Week 7: *Utility Models*

EMSE 6035: Marketing Analytics for Design Decisions

John Paul Helveston

October 12, 2022

Week 7: *Utility Models*

1. Utility models

2. Exploring choice data

3. Linear & discrete parameters

BREAK

4. Outside good

5. Team project utility models

Week 7: *Utility Models*

1. Utility models

2. Exploring choice data

3. Linear & discrete parameters

BREAK

4. Outside good

5. Team project utility models

Random utility model

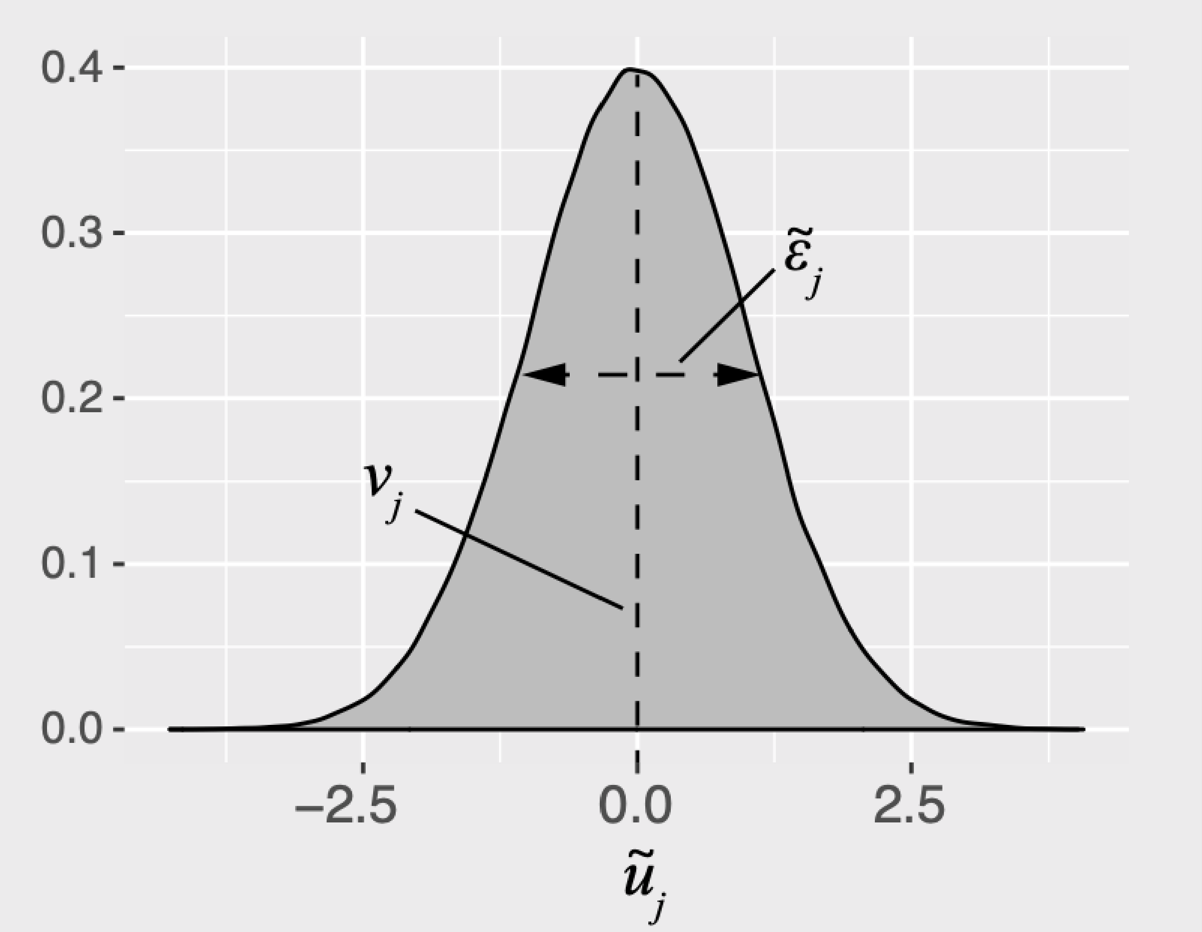
The utility for alternative j is

$$\tilde{u}_j = v_j + \tilde{\varepsilon}_j$$

v_j = Things we observe (non-random variables)

$\tilde{\varepsilon}_j$ = Things we *don't* observe (random variable)

$$\tilde{u}_j = v_j + \tilde{\varepsilon}_j$$



Practice Question 1

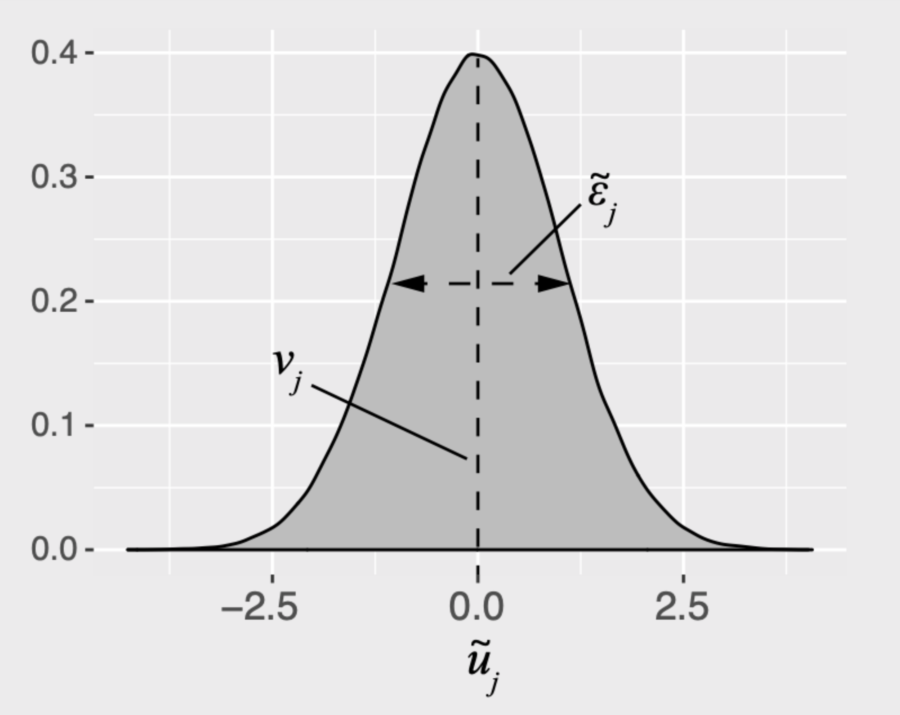
a) A random variable, $\tilde{x}$, has the PDF, $f_{\tilde{x}}(x)$. Write the equation to compute its total probability (hint: think area under the curve!). What is the answer to the equation?

b) A random variable, $\tilde{x}$, has a uniform distribution between the values 0 and 1. Draw the probability density function (PDF) and Cumulative Density Function (CDF) of $\tilde{x}$.

c) The value of a random variable, $\tilde{x}$, is determined by rolling one fair, 6-sided dice. Draw the PDF and CDF of $\tilde{x}$.

Logit model: Assume that $\tilde{\varepsilon}_j$ ~ Gumbel Distribution

$$\tilde{u}_j = v_j + \tilde{\varepsilon}_j$$



Probability of choosing alternative j:

$$P_j = \frac{e^{v_j}}{\sum_k e^{v_k}}$$

Practice Question 2

a) A consumer is making a choice between two bars of chocolate:

- Milk chocolate (m)
- Dark chocolate (d)

Assume that we know the observed utility of each bar to be $v_m = 3$ and $v_d = 4$. Using a logit model, compute the probabilities of choosing each bar: P_m and P_d.

b) A third bar of chocolate is now added to the choice set. It is the exact same as the milk chocolate bar, but it has a slightly different wrapper (which has no effect on the consumer's utility). Now, $v_{m1} = v_{m2} = 3$, and $v_d = 4$. Based on the probabilities from question a), what would we expect the probabilities of choosing each bar to be? What probabilities does the logit model produce?

"Observed utility" (v_j) is a weighted sum of attribute values

$$v_j = \beta_1 x_j^{\mathrm{A}} + \beta_2 x_j^{\mathrm{B}} + \dots$$

Each x_j is an observable attribute (*price*, etc.)

We know $x_j^{\mathrm{A}}, x_j^{\mathrm{B}}, \dots$,

we want to *estimate* $\beta_1, \beta_2, \dots$

Notation Convention

Continuous: x_j

$$u_j = \beta_1 x_j^{\text{price}} + \dots$$

```
#>   price
#> 1     1
#> 2     2
#> 3     3
```

Discrete: δ_j

$$u_j = \beta_1 \delta_j^{\text{ford}} + \beta_2 \delta_j^{\text{gm}} \dots$$

```
#>   brand brand_BMW brand_Ford brand_GM
#> 1  Ford         0          1        0
#> 2    GM         0          0        1
#> 3   BMW         1          0        0
```

Practice Question 3

Attribute	Bar 1	Bar 2	Bar 3
Price	$1.20	$1.50	$3.00
% Cacao	10%	60%	80%

a) Write out a model for the *observed* utility of each chocolate bar in the above set.

b) If the coefficient for the *price* attribute was -0.1 and the coefficient for *% Cacao* attribute was 0.1, what is the difference in the observed utility between bars 3 and 1?

c) With the addition of the *brand* attribute, repeat part a.

Attribute	Bar 1	Bar 2	Bar 3
Price	$1.20	$1.50	$3.00
% Cacao	10%	60%	80%
Brand	Hershey	Lindt	Ghirardelli

Your Turn

20:00

Let's say our utility function is:

$$v_j = \beta_1 x_j^{\text{price}} + \beta_2 x_j^{\text{cacao}} + \beta_3 \delta_j^{\text{hershey}} + \beta_4 \delta_j^{\text{lindt}}$$

And we estimate the following coefficients:

Parameter	Coefficient
β_1	-0.1
β_2	0.1
β_3	-2.0
β_4	-0.1

a) What are the expected probabilities of choosing each of these bars using a logit model?

Attribute	Bar 1	Bar 2	Bar 3
Price	\$1.20	\$1.50	\$3.00
% Cacao	10%	60%	80%
Brand	Hershey	Lindt	Ghirardelli

b) What price would Bar 2 have to be to get a 50% market share?

Week 7: *Utility Models*

1. Utility models

2. Exploring choice data

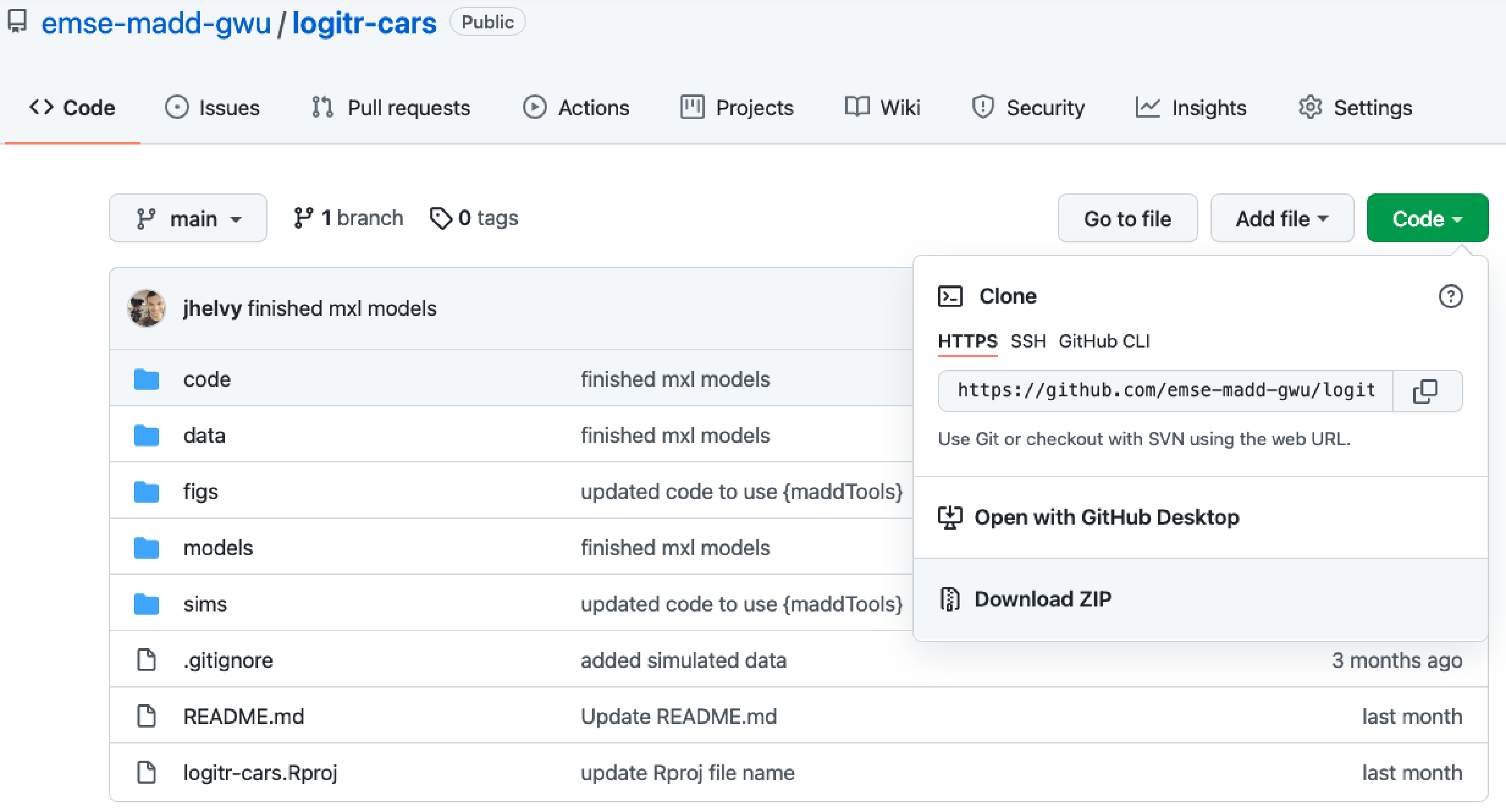
3. Linear & discrete parameters

BREAK

4. Outside good

5. Team project utility models

Download the logitr-cars repo from GitHub



Exploring choice data

1. Open `logitr-cars.Rproj`
2. Open `code/2.1-explore-data.R`

Week 7: *Utility Models*

1. Utility models
2. Exploring choice data
3. Linear & discrete parameters

BREAK

4. Outside good
5. Team project utility models

Dummy-coded variables

Dummy coding: 1 = "Yes", 0 = "No"

Data frame with one variable: *price*

```r
data <- data.frame(price = c(10, 20, 30))

data
```

```
#>   price
#> 1    10
#> 2    20
#> 3    30
```

Add dummy columns for each price "level"

```r
library(fastDummies)

dummy_cols(data, "price")
```

```
#>   price price_10 price_20 price_30
#> 1    10        1        0        0
#> 2    20        0        1        0
#> 3    30        0        0        1
```

Model *price* as *continuous*

$$v_j = \beta_1 x^{\text{price}}$$

```r
model <- logitr(
    data   = data,
    choice = "choice",
    obsID  = "obsID",
    pars   = "price"
)
```

Coef.	Interpretation
β1	how utility changes with increasing *price*

Model *price* as *discrete*

$$v_j = \beta_1 \delta^{\text{price}=20} + \beta_2 \delta^{\text{price}=30}$$

```r
model <- logitr(
    data   = data,
    choice = "choice",
    obsID  = "obsID",
    pars   = c("price_20", "price_30")
)
```

Reference level: *price=10*

Coef.	Interpretation
β1	utility for *price=20* relative to *price=10*
β2	utility for *price=30* relative to *price=10*

Estimating utility models

1. Open `logitr-cars.Rproj`
2. Open `code/3.1-model-mnl.R`

mnl_dummy

All dummy-code variables

```
pars = c(
  "price_20", "price_25",
  "fuelEconomy_25", "fuelEconomy_30",
  "accelTime_7", "accelTime_8",
  "powertrain_Electric")
```

Reference Levels:

- Price: 15
- Fuel Economy: 20
- Accel. Time: 6
- Powertrain: "Gasoline"

mnl_linear

All continuous (linear), except for `powertrain_Electric`

```
pars = c(
  'price', 'fuelEconomy', 'accelTime',
  'powertrain_Electric')
```

Reference Levels:

- Powertrain: "Gasoline"

Your Turn

20:00

1) Run the code chunk to read in the `data.csv` file in the "data" folder, which contains choice observations from chocolate bars with the following attributes:

Attribute	Description
`price`	Price in $
`percent_cacao`	% Cacao (how "dark" the chocolate is)
`crispy_rice`	0 or 1 for if the bar contains crispy rice
`brand`	"Hershey", "Lindt", or "Ghirardelli"

2) Write code to estimate the following utility model
(HINT: you may need to make some dummy-coded variables!):

$$u_j = \beta_1 x_j^{\text{price}} + \beta_2 x_j^{\%\text{cacao}} + \beta_3 \delta_j^{\text{crispy}} + \beta_4 \delta_j^{\text{hershey}} + \beta_5 \delta_j^{\text{lindt}} + \varepsilon_j$$

3) Write code to plot the change in utility for the *price* attribute.

Week 7: *Utility Models*

1. Utility models
2. Exploring choice data
3. Linear & discrete parameters

BREAK

4. Outside good
5. Team project utility models

Estimating utility models with an *Outside Good*

1. Open `logitr-cars.Rproj`
2. Open `code/4.1-model-og.R`

Week 7: *Utility Models*

1. Utility models

2. Exploring choice data

3. Linear & discrete parameters

BREAK

4. Outside good

5. Team project utility models

Simulating choice data

Random choices

```r
data <- simulateChoices(
  survey,
  altID = "altID",
  obsID = "obsID"
)
```

Choices according to assumed model

$$v_j = -0.1 x_j^{\text{price}} + 0.1 x_j^{\text{fuelEconomy}} + 0.1 x_j^{\text{accelTime}} - 4 \delta_j^{\text{electric}}$$

```r
data <- simulateChoices(
  survey,
  altID = "altID",
  obsID = "obsID",
  pars = list(
    price = -0.1,
    fuelEconomy = 0.1,
    accelTime = 0.1,
    powertrain_Electric = -4
  )
)
```

Estimate a choice model

$$v_j = \beta_1 x_j^{\text{price}} + \beta_2 x_j^{\text{fuelEconomy}} + \beta_3 x_j^{\text{accelTime}} + \beta_4 \delta_j^{\text{electric}}$$

```r
model <- logitr(
  data   = data,
  choice = "choice",
  obsID  = "obsID",
  pars   = c("price", "fuelEconomy", "accelTime", "powertrain_Electric")
)
```

Your Turn

20:00

As a team:

1. Go back to your code from last week where you created your choice questions.
2. Write out a utility model for your project.
3. Write code to simulate data according to your utility model - pick some fake parameter values.
4. Write code to estimate a model using your simulated data.