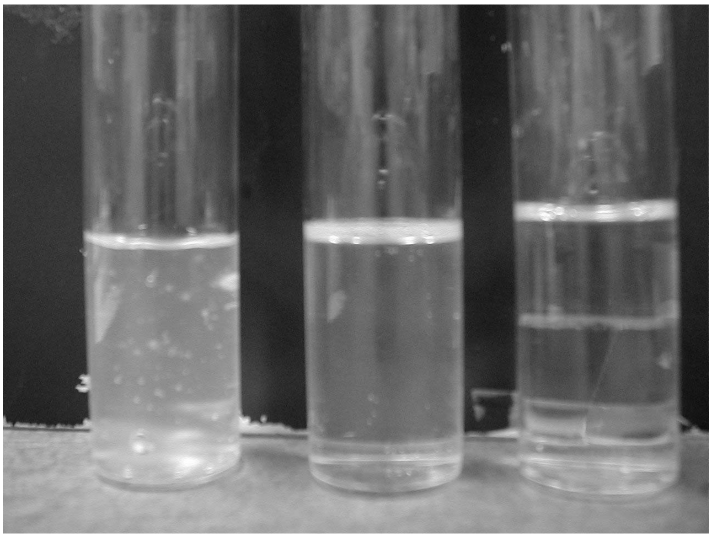


Discrete mixtures of GEV models

Stephane Hess, Imperial College London & RAND Europe
Michel Bierlaire, EPFL
John W. Polak, Imperial College London

Conference paper STRC 2005
Session: Transport Modelling II

STRC

5th Swiss Transport Research Conference
Monte Verità / Ascona, March 9-11, 2005

Discrete mixtures of GEV models

Stephane Hess
Centre for Transport Studies
Imperial College London
Phone: +44(0)20 7594 6105
Fax: +44(0)20 7594 6102
stephane.hess@imperial.ac.uk

Michel Bierlaire
Institute of Mathematics
EPFL
Phone: +41(0)21 693 25 37
Fax: +41(0)21 693 55 70
michel.bierlaire@epfl.ch

John W. Polak
Centre for Transport Studies
Imperial College London
Phone: +44(0)20 7594 6089
Fax: +44(0)20 7594 6102
j.polak@imperial.ac.uk

March 2005

Abstract

Allowing for variations in behaviour across respondents is one of the most fundamental principles in discrete choice modelling, given that the assumption of a purely homogeneous population cannot in general (or ever) be seen to be valid. Two approaches have classically been used to address this problem; the use of deterministic segmentations of the population, and the use of a random continuous representation of variations in tastes across respondents. In this paper, we discuss an alternative approach, based on the use of discrete mixtures of underlying GEV models over a finite set of distinct support points. The paper presents two applications; one illustrating the performance of the model with the help of simulated data, and one showing, on real data, how the model can be used to test the validity of hypotheses such as the presence of individuals with zero valuations of travel-time changes.

Keywords: Mixed logit – discrete distributions – latent classes
Swiss Transport Research Conference – STRC 2005 – Monte Verita

1 Introduction

Allowing for variations in behaviour across respondents is one of the most fundamental principles in discrete choice modelling, given that the assumption of a purely homogeneous population cannot in general (or ever) be seen to be valid. The most basic approach for representing such variations is through a segmentation of the population into mutually exclusive subsets, either in the form of separate models for different population segments, or separate coefficients within the same model for different population segments. These approaches can for example be used to differentiate between different journey purposes, or different income classes. In the case of continuous attributes, such as income, such segmentations can however be seen to be very arbitrary, and it is in this case preferable (though computationally more expensive) to use a continuous variation in tastes as a function of the continuous attribute.

Deterministic variations in tastes, such as those described above, can be implemented within the standard random utility framework, and are applicable for all known model structures. However, although the use of such deterministic variation is appealing from the point of view of interpretation (and especially for forecasting), it is often not preferable to represent all variations in tastes in a deterministic fashion, for reasons of data quality, but also due to inherent randomness in choice behaviour. For this reason, models allowing for random variations in behaviour across respondents have an important advantage in terms of flexibility. Two such model structures have been discussed in detail in the existing literature; the Multinomial Probit model (c.f. Daganzo 1979), and the Mixed Multinomial Logit (c.f. McFadden & Train 2000), where the latter has the advantage of allowing the use of statistical distributions other than the Normal for representing the variations in tastes. Both models additionally allow for the representation of correlation across alternatives in the unobserved utility terms.

The Multinomial Probit (MNP) model and the Mixed Multinomial Logit (MMNL) model have the disadvantage that their choice probabilities take on the form of integrals that do not possess a closed-form solution, such that numerical processes, typically simulation, are required during estimation and application of the models. This greatly limited the use of these structures for many years after their initial developments. Over recent years, gains in computer speed and the efficiency of simulation-based estimation processes have however led to increased interest in the MMNL model in particular, by researchers and, to a lesser degree also practitioners.

Despite the improvements in estimation capability, the cost of using the MMNL model remains high. While this might be acceptable in many cases, two other issues arise with the use of the MMNL model.

The first issue is that of the choice of distribution to be used for representing the random variations in tastes across respondents. This issue can be divided into several sub-issues. Firstly, it is important to reconcile the choice of distribution with the intuitive understanding of the true distribution. As such, a strictly positive distribution would not be used for a coefficient where positive as well as negative values are expected in the population. On the other hand, in the case of a coefficient with a strong sign assumption (such as a negative cost coefficient), the use of strictly bounded distributions can be unable to signal problems with the data or utility specification that would manifest themselves as counter-intuitively signed coefficients for part of the population (c.f. Hess, Bierlaire & Polak 2005*b*). Secondly, even in the case where a choice of distribution can be made that avoids all the issues discussed above, it is highly probable that some bias between the true and postulated distributions will remain; cases will arise in which real-world behaviour cannot be characterised adequately by one of a set of standard statistical distributions. One case in point arises in the modelling of tastes which may theoretically have a significant mass at zero but be exclusively positively or negatively signed elsewhere. The situation becomes even more complicated in the case of an attribute which some individuals value positively and some individuals value negatively, with a remaining part of the population being indifferent to the attribute. Representing this situation is not generally possible with the use of standard continuous distributions.

The other remaining issue with the MMNL model formulation is that of interpretation. A great deal has been said about this in the existing literature, notably by Hensher & Greene (2003). While the advantages of the MMNL model over more basic model structures are very significant, the pitfalls are similarly important. The main issue in interpretation is heavily correlated with the choice of distribution, as highlighted in the previous paragraph, and discussed in detail by Hess, Bierlaire & Polak (2005*b*). If an inappropriate choice of distribution is made, the results may be accurate in terms of the mean value, and possibly also in terms of the measure of dispersion, but it is highly likely that conclusions inferred on the basis of the implied tail behaviour of the distribution are misleading. A case in point is that where an unbounded distribution indicates the presence of a significant share of the population with a counter-intuitively signed value of time; such results can equally well be seen as a result of the shape of the distribution, and need thus not signal the actual existence of such values in the population.

Even though the MMNL model does, despite these problems, remain an appealing model structure, it is of interest to explore alternative ways of representing random variations in tastes across respondents, avoiding the issues of specification, estimation, and interpretation discussed above. This forms the motivation for the present paper, in which we replace the continuous distribution functions by discrete distributions, spreading the mass among several discrete values. This is similar to the case of a latent-class model (c.f. Kamakura & Russell 1989, Chintagunta et al. 1991), assigning different coefficient values to different parts of the population of respondents. In theory, existing discrete distributions (e.g. Poisson) could be used; however, this comes at the cost of flexibility and again leads to the problem of reconciling the assumptions about the shape of the theoretical distribution with the actual shape of the true distribution. This problem does not exist in the case where a fixed set of coefficient values are used that each have an associated probability, but where the values and associated probabilities are free from any a priori constraints. The authors of this paper are aware of only one previous application of this approach, namely by Dong & Koppelman (2003), who use it in the context of mode choice for work trip in New-York, referring to the model as the “Mass Point Mixed Logit model”. They propose a Bayesian estimation procedure, based on Gibbs sampling and data augmentation to accommodate the discrete nature of the dependent variable.

The remainder of this paper is organised as follows. The next section discusses the theory behind discrete mixture models. Section 3 presents two brief applications, and section 4 summarises the contents of the paper and presents the conclusions of the study.

2 Methodology

The choice probability of an alternative in a discrete choice models is given as a function of the utility of all alternatives. The utility is divided into an observed and an unobserved part, where the role of the unobserved part differs between model structures. The observed utility of an alternative i is a function of the attributes of the alternative as faced by respondent n, given by x_{ni}, and the tastes of the decision-maker, given by the vector β. The vector x_{ni} can additionally include socio-demographic characteristics of decision-maker n that allow for deterministic variations in the utility of alternative i across respondents.

In the purely deterministic model, the vector β is constant across respondents, and the observed utility for alternative i and individual n is given by $V(x_{ni}, \beta)$. Let x_n be a vector grouping together the individual vectors x_{nj} across the alternatives contained in the choice-set of respondent n. Let λ represent an additional set of parameters, which can for example contain the structural parameters used in a Nested Logit (NL) model to represent inter-alternative correlation. In a very general form, we can then define $P(i \mid x_n, C_n, \lambda, \beta)$ to give the choice probability of alternative i for individual n, with a choice-set C_n, conditional on the observed vector x_n, and for given values for the vectors of parameters β and λ (to be estimated). Due to the potential inclusion of socio-demographic attributes in x_n, this notation allows for deterministic variations in tastes across respondents.

As an example, in the Multinomial Logit (MNL) model, we have:

$$P(i \mid x_n, C_n, \lambda, \beta) = \frac{e^{V(x_{ni}, \beta)}}{\sum_{j \in C_n} e^{V(x_{nj}, \beta)}}, \tag{1}$$

where C_n gives the choice set of individual n, and where the well-known *IIA* assumption of the MNL model means that the vector of parameters λ is not required.

As described above, the notation $P(i \mid x_n, C_n, \lambda, \beta)$ allows for the possibility of deterministic variations in tastes across respondents. With the help of the vector λ, it can be used to represent the choice probabilities in any closed-form discrete choice model, such as Nested Logit (NL) or Cross-Nested Logit (CNL).

This notation can now be used as the building block for models allowing for a distribution of tastes across respondents. The best known model of this type is the so-called *Mixed Multinomial model*. In fact, a more appropriate notation would be to describe the MMNL model as a *mixture of Multinomial Logit models*.

In statistics, a mixture density is a probability density function (*pdf*) which is a convex linear combination of other *pdf*'s. If $f(\varepsilon, \theta)$ is a *pdf*, and if $w(\theta)$ is a nonnegative function such that

$$\int_a w(a)\mathrm{d}a = 1 \tag{2}$$

then

$$g(\varepsilon) = \int_a w(a) f(\varepsilon, \theta)\mathrm{d}a \tag{3}$$

is also a *pdf*. We say that g is a mixture of f. Discrete mixtures are also possible. If $f(\varepsilon, \theta)$ is a *pdf*, and if w_i, $i = 1, \ldots, n$ are nonnegative weights such that

$$\sum_{i=1}^{n} w_i = 1 \tag{4}$$

then

$$g(\varepsilon) = \sum_{i=1}^{n} w_i f(\varepsilon, \theta_i) \tag{5}$$

is also a *pdf*. We say that g is a discrete mixture of f.

With this in mind, the MMNL model is indeed a mixture of MNL models over the continuous distribution of the vector of tastes β. The choice probabilities for this model are given by:

$$P_n \left(i \mid x_n, C_n, \lambda, \Omega \right) = \int_{\beta} \left[P \left(i \mid x_n, C_n, \lambda, \beta \right) f \left(\beta, \Omega \right) \right] \mathrm{d}\beta, \tag{6}$$

where the vector β is distributed according to $f(\beta, \Omega)$, with vector of parameters β. With $P(i \mid x_n, C_n, \lambda, \beta)$ given by equation (1), equation (6) represents the choice probabilities in an MMNL model; however, any other GEV-type choice probability can be used for $P(i \mid x_n, C_n, \lambda, \beta)$, with an explicit role for the vector λ, leading to a mixture of a more general GEV model, as described by Hess, Bierlaire & Polak (2005*a*).

As mentioned in the introduction, the MMNL model is well documented in the existing literature, notably by Train (2003). The estimation of the model is discussed amongst others by Bhat (1999) and Hess, Train & Polak (2005), while the issues of specification and interpretation are highlighted by Hensher & Greene (2003) and Hess, Bierlaire & Polak (2005*b*). Recent applications of MMNL model (for random taste heterogeneity) are given by Revelt & Train (1998), Bhat (2000), Hess & Polak (2004, 2005) and Hess, Train & Polak (2005).

We will now turn our attention to the development of structures using discrete mixtures of Logit models, and GEV models by extension. We continue to use the same general notation for the choice probabilities as before. However, we now divide the set of parameters β into two sets; $\bar{\beta}$ represents a part of β containing deterministic parameters, while $\widehat{\beta}$ is a set of K random parameters that have a discrete distribution. Within this set, the parameter $\widehat{\beta}_k$ has m_k mass points $\widehat{\beta}_k^j$, $j = 1, \ldots, m_k$, each of them associated with a probability π_k^j, where for each $k = 1, \ldots, K$, we impose the conditions that

$$0 \leq \pi_k^j \leq 1, \quad k = 1, \ldots, K, j = 1, \ldots, m_k, \tag{7}$$

and

$$\sum_{j=1}^{m_k} \pi_k^j = 1. \tag{8}$$

For each realisation of $\beta_1^{j_1}, \ldots, \beta_K^{j_K}$ of $\widehat{\beta}$, the choice probability is given by

$$P\left(i \mid x_n, C_n, \lambda, \beta = \langle \bar{\beta}, \widehat{\beta}_1^{j_1}, \ldots, \widehat{\beta}_K^{j_K} \rangle \right), \tag{9}$$

where the deterministic part of β, $\bar{\beta}$ stays constant across realisations of the vector $\widehat{\beta}$.

The unconditional choice probability for alternative i and decision-maker n can now be written straightforwardly as a mixture over the discrete distributions of the various elements contained in $\widehat{\beta}$ as:

$$P_n\left(i \mid x_n, C_n, \lambda, \bar{\beta}, \widehat{\beta}, \pi\right) = \sum_{j_1=1}^{m_1} \cdots \sum_{j_K=1}^{m_K} P\left(i \mid x_n, C_n, \lambda, \beta = \langle \bar{\beta}, \widehat{\beta}_1^{j_1}, \ldots, \widehat{\beta}_K^{j_K} \rangle\right) \cdot \pi_1^{j_1} \cdot \ldots \cdot \pi_K^{j_K}, \quad (10)$$

where $\bar{\beta}$, $\widehat{\beta}$ and π are vector of parameters to estimate in a regular maximum likelihood estimation procedure. The obvious advantage of this approach is that, if the probability model (9) used inside the mixture has a closed form, then so does the discrete mixture itself.

In this paper, we mainly focus on the simple case where the underlying choice model is of MNL form; however, the form given in equation (10) is appropriate for any underlying GEV model. The approach can easily be extended to the case of combined discrete and continuous random taste variation, by partitioning β into three parts; the above defined parts $\bar{\beta}$ and $\widehat{\beta}$, and an additional part $\widetilde{\beta}$, whose elements follow continuous distributions. This however leads to a requirement to use simulation, as with every continuous mixture models. The choice probabilities for such a model are then given either by replacing the conditional probability inside equation (10) by an integral over the distribution of the terms in $\widehat{\beta}$, or by integration of the whole of the unconditional probability in equation (10) over the distribution of these terms. The position of the integral is in fact of no importance, but, for consistency with equation (10), it can be seen as preferable to have the integration inside.

Allowing for random as well as discrete random terms not only increases flexibility from the point of view of random taste heterogeneity, but allows for the use of error-components to represent heteroscedasticity and inter-alternative correlation (c.f. Walker 2001), where the latter is however also possible with the use of an underlying GEV structure.

The approach we use in this paper can be seen either as a generalisation of a model using fixed-point estimates, or as a simplified version of a model allowing for a continuous distribution of tastes across respondents. As a generalisation of the fixed-point estimates model, our approach offers more modelling flexibility, by allowing for random as well as deterministic variations in tastes.

Compared to a model with a continuous distribution, the use of a fixed number of support-points clearly leads to a reduction in flexibility. On the other side, the model is free from any assumptions resulting from the choice of a specific statistical distribution in the continuous case. Furthermore, the disadvantages caused by relying on a fixed number of support-points should be expected to decrease rapidly with increases in the number of points used.

Another major advantage of our approach is the lack of need for simulation processes. In this spirit, it is tempting to consider discrete distributions models with many support points as a way to approximate a continuous distribution, and prefer this technique over simulated maximum likelihood. Indeed, simulated maximum likelihood is actually also based on a discrete sum based on values drawn from the distribution. However, we must keep in mind an important difference. In order to improve the accuracy of the approximation of the continuous distribution, it is sufficient in the simulated maximum likelihood estimation to increase the number of draws. This does not affect the number of parameters to be estimated. It just takes more time to evaluate the log-likelihood function. In the discrete distribution context, adding one more point increases the number of parameters to estimate by two. When the number of points is high, the estimation can become very difficult, not only because of the size of the optimisation problem, but also because of numerical problems related to the nature of the log-likelihood function discussed below.

A good example showing the advantages of the discrete approach is that of the analysis of value of travel-time savings (VTTS). Random variations across respondents have repeatedly been found for this much-used willingness-to-pay indicator, for example by Algers et al. (1998), such that the use of a fixed-point model is not appropriate. However, with the coefficients used in this trade-off, notably the marginal travel-time coefficient, important issues arise in the case of a continuous mixture model, as discussed by Hess, Bierlaire & Polak (2005*b*). Firstly, there is a possibility of a highly assymetrical distribution, with a mean close to zero, which some bounded distributions might not be able to replicate, but where unbounded distributions can falsely indicate the presence of respondents with a positive marginal utility of travel-time, and hence a negative value of time. Secondly, there is a distinct possibility of a multi-modal distribution, in the case where an appropriate segmentation of the population has not been possible. Finally, there is a possibility that the time constraint of some individuals is not binding, such that their VTTS is zero, while the remaining individuals have a positive VTTS (negative marginal utility of travel-time). While the first two issues can be addressed through appropriate specification, standard continuous distributions are clearly not adequate in the case of a part of the population with a zero VTTS. Discrete mixtures on the other hand can deal with all three issues.

There are several estimation issues related to discrete mixtures of choice models. Firstly, the non concavity of the log-likelihood functions does not allow the identification of a global maximum, even for discrete mixtures of MNL. The experiments conducted as part of this analysis revealed the presence of multiple local maxima. Therefore, performing several estimations from various starting points is advisable. Also, it is good practice to use starting values other than 0 or 1 for the π_k^j parameters.

Secondly, constrained maximum likelihood must be considered to account for constraints (7) and (8). It should be noted that eliminating (8) by replacing π_k^1 with

$$\pi_k^1 = 1 - \sum_{j=2}^{m_k} \pi_k^j \tag{11}$$

does not help, as the constraint $0 \leq \pi_k^1 \leq 1$ now leads to the new condition $0 \leq \sum_{j=2}^{m_k} \pi_k^j \leq 1$.

Thirdly, clustering of mass points (for example around the mode of the true distribution) is a frequent phenomenon which must be handled specifically. Indeed, if may be a sign that the number of mass points is too large, and that some of them should be grouped together. This should be explicitly tested. But it may also be a feature of the optimisation algorithm, which may not be desirable in practice. In this case, the use of additional bounds on the mass points is useful, based on the definition of (potentially mutually exclusive) a priori intervals for the individual mass points.

The model is available in the BIOGEME software package (c.f. Bierlaire 2003), where various constraints on the parameters can be imposed to address the issues described above. This also allows researchers to test the validity of specific assumptions, such as a mass at zero for the VTTS.

3 Experiments

In this paper, we report results from two sets of preliminary experiments. The first experiment makes use of real data, where we compare an MNL model with a normal mixture of MNL and a simple discrete mixture of MNL. We then make use of synthetic data in order to get more insights into the model's performance.

3.1 Real data

We use stated preferences data collected in the context of the analysis of the Swiss Value of Travel Time Savings (VTTS) (Koenig et al. 2004, Axhausen et al. 2004). For the present analysis, we use data from a binomial route-choice survey, for rail travellers on leisure trips. This leads to a total of 1,881 observations, collected from 209 respondents.

The final utility specification uses travel-cost, travel-time, frequency, and the number of interchanges as explanatory variables. Linear specifications are used for all attributes, except for cost, where the distance-elasticity was taken into account. For this, the following approach was used:

$$\ldots + \beta_c \cdot \left(\frac{\delta_n}{43}\right)^{\lambda} \cdot c_i, \tag{12}$$

where c_i is the cost of alternative i, with associated taste-coefficient β_c. The variable δ_n gives the distance observed for the associated *RP* trip for respondent n, with mean value 43, and λ gives the distance-elasticity. Attempts to model the income-elasticity were unsuccessful, with the associated elasticity parameter taking on values indistinguishable from zero.

We first estimate a simple Multinomial Logit model. The estimated parameters are

Parameter	Estimate	std. err.	t-test
`changes`	-1.1699	0.0623	-18.76
`cost`	-0.2689	0.0361	-7.46
λ	-0.6783	0.1117	-6.07
`freq`	0.4249	0.0346	12.27
`ttrail`	-0.0296	0.0040	-7.40
Sample size:	1,881		
Null log-likelihood:	-1303.81		
Final log-likelihood:	-917.499		
Likelihood ratio test:	772.62		
Rho-square:	0.2963		
Adjusted rho-square:	0.2925		

The mean VTTS from this model (for individuals with an observed distance of 43km) is 6.60 CHF/hour. The results also show decreasing cost-sensitivity with increasing distance.

In order to test for the prevalence of random variations across respondents in the VTTS, we next estimated an MMNL model using a normally distributed travel-time coefficient. The model fitting exercise takes into account the repeated choice nature of the dataset. In the next table, `ttrail` is the mean of the distribution, and `ttrails` its standard deviation. The standard deviation for the travel time coefficient is highly significant. With one additional parameter, the model offers a very significant improvement in LL by 34.30 units. The mean VTTS for train is very close to the MNL results, at 7.58CHF/hour. However, the model indicates significant variation around this mean, and indeed suggests a probability of 16.95% of a negative VTTS, with lower and upper 95% confidence intervals of -7.97CHF/hour and 23.15CHF/hour respectively. This does however not imply the presence of individuals with positive valuations of increases in travel-time, but should rather be seen as an artefact of the use of the Normal distribution (c.f. Hess, Bierlaire & Polak 2005*b*). Indeed, the conditions imposed by microeconomic time-allocation theory stating that the valuation of travel-time savings should be non-negative apply also in the case of leisure travellers. Nevertheless, the results show the existence of significant levels of taste heterogeneity.

Parameter	Estimate	std. err.	t-test
`changes`	-1.3312	0.0985	-13.52
`cost`	-0.3560	0.0583	-6.11
λ	-0.6107	0.1214	-5.03
`freq`	0.4952	0.0481	10.28
`ttrail`	-0.0450	0.0074	-6.08
`ttrails`	0.0471	0.0087	5.41
Sample size:	1,881		
Null log-likelihood:	-1303.81		
Final log-likelihood:	-883.201		
Likelihood ratio test:	841.218		
Rho-square:	0.3226		
Adjusted rho-square:	0.3180		

We next estimate a discrete mixture model, with two support points. In this model, the first support point is fixed at zero, in order to model the share of the population with a zero VTTS, that is people for whom the time constraint is not binding. The second support point is estimated from the data.

We observe a final LL of -897.31, with one additional estimated parameter when compared to the MNL model (only one mass-point needs to be estimated), and the same number of parameters as the MMNL model. The model offers statistically significant improvements over the MNL model, but offers lower performance than the MMNL model. This is a sign that the model is not able to fully account for all variations in tastes; this is however not the aim in the present analysis, which simply looks for the prevalence of zero valuations of travel-time savings.

The results indicate a very high share of 67.10% of respondents with a zero VTTS. Tests showed that the data did not include any *non-traders*, such that the results should not be seen as a simple effect of estimation bias due to captivity. In any case, in a binomial route-choice survey which does not include the *RP* alternative as an option, captivity should not play a significant role.

Although this share is very high, it is realistic to assume that a non-trivial part of respondents travelling for leisure purposes are indeed indifferent to travel-time changes (either positive or negative), in the absence of a binding time constraint. It must be stressed that this does not imply that travellers seek increases in travel-time, but are equally indifferent to increases as to decreases.

Parameter	Estimate	std. err.	t-test
`changes`	-1.5579	0.1043	-14.93
`cost`	-0.4316	0.0638	-6.76
λ	-0.5690	0.0920	-6.18
`freq`	0.5540	0.0504	11.00
`ttrail`	-0.1884	0.0397	-4.74
`massZero`	0.6710	0.0393	17.07
Sample size:	1881		
Null log-likelihood:	-1303.81		
Final log-likelihood:	-897.313		
Likelihood ratio test:	812.994		
Rho-square:	0.3118		
Adjusted rho-square:	0.3064		

In the segment of the population with strictly positive values of travel-time savings, the average VTTS is now a high 26.19CHF/hour for rail. This could be seen as an indication of problems with this model (given the much lower values in the MNL model). However, there is an alternative explanation. The MNL model is unable to explicitly represent the presence of a part of the population with a zero VTTS. As such, it can be expected that the fixed-point estimate it produces is biased, and indeed, gives the weighted average of the two values found in the population (assuming for now only two separate values). This notion is supported by a calculation of the weighted average on the basis of the results from the discrete mixture model. Indeed, we have $0.6710 \cdot 0 + 0.3290 \cdot 26.19 = 8.62$CHF/hour, which is acceptably close to the fixed-point VTTS obtained with the simple MNL model.

These results are very promising. Indeed, the model with the discrete mixture offers clear improvements in performance over the MNL model. Although it leads to poorer model fit than the MMNL model, it offers insights into taste variations that are more easily reconcilable with intuition (zero instead of negative VTTS). Nevertheless, the model using the Normal distribution clearly shows the presence of large levels of taste heterogeneity.

It is thus of interest to attempt to retrieve further levels of heterogeneity, at the same time as accounting for the prevalence of zero valuations of travel-time savings.

One possible approach is to use discrete mixtures with more than two values. However, the estimation becomes significantly more complicated, with the presence of several local maxima, and possible degeneracy, that is convergence of two points toward a common value. In this context, more investigation is necessary. Also, the estimation algorithm should be tailored to the structure of this problem, and a heuristic is needed to determine the optimal number of support points.

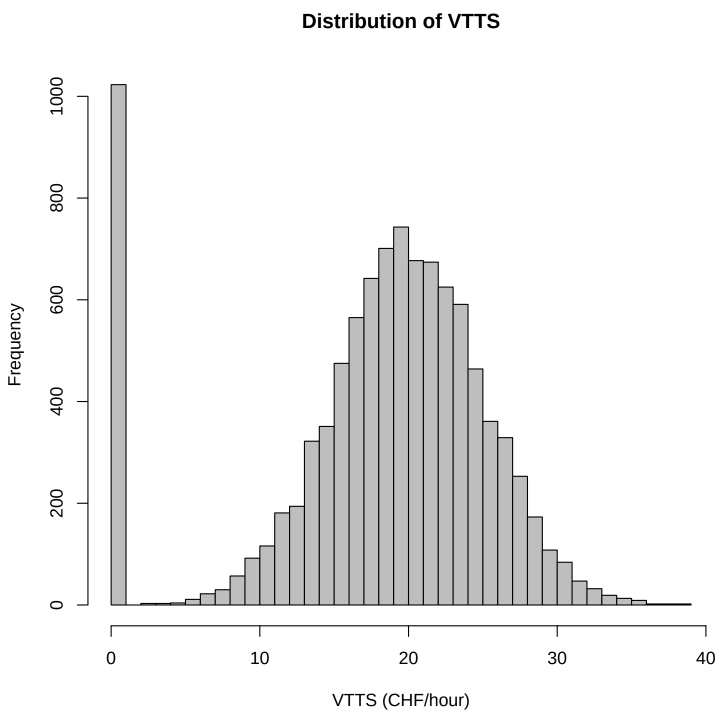
An alternative approach is based on combining a continuous distribution with an approach that allocates a specific share to a zero valuation of travel-time savings, as discussed by Cirillo & Axhausen (2004).

In the case where the main distribution is to be continuous, with an inflated mass at one point, only zero should really be seen as a candidate for such a point, given its specific role. As such, the location of the mass need not be estimated. Fixing the location of the support point also greatly facilitates the estimation of the model. In this case, the following approach can be used.

Let us assume we want to use a Normal distribution, with a mass at zero. In this case, we estimate three parameters; the mean of the Normal distribution, μ, its standard deviation, σ, and the mass at zero, γ. In the estimation code, simulation over the distribution is used at each iteration of the optimisation algorithm. In each iteration of the simulation process, a random draw (or quasi-random draw), say r, is used, where r is contained between 0 and 1. With this draw r, a draw from the $N(\mu,\sigma)$ distribution with a mass of γ at zero is produced as follows. If the draw r is smaller than γ, the value of the draw from the distribution at this iteration of the simulation algorithm is set to 0. Otherwise, it is set to:

$$\mu + \sigma \cdot \Phi^{-1}\left(\frac{r-\gamma}{1-\gamma}\right), \tag{13}$$

where Φ^{-1} is the Inverse Cumulative Normal distribution. As an example, with $\mu = 20$, $\sigma = 5$, and $\gamma = 0.1$, the distribution obtained with this approach has the shape shown below (using 10,000 draws). A similar approach can be used with any other underlying continuous distribution, by replacing Φ^{-1} by the appropriate inverse transform.



3.2 Simulated data

In order to investigate the validity of the discrete mixture model, we generate synthetic data such that we know the "true" value of the parameters. We use a mode-choice survey of $1,242$

observations from the Swiss VOT database as the underlying data, but we use simulated choices instead of actual choices as dependent variables. We also augment the sample size from $1,242$ to $5,000$ through minor random variations on the observed attributes.

In order to generate the synthetic choices, we assume that, except for the travel time coefficient for the car alternative, the true parameters are fixed as follows;

Parameter	Value
`car`	4
`changes`	-1.15
`cost`	-0.3
`freq`	0.9
`ttrail`	-0.07,

giving a true VTTS for rail-travel of 14CHF/hour.

In the first experiment, we assume that the population is divided into two segments. The VTTS of the first segment, composed of 50% of the sample, is assumed to be 16CHF/hour (the travel time coefficient being -0.08), while it is 6CHF/hour for the second segment (the travel time coefficient being -0.03).

As a first test, we estimate a discrete mixture model where the support points are fixed to the true values, and the probability mass distribution is estimated from the data. The 50%-50% share is reproduced quite closely, and the VTTS measures are also closely reproduced.

Parameter	Estimate	std. err.	t-test
`car`	4.0776	0.2489	16.38
`changes`	-1.2531	0.0942	-13.30
`cost`	-0.3198	0.0156	-20.52
`freq`	0.9436	0.0639	14.76
`massCar1`	0.4274	0.1413	3.02
`massCar2`	0.5726	0.1413	4.05
`ttrail`	-0.0792	0.0042	-18.94
Sample size:	5000		
Null log-likelihood:	-3465.74		
Final log-likelihood:	-868.984		
Likelihood ratio test:	5193.5		
Rho-square:	0.7493		
Adjusted rho-square:	0.7472		

In the second step, we estimate the actual support points from the data, in addition to the assigned mass.

Parameter	Estimate	std. err.	t-test
`car`	4.0265	0.2554	15.76
`changes`	-1.2306	0.0969	-12.70
`cost`	-0.3138	0.0160	-19.62
`freq`	0.9282	0.0656	14.15
`massCar1`	0.5149	0.2019	2.55
`massCar2`	0.4851	0.2019	2.40
`ttcar1`	-0.0770	0.0134	-5.73
`ttcar2`	-0.0373	0.0105	-3.54
`ttrail`	-0.0776	0.0057	-13.58
Sample size:	5000		
Null log-likelihood:	-3465.74		
Final log-likelihood:	-868.446		
Likelihood ratio test:	5194.58		
Rho-square:	0.7494		
Adjusted rho-square:	0.7468		

In this model, the recovery of the $50-50$ market share is better than in the model using fixed support points. The upper VTTS is slightly underestimated, at 14.72CHF/hour, while the lower one is overestimated, at 7.13CHF/hour. The VTTS for rail is also slightly overestimated, at 14.84CHF/hour. These slight biases are however well within the acceptable bounds. sampling.

In the second experiment, we assume that the segment with the lower VTTS represents only 30% of the population. This time, we immediately estimate both the support points and their masses from the data.

When we estimate the support points, we recover the 70/30 share almost perfectly. Both VTTS measures are slightly underestimated, at 4.70CHF/hour and 13.75CHF/hour, instead of 6CHF/hour and 16CHF/hour respectively. The rail VTTS is estimated at 13.73CHF/hour, instead of 14CHF/hour. Again, these biases are acceptable.

Parameter	Estimate	std. err.	t-test
`car`	4.1307	0.2522	16.38
`changes`	-1.2055	0.0934	-12.90
`cost`	-0.3203	0.0159	-20.17
`freq`	0.9600	0.0644	14.91
`massCar1`	0.2704	0.1234	2.19
`massCar2`	0.7296	0.1234	5.91
`ttcar1`	-0.0251	0.0107	-2.35
`ttcar2`	-0.0734	0.0101	-7.25
`ttrail`	-0.0733	0.0052	-14.04
Sample size:	5000		
Null log-likelihood:	-3465.74		
Final log-likelihood:	-906.999		
Likelihood ratio test:	5117.47		
Rho-square:	0.7383		
Adjusted rho-square:	0.7357		

3.3 Discussion

The two experiments conducted in this analysis have shown that the discrete mixture models are able to recover the values and market-shares of discretely distributed coefficients. They have also shown that the approach can be a useful tool in the testing of specific hypotheses, such as the presence of travellers with zero-valuations of travel-time changes. The extension to cases with more than two mass-points is straightforward, although it involves more issues with identification of an acceptable maximum, and potentially requires the use of constrained estimation.

4 Summary & Conclusions

In this paper, we have discussed an alternative approach for representing inter-agent variations in tastes, and by extension, choice behaviour. The approach is based on the use of discrete mixtures of GEV models, replacing the fixed-parameter choice probabilities by a weighted sum of GEV choice-probabilities calculated on the basis of different values for the specific coefficients for which taste heterogeneity is to be introduced. The weights or mass-points associated with the different values reflect the market share of this coefficient-value in the sample population.

Our theoretical discussions have highlighted that the model can be seen either as an extension of fixed coefficients models, or as a simplification of continuous mixtures of discrete choice models. Compared to the latter, the model has certain advantages and certain disadvantages, which have been described in detail in the paper. One of the major advantages of discrete mixtures is the ability to test the validity of hypotheses about the presence of a specific coefficient-value in the sample population, such as a zero VTTS.

The paper has presented two brief applications. The first application, which uses real data, has shown how discrete mixtures can be used to reveal the presence of respondents with zero valuations of travel-time changes, something that is not easily possible with continuous mixtures (advanced distributions are required). A second application made use of simulated data to show that the model is able to correctly reveal the presence of separate coefficient values in a sample population, and to reveal the exact market shares with a high level of precision.

Further work is required, notably in estimation with multiple support-points, but the preliminary tests have revealed that the use of discrete mixtures presents an interesting and promising alternative to continuous mixtures.

References

Algers, S., Bergstroem, P., Dahlberg, M. & Dillen, J. L. (1998), *Mixed Logit Estimation of the Value of Travel Time*, Working Paper, Department of Economics, Uppsala University, Uppsala, Sweden.

Axhausen, K., König, A., Abay, G., Bates, J. & Bierlaire, M. (2004), Swiss value of travel time savings, *in* 'Proceedings of the European Transport Conference, Strasbourg, France'.

Bhat, C. R. (1999), 'Quasi-random maximum simulated likelihood estimation of the mixed multinomial Logit model', *Transportation Research* **35B**(7), 677–693.

Bhat, C. R. (2000), 'Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling', *Transportation Science* **34**(2), 228–238.

Bierlaire, M. (2003), *BIOGEME: a free package for the estimation of discrete choice models*, Proceedings of the 3^{rd} Swiss Transport Research Conference, Monte Verita, Ascona, Switzerland.

Chintagunta, P., Jain, D. & Vilcassim, N. (1991), 'Investigating heterogeneity in brand preference in logit models for panel data', *Journal of Marketing Research* **28**, 417–428.

Cirillo, C. & Axhausen, K. W. (2004), *Evidence on the distribution of values of travel time savings from a six-week diary*, Arbeitsbericht Verkehrs- und Raumplanung 212, IVT, ETH Zurich.

Daganzo, C. F. (1979), *Multinomial Probit: The Theory and its Applications to Demand Forecasting*, Academic Press, New York.

Dong, X. & Koppelman, F. S. (2003), *Mass Point Mixed Logit Model: Development and Application*, paper presented at the 10^th triennial IATBR conference, Lausanne, Switzerland.

Hensher, D. & Greene, W. H. (2003), 'The Mixed Logit Model: The State of Practice', *Transportation* **30**(2), 133–176.

Hess, S., Bierlaire, M. & Polak, J. W. (2005*a*), Capturing taste heterogeneity and correlation structure with mixed gev models, *in* R. Scarpa & A. Alberini, eds, 'Applications of simulation methods in environmental and resource economics', Kluwer Academic Publisher, Boston, MA.

Hess, S., Bierlaire, M. & Polak, J. W. (2005*b*), 'Estimation of value of travel-time savings using mixed logit models', *Transportation Research* **39A**(3), 221–236.

Hess, S. & Polak, J. W. (2004), *Mixed Logit estimation of parking type choice*, paper presented at the 83^{rd} Annual Meeting of the Transportation Research Board, Washington, DC.

Hess, S. & Polak, J. W. (2005), 'Mixed logit modelling of airport choice in multi-airport regions', *Journal of Air Transport Management* **11**(2), 59–68.

Hess, S., Train, K. & Polak, J. W. (2005), *On the use of a modified latin hypercube sampling (mlhs) method in the estimation of a mixed logit model for vehicle choice*, Transportation Research B, forthcoming.

Kamakura, K. W. & Russell, G. (1989), 'A probabilistic choice model for market segmentation and elasticity structure', *Journal of Marketing Research* **26**, 379–390.

Koenig, A., Abay, G. & Axhausen, K. (2004), Zeitkostenansaetze im personenverkehr, Schriftenreihe 1065, Bundesamt für Strassen, UVEK, Bern. final report for SVI 2001/534.

McFadden, D. & Train, K. (2000), 'Mixed MNL Models for discrete response', *Journal of Applied Econometrics* **15**, 447–470.

Revelt, D. & Train, K. (1998), 'Mixed Logit with repeated choices: households' choices of appliance efficiency level', *Review of Economics and Statistics* **80**(4), 647–657.

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

Walker, J. L. (2001), Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables, PhD thesis, MIT, Cambridge, MA.