Quantitative Social Science Methods, I, Lecture Notes: Discrete and Limited Outcome Models

Gary King[1]
Institute for Quantitative Social Science
Harvard University

August 17, 2020

[1] GaryKing.org

Ordered Dependent Variable Models

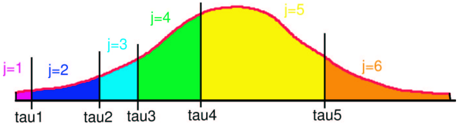
Grouped Binary Variable Models

Count Models

Duration Models and Censoring

Ordered Dependent Variable Models

- Ordered probit: $Y_i^* \sim \text{STN}(y_i^*|\mu_i), \quad \mu_i = x_i\beta, \quad Y_i \perp\!\!\!\perp Y_{i'}|X$
- Quiz: what's the model if Y_i^* is observed? linear regression?
- Quiz: What about the assumption $\sigma^2 = 1$
- Observation mechanism: $y_i = \begin{cases} j & \text{if } \tau_{j-1,i} \le y_i^* \le \tau_{j,i} \\ 0 & \text{otherwise} \end{cases}$



- Quiz: Where would you apply the ordered probit model?
- With observation mechanism: same β as regression
- Quiz: How to interpret β?
- If instead $Y_i^* \sim \text{STL}(y_i^*|\mu_i)$: ordered logit model

Deriving the Ordered Probit Likelihood Function

- Probability of one observation

$$\Pr(Y_i = j) = \Pr(\tau_{j-1} \le Y_i^* \le \tau_j) = \int_{\tau_{j-1}}^{\tau_j} \mathrm{STN}(y_i^* | \mu_i) dy_i^*$$
$$= F_{stn}(\tau_j | \mu_i) - F_{stn}(\tau_{j-1} | \mu_i)$$
$$= F_{stn}(\tau_j | x_i\beta) - F_{stn}(\tau_{j-1} | x_i\beta)$$

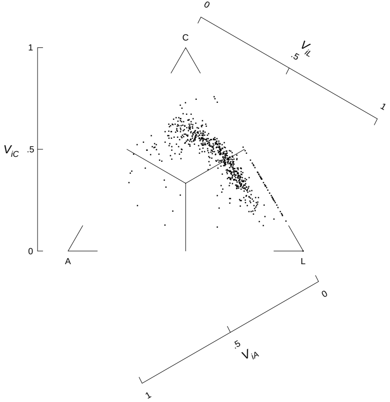
- Joint probability: $P(Y) = \prod_{i=1}^{n} [\Pr(Y_i = j)]$
- Log-likelihood

$$\ln L(\beta, \tau | y) = \sum_{i=1}^{n} \ln \Pr(Y_i = j)$$
$$= \sum_{i=1}^{n} \ln \left[F_{stn}(\tau_j | x_i\beta) - F_{stn}(\tau_{j-1} | x_i\beta) \right]$$

- Careful of optimization constraints: $\tau_{j-1} < \tau_j, \quad \forall j$

Ordinal Probit Interpretation

- β: linear effect of X on Y^* (in SD units)
- $\widehat{\Pr(Y_i|X)}$: on the simplex, J probabilities sum to 1
- One first difference: effects all J probabilities
- When one probability goes up: ≥ 1 must go down
- Ternary diagram: UK Cons., Labour, Alliance vote



Grouped Uncorrelated Binary Model

- Same as binary logit, but only observe sum of iid Bernoulli trials
- Quiz: Where would you apply this?
- Model: $Y_i \sim \text{Binomial}(y_i|\pi_i),\ \pi_i = [1 + e^{-x_i\beta}]^{-1},\ E(Y_i) = N_i\pi_i$
- $L(\pi|y) \propto \prod_{i=1}^n \text{Binomial}(y_i|\pi_i) = \prod_{i=1}^n \binom{N_i}{y_i}\pi_i^{y_i}(1-\pi_i)^{N_i - y_i}$
- Log-likelihood

$$
\begin{aligned}
\ln L(\pi|y) &= \sum_{i=1}^n \left\{ \ln \binom{N_i}{y_i} + y_i \ln \pi_i + (N_i - y_i)\ln(1-\pi_i) \right\} \\
&\doteq \sum_{i=1}^n \left\{ -y_i \ln[1 + e^{-x_i\beta}] + (N_i - y_i)\ln\left(1 - [1 + e^{-x_i\beta}]^{-1}\right) \right\} \\
&= \sum_{i=1}^n \left\{ (N_i - y_i)\ln(1 + e^{x_i\beta}) - y_i \ln(1 + e^{-x_i\beta}) \right\}
\end{aligned}
$$

- Similar to the binary logit log-likelihood

Grouped Uncorrelated Binary Model Interpretation

- Inferential goal: the same π as in binary logit
- Draw one simulation
  - Maximize log-likelihood; save $\hat{\beta}$ and $\hat{V}(\hat{\beta})$
  - Draw $\tilde{\beta} \sim N[\hat{\beta}, \hat{V}(\hat{\beta})]$ from multivariate normal
  - Set X to your choice of values, X_c
  - Calculate simulations of the probability that any of the component binary variables is a one: $\tilde{\pi}_c = \left[1 + e^{-x_c \tilde{\beta}}\right]^{-1}$
  - If π is of interest, stop
  - If simulations of y are needed, draw $\tilde{y}$ from Binomial$(y_i|\pi_i)$
- Compute QOIs: mean, SD, CI's, histogram, etc.

Grouped Correlated Binary Model

- Modeling issues with grouped uncorrelated model:
  - Is iid assumption reasonable (within observation i)?
  - $V(Y) = \pi_i(1-\pi_i)/N_i$, with no σ^2-like parameter to take up slack
  - These are the same issue!
- Extended Beta-Binomial Model:
  $Y_i \sim f_{\text{ebb}}(y_i|\pi_i, \gamma), \quad \pi_i = [1 + e^{-x_i\beta}]^{-1}$
- The EBB pdf:

$$f_{\text{ebb}}(y_i|\pi_i,\gamma) = \Pr(Y_i = y_i|\pi_i,\gamma,N)$$
$$= \frac{N!}{y_i!(N-y_i)!}\prod_{j=0}^{y_i-1}(\pi_i+\gamma j)\prod_{j=0}^{N-y_i-1}(1-\pi_i+\gamma j)/\prod_{j=0}^{N-1}(1+\gamma j)$$

- Math *looks* complicated, but *is* conceptually simple
- Role of γ: soaks up binomial misspecification
- Assuming binomial when EBB is right: se's & fit wrong

Simulating QOIs from Correlated Binary Model

- Draw one simulation
  - Run `optim`, get $\hat{\eta} = \{\hat{\beta}, \hat{\gamma}\}$ and $\hat{V}(\hat{\eta})$
  - Draw $\tilde{\eta}$ from multivariate normal: $N[\eta \mid \hat{\eta}, \hat{V}(\hat{\eta})]$
  - Set X to your choice of values, X_c
  - Calculate sims of the probability that any of the component binary variables is one: $\tilde{\pi}_c = [1 + e^{-x_c\tilde{\beta}}]^{-1}$
  - If π is of interest, stop
  - If simulations of y are needed, draw $\tilde{y}$ from $f_{\text{ebb}}(y_i|\pi_i)$
- Compute QOIs: mean, SD, CI's, histogram, etc.

Ordered Dependent Variable Models

Grouped Binary Variable Models

Count Models

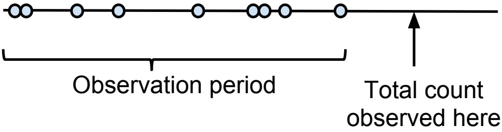
Duration Models and Censoring

Event Count Applications

- Units of analysis: over time (count per year), across areas (count per state), or both
- Event count: Number of events in a time period for some unit
- Upper limit on number of events: none
- Quiz: Can you think of examples from your field?
- Some examples from real research
  - Number of cooperative and conflictual international incidents
  - Number of triplets born in Norway in each half-decade
  - Annual number of appointments to the Supreme Court
  - Number of coups d'etat in African states
  - Number of medical consultations for each survey respondent

Recall Poisson distribution's first principles

- Begin with a (black box) observation period and count point:



- Assumptions concern: events occurring during black box period
- 0 events occur at the start of the period
- No 2 events can occur at the same time
- Markov Independence:

$$\Pr(\text{event at time } t \mid \text{events up to } t-1) \text{ constant } \forall t.$$

- Quiz: When will these assumptions be violated?

The Poisson regression model

- Model: $Y_i \sim \text{Poisson}(y_i|\lambda_i), \quad \lambda_i = \exp(x_i\beta), \quad Y_i \perp\!\!\!\perp Y_j|X$
- The probability density of all the data: $\text{P}(y|\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$
- The log-likelihood:

$$\begin{aligned}
\ln L(\beta|y) &= \sum_{i=1}^{n} \{y_i \ln(\lambda_i) - \lambda_i - \ln y_i!\} \\
&= \sum_{i=1}^{n} \left\{(x_i\beta)y_i - e^{x_i\beta} - \ln y_i!\right\} \\
&\doteq \sum_{i=1}^{n} \left\{(x_i\beta)y_i - e^{x_i\beta}\right\}
\end{aligned}$$

- Modeling issues
  - Like EBB, Markov Independence
  - No extra parameter like σ^2 in regression to take up slack
  - Consequence of violation: SEs and fit wrong

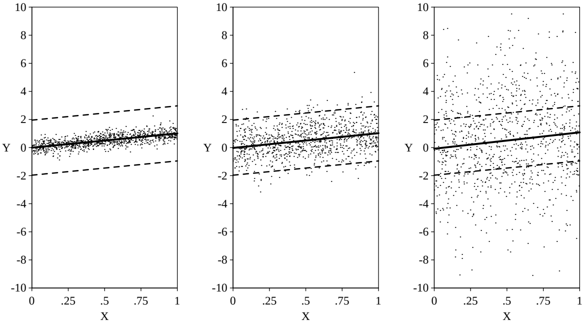
Poisson Model Interpretation

- Derivative method:

$$\frac{\partial \lambda_i}{\partial X_i^1} = \exp(x_i \beta)\beta_1 = \lambda_i \beta_1$$

  so we could use $\bar{y}\beta$ for an approximate linearized effect.
- To simulate
  - Set X_c
  - Draw $\tilde{\beta} \sim N\left(\hat{\beta}, \hat{V}(\hat{\beta})\right)$
  - Compute $\tilde{\lambda}_c = \exp(X_c \tilde{\beta})$
  - Draw fundamental variability: $Y_c \sim \text{Poisson}(y|\tilde{\lambda})$
- Variance under misspecification
  - Under Poisson model: $V(Y_i|X_i) = E(Y_i|X_i)$, heteroskedastic & fixed
  - $V(Y_i|X_i) > E(Y_i|X_i)$ is overdispersion: SEs will be too small (very common)
  - $V(Y_i|X_i) < E(Y_i|X_i)$ is underdispersion: SEs too big
  - Variance conditional on X, dispersion changes with specification

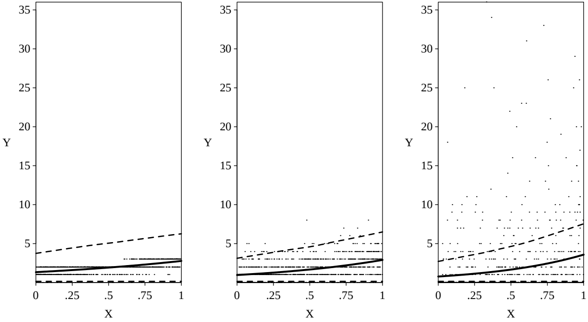
Problems without an extra parameter? Stylized Normal



(a) (b) (c)

$E(Y|X)$ and 95% CI.

Problems without an extra parameter? Poisson



(a) (b) (c)

$E(Y|X)$ and 95% CI.

Negative Binomial Event Count Model

- For overdispersed data (conditional on X)
- The model:
  $Y_i \sim \text{NegBin}(y_i|\phi, \sigma^2), E(Y_i) \equiv \phi = e^{x_i\beta}, Y_i \perp\!\!\!\perp Y_j \mid X$
- Likelihood

$$L(\phi, \sigma^2|y) \propto P(y|\phi, \sigma^2) = \prod_{i=1}^{n} \frac{\Gamma\left(\frac{\phi}{\sigma^2-1} + y_i\right)}{y_i!\Gamma\left(\frac{\phi}{\sigma^2-1}\right)} \left(\frac{\sigma^2 - 1}{\sigma^2}\right)^{y_i} \left(\sigma^2\right)^{\frac{-\phi}{\sigma^2-1}}$$

- Computational Issues
  - $\ln\Gamma(a)$ with large a is hard to compute in 2 steps (since $\Gamma(a) \approx a!$ is immense) but easy in one. In R, see `lgamma`
  - β is unbounded; no need to reparameterize
  - $\sigma^2 > 1$; so reparameterize $\sigma^2 = e^{\gamma} + 1$ and estimate γ
- Interpretation
  - $V(Y|X) = \phi\sigma^2$ and $\sigma^2 > 1$
  - Recall: $\lim_{\sigma^2 \to 1} \text{Negbin}(y_i|\phi_i, \sigma^2) = \text{Poisson}(y_i|\phi_i)$
  - Test of Poisson vs NegBin: look at σ^2 (likelihood ratio doesn't work since Poisson doesn't nest within Negbin)
  - Careful of off-the-shelf programs: maybe $V(Y|X) = \phi(1 + \sigma^2\phi)$

A Generalized Event Count (GEC) Model

An event count model with under-, Poisson, and over-dispersion

Stochastic component:

$$
\begin{aligned}
Y_i \sim \mathrm{GEC}(y_i|\lambda_i, \sigma^2) &\equiv \mathrm{P}(Y = y_i|\lambda_i, \sigma^2) \\
&= \frac{1}{y_i!}\left(\frac{\lambda_i}{\sigma^2}\right)^{\left(y_i, 1-\frac{1}{\sigma^2}\right)} \left[\sum_{j=0}^{y_i^{\max}} \frac{1}{j!}\left(\frac{e^{\lambda_i}}{\sigma^2}\right)^{\left(j, 1-\frac{1}{\sigma^2}\right)}\right]^{-1},
\end{aligned}
$$

where $y_i^{\max} = \infty$ for $\sigma^2 \geq 1$, $n_i = \lambda_i/(1-\sigma^2)$, $y_i^{\max} = [n_i + 1)$ for $0 < \sigma^2 < 1$, and $[x) = x - 1$ for integer x and $\mathrm{floor}(x)$ for non-integer x.

$$
x^{(m,\delta)} = \begin{cases} \prod_{i=0}^{m-1}(x + \delta i) = x(x+\delta)(x+2\delta)\cdots[x+\delta(m-1)] & m \geq 1 \\ 1 & m = 0 \end{cases}
$$

Systematic component:

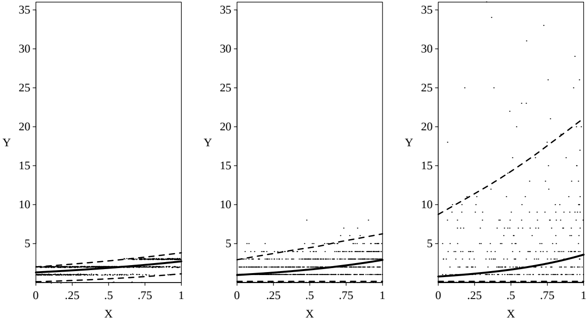
$$
E(Y_i|X_i) \equiv \lambda_i = \exp(X_i\beta)
$$

Crazy math, but same logical structure as the other models today!

GEC Interpretation:

- Special cases of the GEC
  - Negative Binomial, $\sigma^2 > 1$, the over-dispersed case.
  - Poisson, $\sigma^2 = 1$
  - Continuous Parameter Binomial, $0 < \sigma^2 < 1$, the underdispersed case. (This special case itself reduces to an even more special case, the *Binomial*, when $\lambda_i/(1 - \sigma^2)$ is an integer.)
- Can simulate in three parts, indexed by σ^2
- References: King and Signorino, *Political Analysis*, 1996. King *American Journal of Political Science*, 1989

What happens with the extra parameter? GEC



(a) (b) (c)

$E(Y|X)$ and 95% CI.

Ordered Dependent Variable Models

Grouped Binary Variable Models

Count Models

Duration Models and Censoring

The Exponential Model

- The density: Same first principles as the Poisson, except that we observe the duration between events.
- The Model:
  $Y_i \sim \text{expon}(\lambda_i) = \lambda_i e^{-\lambda_i y_i}, \quad E(Y_i) \equiv \frac{1}{\lambda_i} = \frac{1}{e^{-x_i\beta}} = e^{x_i\beta}$
- Quiz: When would you apply this model?
- Log-likelihood

$$\begin{aligned}\ln L(\beta|y) &= \sum_{i=1}^{n} \{\ln \lambda_i - \lambda_i y_i\} \\ &= \sum_{i=1}^{n} \left\{-X_i\beta - e^{-X_i\beta} y_i\right\}\end{aligned}$$

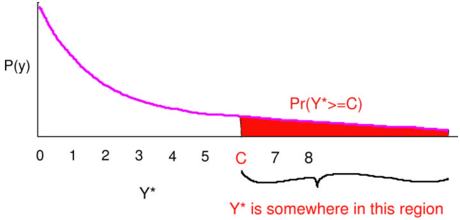
- Reference: King, Alt, Burns, Laver, "A Unified Model of Cabinet Dissolution in Parliamentary Democracies," *American Journal of Political Science*, 1990

What to do about censoring?

- Examples
  - Parliamentary coalition duration; some still in office
  - Duration of unemployment spells; some people still unemployed
  - Duration in graduate school. (What will we do with you?)
  - Longevity
  - Time since you called home
- What to do with the unfinished observations?
  - Drop them: $\rightsquigarrow$ Selection bias.
  - Set duration = observed: $\rightsquigarrow$ Underestimate duration$\rightsquigarrow$bias
  - Guess: Even a good guess $\rightsquigarrow$ biased SEs
  - Better: Include censoring information in the likelihood

Incorporating censoring information in the likelihood

- Observation mechanism $y_i = \begin{cases} y_i^* & \text{if } y_i^* < C \\ C & \text{if } y_i^* \geq C \end{cases}$
- Likelihood for censored observations. All we know is:



- $\Pr(Y_i = C) = \Pr(Y_i^* \geq C) = \int_C^\infty \text{expon}(y_i|\lambda_i)dy_i$
  $= \int_C^\infty \lambda_i e^{-\lambda_i y_i} dy_i = e^{-\lambda_i C}$
- Full likelihood:

$$L(\beta|y) = \left[\prod_{y_i^* < C} \text{expon}(y_i|\lambda_i)\right]\left[\prod_{y_i^* \geq C} \Pr(Y_i^* \geq C)\right]$$