Advanced Quantitative Methods: Limited Dependent Variables I

Johan A. Elkink

University College Dublin

2 April 2013

Outline

Components

Two components of the model:

$\mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2)$	Stochastic
$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$	Systematic

Components

Two components of the model:

$\mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2)$	Stochastic
$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$	Systematic

Generalised version (not necessarily linear):

$\mathbf{y} \sim f(\boldsymbol{\mu}, \boldsymbol{\alpha})$	Stochastic
$\boldsymbol{\mu} = g(\mathbf{X}, \boldsymbol{\beta})$	Systematic

(King 1998, 8)

Components

$$\begin{array}{l|l} \mathbf{y} \sim f(\boldsymbol{\mu}, \boldsymbol{\alpha}) & \text{Stochastic} \\ \boldsymbol{\mu} = g(\mathbf{X}, \boldsymbol{\beta}) & \text{Systematic} \end{array}$$

Stochastic component: varies over repeated (hypothetical) observations on the same unit.

Systematic component: varies across units, but constant given **X**.

(King 1998, 8)

Uncertainty

$$\begin{array}{l|l} \mathbf{y} \sim f(\boldsymbol{\mu}, \boldsymbol{\alpha}) & \text{Stochastic} \\ \boldsymbol{\mu} = g(\mathbf{X}, \boldsymbol{\beta}) & \text{Systematic} \end{array}$$

Two types of uncertainty:

Estimation uncertainty: lack of knowledge about $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$; can be reduced by increasing n.

Uncertainty

$\mathbf{y} \sim f(\boldsymbol{\mu}, \boldsymbol{\alpha})$	Stochastic
$\boldsymbol{\mu} = g(\mathbf{X}, \boldsymbol{\beta})$	Systematic

Two types of uncertainty:

Estimation uncertainty: lack of knowledge about $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$; can be reduced by increasing *n*.
Fundamental uncertainty: represented by stochastic component and exists independent of researcher.

Levels of measurement

	Discreet	Continuous
Nominal Ordinal	party choice	- - -
Interval Ratio		

Categories in no particular order

(examples in cells)

Levels of measurement

	Discreet	Continuous
Nominal	party choice	-
Ordinal	education level	-
		-
Interval		
Ratio		

Categories in a specific order

(examples in cells)

Levels of measurement

	Discreet	Continuous
Nominal	party choice	-
Ordinal	education level	-
		-
Interval	how likely to vote ...	temperature
Ratio		

All values possible

(examples in cells)

Levels of measurement

	Discreet	Continuous
Nominal	party choice	-
Ordinal	education level	-
		-
Interval	how likely to vote ...	temperature
Ratio	deaths in war	ideological distance

All values possible, with a meaningful zero point

(examples in cells)

Levels of measurement

	Discreet	Continuous
Nominal	party choice	-
Ordinal	education level	-
Binary	turnout	-
Interval	how likely to vote ...	temperature
Ratio	deaths in war	ideological distance

Two categories, coded as 0 and 1

(examples in cells)

Limited dependent variables

When a dependent variable is not continuous, or is truncated for some reason, a linear model would lead to implausible predictions.

Limited dependent variables

When a dependent variable is not continuous, or is truncated for some reason, a linear model would lead to implausible predictions.

E.g. regressing whether someone voted on a set of independent variables will give somewhat reasonable estimates (see previous homework), but using these estimates to calculate predictions leads to **meaningless predictions**.

Limited dependent variables

When a dependent variable is not continuous, or is truncated for some reason, a linear model would lead to implausible predictions.

E.g. regressing whether someone voted on a set of independent variables will give somewhat reasonable estimates (see previous homework), but using these estimates to calculate predictions leads to **meaningless predictions**.

Furthermore, estimating limited dependent variable data with a linear model implies serious **heteroscedasticity**.

Outline

1. Introduction
2. Binary
3. Ordinal
4. Nominal
5. Count

Binary models

Binary models have a dependent variable consisting of two categories.

Binary models

Binary models have a dependent variable consisting of two categories.

For example,

- Vote on a particular law
- Turning out in an election
- Approval in a referendum
- Bankrupt or not

Logistic regression

Stochastic:

$$y_i = Bernoulli(\pi_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

Logistic regression

Stochastic:

$$y_i = Bernoulli(\pi_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}$$

Variance of stochastic component:

$$V(y_i) = \pi_i(1 - \pi_i)$$

Logistic regression

Stochastic:

$$y_i = Bernoulli(\pi_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

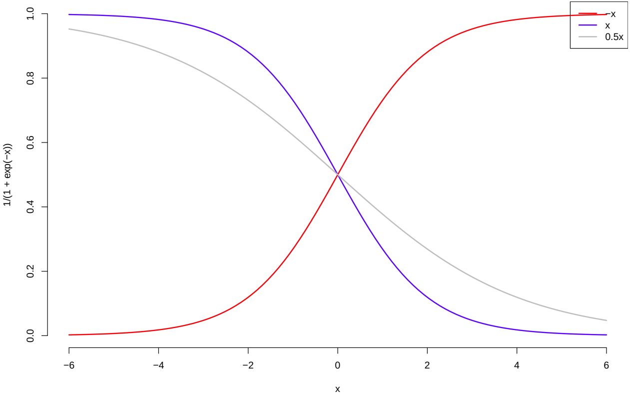
Variance of stochastic component:

$$V(y_i) = \pi_i(1-\pi_i)$$

Systematic:

$$\pi_i = g(\mathbf{x}_i\boldsymbol{\beta}) = \frac{1}{1+e^{-\mathbf{x}_i\boldsymbol{\beta}}}$$

Logistic distribution



Logistic regression

Resulting predictions:

- are bounded between 0 and 1

Logistic regression

Resulting predictions:

- are bounded between 0 and 1
- show limited effect at extremes

Logistic regression

Resulting predictions:

- are bounded between 0 and 1
- show limited effect at extremes
- are a smooth, monotone translation from linear prediction $X_i\beta$

Loglikelihood function

Systematic part:

$$\pi_i = \frac{1}{1 + e^{-\mathbf{x}_i \boldsymbol{\beta}}}$$

Stochastic part:

$$P(y_i = 1) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}$$

$$f(\mathbf{y}|\pi_\mathbf{i}) = \prod_{i=1}^{n} \pi_i^{y_i}(1 - \pi_i)^{1-y_i}$$

$$\ell(\mathbf{y}) = \log \prod_{i=1}^{n} \pi_i^{y_i}(1 - \pi_i)^{1-y_i}$$

$$= \sum_{i=1}^{n} y_i \log \pi_i + \sum_{i=1}^{n} (1 - y_i) \log(1 - \pi_i)$$

Logistic regression in R

```
model <- glm(y ~ x1 + x2 + x3,
     data=data,
     family=binomial(link="logit"))
summary(model)
```

Exercise

Using the `asiabaro.dta` data set, estimate a logistic regression explaining abstention in elections by trust in the government, satisfaction with democracy, gender, urbanisation, and preference for democracy.

Graphical interpretation

Useful for plotting relationship between one x and π, holding the other values of $\mathbf{X}$ constant

Graphical interpretation

Useful for plotting relationship between one x and π, holding the other values of $\mathbf{X}$ constant (e.g. at the mean, median, etc.).

Graphical interpretation

Useful for plotting relationship between one x and π, holding the other values of $\mathbf{X}$ constant (e.g. at the mean, median, etc.).

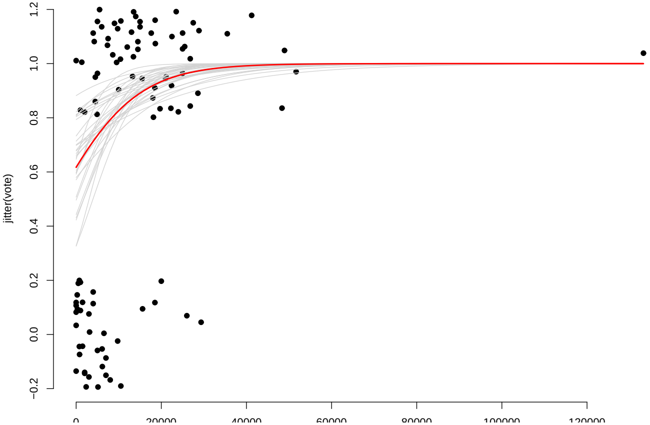
Because the **link function** $g(\mathbf{X}\beta)$ is not linear in this case (but instead $g(\mathbf{X}\beta) = \frac{1}{1+e^{-\mathbf{X}\beta}}$), the effect of $\mathbf{X}$ on $\mathbf{y}$ depends on all $\mathbf{X}$.

Graphical: example in R

```
m <- glm(y ~ x1 + x2 + x3,
   family=binomial(link="logit"))

curve(1/(1+exp(-(coef(m)[1]
           + coef(m)[2]*x
           + coef(m)[3]*mean(x2)
           + coef(m([4]*mean(x3)))),
       from=0, to=1)
```

Graphical: example



Fitted values

A second useful way of interpreting **logit** regression coefficients is by describing typical cases or interesting examples.

Fitted values

A second useful way of interpreting **logit** regression coefficients is by describing typical cases or interesting examples.

Party	Amount	$P(Vote = 1)$	95% C.I.
Republican	$10000	.83	[.69,.95]
Democrat	$10000	.44	[.27,.65]
Republican	$20000	.93	[.84,.99]
Democrat	$20000	.69	[.44,.95]

First differences

Basic idea is to calculate and present:

$$\Delta_{\hat{\pi}} = g(\mathbf{X}\beta) - g(\mathbf{X}^*\beta),$$

whereby $\mathbf{X}^*$ differs only in one variable from $\mathbf{X}$.

First differences

Basic idea is to calculate and present:

$$\Delta_{\hat{\pi}} = g(\mathbf{X}\beta) - g(\mathbf{X}^*\beta),$$

whereby $\mathbf{X}^*$ differs only in one variable from $\mathbf{X}$.

Variable	Values	Diff	95% C.I.
Party	Rep, Dem	-.34	[-.53,-.11]
Amount	\$10000, \$20000	.11	[.03,.20]

Derivatives

$$\frac{\delta \hat{\pi}}{\mathbf{X}_j} = \beta_j \hat{\pi}(1 - \hat{\pi})$$

Derivatives

$$\frac{\delta\hat{\pi}}{\mathbf{X}_j} = \beta_j\hat{\pi}(1-\hat{\pi})$$

Hence a quick method to interpret logit coefficients is to divide them by 4 to get the slope at $\hat{\pi} = .5$.

Presentation

Bottom line: it is much better to present *interpretable and understandable* inferences, with an indication of the level of *uncertainty*, than to present simply estimated coefficients.

Presentation

E.g. "An increase in automobile support for a Republican senator from \$10000 to \$20000 in total increases his or her probability to vote for the Corporate Average Fuel Economy standard bill by 11%, give or take 7%, all else equal."

Predictions check

One method of checking the predictions is to

- Group the cases in bins based on their predicted probabilities

Predictions check

One method of checking the predictions is to

- Group the cases in bins based on their predicted probabilities, e.g. 0-10%, 10-20%, etc.

Predictions check

One method of checking the predictions is to

- Group the cases in bins based on their predicted probabilities, e.g. 0-10%, 10-20%, etc.
- Plot the bins against the number of ones found in this bin.

Predictions check

One method of checking the predictions is to

- Group the cases in bins based on their predicted probabilities, e.g. 0-10%, 10-20%, etc.
- Plot the bins against the number of ones found in this bin.
- The closer this is to the 45 degree line, the more accurate the predictions of the probabilities.

Predictions check

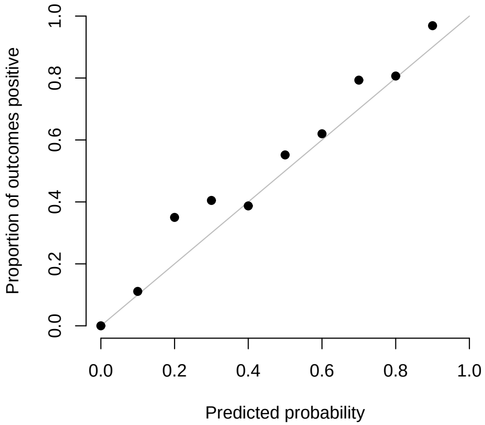
One method of checking the predictions is to

- Group the cases in bins based on their predicted probabilities, e.g. 0-10%, 10-20%, etc.
- Plot the bins against the number of ones found in this bin.
- The closer this is to the 45 degree line, the more accurate the predictions of the probabilities.

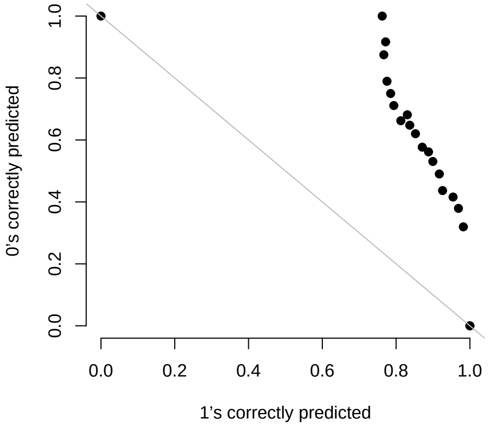
In R, you can use the custom function **plot.predicted()**.

```
m <- glm(..., family = binomial(link = "logit"))
phat <- predict(m, type = "response")
pgroups <- cut(phat, seq(0, 1, .1))
hist(phat, xlim = c(0,1))
barplot(prop.table(table(pgroups)), ylim = c(0,1))
```

Predictions check



Predictions check



Exercise

Using the logistic previous logistic regression:

1. What is the slope of the regression line with respect to preference for democracy at $\hat{\pi} = .5$?
2. Calculate predicted differences for urban vs rural respondent.
3. Calculate first differences for urban vs rural and female vs male.
4. Plot predicted probabilities as a function of preference for democracy.

Threshold models

Imagine we have a latent, unobserved variable:

$$\mathbf{y}^* \sim f(\boldsymbol{\mu})$$
$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$$

Threshold models

Imagine we have a latent, unobserved variable:

$$\mathbf{y}^* \sim f(\boldsymbol{\mu})$$
$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$$

With the following observation mechanism:

$$y_i = \begin{cases} 1 & \text{if} \quad y_i^* > 0 \\ 0 & \text{if} \quad y_i^* \leq 0. \end{cases}$$

Threshold models

If $f(\mu_i)$ is the **standardised logistic distribution**, we replicate the logit model.

Threshold models

If $f(\mu_i)$ is the **standardised logistic distribution**, we replicate the logit model.

$$f(\mathbf{x}_i\beta) = \frac{e^{y_i^* - \mu_i}}{(1 + e^{y_i^* - \mu_i})^2}$$

Threshold models

Alternatively, if $f(\mu_i)$ is the **cumulative standard normal distribution** ($\sigma^2 = 1$), we have the **probit** model.

Threshold models

Alternatively, if $f(\mu_i)$ is the **cumulative standard normal distribution** ($\sigma^2 = 1$), we have the **probit** model.

Hence $\boldsymbol{\beta}$ can now be interpreted as the effect of an increase by 1 in $\mathbf{x}$ on $\mathbf{y}^*$, whereby the unit of $\mathbf{y}^*$ is one standard deviation.

Threshold models

Alternatively, if $f(\mu_i)$ is the **cumulative standard normal distribution** ($\sigma^2 = 1$), we have the **probit** model.

Hence $\boldsymbol{\beta}$ can now be interpreted as the effect of an increase by 1 in $\mathbf{x}$ on $\mathbf{y}^*$, whereby the unit of $\mathbf{y}^*$ is one standard deviation.

$$P(y_i = 1|\boldsymbol{\beta}, \mathbf{x}_i) = \Phi(\mathbf{x}_i\boldsymbol{\beta}),$$

whereby $\Phi(x)$ is the cumulative standard normal distribution (i.e. the surface between $-\infty$ and x under a normal distribution with $\mu = 0$ and $\sigma^2 = 1$).

Threshold models

Alternatively, if $f(\mu_i)$ is the **cumulative standard normal distribution** ($\sigma^2 = 1$), we have the **probit** model.

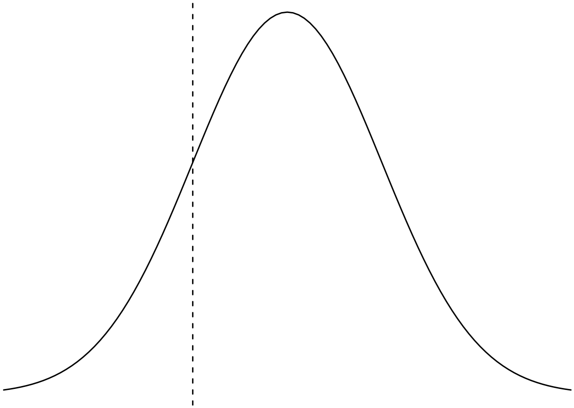
Hence $\boldsymbol{\beta}$ can now be interpreted as the effect of an increase by 1 in $\mathbf{x}$ on $\mathbf{y}^*$, whereby the unit of $\mathbf{y}^*$ is one standard deviation.

$$P(y_i = 1|\boldsymbol{\beta}, \mathbf{x}_i) = \Phi(\mathbf{x}_i\boldsymbol{\beta}),$$

whereby $\Phi(x)$ is the cumulative standard normal distribution (i.e. the surface between $-\infty$ and x under a normal distribution with $\mu = 0$ and $\sigma^2 = 1$).

```
pnorm(Xb);  predict(m, type="response")
```

Threshold models



Loglikelihood function

Systematic part:

$$\pi_i = \Phi(\mathbf{x}_i \boldsymbol{\beta}),$$

where $\Phi(x)$ is the **cumulative standard normal distribution**.

Stochastic part:

$$P(y_i = 1) = \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

$$f(\mathbf{y}|\pi_{\mathbf{i}}) = \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

$$\ell(\mathbf{y}) = \sum_{i=1}^{n} y_i \log \pi_i + \sum_{i=1}^{n}(1-y_i)\log(1-\pi_i)$$

Exercise

Repeat the previous estimation using probit instead of logit.

1. Plot predicted probabilities with respect to preference for democracy.
2. Evaluate the predictive performance of this model.
3. Add `suitdemoc` as a dependent variable to the model. Does the predictive performance improve?

Outline

Ordered data

Models where the dependent variable is categorical, and the categories are in a particular order.

Ordered data

Models where the dependent variable is categorical, and the categories are in a particular order.

We can thus take a similar latent variable approach.

Ordered probit

$$y_i^* \sim N(\mu_i, 1)$$
$$\mu_i = \mathbf{x}_i \boldsymbol{\beta}$$

Ordered probit

$$y_i^* \sim N(\mu_i, 1)$$
$$\mu_i = \mathbf{x}_i \boldsymbol{\beta}$$

With the following observation mechanism:

$$y_i = j \quad \text{if} \quad \tau_{j-1} \leq y_i^* \leq \tau_j$$

Ordered probit

$$y_i^* \sim N(\mu_i, 1)$$
$$\mu_i = \mathbf{x}_i \boldsymbol{\beta}$$

With the following observation mechanism:

$$y_i = j \quad \text{if} \quad \tau_{j-1} \leq y_i^* \leq \tau_j$$

Or, alternatively, when we use dummy variables for each category:

$$y_{ij} = \begin{cases} 1 & \text{if} \quad \tau_{j-1} \leq y_i^* \leq \tau_j \\ 0 & \text{otherwise.} \end{cases}$$

Ordered probit

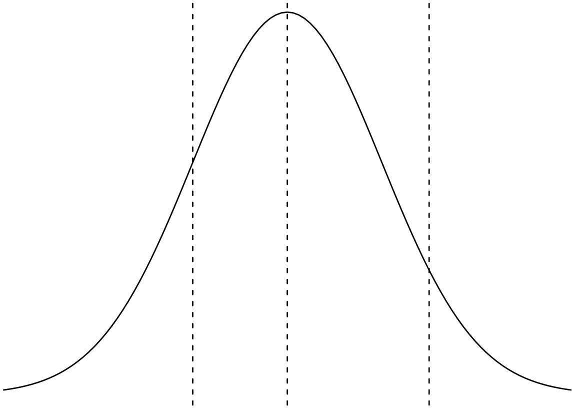
$$y_{ij} = \begin{cases} 1 & \text{if} \quad \tau_{j-1} \leq y_i^* \leq \tau_j \\ 0 & \text{otherwise.} \end{cases}$$

Ordered probit

$$y_{ij} = \begin{cases} 1 & \text{if} \quad \tau_{j-1} \leq y_i^* \leq \tau_j \\ 0 & \text{otherwise.} \end{cases}$$

$$P(y_i = j|\beta, \mathbf{x}_i) = \Phi(\tau_j - \mathbf{x}_i\boldsymbol{\beta}) - \Phi(\tau_{j-1} - \mathbf{x}_i\boldsymbol{\beta})$$

Ordered probit



Ordered probit

To estimate in R:

```
library(MASS)
m <- polr(y ~ x1 + x2, method="probit")
```

Ordered probit

To estimate in R:

```
library(MASS)
m <- polr(y ~ x1 + x2, method="probit")
```

To get predicted probabilities:

```
pnorm(m$zeta[2] - xb) - pnorm(m$zeta[1] - xb)
predict(m, type="probs")
```

Loglikelihood function

Assuming y_i can take the values $0, 1, 2, ..., p$.

$$P(y_i = 0) = \Phi(\tau_1 - \mathbf{x}_i\boldsymbol{\beta})$$
$$P(y_i = p) = \Phi(\mathbf{x}_i\boldsymbol{\beta} - \tau_p)$$
$$P(y_i = j | 0 < y_i < p) = \Phi(\tau_{j+1} - \mathbf{x}_i\boldsymbol{\beta}) - \Phi(\tau_j - \mathbf{x}_i\boldsymbol{\beta})$$

Loglikelihood function

Assuming y_i can take the values $0, 1, 2, ..., p$.

$$P(y_i = 0) = \Phi(\tau_1 - \mathbf{x}_i\boldsymbol{\beta})$$
$$P(y_i = p) = \Phi(\mathbf{x}_i\boldsymbol{\beta} - \tau_p)$$
$$P(y_i = j | 0 < y_i < p) = \Phi(\tau_{j+1} - \mathbf{x}_i\boldsymbol{\beta}) - \Phi(\tau_j - \mathbf{x}_i\boldsymbol{\beta})$$

$$\ell(\mathbf{y}) = \sum_{y_i=0} \log(\Phi(\tau_1 - \mathbf{x}_i\boldsymbol{\beta})) + \sum_{y_i=p} \log(\Phi(\mathbf{x}_i\boldsymbol{\beta} - \tau_p)) + \sum_{0<y_i<p} \log(\Phi(\tau_{j+1} - \mathbf{x}_i\boldsymbol{\beta}) - \Phi(\tau_j - \mathbf{x}_i\boldsymbol{\beta}))$$

Exercise

Using the `asiabaro.dta` data set, estimate a model explaining interest in politics by gender, education, and reliance on fate.

1. Perform *t*-tests for each of the independent variables.
   Use: `se <- sqrt(diag(vcov(m)))`
2. Check the standard errors of the τ values – are the categories clearly separated?
3. Calculate the predicted probability for males and females of the respondent being very interested in politics.

Outline

Multinomial model

Here the dependent variable consists of multiple categories, without particular order.

Multinomial model

Here the dependent variable consists of multiple categories, without particular order.

A latent variable approach will thus not work.

Multinomial model

Here the dependent variable consists of multiple categories, without particular order.

A latent variable approach will thus not work.

The probabilities for a particular case to be in any of those categories still has to be one.

Multinomial logit

R function: *multinom*, in package *nnet*.

Multinomial logit

R function: *multinom*, in package *nnet*.

```
m <- multinom(y ~ x1 + x2)
summary(m)
```

Multinomial logit

R function: *multinom*, in package *nnet*.

```
m <- multinom(y ~ x1 + x2)
summary(m)
```

```
predict(m, type="probs")
```

Multinomial logit

$$P(y_i = j|\beta, \mathbf{x}_i) = \begin{cases} \frac{1}{\sum_{k=1}^{p} e^{\mathbf{x}_i \boldsymbol{\beta}_k}} & \text{if} \quad j = 1 \\ \frac{e^{\mathbf{x}_i \boldsymbol{\beta}_j}}{\sum_{k=1}^{p} e^{\mathbf{x}_i \boldsymbol{\beta}_k}} & \text{if} \quad j > 1 \end{cases}$$

Multinomial logit

$$P(y_i = j|\beta, \mathbf{x}_i) = \begin{cases} \frac{1}{\sum_{k=1}^{p} e^{\mathbf{x}_i \boldsymbol{\beta}_k}} & \text{if} \quad j = 1 \\ \frac{e^{\mathbf{x}_i \boldsymbol{\beta}_j}}{\sum_{k=1}^{p} e^{\mathbf{x}_i \boldsymbol{\beta}_k}} & \text{if} \quad j > 1 \end{cases}$$

To get predicted probabilities for a new dataset ($\mathbf{X}^*$) - e.g. one that is constant on all variables but one:

Multinomial logit

$$P(y_i = j|\beta, \mathbf{x}_i) = \begin{cases} \frac{1}{\sum_{k=1}^{p} e^{\mathbf{x}_i \boldsymbol{\beta}_k}} & \text{if} \quad j = 1 \\ \frac{e^{\mathbf{x}_i \boldsymbol{\beta}_j}}{\sum_{k=1}^{p} e^{\mathbf{x}_i \boldsymbol{\beta}_k}} & \text{if} \quad j > 1 \end{cases}$$

To get predicted probabilities for a new dataset ($\mathbf{X}^*$) - e.g. one that is constant on all variables but one:

```
m <- multinom(y ~ x1 + x2 + x3 + x4)

Xb <- X %*% t(coef(m))
denominator <- 1 - rowSums(exp(Xb))
probs <- exp(Xb) / denominator

baseline <- 1 - rowSums(p)
p <- cbind(baseline, p)
```

Loglikelihood function

$$P(y_i = j|\beta, \mathbf{x}_i) = \begin{cases} \frac{1}{\sum_{k=1}^{p} e^{\mathbf{x}_i \boldsymbol{\beta}_k}} & \text{if} \quad j = 1 \\ \frac{e^{\mathbf{x}_i \boldsymbol{\beta}_j}}{\sum_{k=1}^{p} e^{\mathbf{x}_i \boldsymbol{\beta}_k}} & \text{if} \quad j > 1 \end{cases}$$

Loglikelihood function

$$P(y_i = j|\beta, \mathbf{x}_i) = \begin{cases} \frac{1}{\sum_{k=1}^{p} e^{\mathbf{x}_i \boldsymbol{\beta}_k}} & \text{if} \quad j = 1 \\ \frac{e^{\mathbf{x}_i \boldsymbol{\beta}_j}}{\sum_{k=1}^{p} e^{\mathbf{x}_i \boldsymbol{\beta}_k}} & \text{if} \quad j > 1 \end{cases}$$

$$\ell(\mathbf{y}) = \sum_{i=1}^{n} \sum_{j=0}^{p} \left[I(y_i = j)\mathbf{x}_i \boldsymbol{\beta}_j - \log \left(\sum_{j=0}^{p} e^{\mathbf{x}_i \boldsymbol{\beta}_j} \right) \right]$$

Exercise

Using data set `asiabaro.dta`, selecting only data from Taiwan, explain party choice by urbanisation, gender, age, and trust in the police. Estimate a multinomial model and interpret the results.

Outline

Count models

Here the dependent variable is a count (of events).

For example,

- The number of conflicts in a particular period
- The number of coups d'état in the 1980s
- The number of visits to a psychologist for a respondent

Count models

Here the dependent variable is a count (of events).

For example,

- The number of conflicts in a particular period
- The number of coups d'état in the 1980s
- The number of visits to a psychologist for a respondent

Note:

- Truncated at zero (no negative outcome possible)

Count models

Here the dependent variable is a count (of events).

For example,

- The number of conflicts in a particular period
- The number of coups d'état in the 1980s
- The number of visits to a psychologist for a respondent

Note:

- Truncated at zero (no negative outcome possible)
- No upper limit

Count models

Here the dependent variable is a count (of events).

For example,

- The number of conflicts in a particular period
- The number of coups d'état in the 1980s
- The number of visits to a psychologist for a respondent

Note:

- Truncated at zero (no negative outcome possible)
- No upper limit
- Limited by time, place, or both

Poisson regression

Stochastic: $y_i \sim Poisson(\lambda_i)$

Poisson regression

Stochastic: $y_i \sim Poisson(\lambda_i)$

Systematic: $\lambda_i = e^{\mathbf{x}_i \boldsymbol{\beta}}$

Poisson regression

Stochastic: $y_i \sim Poisson(\lambda_i)$

Systematic: $\lambda_i = e^{\mathbf{x}_i \boldsymbol{\beta}}$

Poisson distribution:

$$f(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Poisson regression

Stochastic: $y_i \sim Poisson(\lambda_i)$

Systematic: $\lambda_i = e^{\mathbf{x}_i \boldsymbol{\beta}}$

Poisson distribution:

$$f(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Note that there is no parameter for the variance. This leads to regular underestimation of the variance (**overdispersion**, most common) or overestimation of the variance (**underdispersion**).

Poisson regression

Stochastic: $y_i \sim Poisson(\lambda_i)$

Systematic: $\lambda_i = e^{\mathbf{x}_i \boldsymbol{\beta}}$

Poisson distribution:

$$f(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Note that there is no parameter for the variance. This leads to regular underestimation of the variance (**overdispersion**, most common) or overestimation of the variance (**underdispersion**).

```
glm(y ~ x1 + x2 + x3, family=poisson)
```

Poisson regression

The exponential of the coefficients can be interpreted in a multiplicative sense.

E.g. if the coefficient of $\mathbf{x}_1$ is 0.012, then $e^{0.012} = 1.012$ implies that an increase of $\mathbf{x}_1$ by 1 increases $\mathbf{y}$ by 1.2%.

Estimating overdispersion

The estimated overdispersion of a Poisson model is

$$\frac{1}{n-k}\sum_{i}^{n}(\frac{y_i-\hat{y}_i}{sd(\hat{y}_i)})^2 = \frac{1}{n-k}\sum_{i}^{n}(\frac{y_i-e^{X_i\beta}}{\sqrt{e^{X_i\beta}}})^2,$$

which has a χ^2_{n-k} distribution.

Negative binomial

The negative binomial model is useful for **overdispersed** data:

Stochastic: $y_i \sim NegBin(\phi, \sigma^2)$

Negative binomial

The negative binomial model is useful for **overdispersed** data:

Stochastic: $y_i \sim NegBin(\phi, \sigma^2)$

Systematic: $\phi = e^{\mathbf{x}_i \boldsymbol{\beta}} = E(y_i)$

Negative binomial

The negative binomial model is useful for **overdispersed** data:

Stochastic: $y_i \sim NegBin(\phi, \sigma^2)$

Systematic: $\phi = e^{\mathbf{x}_i \boldsymbol{\beta}} = E(y_i)$

Variance: $V(\mathbf{y}|\mathbf{X}) = \phi\sigma^2$

Negative binomial

The negative binomial model is useful for **overdispersed** data:

Stochastic: $y_i \sim NegBin(\phi, \sigma^2)$

Systematic: $\phi = e^{\mathbf{x}_i \boldsymbol{\beta}} = E(y_i)$

Variance: $V(\mathbf{y}|\mathbf{X}) = \phi\sigma^2$

When σ^2 approaches 1, the negative binomial approaches the Poisson distribution.

Negative binomial

The negative binomial model is useful for **overdispersed** data:

Stochastic: $y_i \sim NegBin(\phi, \sigma^2)$

Systematic: $\phi = e^{\mathbf{x}_i \boldsymbol{\beta}} = E(y_i)$

Variance: $V(\mathbf{y}|\mathbf{X}) = \phi\sigma^2$

When σ^2 approaches 1, the negative binomial approaches the Poisson distribution.

```
library(MASS)
glm.nb(y ~ x1 + x2 + x3)
```