Binary choice

Model specification: the error term

Virginie Lurkin

Introduction to choice models

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Binary choice model

Two alternatives: $\mathcal{C}_n = \{i, j\}$

$$\begin{aligned} U_{in} &= V_{in} + \varepsilon_{in} \\ U_{jn} &= V_{jn} + \varepsilon_{jn} \end{aligned}$$

Choice model

$$P_n(i|\{i,j\}) = \Pr(U_{in} \geq U_{jn})$$

Binary choice model

Two alternatives: $\mathcal{C}_n = \{i, j\}$

$$\begin{aligned} U_{in} &= V_{in} + \varepsilon_{in} \\ U_{jn} &= V_{jn} + \varepsilon_{jn} \end{aligned}$$

Choice model

$$\begin{aligned} P_n(i|\{i,j\}) &= \Pr(U_{in} \geq U_{jn}) \\ &= \Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}) \end{aligned}$$

Binary choice model

Two alternatives: $\mathcal{C}_n = \{i, j\}$

$$\begin{aligned} U_{in} &= V_{in} + \varepsilon_{in} \\ U_{jn} &= V_{jn} + \varepsilon_{jn} \end{aligned}$$

Choice model

$$\begin{aligned} P_n(i|\{i,j\}) &= \Pr(U_{in} \geq U_{jn}) \\ &= \Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}) \\ &= \Pr(V_{in} - V_{jn} \geq \varepsilon_{jn} - \varepsilon_{in}) \end{aligned}$$

Binary choice model

Two alternatives: $\mathcal{C}_n = \{i, j\}$

$$
\begin{aligned}
U_{in} &= V_{in} + \varepsilon_{in} \\
U_{jn} &= V_{jn} + \varepsilon_{jn}
\end{aligned}
$$

Choice model

$$
\begin{aligned}
P_n(i|\{i,j\}) &= \Pr(U_{in} \geq U_{jn}) \\
&= \Pr(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}) \\
&= \Pr(V_{in} - V_{jn} \geq \varepsilon_{jn} - \varepsilon_{in}) \\
&= \Pr\left(\varepsilon_n \leq V_{in} - V_{jn}\right)
\end{aligned}
$$

where $\varepsilon_n = \varepsilon_{jn} - \varepsilon_{in}$.

Error term

Three assumptions about the random variables ε_{in} and ε_{jn}

1. What's their mean?
2. What's their variance?
3. What's their distribution?

Note

- For binary choice, it would be sufficient to make assumptions about $\varepsilon_n = \varepsilon_{jn} - \varepsilon_{in}$.
- But we want to generalize later on.

The mean

Change of variables

- Define $\mathrm{E}[\varepsilon_{in}] = \beta_{in}$ and $\mathrm{E}[\varepsilon_{jn}] = \beta_{jn}$.
- Define $\varepsilon'_{in} = \varepsilon_{in} - \beta_{in}$ and $\varepsilon'_{jn} = \varepsilon_{jn} - \beta_{jn}$,
- so that $\mathrm{E}[\varepsilon'_{in}] = \mathrm{E}[\varepsilon'_{jn}] = 0$.

Choice model

$P_n(i|\{i,j\}) =$

$$\Pr(V_{in} - V_{jn} \geq \varepsilon_{jn} - \varepsilon_{in})$$

The mean

Change of variables

- Define $\mathrm{E}[\varepsilon_{in}] = \beta_{in}$ and $\mathrm{E}[\varepsilon_{jn}] = \beta_{jn}$.
- Define $\varepsilon'_{in} = \varepsilon_{in} - \beta_{in}$ and $\varepsilon'_{jn} = \varepsilon_{jn} - \beta_{jn}$,
- so that $\mathrm{E}[\varepsilon'_{in}] = \mathrm{E}[\varepsilon'_{jn}] = 0$.

Choice model

$P_n(i|\{i,j\}) =$

$$\Pr(V_{in} - V_{jn} \geq \varepsilon_{jn} - \varepsilon_{in}) =$$
$$\Pr(V_{in} - V_{jn} \geq \varepsilon'_{jn} + \beta_{jn} - (\varepsilon'_{in} + \beta_{in}))$$

The mean

Change of variables

- Define $\mathrm{E}[\varepsilon_{in}] = \beta_{in}$ and $\mathrm{E}[\varepsilon_{jn}] = \beta_{jn}$.
- Define $\varepsilon'_{in} = \varepsilon_{in} - \beta_{in}$ and $\varepsilon'_{jn} = \varepsilon_{jn} - \beta_{jn}$,
- so that $\mathrm{E}[\varepsilon'_{in}] = \mathrm{E}[\varepsilon'_{jn}] = 0$.

Choice model

$P_n(i|\{i,j\}) =$

$$\begin{aligned}
&\Pr(V_{in} - V_{jn} \geq \varepsilon_{jn} - \varepsilon_{in}) = \\
&\Pr(V_{in} - V_{jn} \geq \varepsilon'_{jn} + \beta_{jn} - (\varepsilon'_{in} + \beta_{in})) = \\
&\Pr(V_{in} + \beta_{in} - (V_{jn} + \beta_{jn}) \geq \varepsilon'_{jn} - \varepsilon'_{in})
\end{aligned}$$

The mean

Change of variables

- Define $E[\varepsilon_{in}] = \beta_{in}$ and $E[\varepsilon_{jn}] = \beta_{jn}$.
- Define $\varepsilon'_{in} = \varepsilon_{in} - \beta_{in}$ and $\varepsilon'_{jn} = \varepsilon_{jn} - \beta_{jn}$,
- so that $E[\varepsilon'_{in}] = E[\varepsilon'_{jn}] = 0$.

Choice model

$$
\begin{aligned}
P_n(i|\{i,j\}) = \\
& \Pr(V_{in} - V_{jn} \geq \varepsilon_{jn} - \varepsilon_{in}) = \\
& \Pr(V_{in} - V_{jn} \geq \varepsilon'_{jn} + \beta_{jn} - (\varepsilon'_{in} + \beta_{in})) = \\
& \Pr(V_{in} + \beta_{in} - (V_{jn} + \beta_{jn}) \geq \varepsilon'_{jn} - \varepsilon'_{in}) = \\
& \Pr(V_{in} + \beta_{in} - (V_{jn} + \beta_{jn}) \geq \varepsilon'_n)
\end{aligned}
$$

where $\varepsilon'_n = \varepsilon'_{jn} - \varepsilon'_{in}$.

Alternative specific constant

- The mean of each error term can be moved to the deterministic part.
- It is captured by a parameter, to be estimated from data.
- It is called the alternative specific constant (ASC).
- Only the mean of the difference of the error terms is identified.

Shift invariance

The choice model is not affected by a uniform shift of all utility functions

$$P_n(i|\{i,j\}) = \Pr(U_{in} \geq U_{jn}) = \Pr(U_{in} + K \geq U_{jn} + K) \quad \forall K.$$

Alternative Specific Constant

Many equivalent specifications

$$
\begin{array}{rcll}
U_{in} & = & V_{in} + \beta_{in} & +\varepsilon'_{in} \\
U_{jn} & = & V_{jn} + \beta_{jn} & +\varepsilon'_{jn} \\
 & & \text{or} & \\
U_{in} & = & V_{in} & +\varepsilon'_{in} \\
U_{jn} & = & V_{jn} + \beta_{jn} - \beta_{in} & +\varepsilon'_{jn} \\
 & & \text{or} & \\
U_{in} & = & V_{in} + \beta_{in} - \beta_{jn} & +\varepsilon'_{in} \\
U_{jn} & = & V_{jn} & +\varepsilon'_{jn}
\end{array}
$$

In practice

Normalize one constant to zero and estimate $\beta_n = \beta_{jn} - \beta_{in}$

Scale invariance

The choice model is not affected by a uniform scaling of all utility functions

$$P_n(i|\{i,j\}) = \Pr(U_{in} \geq U_{jn}) = \Pr(\alpha U_{in} \geq \alpha U_{jn}) \quad \forall \alpha > 0.$$

The variance

$$\begin{aligned}\mathrm{Var}(\alpha U_{in}) &= \alpha^2 \mathrm{Var}(U_{in}) \\ \mathrm{Var}(\alpha U_{jn}) &= \alpha^2 \mathrm{Var}(U_{jn})\end{aligned}$$

The variance is not identified

- As any α can be selected arbitrarily, any variance can be assumed.
- No way to identify the variance of the error terms from data.
- The scale has to be arbitrarily decided.

In practice

The scale parameter of the assumed distribution is normalized to 1.

The distribution

Assumption

ε_{in} and ε_{jn} are the maximum of many r.v. capturing unobservable attributes (e.g. mood, experience), measurement and specification errors.

Gumbel theorem

The maximum of many i.i.d. random variables approximately follows an Extreme Value distribution: $\text{EV}(\eta, \mu)$, with $\mu > 0$.

The Extreme Value distribution EV(η, μ)

Probability density function (pdf)

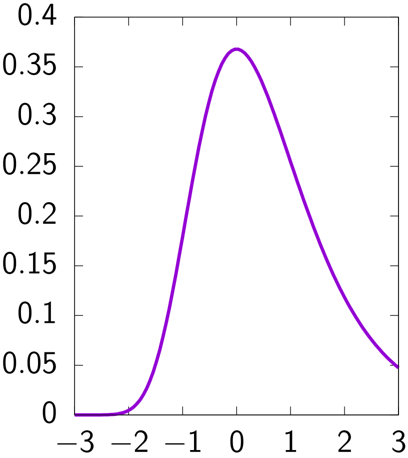
$$f(t) = \mu e^{-\mu(t-\eta)} e^{-e^{-\mu(t-\eta)}}$$

Cumulative distribution function (CDF)

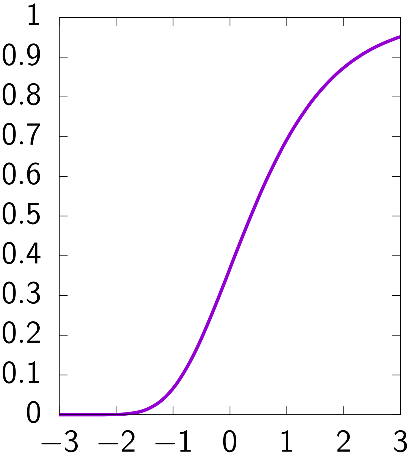
$$\begin{aligned} P(c \geq \varepsilon) = F(c) &= \int_{-\infty}^{c} f(t)dt \\ &= e^{-e^{-\mu(c-\eta)}} \end{aligned}$$

The Extreme Value distribution

pdf EV(0,1)



CDF EV(0,1)



The Extreme Value distribution

Properties

If

$$\varepsilon \sim \mathsf{EV}(\eta, \mu)$$

then

$$\mathsf{E}[\varepsilon] = \eta + \frac{\gamma}{\mu} \quad \text{and} \quad \mathsf{Var}[\varepsilon] = \frac{\pi^2}{6\mu^2}$$

where γ is Euler's constant.

Euler's constant

$$\gamma = -\int_0^\infty e^{-x} \ln x \, dx \approx 0.5772$$

The distribution

Assumptions

- ε_{in} and ε_{jn} are i.i.d. Extreme Value.
- If an alternative specific constant is in the model, their mean can be assumed to be any constant.
- It is convenient to set the location parameter to 0, so that $\mathrm{E}[\varepsilon_{in}] = \mathrm{E}[\varepsilon_{jn}] = \gamma/\mu$.

Distributions

$$\varepsilon_{in} \sim EV(0,\mu), \qquad \varepsilon_{jn} \sim EV(0,\mu)$$

Problem

We need the distribution of $\varepsilon_n = \varepsilon_{jn} - \varepsilon_{in}$

Logistic distribution

From the properties of the extreme value distribution, we have

$$\begin{aligned} \varepsilon_{in} &\sim \text{EV}(0, \mu) \\ \varepsilon_{jn} &\sim \text{EV}(0, \mu) \\ \varepsilon_n = \varepsilon_{jn} - \varepsilon_{in} &\sim \text{Logistic}(0, \mu) \end{aligned}$$

The Logistic distribution: Logistic(η,μ)

Probability density function (pdf)

$$f(t) = \frac{\mu e^{-\mu(t-\eta)}}{(1 + e^{-\mu(t-\eta)})^2}$$

Cumulative distribution function (CDF)

$$P(c \geq \varepsilon) = F(c) = \int_{-\infty}^{c} f(t)dt = \frac{1}{1 + e^{-\mu(c-\eta)}}$$

with $\mu > 0$.

The binary logit model

Choice model

$$P_n(i|\{i,j\}) = \Pr(\varepsilon_n \leq V_{in} - V_{jn}) = F_\varepsilon(V_{in} - V_{jn})$$

The binary logit model

$$P_n(i|\{i,j\}) = \frac{1}{1 + e^{-\mu(V_{in} - V_{jn})}} = \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + e^{\mu V_{jn}}}$$