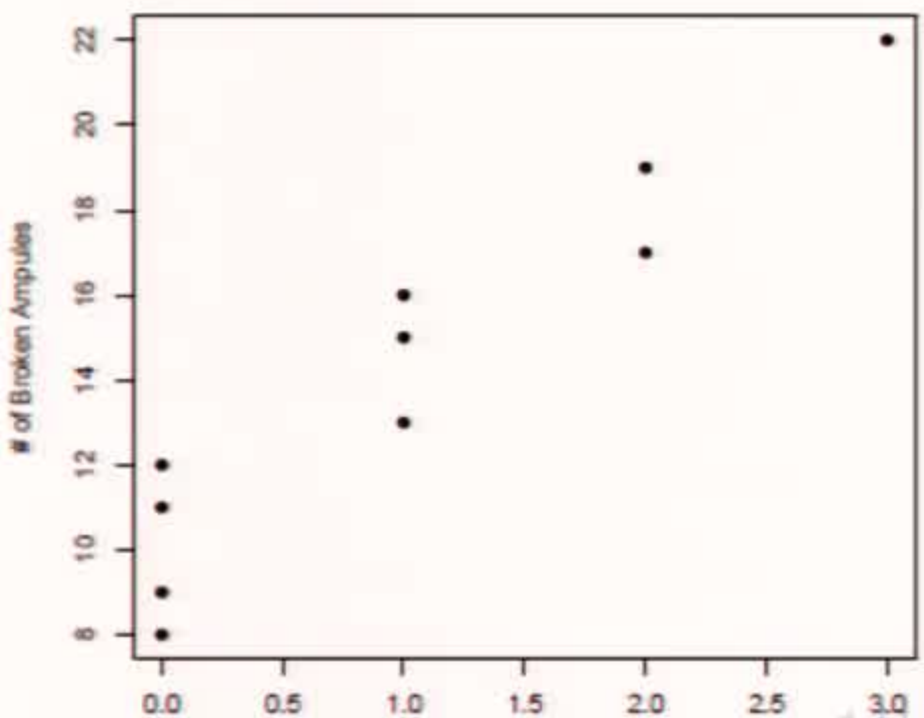
Introduce Conway-Maxwell-Poisson (COM-Poisson) Regression

Determining Maximum Likelihood Estimates (MLEs) (cont.)

- Determined iteratively via Newton-type algorithm (**nlminb** or **optim** in *R*)
  - Starting values: $\hat{\nu} = 1$ and Poisson estimates, $\widehat{\boldsymbol{\beta}}$
- Standard errors derived via Fisher Information matrix
- Hypothesis testing procedure developed for dispersion

Airfreight Breakage: Underdispersion

Data stem from 10 air shipments, each carrying 1000 ampules on the flight. For each shipment, we have the number of times the carton was transferred from one aircraft to another (Y) and the number of ampules found broken upon arrival (X)



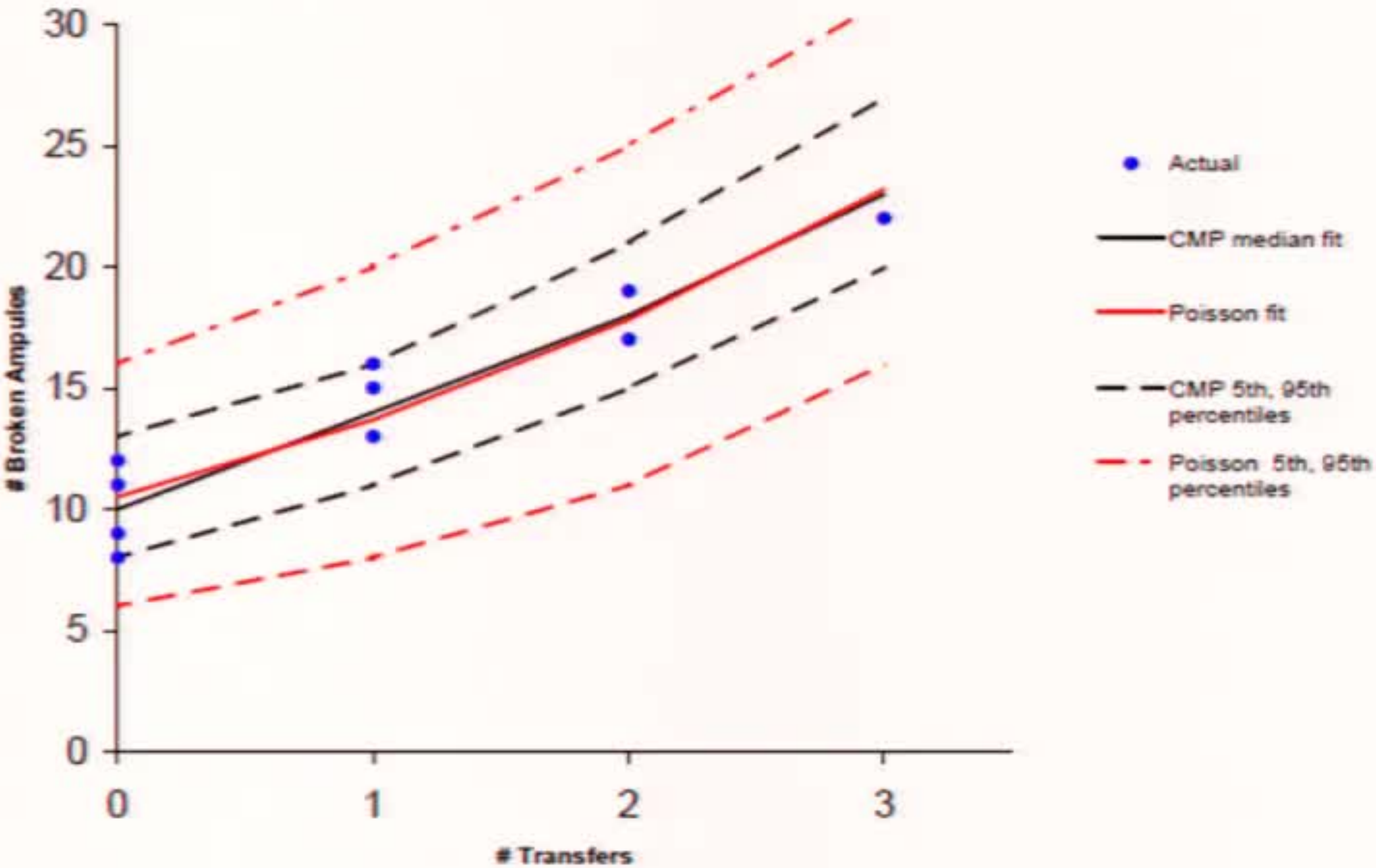
Airfreight Breakage Example (cont.)

Estimated coefficients and standard errors (in parentheses) for Airfreight example, for five types of regression models

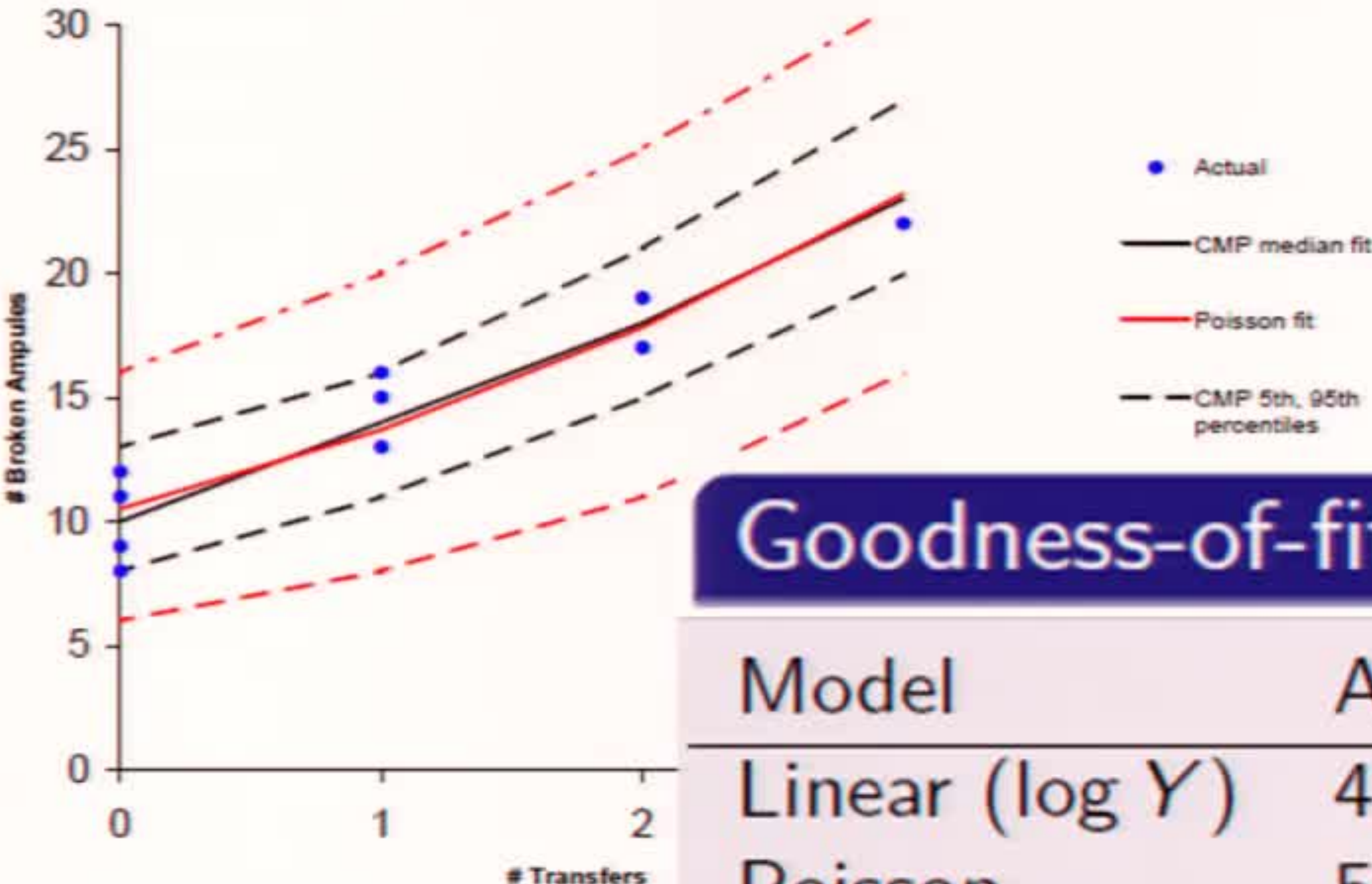
Model	$\hat{\beta}_0\,(\hat{\sigma}_{\hat{\beta}_0})$	$\hat{\beta}_1\,(\hat{\sigma}_{\hat{\beta}_1})$
Linear ($\log Y$) $\hat{\sigma} = 0.141$	2.3273 (0.0631)	0.2800 (0.0446)
Poisson	2.3529 (0.1317)	0.2638 (0.0792)
CMP ($\hat{\nu} = 5.7818$, $\hat{\sigma}_{\hat{\nu}} = 2.597$)	13.8247 (6.2369)	1.4838 (0.6888)

- Negative Binomial regression produces Poisson estimates
- RGPR does not converge
- Test statistic for dispersion: C=9.10 (p-val =0.003)
- 90% bootstrap CI for ν: (4.414, 20.643)

Airfreight Breakage Example (cont.)



Airfreight Breakage Example (cont.)



Goodness-of-fit statistics

Model	AIC_C	MSE
Linear ($\log Y$)	49.37	2.363
Poisson	52.11	2.210
COM-Poisson	**47.29**	**1.900**

Art Book Purchases (cont.)

Estimated coefficients and standard errors (in parentheses) for four regression models.

Model	$\hat{\beta}_0\,(\hat{\sigma}_{\hat{\beta}_0})$	$\hat{\beta}_{Mths}\,(\hat{\sigma}_{\hat{\beta}_{Mths}})$	$\hat{\beta}_{Books}\,(\hat{\sigma}_{\hat{\beta}_{Books}})$
Poisson/NB	-2.29 (0.18)	-0.06 (0.02)	0.73 (0.05)
Logistic	-2.23 (0.24)	-0.07 (0.02)	0.99 (0.14)
COM-Poisson	-2.23 (0.24)	-0.07 (0.02)	0.99 (0.14)
$(\hat{\nu} = 30.4, \hat{\sigma}_{\hat{\nu}} = 10123)$			

Toronto Accidents Example: Overdispersion

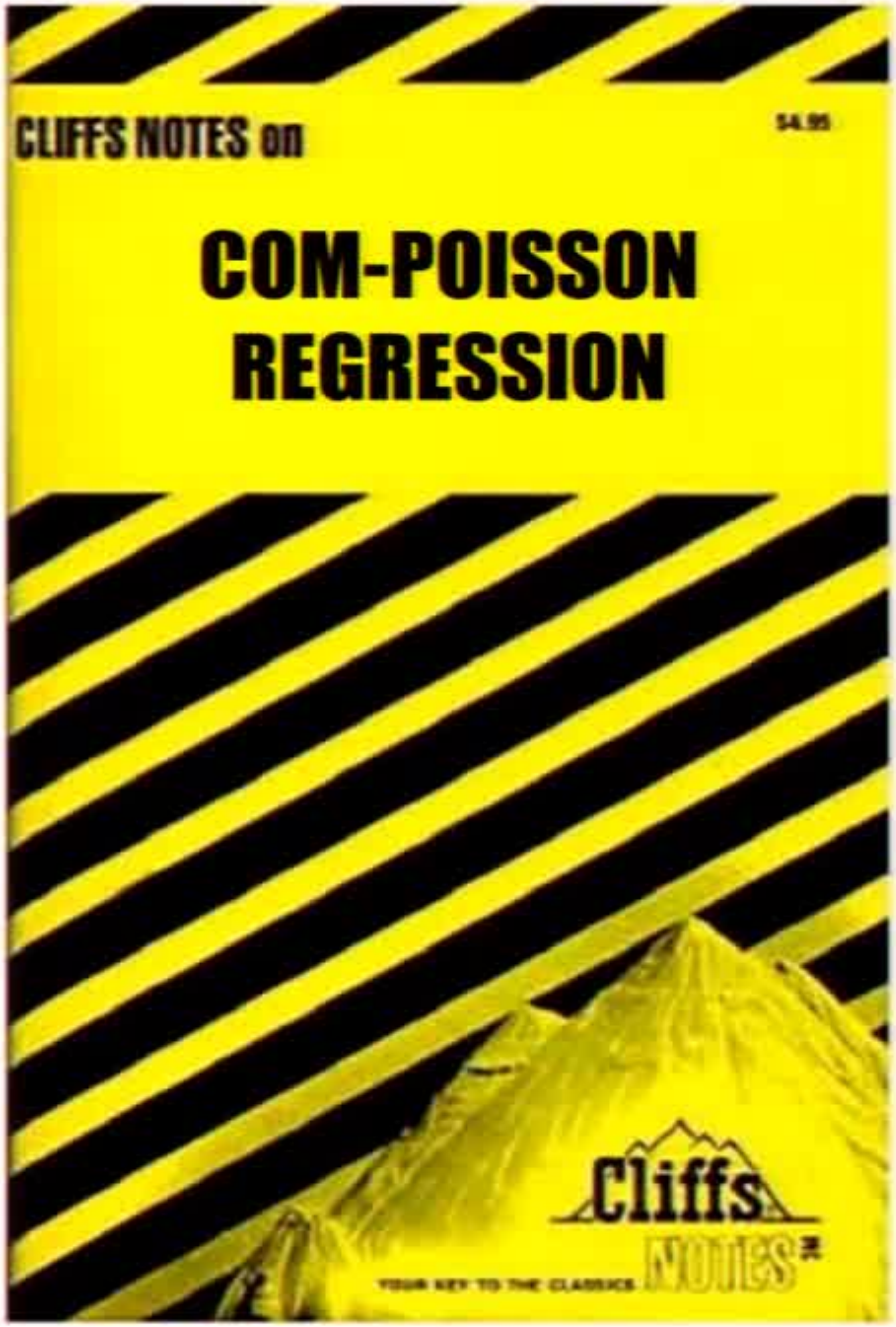
- Lord et al. (2008) independently established Bayesian formulation of COM-Poisson regression
- Used link function, $\log(\lambda^{1/\nu})$
- Estimate parameters using Markov-chain Monte Carlo (MCMC)
- Dataset: 1995 crash data at 868 signalized intersections located in Toronto, Ontario

Example: Toronto Accident Data (Lord et al. 2008) (cont.)

Estimated models: comparing two COM-Poisson formulations (ours and Lord et al. 2008), and four alternative models for the Toronto crash data.

Model	Extra parameter	$\hat{\beta}_0$ ($\hat{\sigma}_{\hat{\beta}_0}$)	$\hat{\beta}_1$ ($\hat{\sigma}_{\hat{\beta}_1}$)	$\hat{\beta}_2$ ($\hat{\sigma}_{\hat{\beta}_2}$)
Our formulation	$\hat{\nu}$=0.3492 (0.0208)	-11.7027$\hat{\nu}$ (0.7501$\hat{\nu}$)	0.6559$\hat{\nu}$ (0.0619$\hat{\nu}$)	0.7911$\hat{\nu}$ (0.0461$\hat{\nu}$)
Lord et al. 2008	$\hat{\nu}$=0.3408 (0.0208)	-11.53 (0.4159)	0.6350 (0.0474)	0.7950 (0.0310)
Linear Reg (log Y)	$\hat{\sigma}$=0.3491	-9.8132 (0.5161)	0.5966 (0.0512)	0.6566 (0.0226)
Poisson		-10.2342 (0.2838)	0.6029 (0.0288)	0.7038 (0.0140)
Neg-Bin	$\hat{r}$=7.154 (0.625)	-10.2458 (0.4626)	0.6207 (0.0456)	0.6853 (0.0215)
RGPR	$\hat{\alpha}$=0.050 (0.004)	-10.2357 (0.4640)	0.6205 (0.0451)	0.6843 (0.0215)

- Two models produce nearly identical coefficients, standard errors, and fitted values
- Significant runtime difference!



- Powerful regression tool for count data given any form of dispersion
- Encompasses common regressions
- **COMPoissonReg**: R package for constant dispersion available on CRAN (Sellers and Lotze, 2010, 2011, 2015)
- Sellers and Raim (2015) address zero-inflation

Who Knew?!

Abstract

Title of Thesis: Iterative Methods for Computing Mean First Passage Times of Markov Chains

Name of degree candidate: Kimberly Ann Flagg Sellers

Degree and year: Master of Arts, 1998

Thesis directed by: Dr. Dianne P. O'Leary, Professor
Department of Computer Science

Much of the literature involving the numerical analysis of Markov chains focuses on computing stationary distributions. Although a great deal of information can be gained from such analyses, we can still learn more from investigating other parameters associated with the chain (e.g., mean first passage times). Finding stationary distributions and mean first passage times involves solving linear systems of the form $AX = B$, where A is an $n \times n$ nonsymmetric matrix, but X and B are matrices of size $n \times s$ ($s \geq 1$). In many practical applications involving queueing networks, the number of states n is enormous and direct methods cannot be used. This research compares the Block-GMRES algorithm to the Block Quasi-Newton method with appropriate preconditioning to solve such systems and calculate the mean first passage time from one state to another.