Nested logit models

Michel Bierlaire

michel.bierlaire@epfl.ch

Transport and Mobility Laboratory

Red bus/Blue bus paradox

- Mode choice example
- Two alternatives: car and bus
- There are red buses and blue buses
- Car and bus travel times are equal: T

Red bus/Blue bus paradox

Model 1

$$
\begin{aligned}
U_{\text{car}} &= \beta T + \varepsilon_{\text{car}} \\
U_{\text{bus}} &= \beta T + \varepsilon_{\text{bus}}
\end{aligned}
$$

Therefore,

$$
P(\text{car}|\{\text{car}, \text{bus}\}) = P(\text{bus}|\{\text{car}, \text{bus}\}) = \frac{e^{\beta T}}{e^{\beta T} + e^{\beta T}} = \frac{1}{2}
$$

Red bus/Blue bus paradox

Model 2

$$\begin{aligned} U_{\text{car}} &= \beta T + \varepsilon_{\text{car}} \\ U_{\text{blue bus}} &= \beta T + \varepsilon_{\text{blue bus}} \\ U_{\text{red bus}} &= \beta T + \varepsilon_{\text{red bus}} \end{aligned}$$

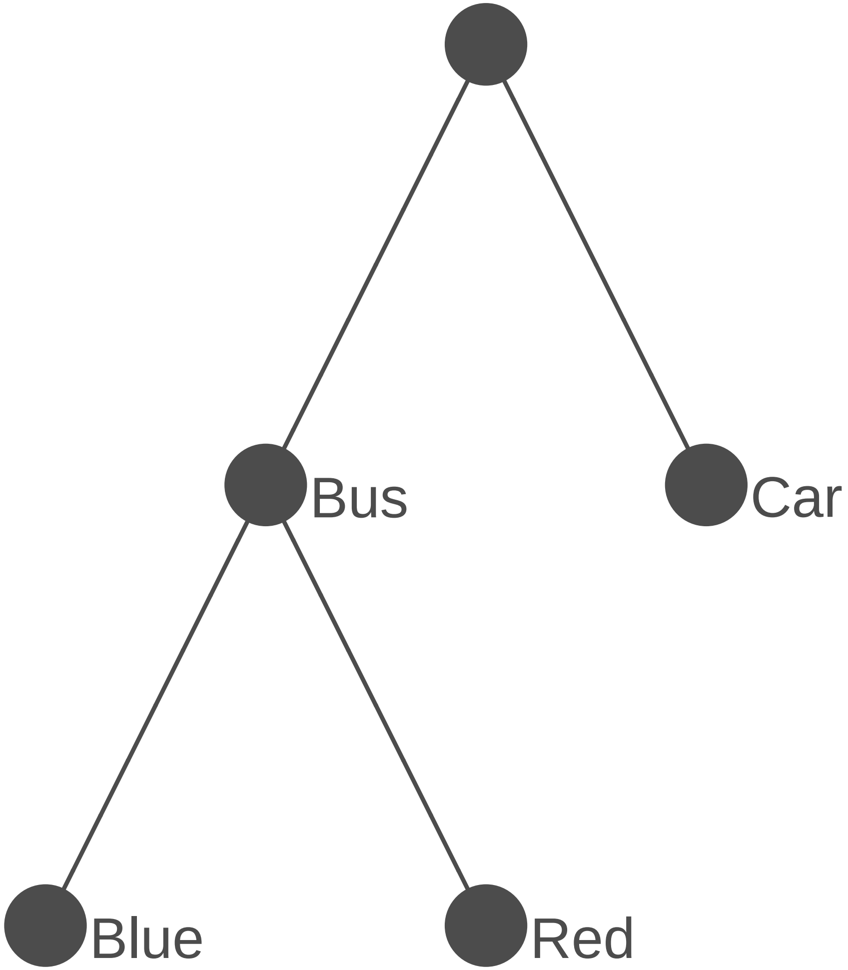
$$P(\text{car}|\{\text{car}, \text{blue bus}, \text{red bus}\}) = \frac{e^{\beta T}}{e^{\beta T} + e^{\beta T} + e^{\beta T}} = \frac{1}{3}$$

$$\left.\begin{array}{r} P(\text{car}|\{\text{car}, \text{blue bus}, \text{red bus}\}) \\ P(\text{blue bus}|\{\text{car}, \text{blue bus}, \text{red bus}\}) \\ P(\text{red bus}|\{\text{car}, \text{blue bus}, \text{red bus}\}) \end{array}\right\} = \frac{1}{3}.$$

Red bus/Blue bus paradox

- Assumption of logit: ε i.i.d
- $\varepsilon_{\text{blue bus}}$ and $\varepsilon_{\text{red bus}}$ contain common unobserved attributes:
  - fare
  - headway
  - comfort
  - convenience
  - etc.

Capturing the correlation



Capturing the correlation

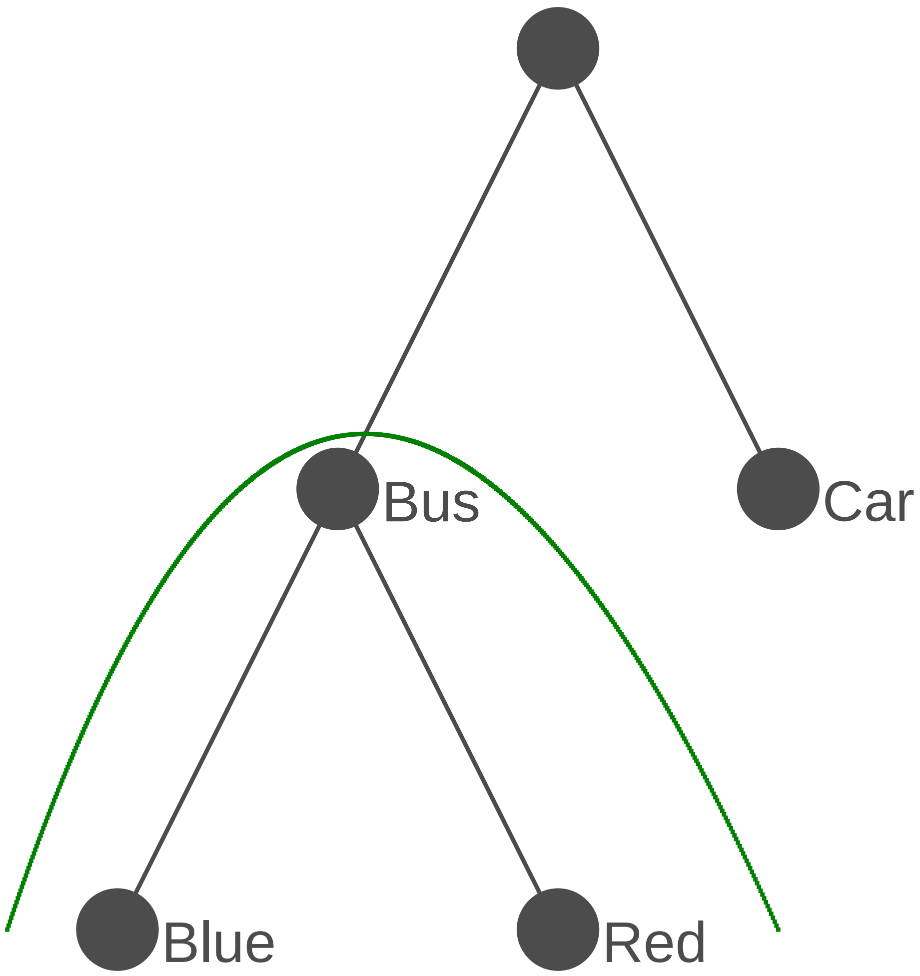
If bus is chosen then

$$\begin{aligned} U_{\text{blue bus}} &= V_{\text{blue bus}} + \varepsilon_{\text{blue bus}} \\ U_{\text{red bus}} &= V_{\text{red bus}} + \varepsilon_{\text{red bus}} \end{aligned}$$

where $V_{\text{blue bus}} = V_{\text{red bus}} = \beta T$

$$P(\text{blue bus}|\{\text{blue bus}, \text{red bus}\}) = \frac{e^{\beta T}}{e^{\beta T} + e^{\beta T}} = \frac{1}{2}$$

Capturing the correlation



Capturing the correlation

What about the choice between bus and car?

$$\begin{aligned} U_{\text{car}} &= \beta T + \varepsilon_{\text{car}} \\ U_{\text{bus}} &= V_{\text{bus}} + \varepsilon_{\text{bus}} \end{aligned}$$

with

$$\begin{aligned} V_{\text{bus}} &= V_{\text{bus}}(V_{\text{blue bus}}, V_{\text{red bus}}) \\ \varepsilon_{\text{bus}} &= ? \end{aligned}$$

Define V_{bus} as the expected maximum utility of red bus and blue bus

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Expected maximum utility

For a set of alternative $\mathcal{C}$, define

$$U_{\mathcal{C}} = \max_{i \in \mathcal{C}} U_i = \max_{i \in \mathcal{C}} (V_i + \varepsilon_i)$$

and

$$V_{\mathcal{C}} = E[U_{\mathcal{C}}]$$

For logit

$$E[\max_{i \in \mathcal{C}} U_i] = \frac{1}{\mu} \ln \sum_{i \in \mathcal{C}} e^{\mu V_i} + \frac{\gamma}{\mu}$$

Expected maximum utility

$$
\begin{aligned}
V_{\text{bus}} &= \frac{1}{\mu_b} \ln(e^{\mu_b V_{\text{blue bus}}} + e^{\mu_b V_{\text{red bus}}}) \\
&= \frac{1}{\mu_b} \ln(e^{\mu_b \beta T} + e^{\mu_b \beta T}) \\
&= \beta T + \frac{1}{\mu_b} \ln 2
\end{aligned}
$$

where μ_b is the scale parameter for the logit model associated with the choice between red bus and blue bus

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Nested Logit Model

Probability model:

$$P(\text{car}) = \frac{e^{\mu V_{\text{car}}}}{e^{\mu V_{\text{car}}} + e^{\mu V_{\text{bus}}}} = \frac{e^{\mu \beta T}}{e^{\mu \beta T} + e^{\mu \beta T + \frac{\mu}{\mu_b} \ln 2}} = \frac{1}{1 + 2^{\frac{\mu}{\mu_b}}}$$

If $\mu = \mu_b$, then P(car) = $\frac{1}{3}$ (Model 2)

If $\mu_b \to \infty$, then $\frac{\mu}{\mu_b} \to 0$, and P(car) $\to \frac{1}{2}$ (Model 1)

Note for $\mu_b \to \infty$

$$e^{\mu V_{\text{bus}}} = \frac{1}{2} e^{\mu V_{\text{red bus}}} + \frac{1}{2} e^{\mu V_{\text{blue bus}}}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

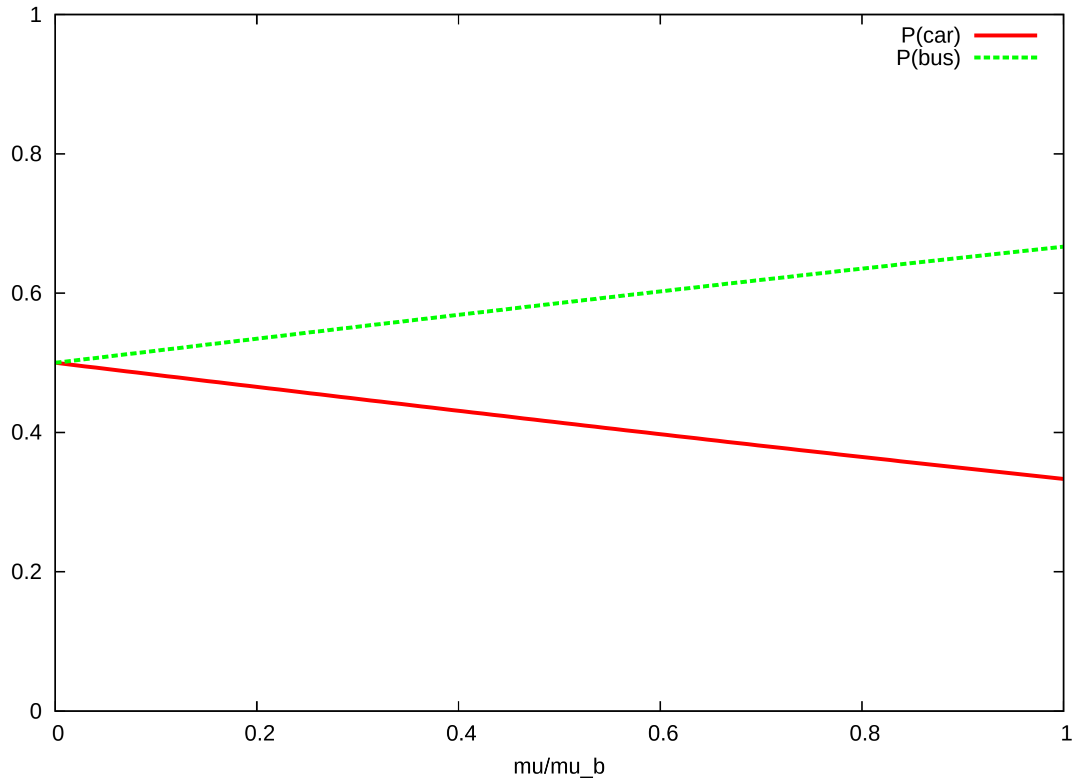
Nested Logit Model

Probability model:

$$P(\text{bus}) = \frac{e^{\mu V_{\text{bus}}}}{e^{\mu V_{\text{car}}} + e^{\mu V_{\text{bus}}}} = \frac{e^{\mu\beta T + \frac{\mu}{\mu_b}\ln 2}}{e^{\mu\beta T} + e^{\mu\beta T + \frac{\mu}{\mu_b}\ln 2}} = \frac{1}{1 + 2^{-\frac{\mu}{\mu_b}}}$$

If $\mu = \mu_b$, then P(bus) = $\frac{2}{3}$ (Model 2)
If $\frac{\mu}{\mu_b} \to 0$, then P(bus) $\to \frac{1}{2}$ (Model 1)

Nested Logit Model



$$\frac{\mu}{\mu_b}$$

Solving the paradox

If $\frac{\mu}{\mu_b} \rightarrow 0$, we have

$$\begin{aligned}
P(\text{car}) &= & & & 1/2 \\
P(\text{bus}) &= & & & 1/2 \\
P(\text{red bus}|\text{bus}) &= & & & 1/2 \\
P(\text{blue bus}|\text{bus}) &= & & & 1/2 \\
P(\text{red bus}) &= & P(\text{red bus}|\text{bus})P(\text{bus}) & = & 1/4 \\
P(\text{blue bus}) &= & P(\text{blue bus}|\text{bus})P(\text{bus}) & = & 1/4
\end{aligned}$$

Comments

- A group of similar alternatives is called a nest
- Each alternative belongs to exactly one nest
- The model is named Nested Logit
- The ratio μ/μ_b must be estimated from the data
- $0 < \mu/\mu_b \leq 1$ (between models 1 and 2)

Derivation from random utility

- Let $\mathcal{C}$ be the choice set.
- Let $\mathcal{C}_1, \ldots, C_M$ be a partition of $\mathcal{C}$.
- The model is derived as

$$P(i|\mathcal{C}) = \sum_{m=1}^{M} \Pr(i|m, \mathcal{C}) \Pr(m|\mathcal{C}).$$

- Each i belongs to exactly one nest m.

$$P(i|\mathcal{C}) = \Pr(i|m) \Pr(m|\mathcal{C}).$$

- Utility: error components

$$U_i = V_i + \varepsilon_i = V_i + \varepsilon_m + \varepsilon_{im}.$$

Derivation: $\Pr(i|m)$

$$
\begin{aligned}
\Pr(i|m) &= \Pr(U_i \geq U_j, j \in \mathcal{C}_m) \\
&= \Pr(V_i + \varepsilon_m + \varepsilon_{im} \geq V_j + \varepsilon_m + \varepsilon_{jm}, j \in \mathcal{C}_m) \\
&= \Pr(V_i + \varepsilon_{im} \geq V_j + \varepsilon_{jm}, j \in \mathcal{C}_m)
\end{aligned}
$$

Assumption: ε_{im} i.i.d. EV$(0, \mu_m)$

$$
\Pr(i|m) = \frac{e^{\mu_m V_i}}{\sum_{j \in \mathcal{C}_m} e^{\mu_m V_j}}.
$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Derivation: $\Pr(m|\mathcal{C})$

$$\Pr(m|\mathcal{C}) = \Pr\left(\max_{i\in\mathcal{C}_m} U_i \geq \max_{j\in\mathcal{C}_\ell} U_j, \forall \ell \neq m\right)$$

$$= \Pr\left(\varepsilon_m + \max_{i\in\mathcal{C}_m}(V_i + \varepsilon_{im}) \geq \varepsilon_\ell + \max_{j\in\mathcal{C}_\ell}(V_j + \varepsilon_{j\ell}), \forall \ell \neq m\right),$$

As ε_{im} are i.i.d. EV$(0, \mu_m)$,

$$\max_{i\in\mathcal{C}_m}(V_i + \varepsilon_{im}) \sim \text{EV}(\tilde{V}_m, \mu_m),$$

where

$$\tilde{V}_m = \frac{1}{\mu_m} \ln \sum_{i\in\mathcal{C}_m} e^{\mu_m V_i}.$$

Derivation: $\Pr(m|\mathcal{C})$

Denote

$$\max_{i\in\mathcal{C}_m}(V_i+\varepsilon_{im})=\tilde{V}_m+\varepsilon'_m,$$

to obtain

$$\Pr(m|\mathcal{C})=\Pr(\tilde{V}_m+\varepsilon'_m+\varepsilon_m\geq\tilde{V}_\ell+\varepsilon'_\ell+\varepsilon_\ell,\forall\ell\neq m).$$

where

$$\varepsilon'_m\sim \text{EV}(0,\mu_m).$$

Define

$$\tilde{\varepsilon}_m=\varepsilon'_m+\varepsilon_m,$$

to obtain

$$\Pr(m|\mathcal{C})=\Pr(\tilde{V}_m+\tilde{\varepsilon}_m\geq\tilde{V}_\ell+\tilde{\varepsilon}_\ell,\forall\ell\neq m).$$

Derivation: $\Pr(m|\mathcal{C})$

Assumption: $\tilde{\varepsilon}_m$ i.i.d. EV$(0, \mu)$

$$\Pr(m|\mathcal{C}) = \Pr(\tilde{V}_m + \tilde{\varepsilon}_m \geq \tilde{V}_\ell + \tilde{\varepsilon}_\ell, \forall \ell \neq m)$$

$$= \frac{e^{\mu \tilde{V}_m}}{\sum_{p=1}^{M} e^{\mu \tilde{V}_p}}.$$

We obtain the nested logit model

$$P(i|\mathcal{C}) = \frac{e^{\mu_m V_i}}{\sum_{j \in \mathcal{C}_m} e^{\mu_m V_j}} \frac{e^{\mu \tilde{V}_m}}{\sum_{p=1}^{M} e^{\mu \tilde{V}_p}}$$

$$= \frac{e^{\mu_m V_i}}{\sum_{j \in \mathcal{C}_m} e^{\mu_m V_j}} \frac{\exp\left(\frac{\mu}{\mu_m} \ln \sum_{\ell \in \mathcal{C}_m} e^{\mu_m V_\ell}\right)}{\sum_{p=1}^{M} \exp\left(\frac{\mu}{\mu_p} \ln \sum_{\ell \in \mathcal{C}_p} e^{\mu_p V_{\ell p}}\right)}$$

Nested Logit Model

- If $\frac{\mu}{\mu_m} = 1$, for all m, NL becomes logit.
- Sequential estimation:
  - Estimation of NL decomposed into two estimations of logit
  - Estimator is consistent but not efficient
- Simultaneous estimation:
  - Log-likelihood function is generally non concave
  - No guarantee of global maximum
  - Estimator asymptotically efficient
  - Log likelihood for observation n is

$$\ln P(i_n|\mathcal{C}_n) = \ln P(i_n|\mathcal{C}_{mn}) + \ln P(\mathcal{C}_{mn}|\mathcal{C}_n)$$

  where i_n is the chosen alternative.

Correlation

$$\mathrm{Corr}(U_i, U_j) = \frac{\mathrm{Cov}(U_i, U_j)}{\sqrt{\mathrm{Var}\,U_i \mathrm{Var}\,U_j}} = \frac{\mathrm{Cov}(\varepsilon_m + \varepsilon_{im}, \varepsilon_m + \varepsilon_{jm})}{\sqrt{\mathrm{Var}(\varepsilon_m + \varepsilon_{im})\mathrm{Var}(\varepsilon_m + \varepsilon_{jm})}}.$$

As $\mathrm{Var}\,\varepsilon_{im} = \mathrm{Var}\,\varepsilon_{jm}$, the denominator is $\mathrm{Var}(\varepsilon_m + \varepsilon_{im})$ so that

$$\mathrm{Corr}(U_i, U_j) = \frac{\mathrm{Var}\,\varepsilon_m}{\mathrm{Var}(\varepsilon_m + \varepsilon_{im})}.$$

- ε'_m and ε_{im} have the same variance.
- $\varepsilon_m + \varepsilon'_m$ and $\varepsilon_m + \varepsilon_{im}$ have the same variance.

$$\mathrm{Var}(\varepsilon_m + \varepsilon_{im}) = \mathrm{Var}(\varepsilon_m + \varepsilon'_m) = \mathrm{Var}\,\tilde{\varepsilon}_m = \pi^2/6\mu^2.$$

Correlation

As

$$\mathrm{Var}(\varepsilon_m + \varepsilon_{im}) = \mathrm{Var}\,\varepsilon_m + \mathrm{Var}\,\varepsilon_{im} + 2\,\mathrm{Cov}(\varepsilon_m, \varepsilon_{im}) = \pi^2/6\mu^2,$$

we obtain

$$\mathrm{Var}\,\varepsilon_m = \pi^2/6\mu^2 - \mathrm{Var}\,\varepsilon_{im} - 2\,\mathrm{Cov}(\varepsilon_m, \varepsilon_{im}).$$

Consequently

$$\begin{aligned}\mathrm{Corr}(U_i, U_j) &= \frac{\pi^2/6\mu^2 - \mathrm{Var}\,\varepsilon_{im} - 2\,\mathrm{Cov}(\varepsilon_m, \varepsilon_{im})}{\pi^2/6\mu^2} \\ &= 1 - \frac{\pi^2/6\mu_m^2}{\pi^2/6\mu^2} - 2\frac{\mathrm{Cov}(\varepsilon_m, \varepsilon_{im})}{\pi^2/6\mu^2} \\ &= 1 - \frac{\mu^2}{\mu_m^2} - 2\frac{\mathrm{Cov}(\varepsilon_m, \varepsilon_{im})}{\pi^2/6\mu^2}.\end{aligned}$$

Correlation

$$\mathrm{Corr}(U_i, U_j) = 1 - \frac{\mu^2}{\mu_m^2} - 2\frac{\mathrm{Cov}(\varepsilon_m, \varepsilon_{im})}{\pi^2/6\mu^2}.$$

If $\mu = \mu_m$, we have the logit model, and the correlation is 0:

$$\mathrm{Corr}(U_i, U_j) = -2\frac{\mathrm{Cov}(\varepsilon_m, \varepsilon_{im})}{\pi^2/6\mu^2} = 0.$$

Therefore,

$$\mathrm{Corr}(U_i, U_j) = 1 - \frac{\mu^2}{\mu_m^2}.$$

To obtain $0 \leq \mathrm{Corr}(U_i, U_j) \leq 1$, it is sufficient that

$$0 \leq \mu \leq \mu_m, \quad m = 1, \ldots, M.$$

Correlation

Correlation matrix is block diagonal:

$$\mathrm{Corr}(U_i, U_j) = \begin{cases} 1 & \text{if } i = j, \\ 1 - \dfrac{\mu^2}{\mu_m^2} & \text{if } i \neq j,\ i \text{ and } j \text{ are in the same nest } m, \\ 0 & \text{otherwise.} \end{cases}$$

Variance-covariance matrix is block diagonal:

$$\mathrm{Cov}(U_i, U_j) = \begin{cases} \dfrac{\pi^2}{6\mu^2} & \text{if } i = j, \\ \dfrac{\pi^2}{6\mu^2} - \dfrac{\pi^2}{6\mu_m^2} & \text{if } i \neq j,\ i \text{ and } j \text{ are in the same nest } m, \\ 0 & \text{otherwise.} \end{cases}$$