Political Methodology III: Model Based Inference

Justin Grimmer

Professor
Department of Political Science
Stanford University

April 25th, 2019

Model Based Inference

1) Likelihood inference
2) Logit/Probit
3) Ordered Probit
4) Choice Models:
    - Multinomial Probit
        a) DGP
        b) No IIA, But No Likelihood
        c) Quantities of Interest
        d) Interpretation
- Count Models
    - Poisson Regression
        - DGP
        - Quantities of Interest
        - Interpretation
    - Negative Binomial Regression
        - DGP
        - Quantities of Interest
        - Interpetation

Revisiting The IIA Assumption

- IIA (Trump, Cruz, and Sanders)
- Formally: MNL assumes ϵ_{ij} is i.i.d. $\epsilon_{ij} \perp\!\!\!\perp \epsilon_{ik}$ for $j \neq k$
- This implies that unobserved factors affecting Y_{ij}^* are unrelated to those affecting Y_{ik}^*
- When is this assumption plausible?
- Example: Multiparty election with parties R, L1 and L2.
- Do voters' unobserved ideological preferences affect Pr($Y_i =$L1) independently of their effect on Pr($Y_i =$L2)? Probably not.

Revisiting The IIA Assumption

- IIA (Trump, Cruz, and Sanders)
- Formally: MNL assumes ϵ_{ij} is i.i.d. $\epsilon_{ij} \perp\!\!\!\perp \epsilon_{ik}$ for $j \neq k$
- This implies that unobserved factors affecting Y_{ij}^* are unrelated to those affecting Y_{ik}^*
- When is this assumption plausible?
- Example: Multiparty election with parties R, L1 and L2.
- Do voters' unobserved ideological preferences affect Pr($Y_i =$L1) independently of their effect on Pr($Y_i =$L2)? Probably not.

Revisiting The IIA Assumption

- IIA (Trump, Cruz, and Sanders)
- Formally: MNL assumes ϵ_{ij} is i.i.d. $\epsilon_{ij} \perp\!\!\!\perp \epsilon_{ik}$ for $j \neq k$
- This implies that unobserved factors affecting Y_{ij}^* are unrelated to those affecting Y_{ik}^*
- When is this assumption plausible?
- Example: Multiparty election with parties R, L1 and L2.
- Do voters' unobserved ideological preferences affect Pr($Y_i =$L1) independently of their effect on Pr($Y_i =$L2)? Probably not.

Revisiting The IIA Assumption

- IIA (Trump, Cruz, and Sanders)
- Formally: MNL assumes ϵ_{ij} is i.i.d. $\epsilon_{ij} \perp\!\!\!\perp \epsilon_{ik}$ for $j \neq k$
- This implies that unobserved factors affecting Y_{ij}^* are unrelated to those affecting Y_{ik}^*
- When is this assumption plausible?
- Example: Multiparty election with parties R, L1 and L2.
- Do voters' unobserved ideological preferences affect Pr($Y_i =$L1) independently of their effect on Pr($Y_i =$L2)? Probably not.

Revisiting The IIA Assumption

- IIA (Trump, Cruz, and Sanders)
- Formally: MNL assumes ϵ_{ij} is i.i.d. $\epsilon_{ij} \perp\!\!\!\perp \epsilon_{ik}$ for $j \neq k$
- This implies that unobserved factors affecting Y_{ij}^* are unrelated to those affecting Y_{ik}^*
- When is this assumption plausible?
- Example: Multiparty election with parties R, L1 and L2.
- Do voters' unobserved ideological preferences affect Pr($Y_i =$L1) independently of their effect on Pr($Y_i =$L2)? Probably not.

Revisiting The IIA Assumption

- IIA (Trump, Cruz, and Sanders)
- Formally: MNL assumes ϵ_{ij} is i.i.d. $\epsilon_{ij} \perp\!\!\!\perp \epsilon_{ik}$ for $j \neq k$
- This implies that unobserved factors affecting Y_{ij}^* are unrelated to those affecting Y_{ik}^*
- When is this assumption plausible?
- Example: Multiparty election with parties R, L1 and L2.
- Do voters' unobserved ideological preferences affect Pr($Y_i =$L1) independently of their effect on Pr($Y_i =$L2)? Probably not.

Revisiting The IIA Assumption

- IIA (Trump, Cruz, and Sanders)
- Formally: MNL assumes ϵ_{ij} is i.i.d. $\epsilon_{ij} \perp\!\!\!\perp \epsilon_{ik}$ for $j \neq k$
- This implies that unobserved factors affecting Y_{ij}^* are unrelated to those affecting Y_{ik}^*
- When is this assumption plausible?
- Example: Multiparty election with parties R, L1 and L2.
- Do voters' unobserved ideological preferences affect Pr($Y_i =$L1) independently of their effect on Pr($Y_i =$L2)? Probably not.

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* = X_i^{\prime}\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \boldsymbol{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \ \cdots \ Y_{iJ}^*]^{\prime} \\ X_i = [X_{i1} \ \cdots \ X_{iJ}]^{\prime} \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* = X_i'\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \epsilon_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \ \cdots \ Y_{iJ}^*]' \\ X_i = [X_{i1} \ \cdots \ X_{iJ}]' \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* = X_i^{'}\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \boldsymbol{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \ \cdots \ Y_{iJ}^*]^{'} \\ X_i = [X_{i1} \ \cdots \ X_{iJ}]^{'} \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* \;=\; X_i^{'}\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \boldsymbol{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \;\cdots\; Y_{iJ}^*]^{'} \\ X_i = [X_{i1} \;\cdots\; X_{iJ}]^{'} \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* \;=\; X_i^{'}\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \boldsymbol{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \;\cdots\; Y_{iJ}^*]^{'} \\ X_i = [X_{i1} \;\cdots\; X_{iJ}]^{'} \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* \;=\; X_i^{'}\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \boldsymbol{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \;\cdots\; Y_{iJ}^*]^{'} \\ X_i = [X_{i1} \;\cdots\; X_{iJ}]^{'} \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* = X_i^{'}\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \boldsymbol{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \ \cdots \ Y_{iJ}^*]^{'} \\ X_i = [X_{i1} \ \cdots \ X_{iJ}]^{'} \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* \; = \; X_i^{'}\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \boldsymbol{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \; \cdots \; Y_{iJ}^*]^{'} \\ X_i = [X_{i1} \; \cdots \; X_{iJ}]^{'} \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Limitations of Multinomial Probit

- MNP has no closed-form expression for the likelihood:

$$\pi_{ij} = \int_{-\infty}^{-\ddot{X}_{1j}^\top \beta} \cdots \int_{-\infty}^{-\ddot{X}_{Jj}^\top \beta} \phi(\ddot{\epsilon}_{1j}, ..., \ddot{\epsilon}_{Jj}) d\ddot{\epsilon}_{1j} \cdots d\ddot{\epsilon}_{Jj} \text{ where } \left\{ \begin{array}{lcl} \ddot{X}_{kl} & = & X_{ik} - X_{il} \\ \ddot{\epsilon}_{kl} & = & \epsilon_{ik} - \epsilon_{il} \end{array} \right.$$

- This makes its estimation computationally costly when J large
- Must use numerical approximation (quadratures) or simulation methods (maximum simulated likelihood or MCMC)
- Moreover, # of parameters in Σ_J increases as J gets large, but data contain little information about Σ_J:

J	3	4	5	6	7
# of elements in Σ_J	6	10	15	21	28
# of parameters identified	2	5	9	14	20

- Consequently, MNP is only feasible when J is small
- MNP in R
- Alternative solutions: Nested logit

Limitations of Multinomial Probit

- MNP has no closed-form expression for the likelihood:

$$\pi_{ij} = \int_{-\infty}^{-\ddot{X}_{1j}^\top \beta} \cdots \int_{-\infty}^{-\ddot{X}_{Jj}^\top \beta} \phi(\ddot{\epsilon}_{1j}, ..., \ddot{\epsilon}_{Jj}) d\ddot{\epsilon}_{1j} \cdots d\ddot{\epsilon}_{Jj} \text{ where } \left\{ \begin{array}{lcl} \ddot{X}_{kl} & = & X_{ik} - X_{il} \\ \ddot{\epsilon}_{kl} & = & \epsilon_{ik} - \epsilon_{il} \end{array} \right.$$

- This makes its estimation computationally costly when J large
- Must use numerical approximation (quadratures) or simulation methods (maximum simulated likelihood or MCMC)
- Moreover, # of parameters in Σ_J increases as J gets large, but data contain little information about Σ_J:

J	3	4	5	6	7
# of elements in Σ_J	6	10	15	21	28
# of parameters identified	2	5	9	14	20

- Consequently, MNP is only feasible when J is small
- `MNP` in `R`
- Alternative solutions: Nested logit

Limitations of Multinomial Probit

- MNP has no closed-form expression for the likelihood:

$$\pi_{ij} = \int_{-\infty}^{-\ddot{X}_{1j}^{\top}\beta} \cdots \int_{-\infty}^{-\ddot{X}_{Jj}^{\top}\beta} \phi(\ddot{\epsilon}_{1j}, ..., \ddot{\epsilon}_{Jj}) d\ddot{\epsilon}_{1j} \cdots d\ddot{\epsilon}_{Jj} \text{ where } \begin{cases} \ddot{X}_{kl} & = & X_{ik} - X_{il} \\ \ddot{\epsilon}_{kl} & = & \epsilon_{ik} - \epsilon_{il} \end{cases}$$

- This makes its estimation computationally costly when J large
- Must use numerical approximation (quadratures) or simulation methods (maximum simulated likelihood or MCMC)
- Moreover, # of parameters in Σ_J increases as J gets large, but data contain little information about Σ_J:

J	3	4	5	6	7
# of elements in Σ_J	6	10	15	21	28
# of parameters identified	2	5	9	14	20

- Consequently, MNP is only feasible when J is small

- MNP in R
- Alternative solutions: Nested logit

Limitations of Multinomial Probit

- MNP has no closed-form expression for the likelihood:

$$\pi_{ij} = \int_{-\infty}^{-\ddot{X}_{1j}^{\top}\beta} \cdots \int_{-\infty}^{-\ddot{X}_{Jj}^{\top}\beta} \phi(\ddot{\epsilon}_{1j}, ..., \ddot{\epsilon}_{Jj}) d\ddot{\epsilon}_{1j} \cdots d\ddot{\epsilon}_{Jj} \text{ where } \left\{ \begin{array}{lcl} \ddot{X}_{kl} & = & X_{ik} - X_{il} \\ \ddot{\epsilon}_{kl} & = & \epsilon_{ik} - \epsilon_{il} \end{array} \right.$$

- This makes its estimation computationally costly when J large
- Must use numerical approximation (quadratures) or simulation methods (maximum simulated likelihood or MCMC)
- Moreover, # of parameters in Σ_J increases as J gets large, but data contain little information about Σ_J:

J	3	4	5	6	7
# of elements in Σ_J	6	10	15	21	28
# of parameters identified	2	5	9	14	20

- Consequently, MNP is only feasible when J is small
- `MNP` in `R`
- Alternative solutions: Nested logit

Even When You Choose Not To Decide, You Still Have Made a Choice

Lacy and Burden (1999)

Even When You Choose Not To Decide, You Still Have Made a Choice

Lacy and Burden (1999)
Multinomial Probit Model:

Even When You Choose Not To Decide, You Still Have Made a Choice

Lacy and Burden (1999)
Multinomial Probit Model:
Three Choices: Bush, Perot, and Clinton
Four Choices: Bush, Perot, Clinton, and Abstention

Even When You Choose Not To Decide, You Still Have Made a Choice

Lacy and Burden (1999)
Multinomial Probit Model:
Three Choices: Bush, Perot, and Clinton
Four Choices: Bush, Perot, Clinton, and Abstention
What happens if Perot doesn't run?

Even When You Choose Not To Decide, You Still Have Made a Choice

Lacy and Burden (1999)
Multinomial Probit Model:
Three Choices: Bush, Perot, and Clinton
Four Choices: Bush, Perot, Clinton, and Abstention
What happens if Perot doesn't run?

	Actual	3-choice	4-choice
Bush	32.0	45.7	38.4
Clinton	48.6	54.3	61.6
Abstention	20.9	-	23.7

Even When You Choose Not To Decide, You Still Have Made a Choice

Lacy and Burden (1999)
Multinomial Probit Model:
Three Choices: Bush, Perot, and Clinton
Four Choices: Bush, Perot, Clinton, and Abstention
What happens if Perot doesn't run?

	Actual	3-choice	4-choice
Bush	32.0	45.7	38.4
Clinton	48.6	54.3	61.6
Abstention	20.9	-	23.7

Perot stole from Clinton!

Event Count Models

Event Count Outcomes

- Outcome: number of times an event occurs

$$Y_i \quad \in \quad \{0, 1, 2, 3, \ldots, \}$$

- Examples:
  1) Number of militarized disputes a country is involved in
  2) Number of times a phrase is used
  3) Number of messages into a Congressional office
  4) Number of votes cast for a particular candidate
- Goal:
  - Model the rate at which events occur
  - Understand how an intervention (e.g. country becoming a democracy) affects rate
  - Predict number of future events

Deriving the Poisson Distribution

Suppose that events occur

1) Continuously (no simultaneous events)
2) Independently (occurrence of one event has no effect on occurrence of other event)
3) With constant probability

Poisson Distribution

Poisson Distribution

Definition

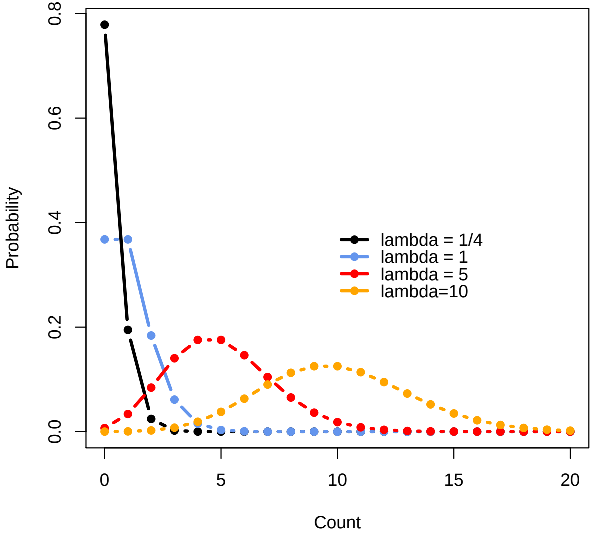
Suppose Y is a random variable that takes on values $Y \in \{0, 1, 2, \ldots, \}$ and that $P(Y = y) = p(y)$ is,

$$p(y) \quad = \quad e^{-\lambda} \frac{\lambda^y}{y!}$$

for $y \in \{0, 1, \ldots, \}$ and 0 otherwise. Then we will say that Y follows a Poisson distribution with rate parameter λ.

$$Y \quad \sim \quad \textit{Poisson}(\lambda)$$

Poisson Distribution



Poisson Distribution

Suppose $Y \sim$ Poisson(λ). Then:

$$\begin{aligned} \mathsf{E}[Y] &= \lambda \\ \mathsf{Var}(Y) &= \lambda \end{aligned}$$

If $Y \sim$ Poisson(λ) then the wait time between events, $W \sim$Exponential$(\frac{1}{\lambda})$

Poisson Distribution: Modeling Number of International Incidents

Suppose we observe N observations with

$$Y_i \sim_{\text{iid}} \text{Poisson}(\lambda)$$

Then:

$$\begin{aligned} L(\lambda|\boldsymbol{Y}) &= f(\boldsymbol{Y}|\boldsymbol{\lambda}) \\ &= \prod_{i=1}^{N} f(Y_i|\lambda) \\ &= \prod_{i=1}^{N} e^{-\lambda}\frac{\lambda^{Y_i}}{Y_i!} \end{aligned}$$

Poisson Distribution: Modeling Number of International Incidents

$$
\begin{aligned}
L(\boldsymbol{\lambda}|\boldsymbol{Y}) &= \prod_{i=1}^{N} e^{-\lambda}\frac{\lambda^{Y_i}}{Y_i!} \\
\log L(\boldsymbol{\lambda}|\boldsymbol{Y}) &= \sum_{i=1}^{N}(-\lambda + Y_i \log\lambda + \log Y_i!) \\
&= -N\lambda + \sum_{i=1}^{N} Y_i \log\lambda + c
\end{aligned}
$$

Poisson Distribution: Modeling Number of International Incidents

Differentiate, set equal to zero and solve:

$$
\begin{aligned}
\frac{\partial \ell(\boldsymbol{\lambda}|\boldsymbol{Y})}{\partial \lambda} &= -N + \sum_{i=1}^{N} \frac{Y_i}{\lambda} \\
0 &= -N + \sum_{i=1}^{N} \frac{Y_i}{\lambda^*} \\
\lambda^* &= \frac{\sum_{i=1}^{N} Y_i}{N}
\end{aligned}
$$

Poisson Distribution: Modeling Number of International Incidents

Uncertainty: inverse of negative expected hessian

$$\frac{\partial^2 \ell(\boldsymbol{\lambda}|\boldsymbol{Y})}{\partial \lambda \partial \lambda} = -\left(\frac{\sum_{i=1}^{N} E[Y_i]}{\lambda^2}\right)^{-1}$$
$$= \left(\frac{N\lambda}{\lambda^2}\right)^{-1}$$
$$= \left(\frac{N}{\lambda}\right)^{-1}$$
$$= \frac{\bar{Y}}{N} \text{ evaluated at MLE}$$

Asymptotically,

$$\lambda^* \longrightarrow^{D} \text{Normal}(\bar{Y}, \frac{\bar{Y}}{N})$$

Modeling the rate with covariates

Poisson Regression

$$
\begin{aligned}
Y_i &\sim \textsf{Poisson}(\lambda_i) \\
\lambda_i &= \exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta})
\end{aligned}
$$

This implies:

$$
\begin{aligned}
L(\boldsymbol{\beta}|\boldsymbol{X}_i, \boldsymbol{Y}) &= f(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{Y}) \\
&= \prod_{i=1}^{N} f(Y_i|\boldsymbol{X}_i, \boldsymbol{\beta}) \\
&= \prod_{i=1}^{N} \exp\left\{-\exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta})\right\} \frac{\exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta})^{Y_i}}{Y_i!}
\end{aligned}
$$

Poisson Regression

$$
\begin{aligned}
L(\boldsymbol{\beta}|\boldsymbol{X}_i, \boldsymbol{Y}) &= \prod_{i=1}^{N} \exp\left\{-\exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta})\right\} \frac{\exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta})^{Y_i}}{Y_i!} \\
\log L(\boldsymbol{\beta}|\boldsymbol{X}_i, \boldsymbol{Y}) &= \sum_{i=1}^{N} \left(-\exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta}) + Y_i \boldsymbol{X}_i \boldsymbol{\beta} - \log Y_i\right)
\end{aligned}
$$

Score: $s(\boldsymbol{\beta}|Y_i, \boldsymbol{X}_i) =$

$$
\left((Y_i - \exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta})), (Y_i - \exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta}))X_{i1}, \ldots, (Y_i - \exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta}))X_{iK}\right)
$$

Hessian:

$$
\frac{\partial^2 \ell(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{X})}{\partial \boldsymbol{\beta} \boldsymbol{\beta}} = -\exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta})\boldsymbol{X}_i \boldsymbol{X}_i^{'}
$$

Poisson Regression

$\boldsymbol{\beta}^* =$ numeric optimization

Poisson Regression

$\boldsymbol{\beta}^* =$ numeric optimization

$\boldsymbol{\beta}^* \longrightarrow^D$ Multivariate Normal$(\boldsymbol{\beta}, I_N(\boldsymbol{\beta}^*)^{-1})$

Poisson Regression

$\boldsymbol{\beta}^* =$ numeric optimization

$\boldsymbol{\beta}^* \longrightarrow^D$ Multivariate Normal$(\boldsymbol{\beta}, I_N(\boldsymbol{\beta}^*)^{-1})$

Quantities of Interest:

Poisson Regression

$\boldsymbol{\beta}^* =$ numeric optimization
$\boldsymbol{\beta}^* \longrightarrow^D$ Multivariate Normal$(\boldsymbol{\beta}, I_N(\boldsymbol{\beta}^*)^{-1})$
Quantities of Interest:

1) Expected Rate of Events, given characteristics:
   $\text{E}[Y|\tilde{\boldsymbol{X}}]$

Poisson Regression

$\boldsymbol{\beta}^* =$ numeric optimization

$\boldsymbol{\beta}^* \longrightarrow^D$ Multivariate Normal$(\boldsymbol{\beta}, I_N(\boldsymbol{\beta}^*)^{-1})$

Quantities of Interest:

1) Expected Rate of Events, given characteristics:
   $\mathsf{E}[Y|\tilde{\boldsymbol{X}}]$
2) Probability of Event at $\tilde{\boldsymbol{X}}$
   $\mathsf{Pr}(Y_i = y|\tilde{\boldsymbol{X}})$

Poisson Regression

$\boldsymbol{\beta}^* =$ numeric optimization

$\boldsymbol{\beta}^* \longrightarrow^D$ Multivariate Normal$(\boldsymbol{\beta}, I_N(\boldsymbol{\beta}^*)^{-1})$

Quantities of Interest:

1) Expected Rate of Events, given characteristics:
   $\mathsf{E}[Y|\tilde{\boldsymbol{X}}]$
2) Probability of Event at $\tilde{\boldsymbol{X}}$
   $\Pr(Y_i = y|\tilde{\boldsymbol{X}})$
3) Treatment effect of intervention T_i
   $\mathsf{E}\left[\mathsf{E}[Y|T_i = 1, \boldsymbol{X}_i] - \mathsf{E}[Y|T_i = 0, \boldsymbol{X}_i]\right]$

Poisson Regression

$\boldsymbol{\beta}^* =$ numeric optimization
$\boldsymbol{\beta}^* \longrightarrow^D$ Multivariate Normal$(\boldsymbol{\beta}, I_N(\boldsymbol{\beta}^*)^{-1})$
Quantities of Interest:

1) Expected Rate of Events, given characteristics:
   $\mathsf{E}[Y|\tilde{\boldsymbol{X}}]$
2) Probability of Event at $\tilde{\boldsymbol{X}}$
   $\mathsf{Pr}(Y_i = y|\tilde{\boldsymbol{X}})$
3) Treatment effect of intervention T_i
   $\mathsf{E}\left[\mathsf{E}[Y|T_i = 1, \boldsymbol{X}_i] - \mathsf{E}[Y|T_i = 0, \boldsymbol{X}_i]\right]$
4) Instantaneous change in X_{ik}:
   $\frac{\partial E[Y|\boldsymbol{X}]}{\partial X_{ik}} = \exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta})\beta_k$

Poisson Regression

$\boldsymbol{\beta}^* =$ numeric optimization

$\boldsymbol{\beta}^* \longrightarrow^D$ Multivariate Normal$(\boldsymbol{\beta}, I_N(\boldsymbol{\beta}^*)^{-1})$

Quantities of Interest:

1) Expected Rate of Events, given characteristics:
   $\mathsf{E}[Y|\tilde{\boldsymbol{X}}]$
2) Probability of Event at $\tilde{\boldsymbol{X}}$
   $\mathsf{Pr}(Y_i = y|\tilde{\boldsymbol{X}})$
3) Treatment effect of intervention T_i
   $\mathsf{E}\left[\mathsf{E}[Y|T_i = 1, \boldsymbol{X}_i] - \mathsf{E}[Y|T_i = 0, \boldsymbol{X}_i]\right]$
4) Instantaneous change in X_{ik}:
   $\frac{\partial E[Y|\boldsymbol{X}]}{\partial X_{ik}} = \exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta})\beta_k$

Uncertainty estimation:

Poisson Regression

$\boldsymbol{\beta}^* =$ numeric optimization

$\boldsymbol{\beta}^* \longrightarrow^D$ Multivariate Normal$(\boldsymbol{\beta}, I_N(\boldsymbol{\beta}^*)^{-1})$

Quantities of Interest:

1) Expected Rate of Events, given characteristics:
   $\mathsf{E}[Y|\tilde{\boldsymbol{X}}]$
2) Probability of Event at $\tilde{\boldsymbol{X}}$
   $\Pr(Y_i = y|\tilde{\boldsymbol{X}})$
3) Treatment effect of intervention T_i
   $\mathsf{E}\left[\mathsf{E}[Y|T_i = 1, \boldsymbol{X}_i] - \mathsf{E}[Y|T_i = 0, \boldsymbol{X}_i]\right]$
4) Instantaneous change in X_{ik}:
   $\frac{\partial E[Y|\boldsymbol{X}]}{\partial X_{ik}} = \exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta})\beta_k$

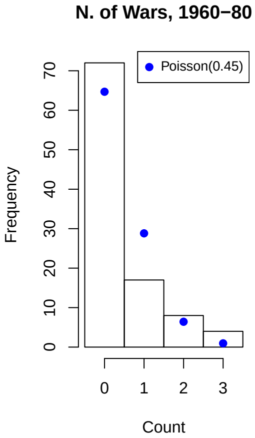
Uncertainty estimation:

1) Bootstrap
2) Delta Method
3) Simulation

Example: Democracy and War Involvement

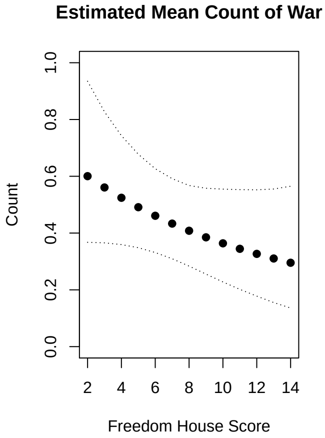
Benoit (1996):

- Y_i: # of involvement in international wars, 1960–80
- X_i: democracy (Freedom House score), population, military capacity, economic interdependence



Coefficients:

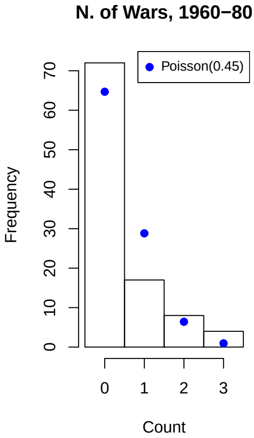
	Est.	s.e.		z	p
(Int.)	-3.97	1.62	*	-2.45	0.014
fh73	-0.06	0.04		-1.49	0.136
lpopln70	0.62	0.30	*	2.07	0.039
lmilwp70	1.29	0.45	**	2.85	0.004
ecintdep	-1.28	1.11		-1.16	0.247



Example: Democracy and War Involvement

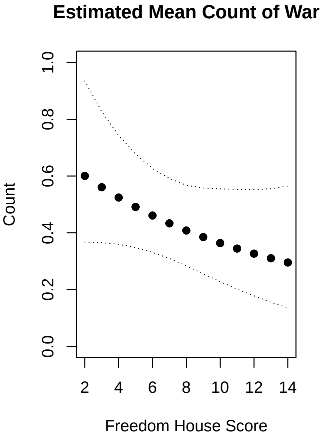
Benoit (1996):

- Y_i: # of involvement in international wars, 1960–80
- X_i: democracy (Freedom House score), population, military capacity, economic interdependence



Coefficients:

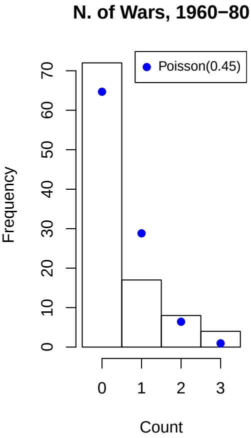
	Est.	s.e.		z	p
(Int.)	-3.97	1.62	*	-2.45	0.014
fh73	-0.06	0.04		-1.49	0.136
lpopln70	0.62	0.30	*	2.07	0.039
lmilwp70	1.29	0.45	**	2.85	0.004
ecintdep	-1.28	1.11		-1.16	0.247



Example: Democracy and War Involvement

Benoit (1996):

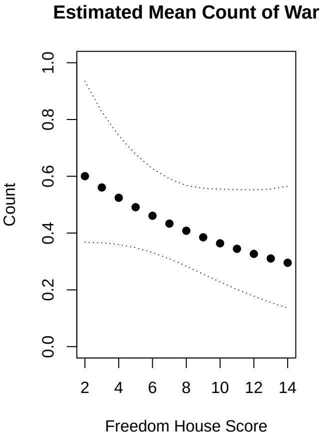
- Y_i: # of involvement in international wars, 1960–80
- X_i: democracy (Freedom House score), population, military capacity, economic interdependence



Coefficients:

	Est.	s.e.	
(Int.)	-3.97	1.62	*
fh73	-0.06	0.04	
lpopln70	0.62	0.30	*
lmilwp70	1.29	0.45	**
ecintdep	-1.28	1.11	

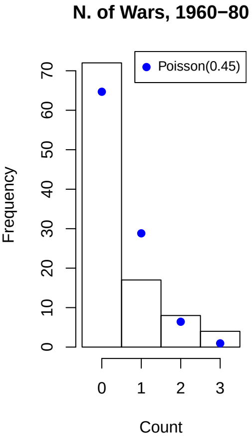
z	p
-2.45	0.014
-1.49	0.136
2.07	0.039
2.85	0.004
-1.16	0.247



Example: Democracy and War Involvement

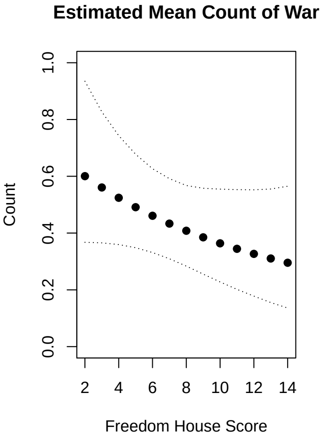
Benoit (1996):

- Y_i: # of involvement in international wars, 1960–80
- X_i: democracy (Freedom House score), population, military capacity, economic interdependence



Coefficients:

	Est.	s.e.		z	p
(Int.)	-3.97	1.62	*	-2.45	0.014
fh73	-0.06	0.04		-1.49	0.136
lpopln70	0.62	0.30	*	2.07	0.039
lmilwp70	1.29	0.45	**	2.85	0.004
ecintdep	-1.28	1.11		-1.16	0.247



Overdispersion in Poisson Model

- The Poisson model assumes $\mathsf{E}(Y_i \mid X_i) = \mathsf{Var}(Y_i \mid X_i)$
- But for many count data, $\mathsf{E}(Y_i \mid X_i) < \mathsf{V}(Y_i \mid X_i)$
- Potential sources of overdispersion:
  1. unobserved heterogeneity
  2. clustering
  3. contagion or diffusion
  4. (classical) measurement error
- Underdispersion could occur, but rare
- One solution to this is to modify the Poisson model by assuming:

$$\mathsf{E}(Y_i \mid X_i) = \lambda_i = \exp(X_i^\top \beta) \quad \text{and} \quad \mathsf{Var}(Y_i \mid X_i) = V_i = \sigma^2 \lambda_i$$

- This is called the overdispersed Poisson regression model
- When $\sigma^2 > 0$, this corresponds to the negative binomial regression model

Overdispersion in Poisson Model

- The Poisson model assumes $\mathsf{E}(Y_i \mid X_i) = \mathsf{Var}(Y_i \mid X_i)$
- But for many count data, $\mathsf{E}(Y_i \mid X_i) < \mathsf{V}(Y_i \mid X_i)$
- Potential sources of overdispersion:
  1. unobserved heterogeneity
  2. clustering
  3. contagion or diffusion
  4. (classical) measurement error
- Underdispersion could occur, but rare
- One solution to this is to modify the Poisson model by assuming:

$$\mathsf{E}(Y_i \mid X_i) = \lambda_i = \exp(X_i^\top \beta) \quad \text{and} \quad \mathsf{Var}(Y_i \mid X_i) = V_i = \sigma^2 \lambda_i$$

- This is called the overdispersed Poisson regression model
- When $\sigma^2 > 0$, this corresponds to the negative binomial regression model

Overdispersion in Poisson Model

- The Poisson model assumes $\mathsf{E}(Y_i \mid X_i) = \mathsf{Var}(Y_i \mid X_i)$
- But for many count data, $\mathsf{E}(Y_i \mid X_i) < \mathsf{V}(Y_i \mid X_i)$
- Potential sources of overdispersion:
  1. unobserved heterogeneity
  2. clustering
  3. contagion or diffusion
  4. (classical) measurement error
- Underdispersion could occur, but rare
- One solution to this is to modify the Poisson model by assuming:

$$\mathsf{E}(Y_i \mid X_i) = \lambda_i = \exp(X_i^\top \beta) \quad \text{and} \quad \mathsf{Var}(Y_i \mid X_i) = V_i = \sigma^2 \lambda_i$$

- This is called the overdispersed Poisson regression model
- When $\sigma^2 > 0$, this corresponds to the negative binomial regression model

Overdispersion in Poisson Model

- The Poisson model assumes $\mathsf{E}(Y_i \mid X_i) = \mathsf{Var}(Y_i \mid X_i)$
- But for many count data, $\mathsf{E}(Y_i \mid X_i) < \mathsf{V}(Y_i \mid X_i)$
- Potential sources of overdispersion:
  1. unobserved heterogeneity
  2. clustering
  3. contagion or diffusion
  4. (classical) measurement error
- Underdispersion could occur, but rare
- One solution to this is to modify the Poisson model by assuming:

$$\mathsf{E}(Y_i \mid X_i) = \lambda_i = \exp(X_i^\top \beta) \quad \text{and} \quad \mathsf{Var}(Y_i \mid X_i) = V_i = \sigma^2 \lambda_i$$

- This is called the overdispersed Poisson regression model
- When $\sigma^2 > 0$, this corresponds to the negative binomial regression model

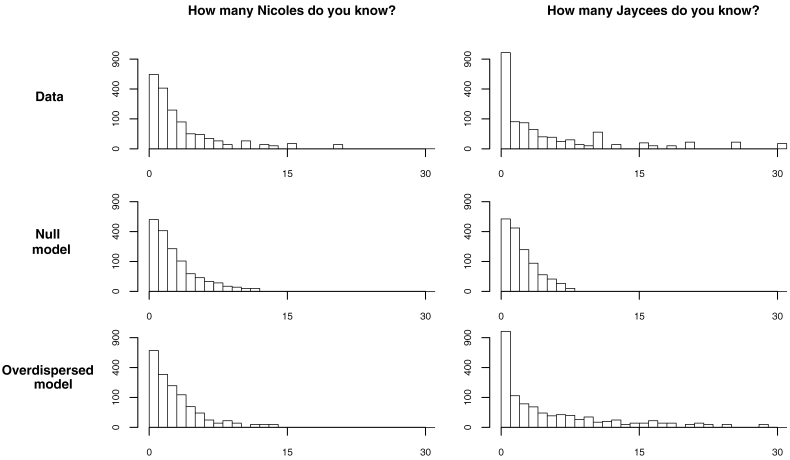
Overdispersion in Poisson Model

- The Poisson model assumes $\mathsf{E}(Y_i \mid X_i) = \mathsf{Var}(Y_i \mid X_i)$
- But for many count data, $\mathsf{E}(Y_i \mid X_i) < \mathsf{V}(Y_i \mid X_i)$
- Potential sources of overdispersion:
  1. unobserved heterogeneity
  2. clustering
  3. contagion or diffusion
  4. (classical) measurement error
- Underdispersion could occur, but rare
- One solution to this is to modify the Poisson model by assuming:

$$\mathsf{E}(Y_i \mid X_i) = \lambda_i = \exp(X_i^\top \beta) \quad \text{and} \quad \mathsf{Var}(Y_i \mid X_i) = V_i = \sigma^2 \lambda_i$$

- This is called the overdispersed Poisson regression model
- When $\sigma^2 > 0$, this corresponds to the negative binomial regression model

Example: Social Network Survey Data



(Zheng, et al., 2006 *JASA*)

Negative Binomial Distribution

Suppose $Y_i \in \{0, 1, 2, \ldots, \}$. If Y_i has pmf

$$p(y_i) = \frac{\Gamma\left(\frac{\lambda}{\sigma^2 - 1} + y_i\right)}{y_i! \Gamma\left(\frac{\lambda}{\sigma^2-1}\right)} \left(\frac{\sigma^2 - 1}{\sigma^2}\right)^{y_i} \left(\sigma^2\right)^{\frac{-\lambda}{\sigma^2-1}}$$

Then we will say

$$\begin{aligned} Y_i &\sim \text{NegBin}(\lambda, \sigma^2) \\ E[Y_i] &= \lambda \\ \text{Var}(Y_i) &= \lambda\sigma^2 \end{aligned}$$

Negative Binomial Regression

Suppose:

$$\begin{aligned} Y_i &\sim \text{Negative Binomial}(\lambda_i, \sigma^2) \\ \lambda_i &= \exp(\boldsymbol{X}_i^{\prime}\boldsymbol{\beta}) \end{aligned}$$

This implies a likelihood of:

$$\begin{aligned} L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) &= f(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\beta}) \\ &= \prod_{i=1}^{N} f(Y_i|\boldsymbol{X}_i, \boldsymbol{\beta}) \\ &= \prod_{i=1}^{N} \frac{\Gamma\left(\frac{\lambda_i}{\sigma^2-1} + y_i\right)}{y_i!\Gamma\left(\frac{\lambda_i}{\sigma^2-1}\right)} \left(\frac{\sigma^2-1}{\sigma^2}\right)^{y_i} \left(\sigma^2\right)^{\frac{-\lambda_i}{\sigma^2-1}} \end{aligned}$$

Optimize numerically. Usual theorems about asymptotic distributions apply.

Negative Binomial Regression

Negative Binomial Regression:

1) Variance is sometimes:

$$\mathsf{Var}(Y_i|\boldsymbol{X}_i) \quad = \quad \lambda_i(1 + \sigma^2 \lambda_i)$$

2) Run in `R` using

```
library(MASS)
out<- glm.nb(Y~X)
```

Clustering and Survival analysis