Preferred citation style

Schuessler, N. (2006) Initial ideas on accounting for similarities between alternatives in route, mode and destination choice, Second Workshop on Applications of Discrete Choice Models, EPF Lausanne, Lausanne, September 2006.

Initial ideas on accounting for similarities between alternatives in route, mode and destination choice

Nadine Schuessler

IVT
ETH
Zurich

September 2006

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Outline

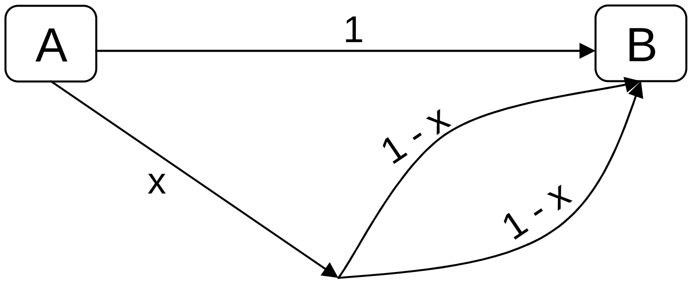
- Problem and objectives
- Approaches restructuring the variance-covariance matrix
- Similarity measures
- Work Program

Independence of irrelevant alternatives (IIA)

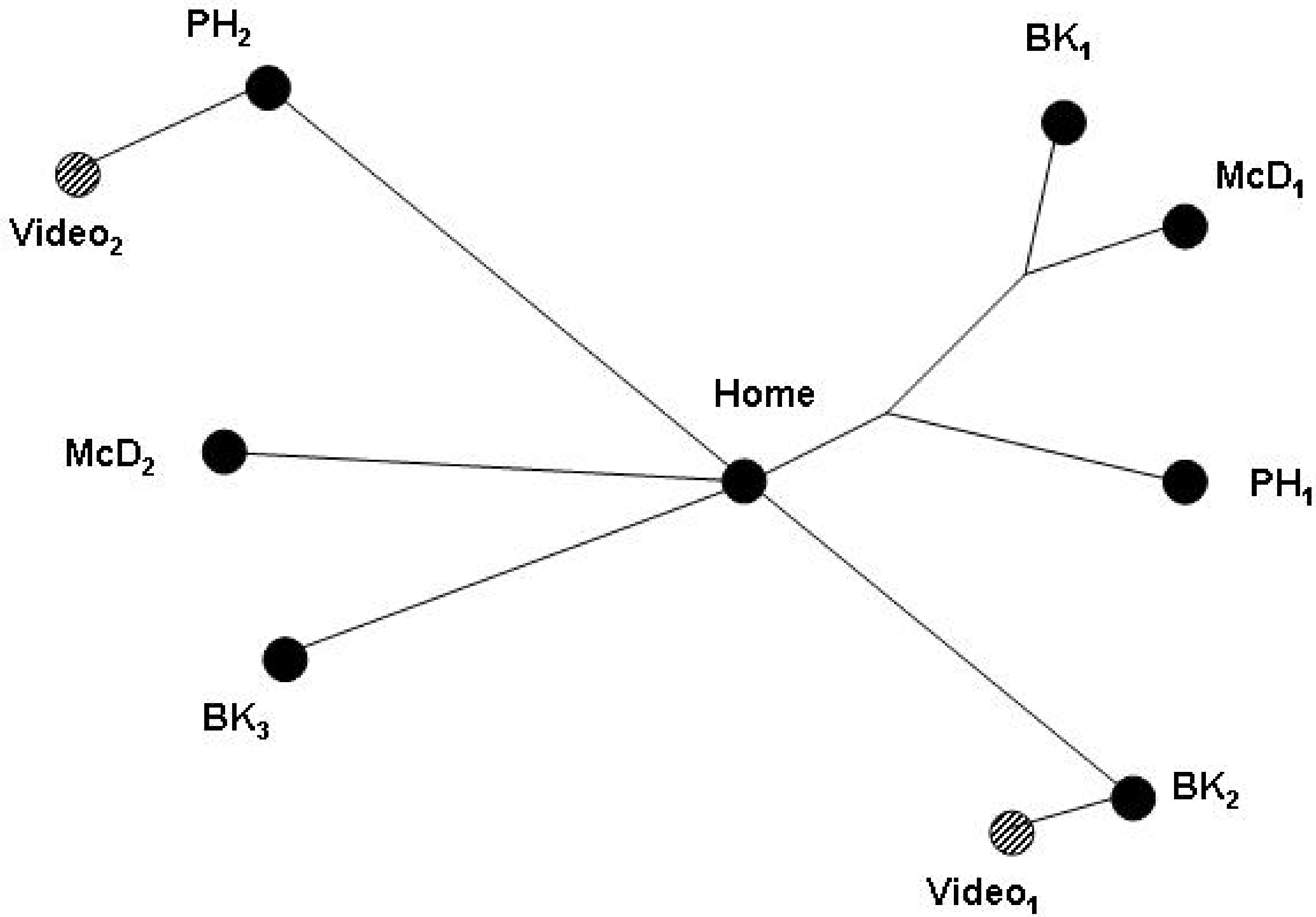
The probability of two alternatives to be chosen is not effected by the presence or the characteristics of any other alternatives.

Prominent examples:

- red bus /blue bus paradox
- route choice problem by Daganzo and Sheffi



Similarities between alternatives



Accounting for similarities in discrete choice models

Similarities can be accounted for by

- restructuring the variance-covariance matrix, or
- introducing a similarity measure in the systematic part of the utility function.

The aimed approach should be:

- usable for practical (i.e. large scale) applications
- easy to compute
- transferable to any choice context

Restructuring the variance-covariance matrix

Allowing for correlations between alternatives by:

- opening the variance-covariance structure
- introducing multivariate error terms

Most prominent and most general approaches are:

Multinomial Probit

Mixed Multinomial Logit

- Random Parameter Logit
- Error Component Logit
- Mixed Spatially Correlated Logit Model

Network GEV models

- Nested Logit
- Cross Nested Logit

Restructuring the variance-covariance matrix

Advantages:

- very flexible
- Many/all correlation structures can be represented
- accounting for taste heterogeneities
- MMNL and GEV models have a closed model formulation

Disadvantages:

- high estimation complexity
- much time and effort needed for specification as well as for identification, i.e. employing constraints to find unique solutions
- not suitable for large size applications such as real transport networks

Implicit availability/perception model (IAP)

Main assumptions

- the decision-maker has an imperfect knowledge of the alternatives, and
- he has limited information processing abilities

=> he does not chose from the universal choice set C but from his individual choice set C_n

The utility of an alternative is extended by a term q_{Cin} representing its degree of membership in the individual choice set C_n:

$$U_{in} = V_{in}^{'} + q_{Cin} + \varepsilon_{in}$$

q_{Cin} as a measure of independence of an alternative

The dependencies of an alternative

- decrease its probability to be perceived as separate alternative.
- decrease its probability to be included in the individual choice set.
- are equivalent to the degree of similarities with other alternatives.

Gower's (1985) "General Coefficient of Similarity":

$$S_{ij} = \frac{\sum_{k=1}^{p} w_k(x_{ik}, x_{jk})\, s(x_{ik}, x_{jk})}{\sum_{k=1}^{p} w_k(x_{ik}, x_{jk})}$$

Existing formulations accounting for similarities

Private transport route choice

- C-Logit	Cascetta et al. (1996)
- Path Size Logit	Bierlaire and Ben-Akiva (1999)
- Generalised Path Size Logit	Ramming (2002)

Public transport Connection Choice

- Autonomy of a connection	Friedrich et al. (2000)

Spatial Correlations

- Competing Destinations	Fotheringham (1988)
- Spatial Dependency Parameter	Mohammadian et al. (2005)
- Field Effect Variable	Dugundji and Walker (2005)

Activity or trip chains

- Sequence Alignment Method	Joh et al. (2002)
- Prospective Utility	Kitamura (1984)

C-Logit and Generalised Path Size Logit

C-Logit

- similarity coefficient CF_{in} is added to the utility function
- CF_{in} gives the logarithmic percentage of route length that a route shares with other routes.

Path Size Logit

- The Path Size factor PS_{in} corrects the length of each route.
- It is based on the length of the shared links of the route i and the length of the routes that share a link with i, relative to the length of the shortest route using that link

Generalised Path Size Logit

- Each link is given the size one, which is allocated among all the routes using that link.
- GPS_{in} is the sum of the link sizes weighted by the contribution of each link to the overall route length.

Autonomy of a connection

There relative influence $e_j(i)$ of a connection *i* on another one *j* is characterised by:

- the time gap between corresponding departure and arrival times
- the difference in speed
- the difference in price

The Autonomy of a Connection is then defined as the reciprocal of one plus the sum of all influences $e_j(i)$ from other connections.

Spatial Correlations

Competing Destinations

A correction term is calculated

- based on the differences between their attributes, or
- based on the spatial distance between them.

Spatial Dependency Parameter

- Spatial dependency parameter represents the influence of one decision-maker has on another by choosing a certain alternative.

Field Effect Variable

- CNL model for the correlations between alternatives.
- Field Effect variable accounts for spatial or social correlations between decision-makers, represented by a graphical network of the interdependencies.

Activity and trip chaining problems

Sequence Alignment Method

- designed for alternatives characterised by multiple attributes with multivariate descriptions and a certain sequential order.
- biological rather than geometrical distance is employed, i.e. the smallest number of attribute changes (mutations) that is necessary to equalise two alternatives

Prospective Utility

- accounts for dependencies not only in spatial but also in temporal and causal dimensions
- integrates the utility of a consequent trip that might be made after the visit to the spatial zone in question
- characteristicis: likelihood of visiting another zone, the spatial distance between the zones and the utility of the activity in the subsequent zone.

Evaluation of the similarity factor formulations

Advantages

- Attenuation of the IIA property
- Error terms remain type I extreme value distributed
- Closed model formulation preserved
- No necessity for a priori assumptions about the correlation structure
- Relatively low effort in terms of specification and estimation

Shortcomings

- Presented similarity measures are only applicable to specific choice situations
- No simultaneous account of route, mode and destination choice

Work Program

1. Identification of appropriate similarity factors
2. Formulation of the Logit model
3. Specification of the similarity factors and model tests
4. Application to a combined route mode and destination choice model
5. General guidance on similarities in discrete choice modelling

Data Sets

GPS data sets

- for Zurich, Geneva and Winterthur
- “on-person” GPS data, including all trips of 4886 man-days
- mode, route and destination choice

Air traffic route choice

- air connections booked by Swiss frequent flyer November 2006 for 75 European ODs
- choice set generation by web robots
- connection choice including fares

Swiss Microcensus 2005

- will be combined with spatial information about the zones
- route, mode and destination choice

References

Ben-Akiva, M.E. and M. Bierlaire (1999) Discrete choice methods and their applications to short-term travel decisions, in R. Hall (ed.) *Handbook of Transportation Science*, 5-34, Kluwer, Dordrecht.

Ben-Akiva, M.E. and D. Bolduc (1996) Multinomial probit with a logit kernel and a general parametric specification of the covariance structure, 3rd Invitational Choice Symposium, Columbia University.

Bhat, C. and J. Guo (2004) A mixed spatially correlated logit model: formulation and application to residential choice modeling, Transportation Research, 38B (2) 147-168.

Bierlaire, M. (2002) The Network GEV model, paper presented at the 2nd Swiss Transport Research Conference, Ascona.

Cascetta, E., A. Nuzzola, F. Russo and A. Vitetta (1996) A modified logit route choice model overcoming path overlapping problems: Specification and some calibration results for interurban networks, *Proceedings of the International Symposium on Transportation and Traffic Theory*, Hrsg. J.B. Lesort 697-711, Lyon.

References

Dugundji, E.R. and J.L. Walker (2005) Discrete Choice With Social and Spatial Network Interdepencies: An Empirical Example Using Mixed GEV Models with Field and “Panel” Effects, 84th Annual Meeting of the Transportation Research Board, Washington, January 2005.

Fotheringham, A.S. (1988) Consumer store choice and choice set definition, *Marketing Science*, 7 (3) 299-309

Friedrich, M., I. Hofsäss and S. Wekeck (2000) *Verbindungssplit im ÖV*: Übersicht zur Modellierung des Verbindungssplits in VISUM, report, Karlsruhe.

Joh, C.H., T.A. Arentze, F. Hofman and H.J.P. Timmermans (2002) Activity pattern similarity: A multidimensional alignment method, Transportation Research B, **36** (5) 385-403.

Kitamura, R. (1984) Incorporating trip chaining into analysis of destination choice, Transportation Research B, 18 (1) 67-81.

References

Mohammadian, A.K., M. Haider and P.S. Kanaroglou (2005) Incorporating Spatial Dependencies in Random Parameter Discrete Choice models, 84th Annual Meeting of the Transportation Research Board, Washington, January 2005.

Ramming, M.S. (2001) Network knowledge and route choice, Dissertation, Massachusetts Institute of Technology, Boston.

Train, K.R. (2003) *Discrete Choice Methods with Simulation*, University Press, Cambridge.

Vrtic, M. (2003) Simultanes Routen- und Verkehrsmittelwahlmodell, Dissertation, Fakultät für Verkehrswissenschaften, TU Dresden.

Multinomial Probit and Mixed Logit

Multinomial Probit

- error terms ε_{in} are multivariate normal distributed
- alternatives can be correlated in any fashion

Mixed (Multinomial) Logit

- Random parameter Logit (RPL)
- Error Component Logit (ECL)
- error terms remain type I extreme value distributed
- introduction of a multivariate distributed term ξ_{in}
- utility function: $U_{in} = V'_{in} + \xi_{in} + \varepsilon_{in}$

Network GEV models

Correlations are represented by a single source network without circles

This can also be interpreted as subdividing alternatives into nests

Nested Logit

- no correlations between alternatives belonging to different nests, only within each nest
- utility function of alternative *i* in nest *j*:

$$U_{in} = V_{in} + V_{jn} + V_{i|jn} + \varepsilon_{in} + \varepsilon_{jn} + \varepsilon_{i|jn}$$

Cross Nested Logit

- alternatives can belong to multiple nests at the same time
- allocation parameter α_{in} specifies the degree of membership for each nest

C-Logit and Path Size Logit

C-Logit (Cascetta et al. 1996)

$$CF_{in} = -\beta_0 \ln \sum_{a \in \Gamma_i} \frac{l_a}{L_i} N_{an}$$

Path Size Logit (Ben-Akiva and Bierlaire, 1999)

$$S_{in} = \sum_{a \in \Gamma_i} \frac{l_a}{L_i} \frac{1}{\sum_{k \in C_n} \delta_{ak} \frac{L^*_{C_n}}{L_k}}$$

Generalised Path Size (Ramming, 2002)

$$PS_{in} = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i} \right) \frac{1}{\sum_{j \in C_n} \frac{L_i^\gamma}{L_j^\gamma} \delta_{aj}}$$

Competing Destinations

Likelihood function formulations for the inclusion of Alternative i in the individual choice set C_n

- for the differences between alternative characteristics

$$l_n(i \in C_n) = \exp\left(\frac{1}{J-1} \sum_j \sum_k \theta_k \left| X_{ik} - X_{jk} \right| \right)$$

- for the spatial distance

$$l_n(i \in C_n) = \left[\frac{1}{K-1} \sum_{\substack{k \in C_n \\ i \neq k}} \frac{w_i}{d_{ik}} \right]^{\theta}$$