Computational Methods in Economics

Maximum Likelihood

January 28, 2020

Overview: Linear Models

- Study the relationship between an outcome variable y and a set of regressors x.
  - Conditional Prediction.
  - Causal inference.
  - Example: propensity to consume.
- Loss function approach

$$L(e) = L(y - \hat{y})$$

where $\hat{y} = E(y \mid x)$ is a predictor of y, and the error $e = y - \hat{y}$

Squared Loss Function

- Squared error loss: $L(e) = e^2$
- Optimization problem

$$\min_{\beta} \sum_{i}^{N} (y_i - f(x_i, \beta))^2$$

Linear Prediction

- $E[y \mid x] = x'\beta$
- OLS

$$y = x\beta + e$$

- Derivation

$$\begin{aligned} L(\beta) &= (y - x\beta)'(y - x\beta) \\ &= y'y - 2y'x\beta + \beta'X'X\beta \end{aligned}$$

Then

$$\frac{\partial L(\beta)}{\partial \beta} = -2x'y + 2x'x\beta = 0$$

- Formula

$$\hat{\beta} = (x'x)^{-1}x'y$$

Properties

see 4.4.4 and 4.4.5.

Properties of an estimator

- Unbiasedness: $E(\hat{\theta}) = \theta$.
- Consistency: $plim\hat{\theta}_n = \theta$.
- Efficiency: Reach Cramer-Rao lower bound asymptotically.

Codes

```
////R
fitR = lm(Y~X)

////Matlab
fitM = fitlm(X,Y,'linear')

////Stata
fitM = reg Y X
```

Codes

```
////R

####  define X matrix and y vector
X = as.matrix(cbind(1,X))
y = as.matrix(Y)

####  estimate the coeficients beta
####  beta = ((X'X)^(-1))X'y
beta  = solve(t(X)%*%X)%*%t(X)%*%y
```

Maximum Likelihood

Applications

- Binary Outcomes
- Multinomial Choices
- Count Data

Introduction to MLE

Consider a parametric model in which the joint distribution of $Y = (Y_1, \ldots, Y_n)$ has a density $\ell(y, \theta)$ with respect to a measure μ. Then consider $P_\theta = \ell(y, \theta)\mu$ where $\theta \in \Theta \in \mathbb{R}^p$. Once $y = (y_1, \ldots, y_n)$ is observed, the maximum likelihood method consists of estimating the parameter θ a value $\hat{\theta}(y)$ that maximizes the likelihood function $\theta \to \ell(y, \theta)$. Formally, a maximum likelihood estimator of θ is a solution to the maximization problem

$$\max_\theta \ell(Y; \theta)$$

or

$$\max_\theta \log(\ell(Y; \theta))$$

Feasible examples: Poisson distribution

Consider a dependent variable that takes only non negative integer values $0, 1, 2, \ldots$, and one assumes that the dependent variable follows a Poisson distribution, and we wishes to estimate the Poisson parameter.

- Given $y_i \sim f(\lambda, y_i) = \frac{\exp(-\lambda)\lambda^{y_i}}{y_i!}$
- Likelihood $\mathcal{L}(y; \lambda) = \prod_{i=1}^{N} \frac{\exp(-\lambda)\lambda^{y_i}}{y_i!} = \frac{\exp(-N\lambda)\lambda^{\sum_i^N y_i}}{\prod_{i=1}^{N} y_i!}$
- Log likelihood
  $\log \mathcal{L}(y; \lambda) = -N\lambda + \sum_i^N y_i \log(\lambda) - \sum_i^N \log(y_i!)$
- Estimate

$$\frac{\partial \log \mathcal{L}(y; \lambda)}{\partial \lambda} = 0 \Longrightarrow \widehat{\lambda} = \frac{\sum_i^N y_i}{N}$$

Feasible examples: Least Squares

- Normality assumption $e \sim \mathbb{N}(0, \sigma^2)$, then $y \sim \mathbb{N}(x\beta, \sigma^2)$.
- Likelihood $L(\beta) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp(-0.5\sigma^{-2}(y - x\beta)'(y - x\beta))$
- log likelihood $\log L(\beta) = -\frac{N}{2} \log \sigma^2 - \frac{1}{2\sigma^2}(y - x\beta)'(y - x\beta)$
- $\beta = (x'x)^{-1}x'y$

Some difficulties

- Non-uniqueness of the Likelihood Function
- Non-existence of a solution to the Maximization Problem
- Multiple Solutions to the Maximization Problem

Asymptotic Properties (1): Convergence

Definition

Under a set of regularity conditions, there exists a sequence of maximum likelihood estimators converging almost surely to the true parameter value θ_0

- The variables $Y_i, i = 1, 2, \ldots$ are independent and identically distributed with density $f(y; \theta), \theta \in \Theta \in \mathbb{R}^p$
- The parameter space Θ is compact.
- The log likelihood function $\mathcal{L}(y, \theta)$ is continuous in θ and is a measurable function of y.
- The log-likelihood function is such that $(1/n)\mathcal{L}_n(y, \theta)$ converges surely to $E_{\theta_0} log(f(Y_i; \theta))$ uniformly in $\theta \in \Theta$. $E_{\theta_0} log(f(Y_i; \theta))$ exists.

Asymptotic Properties (2): Asymptotic Normality

- The log likelihood function $\mathcal{L}_n(\theta)$ is twice continuously differentiable in an open neighborhood of θ_0
- The matrix (Fisher Information Matrix)

$$\mathcal{I}_1(\theta_0) = E_{\theta_0}\left(-\frac{\partial^2 \log f(Y_1;\theta_0)}{\partial\theta\partial\theta'}\right)$$

Definition

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to \mathbb{N}(0, \mathcal{I}_1(\theta_0)^{-1}).$$

Concentrated Likelihood Function

Definition

Let the parameter set $\theta = (\alpha, \beta)$. The solutions $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ to the mazimization problem $\max_{\alpha,\beta} \log \mathcal{L}(y; \alpha, \beta)$ can be obtained via the following two-step procedure:

a) Maximize the log-likelihood function with respect α given β. The maximum value is attained for values of α in a set $A(\beta)$ depending on the parameter β. Thus, if $\alpha \in A(\beta)$, the log-likelihood value is

$$\log \mathcal{L}_c(y; \beta) = \max_{\alpha} \log \mathcal{L}(y; \alpha, \beta)$$

The mapping $\log \mathcal{L}_c$ is called the concentrated (in α) log likelihood function.

b) In a second step, maximize the concentrated log-likelihood function with respect to β.

Application

Consider the likelihood

$$\mathcal{L}(y, \beta, \sigma) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 - \frac{1}{2\sigma^2}(y - x\beta)'(y - x\beta)$$

- First step

$$\frac{\partial \mathcal{L}(y; \beta, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}(y - x\beta)'(y - x\beta) = 0$$

  Then

$$\sigma^2(\beta) = \frac{1}{n}(y - x\beta)'(y - x\beta)$$

- Substituting $\sigma^2(\beta)$ into the likelihood

$$\mathcal{L}_c(y, \beta, \sigma) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \frac{1}{n}(y - x\beta)'(y - x\beta) - \frac{n}{2}$$

Hypothesis Testing

Three procedures to do tests

Likelihood Ratio

- The likelihood ratio statistic is

$$LR = 2(\ell(\theta) - \ell(\tilde{\theta}))$$

where $\hat{\theta}$ and $\tilde{\theta}$ are the restricted and unrestricted maximum likelihood estimates of θ.

- Wilk's theorem shows that

$$LR \sim \chi^2(r)$$

where r is the number of restrictions.

Additional Tests

- Wald Test
- LM test

We will see in GMM.

In practice

- The regularity conditions are strong.
- What happens if we weaken them?

Problem 1

Number of parameters increases with the number of observations

- Convergence holds
- Estimates may be biased

Problem 2

True parameter value θ_0 does not belong to Θ: The model is misspecified

- Convergence holds to a parameter that is not the true parameter.

Problem 3

Correlated Observations

- Convergence does not hold.

Problem 4

Discontinuity of the likelihood function

- Numerical problems.

Problem 5

Known parameter space

- Constrained Optimization

Maximum Likelihood

Applications

- Binary Outcomes
- Multinomial Choices
- Count Data

Maximum Likelihood Estimation: Logit and Probit Models

For binary outcome data, the dependent variable y takes one of two values. We let

$$y = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1 - p \end{cases} \tag{1}$$

Parametrize conditional probabilities:

$$p_i = F_\epsilon(X\beta) \tag{2}$$

And, Marginal Effects

$$\frac{\partial Pr(y_i = 1 \mid x_i)}{\partial x_{ij}} = F'_\epsilon(X\beta)\beta_j \tag{3}$$

Probit Model (1)

The probit model corresponds to the case where $F(x)$ is the cumulative standard normal distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(\tfrac{1}{2}X^2) dX \tag{4}$$

Where $F(X\beta) = \Phi(X\beta)$.

Probit Model (2)

- Consider the latent approach

$$y^\star = X\beta + \epsilon \tag{5}$$

where $\epsilon \sim N(0,1)$. Think of $y^\star$ as the net utility associated with some action. If the action yields positive net utility, it is undertaken otherwise it is not.

- We would care only about the sign of $y^\star$

$$y = \begin{cases} 1, & \text{if } p \\ 0, & \text{with probability } 1-p \end{cases} \tag{6}$$

- Probabilities

$$\begin{aligned} Pr(y=1) &= Pr(y^\star > 0) = Pr(X\beta + \epsilon \geq 0) \\ &= Pr(\epsilon \geq -X\beta) = Pr(\epsilon \leq X\beta) = \Phi(X\beta) \end{aligned}$$

Logit Model

The logit model specifies the cdf function $F(x)$ is now the logistic function

$$\Lambda(x) = \frac{1}{1 + \exp(-x)} = \frac{exp(x)}{1 + \exp(x)} \tag{7}$$

The logit model is most easily derived by assuming that

$$\log\left(\frac{P}{1-P}\right) = X\beta \tag{8}$$

The logarithm of the odds (ratio of two probabilities) is equal to $X\beta$

Maximum Likelihood Estimation

- Likelihood can not be defined as a joint density function.
- Outcome of a Bernouilli trial

$$f(y_i \mid x_i) = p_i^{y_i}(1 - p_i)^{1-y_i} \tag{9}$$

- Given the independence of individuals, the likelihood can be written

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} F(x_i\beta)^{y_i}(1 - F(x_i\beta))^{1-y_i} \tag{10}$$

Log Likelihood

- The log likelihood

$$\log \mathcal{L}(\beta) = \sum_{i=1}^{n} y_i ln F(x_i\beta) + (1 - y_i) ln(1 - F(x_i\beta)) \qquad (11)$$

- First order conditions

$$\sum_{i=1}^{n} \frac{y_i - F(x_i\beta)}{F(x_i\beta)(1 - F(x_i\beta))} F'(x_i\beta) x_i = 0 \qquad (12)$$

Empirical considerations

- Probit and logit yield same outcomes. Only difference is how parameters are scaled.
- The natural metric to compare models is the fitted log-likelihood provided that the models have the same number of parameters.
- Although estimated parameters are different, marginal effects are quite similar.

Pseudo R2

$$R^2_{\text{Binary}} = 1 - \frac{\mathcal{L}(\hat{\beta})}{N[\bar{y} ln \bar{y} + (1 - \bar{y}) ln(1 - \bar{y})]} \tag{13}$$

Predicted Outcomes

- The criterion $\sum_i (y_i - \hat{y}_i)^2$ gives the number of wrong predictions.
  - average rule: let $\hat{y} = 1$ when $\hat{p} = F(X\beta) > 0.5$
  - Receiver Operating Characteristics (ROC) curve plots the fractions of $y = 1$ correctly classified against the fractions of $y = 0$ incorrectly specified as the cutoffs $\hat{p} = F(X\beta) > c$ varies.

Maximum Likelihood

Applications

- Binary Outcomes
- Multinomial Choices
- Count Data

Multinomial Models

Dependent variable has several possible outcomes, that are mutually exclusive

- Commute to work (car, bus, bike, walking)
- Employment status (full time, part time, unemployed)
- Occupation choice, field of study, product choice
- Ordered choices eg. education choices
- Unordered choices eg. fishing mode

Ordered Discrete Response

Suppose that $y^{\star}(= x_i\beta + \epsilon_i)$ is continuously distributed with standard deviation σ but the observed response y_i is an ordered discrete choice (ODR) taking 0,1,. . . ,R-1 determined by fixed threesholds γ_r. Formally, we have

$$y = \begin{cases} 0, & \text{if } x\beta + \epsilon < \gamma_1 \\ 1, & \text{if } \gamma_1 \leq x\beta + \epsilon < \gamma_2 \\ 2, & \text{if } \gamma_2 \leq x\beta + \epsilon < \gamma_3 \\ \dots & \\ R-1, & \text{if } \gamma_{R-1} \leq x\beta + \epsilon \end{cases} \tag{14}$$

Identification

- Not all parameters are identified as in the binary response model.
- Consider $\gamma_r \leq x\beta + \epsilon < \gamma_{r+1}$. Then, we have
$$\frac{\gamma_r - \gamma_1}{\sigma} \leq \frac{x\beta + \epsilon - \gamma}{\sigma} < \frac{\gamma_{r+1} - \gamma_1}{\sigma}$$
- Then, the identified parameters are

$$\alpha = \left(\frac{\beta_1 - \gamma_1}{\sigma}, \frac{\beta_2}{\sigma}, \ldots, \frac{\beta_k}{\sigma} \right) \tag{15}$$

$$\rho_r = \frac{\gamma_r - \gamma_1}{\sigma} \tag{16}$$

for $r = 2, \ldots, R-1$.

Toward the Likelihood

$y = r$

$$
\begin{aligned}
\gamma_r &\leq x\beta + \epsilon &&< \gamma_{r+1} \\
\gamma_r - x\beta &\leq \epsilon &&< \gamma_{r+1} - x\beta \\
\frac{\gamma_r - \gamma_1}{\sigma} + \frac{\gamma_1 - x\beta}{\sigma} &\leq \frac{\epsilon}{\sigma} &&< \frac{\gamma_{r+1} - \gamma_1}{\sigma} + \frac{\gamma_1 - x\beta}{\sigma} \\
\rho_r - x\alpha &\leq \frac{\epsilon}{\sigma} &&< \rho_{r+1} - x\alpha
\end{aligned}
$$

Likelihood

- Choice probabilities

$$\begin{aligned} P(y = r \mid x) &= P(\rho_r - x\alpha \leq \frac{\epsilon}{\sigma} < \rho_{r+1} - x\alpha) \\ &= F(\rho_{r+1} - x\alpha) - F(\rho_r - x\alpha) \end{aligned}$$

- Likelihood

$$\mathcal{L}(a, t) = \prod_{i=1}^{N} \prod_{r=0}^{R-1} [P(y_i = r \mid x)]^{y_{ir}} \tag{17}$$

- Log Likelihood of the probit model

$$\log \mathcal{L}(a, t) = \sum_{i=1}^{N} \sum_{r=0}^{R-1} y_{ir} \log(\Phi(t_{r+1} - xa) - \Phi(t_r - xa)) \tag{18}$$

where a and t are respectively the parameters to be estimated and the cutoffs.

Marginal Effects

- The marginal effects of x on each choice probabilities can be derived as:

$$\frac{\partial P(y = r \mid x)}{\partial x} = -\alpha(\phi(\rho_{r+1} - x\alpha) - \phi(\rho_r - x\alpha)) \tag{19}$$

- Caution

$$\sum_{r=0}^{R} P(y = r \mid x) = 1 \tag{20}$$

Then

$$\sum_{r=0}^{R} \frac{\partial P(y = r \mid x)}{\partial x} = 0 \tag{21}$$

An increase in some choice probability necessarily entails a decrease in some other choice probabilities.

Multinomial Logit

The models differ according to whether or not regressors vary across alternatives.

- Conditional logit model

$$p_{ij} = \frac{\exp(x_{ij}\beta)}{\sum_{l=1}^{m} \exp(x_{il}\beta)} \qquad j = 1, \ldots, m. \tag{22}$$

- Multinomial logit model

$$p_{ij} = \frac{\exp(x_i\beta_j)}{\sum_{l=1}^{m} \exp(x_i\beta_l)} \qquad j = 1, \ldots, m. \tag{23}$$

- Mixed logit model

$$p_{ij} = \frac{\exp(x_{ij}\beta + w_i\gamma_j)}{\sum_{l=1}^{m} \exp(x_{il}\beta + w_i\gamma_l)} \qquad j = 1, \ldots, m. \tag{24}$$

Marginal Effects

- Conditional logit

$$\frac{\partial p_{ij}}{\partial x_{ik}} = p_{ij}(\delta_{ijk} - p_{ik})\beta \tag{25}$$

where δ_{ijk} is an indicator variable equal to 1 if $j = k$ and equal to 0 otherwise.

- Multinomial logit

$$\frac{\partial p_{ij}}{\partial x_{i}} = p_{ij}(\beta_j - \bar{\beta}_i) \tag{26}$$

where $\bar{\beta}_i = \sum_l p_{il}\beta_l$

Independence of Irrelevant Alternatives

- A property of the conditional logit and multinomial logit is that discrimination among the m alternatives reduces to a series of pairwise comparisons that are unaffected by the characteristics of alternatives other than the pair under consideration.
- The choice probabilities must be unaffected by the removal of one alternative.

That is because

$$Pr(y = j \mid y = k) = \frac{p_j}{p_j + p_k} \tag{27}$$

Testing for IIA

- Estimate the model twice
  - On the full set of alternatives and obtain θ_{full}
  - On a subset of alternatives and obtain θ_{subset}
- Compare $\mathcal{L}_{subset}(\theta_{full})$ and $\mathcal{L}_{subset}(\theta_{subset})$. If there is a significant different, then IIA is violated.

It is very (very) rare that IIA is not violated.

Hausman and McFadden Test

$$HM = (\hat{\beta}^r - \hat{\beta}^f)'[var_{\hat{\beta}^r} - var_{\hat{\beta}^f}]^{-1}(\hat{\beta}^r - \hat{\beta}^f) \tag{28}$$

We have that

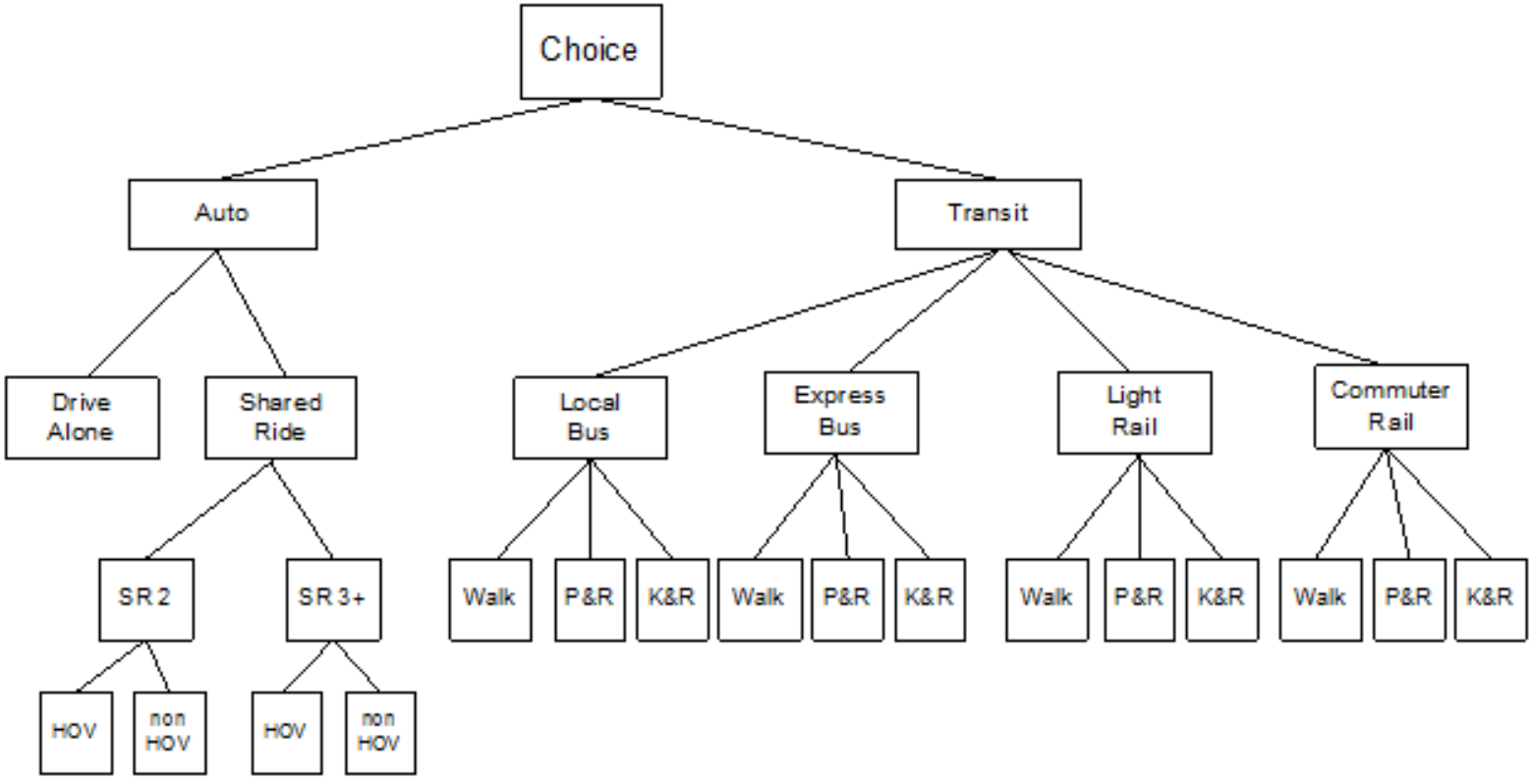
$$HM \sim \chi^2(||\beta^r||) \tag{29}$$

If IIA holds.

Alternatives

- Generalized Extreme Value Model
- Nested Logit Model
- Random Parameters Logit
- Multinomial Probit

Nested Logit



Nested Logit

The nested logit model breaks decision making into groups. The utility for the alternative is given

$$U_{jk} = V_{jk} + \epsilon_{jk} \quad k = 1, 2, \ldots, K_j, \quad j = 1, 2, \ldots, J \tag{30}$$

Utilities are given by:

- $V_{11} + \epsilon_{11}$
- $\ldots$
- $V_{JK_J} + \epsilon_{JK_J}$

Choice Probabilities

- Choice probability

$$p_{jk} = p_j \times p_{k|j}. \tag{31}$$

- This arises from GEV joint distribution

$$F(\epsilon) = exp(-G(e^{-\epsilon_{11}}, \ldots, e^{-\epsilon_{1K_1}}; \ldots; e^{-\epsilon_{J1}}, \ldots, e^{-\epsilon_{JK_J}}))$$

with

$$G(Y) = G(Y_{11}, \ldots, Y_{1K_1}, \ldots, Y_{JK_J}) = \sum_{j=1}^{J} \left(\sum_{k=1}^{K_j} Y_{jk}^{\frac{1}{\rho_j}} \right)^{\rho_j} \tag{32}$$

Model

Consider

$$V_{jk} = z_j\alpha + x_{jk}\beta_j \quad k = 1, \ldots, K_j, \quad j = 1, \ldots, J \qquad (33)$$

The probability of the nested logit model

$$p_{jk} = p_j \times p_{k|j} = \frac{\exp(z_j\alpha + \rho_j I_j)}{\sum_{m=1}^{J} \exp(z_m\alpha + \rho_m I_m)} \times \frac{\exp(x_{jk}\beta/\rho_j + \rho_j I_j)}{\sum_{m=1}^{J} \exp(z_m\alpha + \rho_m I_m)}$$

where

$$I_j = ln\left(\sum_{l=1}^{K_j} \exp(x_{jl}\beta_j/\rho_j)\right)$$

is the inclusive value or the log-sum.

Likelihood

For the *ith* observation, we observe $K_1 + \ldots + K_J$ outcomes y_{ijk}, where $y_{ijk} = 1$ if alternative jk is chosen and is zero otherwise. Then the density of one observation y_i can be expressed

$$f(y_i) = \prod_{j=1}^{J} \prod_{k=1}^{K_J} [p_{ij} \times p_{ik|j}]^{y_{ijk}} = \prod_{j=1}^{J} p_{ij}^{y_{ij}} \left(\prod_{k=1}^{K_J} p_{ik|j}^{y_{ijk}} \right)$$

The log likelihood is given by

$$lnL = \sum_{i=1}^{N} \sum_{y=1}^{J} y_{ij} \log(p_{ij}) + \sum_{i=1}^{N} \sum_{y=1}^{J} \sum_{k=1}^{K_J} y_{ijk} \log(p_{ik|j})$$

Discussions and Limitations

- Nested Logit can be estimated in two steps following the construction of the densities
- Not all choices are easy to nest.

Multinomial Probit

- Consider m-choice with

$$U_j = V_j + \epsilon_j \qquad j = 1, 2, \ldots, m. \tag{34}$$

  where $\epsilon \sim \mathbb{N}(0, \Sigma)$.
- Σ is the matrix of variance-covariance, which can be left unrestricted to capture the correlation between choices.
- Choice probabilities.. For 4 choices, we have

$$P(Y = 1) = \int_{-\infty}^{-V_{41}} \int_{-\infty}^{-V_{31}} \int_{-\infty}^{-V_{21}} f(x, y, z) dz dy dx \tag{35}$$

  where $f(x, y, z)$ is the pdf of the trivariate normal.

Need simulation to compute the choice probabilities and the likelihood.

Maximum Likelihood

Applications

- Binary Outcomes
- Multinomial Choices
- Count Data

Introduction

Suppose y takes integers 0,1,2,. . . , which are cardinal not just ordinal. Then y is a count response.

- Number of accidents
- Number of live births
- Doctor visits
- Prescription drugs
- Recreational trips

Examples

- Number of children over a specified age interval of the mother, depending on mother's schooling, age, household income
- Visits to a museum to characterize the impact of specific attractions (animals, science room)
- Airline safety (number of accidents) to airline profitability

Motivation

- Linear regression on count data
- Discrete choice methods (ordered, multinomial)

Table 20.1. *Proportion of Zero Counts in Selected Empirical Studies*

Study	Variable	Sample Size	Proportion of Zeros
Cameron et al. (1988)	Doctor visits	5,190	0.798
Pohlmeier and Ulrich (1995)	Specialist visits	5,096	0.678
Grootendorst (1995)	Prescription drugs	5,743	0.224
Deb and Trivedi (1997)	Number of hospital stays	4,406	0.806
Gurmu and Trivedi (1996)	Recreational trips	659	0.632
Geil et al. (1997)	Hospitalizations	30,590	0.899
Greene (1997)	Major derogatory reports	1,319	0.803

Poisson Models

- The natural choice is a Poisson Model

$$Pr(Y) = \frac{\exp(-\mu)\mu^y}{y!}$$

  where μ is the mean and the variance of Y.
- In Poisson MLE, we assume that $y \mid x$ follows a poisson distribution with parameter $\lambda(x) > 0$, and we have

$$P(y = r \mid x) = \frac{\lambda(x)^r}{r!} \exp(-\lambda(x))$$

- We specify

$$\lambda(x) = \exp(x_i'\beta) \tag{36}$$

Likelihood

- The likelihood is given by

$$\log \mathcal{L}(\beta) = \sum_i y_i x_i' \beta - \exp(x_i' \beta) - \log y_i! \tag{37}$$

- FOC

$$\frac{\partial \mathcal{L}(\beta)}{\partial \beta} = \sum_i \{y_i - \exp(x_i' \beta)\} x_i \tag{38}$$

- SOC

$$\frac{\partial^2 \mathcal{L}(\beta)}{\partial \beta \partial \beta'} = \sum_i [-\exp(x_i' \beta)] x_i x_i' \tag{39}$$

Marginal Effects

- Standard solution

$$\frac{\partial E(y \mid x)}{\partial x} = \beta \exp(x\beta) \tag{40}$$

- Better marginal effect: Proportional change in $E(y \mid x)$ as x changes by one unit.

$$\frac{\partial E(y \mid x)/\partial x}{E(y \mid x)} = \beta \tag{41}$$

Over-dispersion problem

- The Poisson Model implies that $E(y \mid x) = Var(y \mid x)$.
- Too restrictive, and leads to a problem of excess zeroes.
- Explanations

Negative Binomial Model

- The negative binomial distribution is a mixture between the Poisson and Gamma distributions.
- Parametrization $\lambda_i \equiv \exp(x_i'\beta)$ and $\psi_i \equiv \frac{1}{\alpha}\lambda_i^{\kappa}$ where β, $\alpha > 0$ and κ are parameters.
- Choice probabilities are given by

$$P(y_i \mid x_i) = \frac{\Gamma(y_i + \psi_i)}{\Gamma(\psi_i)\Gamma(y_i + 1)} \left(\frac{\psi_i}{\lambda_i + \psi_i}\right)^{\psi_i} \left(\frac{\psi_i}{\lambda_i + \psi_i}\right)^{y_i} \quad (42)$$

- The moments are given

$$\begin{aligned} E(y \mid x_i) &= \lambda_i \\ Var(y \mid x_i) &= \lambda_i + \alpha\lambda_i^{2-\kappa} \end{aligned}$$

Zero Inflated Model

- Imagine, two densities $f_1(\cdot)$ and $f_2(\cdot)$.

$$g(y) = \begin{cases} f_1(0) + (1 - f_1(0))f_2(0), & \text{if } y = 0 \\ (1 - f_1(0))f_2(y), & \text{if } y \geq 1 \end{cases} \tag{43}$$

- The likelihood can be derived, and the model is able to cope with many zeros.