Discrete Response:
The Logit Model
Advanced Microeconometrics

Anders Munk-Nielsen
2021

Plan for lectures: Helicopter

Part I: Linear methods. ✓

Part II: High-dimensional methods. ✓

Part III: M-estimation, theory ✓

Part IV: M-estimation, examples ←

Where are we in the course?

Part	Topic	Parameterization non-linear	Estimation non-linear	Dimension dim(x)	Numerical optimization	M-estimation (Part III)	Outcome (y_i)	Panel (c_i)
I	OLS	÷	÷	low	÷	✓	$\mathbb{R}$	✓
II	LASSO	÷	✓	high	✓	÷	$\mathbb{R}$	÷
IV	Probit	✓	✓	low	✓	✓	$\{0, 1\}$	÷
	Logit	✓	✓	low	✓	✓	$\{1, 2, ..., J\}$	÷
	Tobit	✓	✓	low	✓	✓	$[0; \infty)$	÷
	Sample selection	✓	✓	low	✓	✓	$\mathbb{R}$ and $\{0, 1\}$	÷
	Simulated Likelihood	✓	✓	low	✓	✓	Any	✓
	Quantile Regression	÷	✓	(low)	✓	✓	$\mathbb{R}$	÷
	Non-parametric	✓	(✓)	∞	÷	÷	$\mathbb{R}$	÷

Outline

Big Big Picture: Modeling Approaches

- **Simplest:** with $q_{jt} = \#$ sold cars and $\mathbf{x}_{jt} =$ car attributes (price, horsepower, km/l, etc.)

$$q_{jt} = \mathbf{x}_{jt}\beta + \text{error}_{jt}$$

  - **Con:** Competitor prices / selection does not affect sales. E.g. zero cross-price elasticity.

- **Big OLS:** Include prices of *all* cars in explaining j's sales

$$q_{jt} = \mathbf{x}_{jt}\beta + \sum_{k \neq j} \mathbf{x}_{kt}\gamma_{jk} + \text{error}_{jt}.$$

  - **Cons:**
    - *many* parameters (e.g. J^2 cross-price coefficients!!),
    - **Too flexible?** Might violate basic logic (Slutsky symmetry, adding up, etc.)

- **Logit:**

$$\frac{q_{jt}}{\sum_{k=1}^{J} q_{kt}} = \frac{\exp(\mathbf{x}_{jt}\beta)}{\sum_{k=1}^{J} \exp(\mathbf{x}_{kt}\beta)}.$$

  - **Pros:** Few parameters ($\dim \beta = K$), cross-product "market stealing" (substitution)
  - **Cons:** Restricted substitution patterns (the "IIA property")

Logit Intuition

- **Choices:** four different cars.
- **Preferences:** $u^{\mathrm{bat}} \succ u^{\mathrm{Delorean}} \succ u^{\mathrm{wagon}} \succ u^{\mathrm{apple}}$.
- **Data:**
  - Characteristics of cars, x_j (e.g. price, horsepower, flux-capacitor dummy, ...)
  - Characteristics of decision maker, x_i (e.g. income, work distance).
  - The final decision (which e.g. says $u^{\mathrm{batmobile}} \succ u^j$ for all other j)
- **Model:** A *utility function*, $u : (\mathbf{x}_{ij}, \theta) \mapsto \mathbb{R}$.
  - **Goal:** Choose θ st. $u(\mathbf{x}_{i\mathrm{bat}}, \theta) > u(\mathbf{x}_{ij}, \theta)$ for $j \neq \mathrm{bat}$.
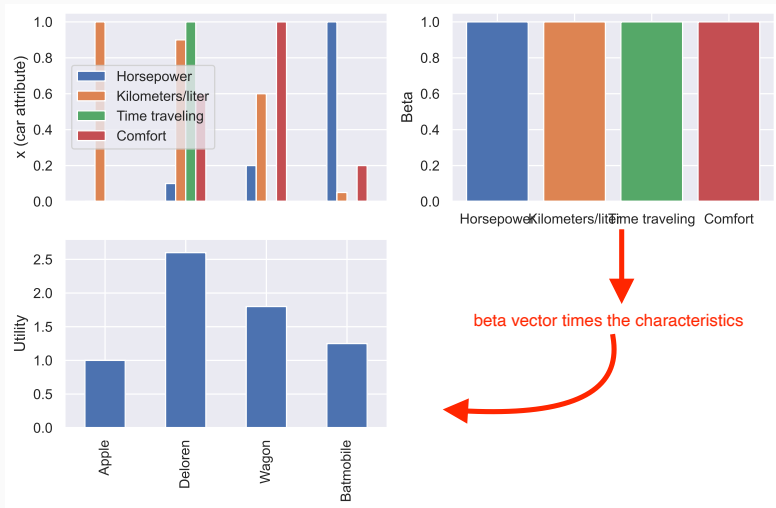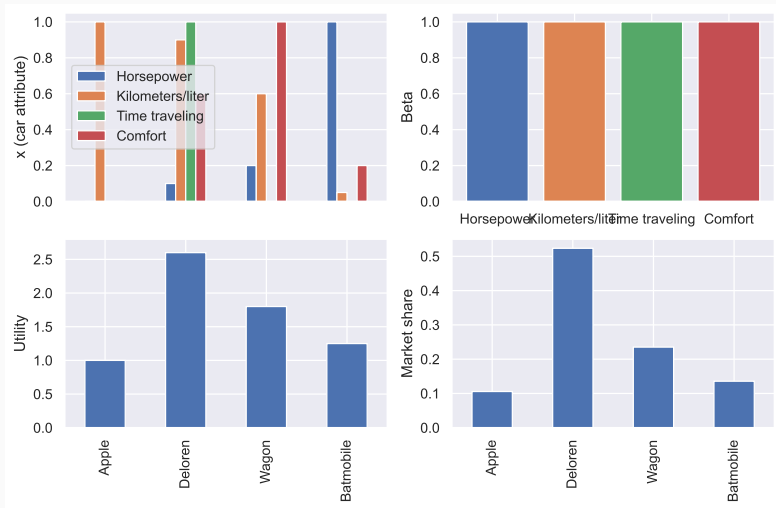   We are formalizing a utility function and estimate it

Logit intuition

beta vector times the characteristics
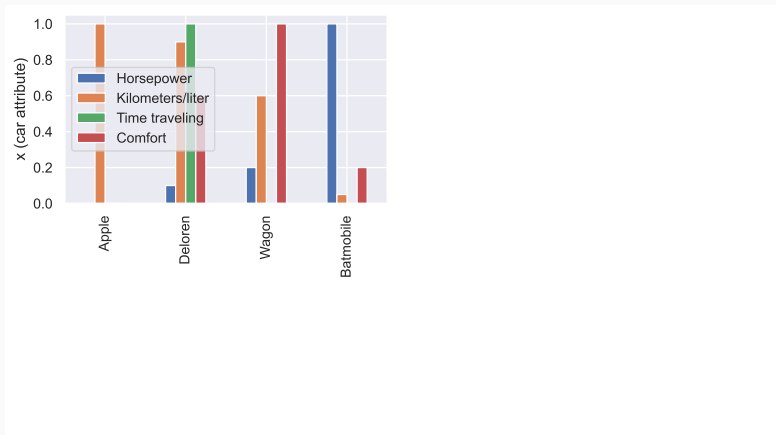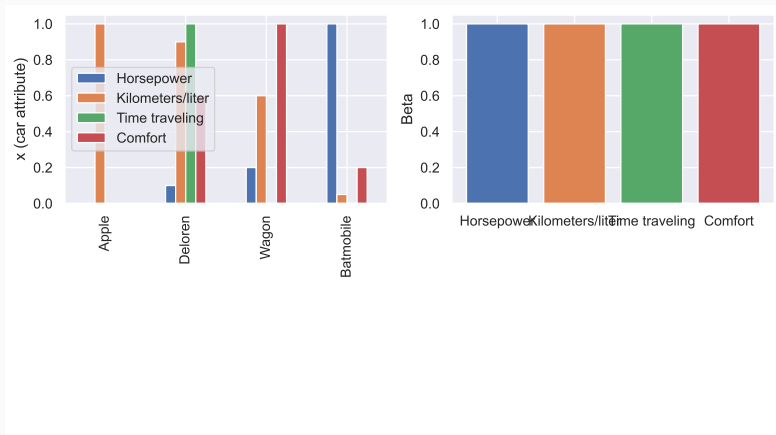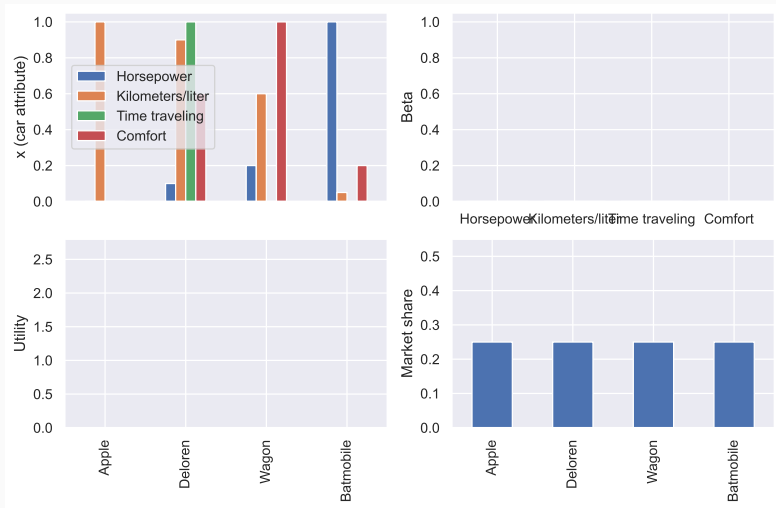
No utility from any characteristics $\Rightarrow$ identical market shares.
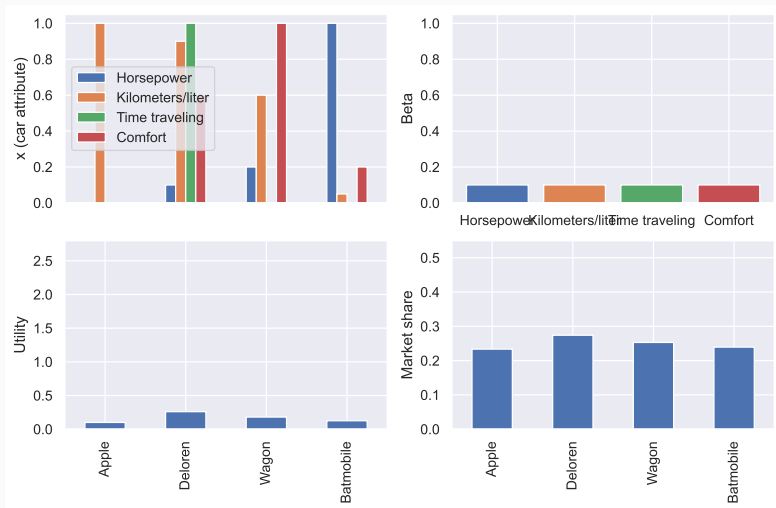
Market shares have the same pattern but smaller coefficients move the markets shares closer to each other

Here, $\beta = \left(\frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}\right)$

Consumers prefer comfort and time traveling capabilities
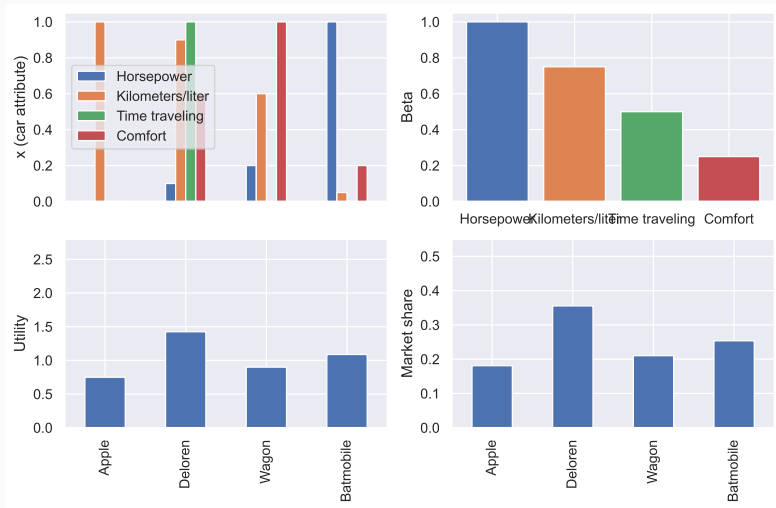
Consumers prefer horsepower and fuel economy.
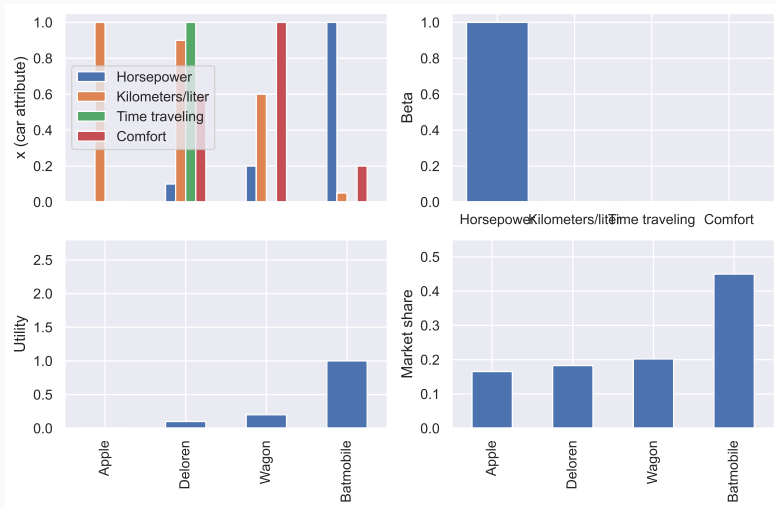
Consumers only care about horsepower (and idiosyncraticies)
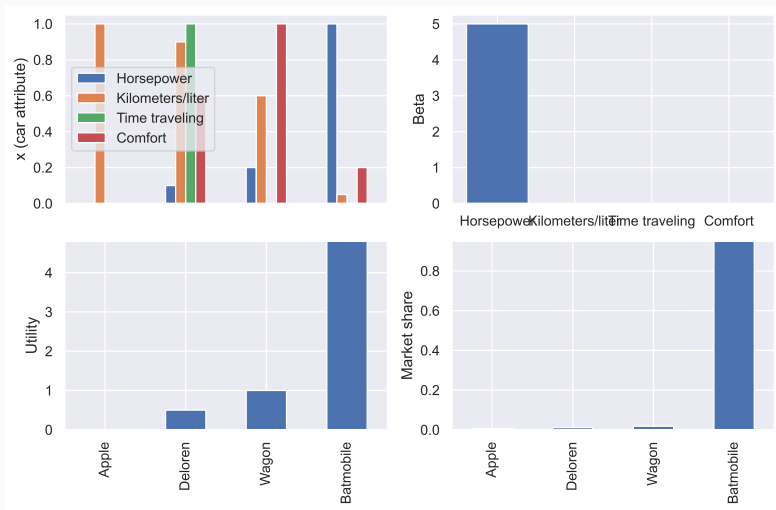
Consumers only care about horsepower, and by a lot!

Model, 1st idea

For cars $j = 1, ..., J$,

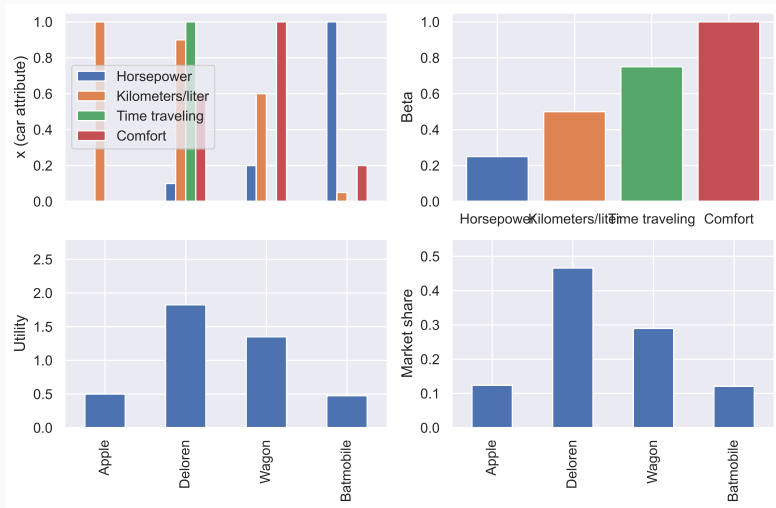$$u_{ij} = \beta_0 + \beta_1 \text{horsepower} + \beta_2 \text{income} + \text{error}, \quad \text{error} \sim \text{IID}(\mu, \sigma^2).$$

Discuss

What can we hope to identify here?

We cannot identify Beta_0: The level of utility is unidentified: Adding 100 to all utilities does not change the answer of the model.

We can identify Beta_1: Changing it will change the market shares

We cannot identify Beta_2: Attribute of the decision maker at not the cars. No variations across cars (alternatives)
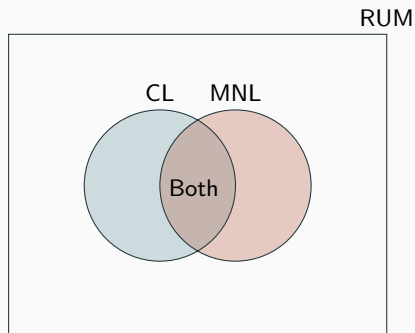
Random Utility Model (RUM)

Dont observue u_ij because eps_ij is unobserved

$$u_{ij} = v_{ij} + \varepsilon_{ij},$$

$$y_i = \underset{j \in \{1,...,J\}}{\operatorname{argmax}} u_{ij}.$$

i: Individual, $i = 1, ..., N$,

j: Alternative, $j = 1, ..., J$,

y_i: Chosen alternative by i,

u_{ij}: Utility from choosing alternative j,

ε_{ij}: Unobserved error term,

v_{ij}: Observed part (regressors and parameters).

- **Primary** two types of RUM:
  1. **Conditional logit:** $v_{ij} = \mathbf{x}_j \beta$,
  2. **Multinomial logit:** $v_{ij} = \mathbf{x}_i \beta_j$,

Overview of models

Random Utility Model

$$u_{ij} = v_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{IID Extreme Value Type I},$$

$$y_i = \operatorname{argmax}_{j \in \{1,\ldots,J\}} u_{ij}.$$

- **Conditional choice probabilities (CCPs):** to be shown...

$$\Pr(d|v_{ij}) = \frac{\exp(v_{id})}{\sum_{j=1}^{J} \exp(v_{ij})}.$$

  - due to Dan McFadden.

Conditional logit (CL)

- $\mathbf{x}_j$: characteristics of the *alternatives*.

- **Example**: car choice, $\mathbf{x}_j = $ (price, horsepower, ...).

- **Normalization**: No intercept in $\mathbf{x}$.

- β_k **interpretation**: marginal utility of characteristic k.

Beta is constant and does not vary with each car -> Prefrences are fixed across decession makers

Multinomial logit (MNL)

- $\mathbf{x}_i$: characteristics of the *individual*.

- **Example**: car choice, $\mathbf{x}_i = $ (income, work distance, ...).

- **Normalization**: $\beta_1 = \mathbf{0}_{K \times 1}$.

- β_{jk} **interpretation**: change in utility of car j (relative to car 1) when (e.g.) income increases (if $\mathbf{x}_{ik}$ is income).

Outline

Random Utility Model

Random Utility Model

$$u_{ij} = v(\mathbf{x}_{ij}, \theta) + \varepsilon_{ij}, \quad \text{IID Extreme Value Type I,}$$

$$y_i = \text{argmax}_{j \in \{1,\dots,J\}} u_{ij}.$$

$v(\cdot, \cdot)$: Known functional form,

$\mathbf{x}_{ij}$: Data (characteristics of individuals and/or alternatives),

- **Functional form:** CL and MNL are two *linear* forms of v
  - **CL:** $v(\mathbf{x}_{ij}, \beta) = \mathbf{x}_{ij}'\beta$
    - Note: no intercept in $\mathbf{x}_{ij}$ (intercept is normalized)
  - **MNL:** $v(\mathbf{x}_i, \beta_j) = \mathbf{x}_i'\beta_j$.
    - Note: $\theta = (\beta_2, \dots, \beta_J)$ ($\beta_1 := 0$ is normalized)

Multinomial Logit

$$u_{ij} = \mathbf{x}_i\boldsymbol{\beta}_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{IID Extreme Value Type I}, \; \boldsymbol{\beta}_1 = \mathbf{0}_{K \times 1},$$

$$y_i = \text{argmax}_{j \in \{1, \dots, J\}} u_{ij}.$$

```python
1  from scipy.stats import genextreme # generalized EV distr.
2  def sim_data(theta, N):
3      K, J_1 = theta.shape # theta must be (K, J-1)
4      J = J_1 + 1
5      xx = np.random.normal(size=(N,K-1))
6      oo = np.ones((N,1)) # constant term
7      x  = np.hstack([oo,xx]) # full x matrix
8      zz = np.zeros((K,1)) # normalized coefficients
9      beta = np.hstack([zz, theta]) # full (K,J) matrix
10     v = x @ beta # (N,J): observable utility
11     uni = np.random.uniform(size=(N,J)) # uniform draws
12     e = genextreme.ppf(uni, c=0) # generalized extreme value errors
13     u = v + e # full (unobserved) utility
14     y = u.argmax(axis=1) # observed, chosen alternative
15     return y, x
```

Conditional Logit

$$u_{ij} = \mathbf{x}'_{ij}\beta + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{IID Extreme Value Type I},$$

$$y_i = \text{argmax}_{j \in \{1,\dots,J\}} \, u_{ij}.$$

```python
1 def sim_data(theta, N, J):
2         K = theta.size
3         assert theta.ndim == 1 # theta should be (K,)
4         x = np.random.normal(size=(N,J,K)) # Explanatory variables
5         uni = np.random.uniform(size=(N,J))
6         e = genextreme.ppf(uni, c=0) # generalized extreme value
7         v = x @ theta # (N,J) matrix of "observable utilities"
8         u = v + e # full utility (unobserved)
9         y = u.argmax(axis=1) # observed chosen alternative
10        return y,x
```

- **Goal:** Derive the likelihood function.
- **How** do we begin?
- **Normally:** we always try to isolate the error terms..
- **This time:** there are J of them and only one y_i is observed...

Conditional Logit Model

$$u_{ij} = \mathbf{x}_j \boldsymbol{\beta}_o + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{Extreme Value Type I.}$$

$$y_i = \text{argmax}_{j \in \{1,\dots,J\}} \ u_{ij}.$$

- **Choice probability:**

$$
\begin{aligned}
\Pr(y = j | \mathbf{X}) &= \Pr(u_{ij} > u_{ik} \ \forall k \neq j | \mathbf{X}) \\
&= \Pr(\mathbf{x}_j \boldsymbol{\beta}_o + \varepsilon_{ij} > \mathbf{x}_k \boldsymbol{\beta}_o + \varepsilon_{ik} \ \forall k \neq j | \mathbf{X}) \\
&= \Pr(\varepsilon_{ik} < \mathbf{x}_j \boldsymbol{\beta}_o - \mathbf{x}_k \boldsymbol{\beta}_o + \varepsilon_{ij} \ \forall k \neq j | \mathbf{X}) \\
&= \cdots \\
&= \frac{\exp(\mathbf{x}_j \boldsymbol{\beta}_o)}{\sum_{k=1}^{J} \exp(\mathbf{x}_k \boldsymbol{\beta}_o)}.
\end{aligned}
$$

All epsilons are IID:
Law iterated expectations to get epsilon outside

▸ Details

Random Utility Model

Conditional Logit Model

$$u_{ij} = v_{ij}(\boldsymbol{\theta}) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{Extreme Value Type I.}$$
$$y_i = \text{argmax}_{j \in \{1,\dots,J\}} u_{ij}.$$

- **Choice probability:**

$$
\begin{aligned}
\Pr(y = j | \mathbf{X}) &= \Pr(u_{ij} > u_{ik} \,\forall k \neq j | \mathbf{X}) \\
&= \Pr(v_{ij} + \varepsilon_{ij} > v_{ik} + \varepsilon_{ik} \,\forall k \neq j | \mathbf{X}) \\
&= \Pr\left[\varepsilon_{ik} < v_{ij} - v_{ik} + \varepsilon_{ij} \,\forall k \neq j | \mathbf{X}\right] \\
&= \cdots \\
&= \frac{\exp(v_{ij})}{\sum_{k=1}^{J} \exp(v_{ik})}.
\end{aligned}
$$

We showed

$$\Pr(y_i = j | \mathbf{X}) = \frac{\exp(v_{ij})}{\sum_{k=1}^{J} \exp(v_{ik})}.$$

Log Likelihood function

$$\ell_i(\boldsymbol{\theta}) = v_{iy_i} - \log\left[\sum_{k=1}^{J} \exp(v_{ik})\right].$$

```
1  def util(theta, x): for conditional logit
2          v = x @ theta # (N,J,K) * (K,) = (N,J)
3          return v
4
5  def loglikelihood(theta, y, x):
6          N,J,K = x.shape
7          v = util(theta, x)
8          v_i = v[np.arange(N), y] # (N,) vector
9          denom = np.exp(v).sum(axis=1) # (N,) vector
10         ll_i = v_i - np.log(denom)
11         return ll_i
```

Max-rescaling

- **Crucial trick:** *Analytically*, the following holds

$$\frac{\exp(v_{ij})}{\sum_{k=1}^{J} \exp(v_{ik})} = \frac{\exp(v_{ij} - K_i)}{\sum_{k=1}^{J} \exp(v_{ik} - K_i)}, \quad \forall K_i \in \mathbb{R}.$$

- **Numerically** however, `exp()` quickly becomes highly imprecise...
  - `np.exp(800) = inf`: overflow
  - `np.exp(-800) = 0.0`: underflow
- **Both** types of error are bad...
  - ... but overflow is much worse.
- **Solution:** *Max-rescaling* our utilities

```
1 def util(theta,x):
2         v = x @ theta
3         vmax = v.max(axis=1, keepdims=True) # (N,1)
4         return v - vmax # subtract columns-wise max
```

Conditional Logit

$$
\begin{aligned}
u_{ij} &= \mathbf{x}_j \boldsymbol{\beta} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{IID Extreme Value Type I}, \\
y_i &= \operatorname*{argmax}_{j \in \{1,\ldots,J\}} u_{ij}.
\end{aligned}
$$

Discuss

- Why can't we allow for an intercept (constant) in x_j?
- Why can't we estimate $\mathrm{Var}(\varepsilon_{ij})$ here? Same logic as in probit
- How does this relate to standard micro with *ordinal* vs. *cardinal* utility?

Normalization in MNL

Multinomial Logit

$$u_{ij} = \mathbf{x}_i \boldsymbol{\beta}_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{IID Extreme Value Type I},$$

$$y_i = \underset{j \in \{1,\dots,J\}}{\operatorname{argmax}} u_{ij},$$

where we normalize $\boldsymbol{\beta}_1 := \mathbf{0}_{K \times 1}$.

Discuss

- Why do we *have* to normalize β for one alternative?
- Why can we allow for an intercept in x_i and estimate β_{j0} for all j except the normalized alternative?
- What is the utility of the normalized alternative?

Outline

Features of Interest

- **Generally:** We are rarely interested in βs directly.
- **Note:** The average choice probability is the *market share*.
  - so average partial effects say how market shares change.
- **Car example:** Interested in how car choice is affected by x'es.
  - I.e. $\frac{\partial}{\partial x_{ijk}} \Pr(j|x)$.
  - Standard errors: via the delta method.
- **Conditional logit:** E.g. the effect raising the price of Volvo V70, holding fixed all other car prices

$$\frac{\partial}{\partial x_{kl}} \Pr(j) = \Pr(j) \left[\mathbf{1}_{\{k=j\}} \beta_l - \Pr(l)\beta_l \right].$$

- **Multinomial logit:** E.g. the effect on $\Pr(\text{V70})$ of raising household income,

$$\frac{\partial}{\partial x_k} \Pr(j) = \Pr(j) \left[\beta_{jk} - \sum_{l=1}^{J} \Pr(l)\beta_{lk} \right].$$

Multinomial Logit: Parameter Interpretation

- **Interpreting** β_{jp}, **warning:** Partial effect is complicated.
  - The sign is not determined by β_{jp} alone!!!

- **Instead:** Consider the log Odds-ratio (and use $\beta_1 = \mathbf{0}$)

$$\log \frac{\Pr(y = j|\mathbf{x}_i)}{\Pr(y = 1|\mathbf{x}_i)} = \log \frac{\frac{\exp(\mathbf{x}_j\beta_j)}{\sum_{k=1}^{J} \exp(\mathbf{x}_k\beta)}}{\frac{\exp(\mathbf{x}_1\beta_1)}{\sum_{k=1}^{J} \exp(\mathbf{x}_k\beta)}} = \mathbf{x}_i\beta_j.$$

- **Linear!** This is why some software reports (log) odds ratios.

More generally

- **Larger choicesets:** If $J > 1000$, partial effects are not interesting.

- **Instead:** Predicted characteristics,

$$\mathbb{E}(\text{horsepower}) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{J} \Pr(j) \times \text{horsepower}_j.$$

- **Elasticities?** Computed numerically;
    1. Compute baseline probabilities,
    2. increase the horsepower of car j by 1%,
    3. Compute new probabilities,
    4. Elasticity is $\cong$ change in probabilities.

Welfare

Discuss

How does measures like *compensating variation* relate to our model?

Definition: The compensating variation is the monetary transfer that would make an individual indifferent between two regimes.

Welfare Analysis

- **Result:** With $u_{ij} = v_{ij} + \varepsilon_{ij}$, it can be shown that

$$\mathbb{E}[\max_j u_{ij}] = \log \sum_{j=1}^{J} \exp(v_{ij}) + \gamma,$$

where $\gamma \cong .5772$ is Euler's constant.

- **Suppose** one regressor is price, e.g. x_{j1}.
- **Then** $\frac{1}{\beta_1}$ converts from utils to money.
- **Hence** we can compare "welfare" under $\{v_{ij}\}$ and $\{\tilde{v}_{ij}\}$.

$$\mathrm{CV} = \frac{1}{\beta_1} \log \sum_{j=1}^{J} \exp(v_{ij}) - \frac{1}{\beta_1} \log \sum_{j=1}^{J} \exp(\tilde{v}_{ij}).$$

- **Model:** Consider the Conditional Logit model,

$$u_{ij}^* = \mathbf{x}_j \boldsymbol{\beta}_o + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{Extreme Value Type I.}$$

- **Property:** Let $\Lambda(\mathbf{X}) \equiv \sum_{j=1}^{J} \exp(\mathbf{x}_j \beta)$.

$$
\begin{aligned}
\frac{\Pr(y=j)}{\Pr(y=k)} &= \frac{\exp(\mathbf{x}_j \beta)/\Lambda(\mathbf{X})}{\exp(\mathbf{x}_j \beta)/\Lambda(\mathbf{X})} \\
&= \exp\left[(\mathbf{x}_j - \mathbf{x}_k)\beta\right].
\end{aligned}
$$

This is called the IIA assumption.

- **I.e.** The "odds-ratio" (relative prob.) *does not depend on*
  - x for any $h \neq j, k$,
  - Size of the choiceset.
- **Implication:** Restricts substitution patterns...
  ... two examples clarify.
  shows that the highest luxury car will change the market share for the shittiesst car - unrealistic

IIA: Bus Example

- **Assume:** $j = 0$ for (red) bus and $j = 1$ for car.
    - With $\mathbf{x} = (\text{price}, \text{time})$, assume $\mathbf{x}_0 = (1, 2)$, $\mathbf{x}_1 = (2, 1)$.
- **Assume further:** That $\beta_o = (1, 1) \Rightarrow$ then $\Pr(y = 1) = .5$.
- **Since** prob.s sum to 1, $\Pr(y = 0) = .5$.
    - ($*$) Therefore $\Pr(y = 1)/\Pr(y = 0) = $.
- **Now:** Add a new (blue) bus; $j = 2$ with $\mathbf{x}_2 = (1, 2)$.
  [identical to red bus]
    - $\Rightarrow \Pr(y = 2) = $
- **The problem:** But by IIA and unchanged β_o, ($*$) must still hold,
  $\Rightarrow \Pr(y = 1) = $.
- **Implication:** The share choosing car would fall to $\frac{1}{3}$...
- **... counterintuitive:** We added an alternative *identical* to an existing.
    - How can that lure people away from cars?

IIA: Car Example

- **Another example:** Suppose $y \in \{1, ..., J\}$ denotes car type, $\mathbf{x}_j = (\text{price}_j, \text{size}_j)$.

- **Suppose** the price of car J increases.
  - i.e. $x_{J1} := x_{J1} + \Delta$

- **Then** the change in probability (i.e. market share) for *all other* cars, k, are changed **similarly**,

$$\forall k \neq J : \Pr(y = k | \mathbf{X}) = \frac{\exp(\mathbf{x}_k \beta)}{\exp(\beta_1 \cdot \Delta) \exp(\mathbf{x}_J \beta) + \sum_{j \neq J} \exp(\mathbf{x}_j \beta)}.$$

- **Implication:**
  - Car j of course loses some of the market...
  - ... but that loss is divided *similarly* across the remaining cars.
  - ... i.e. not just to the closest substitute.

- **Interpretation:** IIA is a restriction on the possible *substitution patterns*.

Without micro data

- **Model:** Conditional logit.
- **Data:** Aggregate market shares and characteristics of choices.
- **Note:** Expected market share $=$ avg. choice probability (!)

$$s_j \quad \equiv \quad \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\{y = j\},$$

$$s_j \quad \overset{N \to \infty}{\Rightarrow} \quad \Pr(y = j).$$

- **Data:** Time-series information on
    - Characteristics of the available cars at time t, x_{jt},
    - Realized market shares during that year, s_{jt}.

Model

$$s_{jt} = \Pr(y = j | \mathbf{X}_t) = \frac{\exp(\mathbf{x}_{jt}\beta)}{\sum_{k=1}^{J} \exp(\mathbf{x}_{kt}\beta)}.$$

- **Relative** market share of j to l

$$\frac{s_{jt}}{s_{lt}} = \frac{\frac{\exp(\mathbf{x}_{jt}\beta)}{\sum_{k=1}^{J} \exp(\mathbf{x}_{kt}\beta)}}{\frac{\exp(\mathbf{x}_{lt}\beta)}{\sum_{k=1}^{J} \exp(\mathbf{x}_{kt}\beta)}} = \frac{\exp(\mathbf{x}_{jt}\beta)}{\exp(\mathbf{x}_{lt}\beta)}.$$

- **In logs**

$$\log s_{jt} - \log s_{lt} = (x_{jt} - x_{lt})\beta.$$

Regression idea

- **We showed:**
$$\log s_{jt} - \log s_{lt} = (\mathbf{x}_{jt} - \mathbf{x}_{lt})\beta.$$

- **Idea:**
    - $\tilde{y}_{jt} := \log s_{jt} - \log s_{1t}$,
    - $\tilde{\mathbf{x}}_{jt} := \mathbf{x}_{jt} - \mathbf{x}_{1t}$.

- **Regression:** Obtain $\hat{\beta}$ by regressing $\tilde{y}_{jt}$ on $\tilde{\mathbf{x}}_{jt}$ for all $j \neq 1$ and all t.

- **Applications:** Requires many markets with much exogenous variation in $\mathbf{x}$...
    - but a good first approximation.

Full Generality

- **Nested logit:** Some cars are more similar than others.
  - **Think:** "luxury car" nest and "green car" nest.
  - If Volvo introduces a luxury car, it steals more from other luxury cars than from shitty cars.
- **Random parameters logit:** $v_{ij} = x_{ij}\beta_i$, where β_i is heterogeneous and related to x_i.
  - **Think:** Interact x_i and z_j, ...
  - ... but also allows $\mathrm{Cov}(\beta_{ik}, \beta_{il}|x_i) \neq 0$.
  - Allows different consumers to value attributes differently.
  - Called *mixed logit* by everyone other than C&T.
- **Pricing:** Suppose one x is car price.
  - Possible to solve FOCs for optimal price in Bertrand-Nash.
  - ... gives markups of firms.

A 2-period Structural Model

2-period car model

$$u_{ij} = \gamma \left(\text{income} - km \cdot p^{km} \right) + \beta_1 km - \beta_2 km^2 + \varepsilon_{ij}.$$

- **Intuition:** Nicer to drive a more luxurious car, but it costs fuel.
  - Utility of driving is increasing at a diminishing rate (if $\beta_2 > 0$).
- **Problem:** Driving is unknown at time t_1, since fuel prices move.
- **Solution:** Backwards induction.
  - t_1 Buy a car, y.
  - t_2 Fuel prices are realised. Drive the car.

In period t_2

- **Easy:** Solve the FOC,

$$
\begin{aligned}
\frac{\partial u_{ij}}{\partial km} &= 0 \\
-\gamma p^{km} + x_j \beta_1 - 2\beta_2 km &= 0 \\
km^* &= \frac{1}{2\beta_2}(\gamma p^{km} - x_j \beta_1).
\end{aligned}
$$

- Linear demand curve for driving in period t_2.

- **Insert** optimal driving to form new RUM

RUM

$$v_{ij} = \gamma \left[\text{income} - km^* \cdot p^{km} \right] + \beta_1 km^* - \beta_2 \left(km^* \right)^2.$$

- **Criterion:**

$$\ell^{km}(\theta) = \frac{1}{\sigma} \phi \left(\frac{km_i - km^*}{\sigma} \right)$$

$$\Pr(y_i | \theta) = \frac{\exp(v_{ij})}{\sum_{j'} \exp(v_{ij'})}.$$

- **Joint likelihood:**

$$\ell_i(\theta) = \ell^{km}(\theta) + \log \Pr(y_i | x, \theta)$$

Why this model?

- **Endogenous selection:** Households with higher driving demand choose differently.

- **Environmental costs:** Change taxes, compute impacts on
  - Choice margin: people shop around,
  - Driving: end in a new car, drive differently.
  - Rebound effect.

- **Munk-Nielsen (2015):** Fuel tax actually *lowers* revenue:
  - People drive less...
  - ... but buy smaller cars, and registration taxes are much larger.

Alternative 2-period models

- **Heating:** Dubin & McFadden: first buy airconditioner, then use it.
  - Endogenous selection on usage.
- **Insurance:** First buy a plan; then utilize it.
  - Endogenous selection based on known "needs" (e.g. exercise).
- **Retirement:** First choose retirement age, then consumption afterwards; the two are obviously linked.

Structural models

- **Often:** Full life-cycle considered.
- **Solution method:** Backwards induction
    - [same as for the 2-period model]
- **Relevance:** Counterfactuals that are far from status quo.
- **Best practice:** Structural + reduced form.
    - Estimate structural model up until a reform occurs,
    - Predict response out of sample,
    - Compare with behavioral response.

Outline

Deriving the likelihood: A few more steps

Choice probability:

$$
\begin{aligned}
\Pr(y = j | z) &= \Pr(u_{ij} > u_{ik} \, \forall k \neq j | z) \\
&= \Pr(x_j \beta_o + \varepsilon_{ij} > x_k \beta_o + \varepsilon_{ik} \, \forall k \neq j | z) \\
&= \Pr\left[\varepsilon_{ik} < (x_j - x_k)\beta_o + \varepsilon_{ij} \, \forall k \neq j | z\right] \\
&= \mathbb{E}_{\varepsilon_{ij}} \left\{ \Pr\left[\varepsilon_{ik} < (x_j - x_k)\beta_o + \varepsilon_{ij} \, \forall k \neq j | z, \varepsilon_{ij}\right] \right\} \\
&= \mathbb{E}_{\varepsilon_{ij}} \left\{ \prod_{k \neq j} G\left[(x_j - x_k)\beta_o + \varepsilon_{ij}\right] \right\} \\
&= \int \left\{ \prod_{k \neq j} G\left[(x_j - x_k)\beta_o + \epsilon\right] \right\} e^{-\epsilon} e^{-e^{-\epsilon}} \, \mathrm{d}\epsilon \\
&= \cdots \\
&= \frac{\exp(x_j \beta_o)}{\sum_{k=0}^{J} \exp(x_k \beta_o)}.
\end{aligned}
$$