Generating Synthetic Populations using Iterative Proportional Fitting (IPF) and Monte Carlo Techniques

Martin Frick, IVT, ETH Zürich
Kay W. Axhausen, IVT ETH Zürich

Conference paper STRC 2003
Session Model & Generation

STRC | 3rd Swiss Transport Research Conference
Monte Verità / Ascona, March 19-21, 2003

Generating Synthetic Populations using Iterative Proportional Fitting (IPF) and Monte Carlo Techniques

Martin Frick
IVT
ETH Zürich
8093 Zürich

Phone: 01-633 37 13
Fax: 01-633 10 57
e-Mail: frick@ivt.baug.ethz.ch

Abstract

The generation of synthetic populations represents a substantial contribution to the acquisition of useful data for large scale multi agent based microsimulations in the field of transport planning. Basically, the observed data is available from censuses (microcensus) in terms of simple summary tables of demographics, such as the number of persons per household for census block group sized areas. Nevertheless, there is a need of more disaggregated personal data and thus another data source is considered. The Public Use Sample (PUS), often used in transportation studies, is a 5% representative sample of almost complete census records for each individual, missing addresses and unique identifiers, including missing items. The problem is, to generate a large number of individual agents (~ 1Mio.) with appropriate characteristic values of the demographic variables for each agent, interacting in the microsimulation. Due to the fact that one is faced with incomplete multivariate data it is useful to consider multiple data imputation techniques. In this paper an overview is given over existing methods such as Iterative Proportional Fitting (IPF), the Maximum Likelihood method (MLE) and Monte-Carlo (MC) techniques. To exemplify the generation of synthetic populations the Swiss PUS's from 1970, 1980 and 1990 as well as Swiss microcensus data is used to generate a large number (~1.5 Mio) of synthetic agents, representing the people in the Glattal, which is a fast developing area north of Zürich. The data sources are fused together using IPF and Monte Carlo technique to estimate the joint distribution of demographic characteristics such as age, sex, car ownership, drivers licence ownership, vehicle miles, car availability, etc. for each agent. In a further step the generation of representative households for the Glattal area is addressed.

Keywords

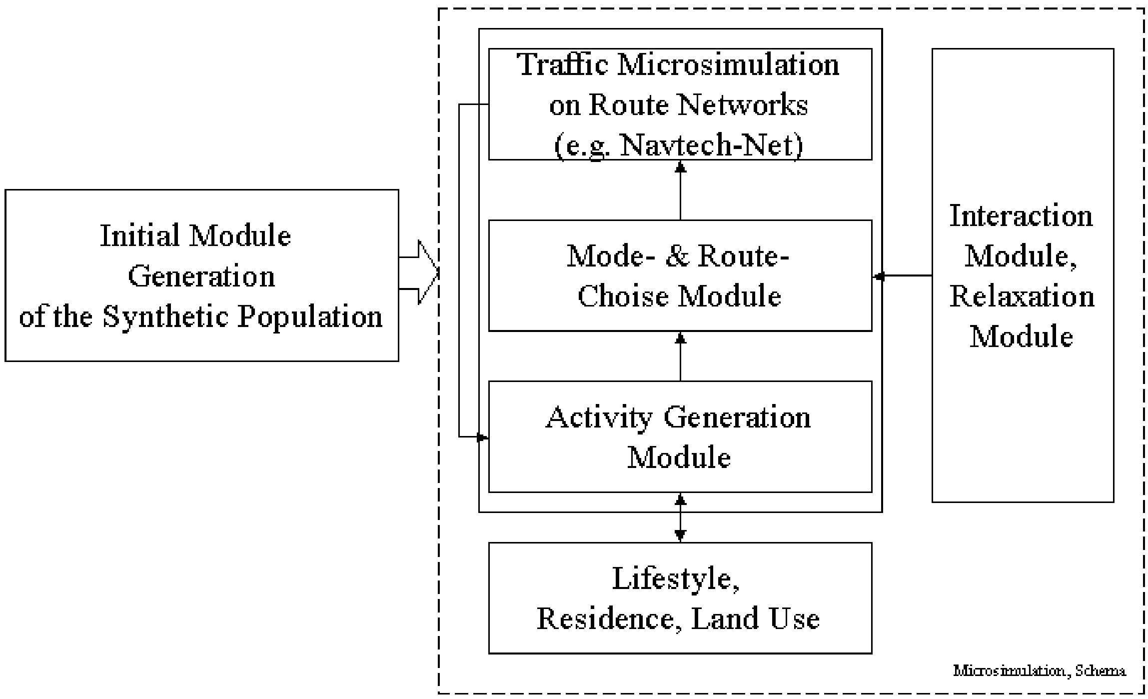
Synthetic Population Generation – IPF – MCMC – Imputing missing Data – Glattal – 3rd Swiss Transport Research Conference – STRC 2003 – Monte Verità

1. Introduction

In recent years, the computing power of distributed computing has increased significantly. Therefore it is now possible to do traffic simulations on a microscopic scale for large metropolitan areas with more than 1 million habitants. Microscopic means that each entity in the simulation is an object, e.g. travellers, roads, vehicles, traffic lights, etc. The TRANSIMS system, e. g. explained in Smith et al., (1995) is an existing implementation. For an implementation for Switzerland see Voellmy et al. (2001). For a 1 day simulation of 1 million intelligent agents, one need about 5 hours computing time on a cluster with 64 CPU's, from scratch.

For the microsimulation one needs different modules for different tasks. Figure 1 shows the most important modules.

Figure 1 Traffic Microsimulation, Schema



The reason why people move from one point in space and time to another is assumed to be that people starts activities due to personal needs. For details about activity based transport models see for example Axhausen and Gärling, 1991 or Gärling et al., 1994. Therefore in the simulation there is a module called activity generation module, which provides each agent with a daily plan of activities witch should be executed. Once the agent has the activities for the day, the mode- and route choice module relates a particular route and vehicle modes to each agent depending on the individual activities. The next step is to execute these choices for all agents simultaneously within the traffic microsimulation module. The algorithm behind the traffic microsimulation module is similar to molecular dynamics method, well known from physics, except for the fact that the agents (particles) are simulated on a road net. If an inconsistency occurs within the daily plans of some agents, the interaction module is responsible for adapting the plans and run the procedure as long as all agent plans are consistent. This is usually the case after about 50 iterations. In a first and important step it is necessary to generate the needed number of agents with individual demographic properties as an initial dataset for the microsimulation. This module is called Synthetic Population Generation Module. The purpose of this paper is to describe in detail the IPF method used for synthetic population generation and to present the obtained first results for the Kanton Zürich in Switzerland.

The task is to generate agents, which reflects the demographic structure and the travel behaviour of the real people. Almost all information, which is easily available about people and their travel behaviour, comes from empirical data obtained by surveys. Due to the fact that there exist laws to protect the privacy of people, not all desired information about people could be obtained. In general the situation is as follows: The more information we get from an individual traveller about his demographics (e.g. age, sex, income, employment, car ownership, etc.) the less we know about his residency, i.e. the spatial resolution of where he or she lives. On the other hand, if we have a good spatial resolution for the e.g. residence of the traveller, i.e. the municipality, the hectare data or the street address of that particular traveller, we do not have information about the other sociodemographics of that individual traveller but usually we have 1D or 2D summary tables for various demographics instead.

To overcome the situation described above, one can use methods for disaggregating and synchronizing the data given to public. One method, which has been used in traffic research before is called Iterative Proportional Fitting (IPF), e.g. (Papacostas and Prevedouros, 1993). The present paper describe some modifications to the original method, and shows how high dimensional contingency tables can be obtained with this method. As an example the method is adapted to the Kanton Zürich, witch is a metropolitan area with about 1 million residents in Switzerland. First results will be presented.

2. Census Data

One of the main problems in generating synthetic populations is the spatial disaggragation and synchronization of the available census data. In this approach the census 2000 from Switzerland as well as the microcensuses from 1989 to 2000 (BFS and ARE, 2001) are used to determine the synthetic population.

First of all, it is necessary to choose a set of demographic variables of interest. For example this may be residence, age, sex, income, car ownership, working place, etc. These demographics are available in different resolutions in each data source. In particular the residence varies from Kanton level down to municipality level and even hectare grid or exact xy-coordinates are possible. To obtain the best spatial resolution for residence or work place the process of generating synthetic agents is divided into two major steps.

Starting from the 1-dimensional (1D) marginal distributions of the demographics for each municipality the data for each demographic variable is aggregated to the low spatial resolution level which is here the Kanton Zürich, to obtain 1D, 2D, etc. distributions for each demographic or a combination of demographics except for the residence. The Tables 1-4 show the 1D distribution of age and sex for the municipality Aeugst am Albis and the aggregated results for the Kanton Zürich, respectively. More tables for other demographics and cross-classified tables are possible but not presented here. Furthermore, Figure 2 and 3 shows a chart of the aggregated distributions for age and sex.

Table 1 Sex Distribution for the Municipality Aeugst am Albis, 2000

Sex	1 = male	2 = female
Frequency	771	773

Table 2 Age distribution for the Municipality Aeugst am Albis, 2000

Age (categories)	1	2	3	4	5	6
Frequency	17	67	104	109	78	75
Age (categories)	7	8	9	10	11	12
Frequency	58	108	132	184	147	121
Age (categories)	13	14	15	16	17	18
Frequency	109	75	59	44	35	10
Age (categories)	19	20	21	22	23	
Frequency	10	2	0	0	0	

Labels 1 = 0 years, 2 = 1-4 years, 3 = 5-9 years,…, 22 = 100-104 years, 23 = >105 years

Table 3 Age distribution for the Kanton Zürich, 2000

Age (categories)	1	2	3	4	5	6
Frequency	11593	51128	64662	64225	64684	77601
Age (categories)	7	8	9	10	11	12
Frequency	93278	110767	111411	96505	86962	86131
Age (categories)	13	14	15	16	17	18
Frequency	79676	61763	53412	46781	38128	24765
Age (categories)	19	20	21	22	23	
Frequency	15942	6984	1403	103	2	

Labels 1 = 0 years, 2 = 1-4 years, 3 = 5-9 years,…, 22 = 100-104 years, 23 = >105 years

Table 4 Sex Distribution for the Kanton Zürich, 2000

Sex	1 = male	2 = female
Frequency	613038	634868

Figure 2 Age Distribution, Kanton Zürich, 2000

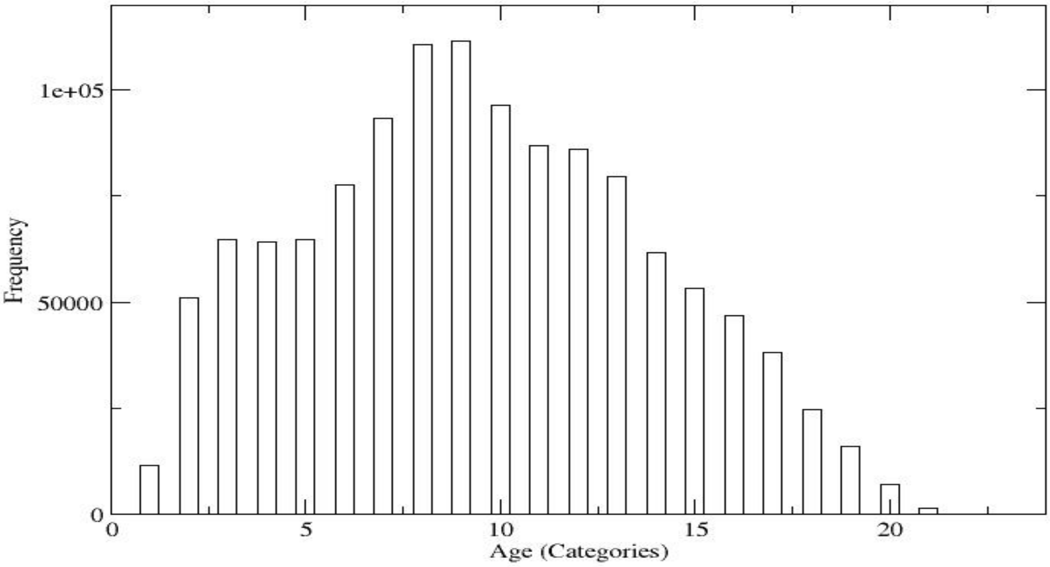
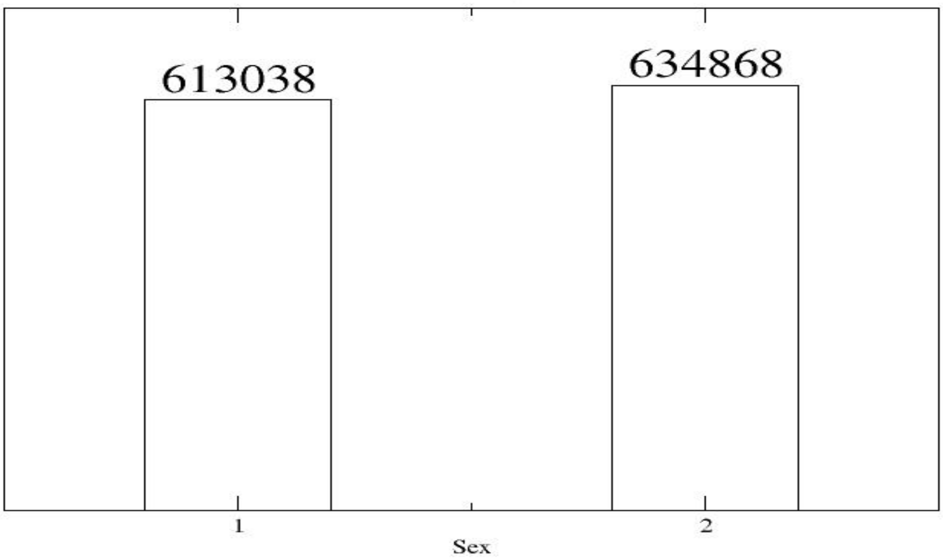


Figure 3 Sex Distribution, Kanton Zürich, 2000



Another data source, we can use for estimation of the multiway table, is usually the Public Use Sample (PUS, 2001) from Switzerland. This is a 5% sample from the census from Switzerland in an anonymized form. Any data imputation techniques for missing items are not performed so far, but will be one of the next steps. When the presented calculation was done, the PUS 2000 was not available for public, but fortunately a preversion for research purposes was. From that data source we can get a 5% multiway table for age x sex sociodemographic variables. We simply need to summarize all people with the same values of the demographic variables. Unfortunately this table is only valid for the 5% sample. Since we know at least the real 1D-distributions of the demographic variables for the entire population and the complete multiway table for the 5% sample, we can use methods like Iterative Proportional Fitting to obtain estimations for the multiway table valid for the entire population. Once the multiway table is generated, we simply have to choose agents from that distribution with the appropriate probability in order to get the complete synthetic population. In the next chapter the IPF method is described in detail.

3. Iterative Proportional Fitting

3.1 The Basic Algorithm

IPF was first established by Deming and Stephan, (1940). This chapter is based on Anderson (1997). Let us consider a multiway table in N dimensions. Each dimension is associated with one sociodemographics. For ease of description, the algorithm is presented for N=3. It's easy to expand the algorithm to N dimensions.

Assume a multiway table $\pi_{ijk} \in R^{+}$ with unknown components and a set of given marginal distributions $\{x_{ij\bullet}, x_{i\bullet k}, x_{\bullet jk}\}$. A dot means that we have summed over that index. π_{ijk} denote the number of observations of category i of the first sociodemographic, of category j of the second demographics and so on. The multiway table π_{ijk} have to obey the following constraints:

$$n\pi_{ij\bullet} = x_{ij\bullet}, \quad n\pi_{i\bullet k} = x_{i\bullet k}, \quad n\pi_{\bullet jk} = x_{\bullet jk} \tag{1.1}$$

$$n = \pi_{\bullet\bullet\bullet} = x_{\bullet\bullet\bullet} \tag{1.2}$$

were n is the total sum of observations. The iteration process begins by letting the multiway table for the 5% sample the 0th estimate $\pi_{ijk}^{(0)}$ for the π_{ijk} . One iteration (in 3 dimensions) is done by executing the following equations in turn.

$$\pi_{ijk}^{(1)} = \frac{1}{n}\frac{x_{ij\bullet}\pi_{ijk}^{(0)}}{\pi_{ij\bullet}^{(0)}} \tag{2.1}$$

$$\pi_{ijk}^{(2)} = \frac{1}{n}\frac{x_{i\bullet k}\pi_{ijk}^{(1)}}{\pi_{i\bullet k}^{(1)}} \tag{2.2}$$

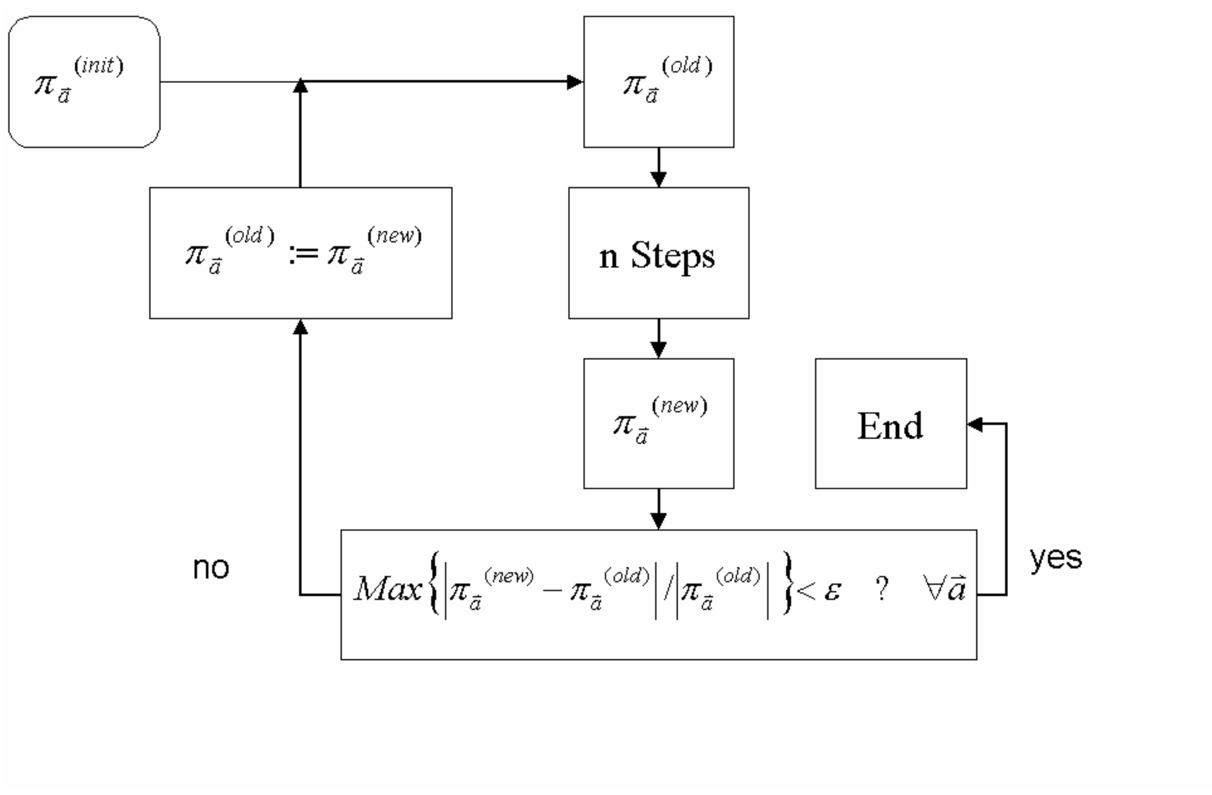
$$\pi_{ijk}^{(3)} = \frac{1}{n}\frac{x_{\bullet jk}\pi_{ijk}^{(2)}}{\pi_{\bullet jk}^{(2)}} \tag{2.3}$$

(2.1)–(2.3): one iteration (in 3 dimensions)

This is necessary to fulfil the constraints given in equation (1.1). For a prove, that the constraints in equation (1.1) are fulfilled, one can for example sum over the index k in equation (2.1) on both sides to make sure, that this leads to the first equation in (1.1).

The iterations continue until the relative change between iterations in each estimated $\pi_{ijk}(t)$ is small. t denotes the number of iterations executed. This procedure converges sufficiently in about 20 iterations. Figure 2 shows a schema for the iteration process. $\vec{a}$ denotes a vector of the demographics.

Figure 4 Iteration Process



3.2 Description of the Enhanced IPF Procedure

The basic algorithm is used in two major steps. In the first step, a multiway table for the whole low spatial resolution area is computed, e.g. a multiway table for the Kanton Zürich. In a second step this estimation is uses together with the marginal distributions of the areas of high spatial resolution, e. g. the municipalities of the Kanton Zürich to obtain the right marginals for each municipality. In both steps, the basic IPF procedure is executed many times, depending on the number of demographics used for the estimation. In this chapter we will illustrate which IPF procedures are necessary.

Consider the case of four demographic variables. And let (1,0,1,1) denote a vector with elements 0 or 1. Each component of the vector corresponds to a different demographic variable. A '1' means that this demographic variable exists whereas a '0' means that the demographic variables doesn't exists. And thus (1,0,1,1) describe the 3d multiway table: demographic1 x demographic3 x demographic4. The demographic2 doesn't exist. With that notation a complete basic IPF run can be described by giving the known marginal tables on the left hand side and the estimated table on the right side separated by a '=>' character. An example is shown in equation (3.1).

$$(1,1,1,0);\ (1,1,0,1);\ (1,0,1,1);\ (0,1,1,1) \quad \Rightarrow \quad (1,1,1,1) \qquad (3.1)$$

The complete IPF runs for a major step can be stated as a list of equations like the equation (3.1). All possible runs for three demographics are shown in equation (3.2).

$$\begin{array}{lcl} (1,0,0);\ (0,1,0) & \Rightarrow & (1,1,0) \\ (1,0,0);\ (0,0,1) & \Rightarrow & (1,0,1) \\ (0,1,0);\ (0,0,1) & \Rightarrow & (0,1,1) \\ (1,1,0);\ (1,0,1);\ (0,1,1) & \Rightarrow & (1,1,1) \end{array} \qquad (3.2)$$

As one can see, if a particular marginal distribution is given, it can be used as the resulting estimation of this particular IPF run and thus the run need not to be executed. At least the information about the 1D distributions is needed to obtain a final result.

3.3 Properties

For mathematical tractability purposes it is assumed that all of the areas considered for the IPF runs have the same correlation structure. This assumption is probably not a serious problem, due to relationships between the different areas under consideration.

The reason, why it is necessary to make a second IPF step for the spatial disaggrgation is the correlation structure of the demographics in the population. When we use IPF, the correlation structure in each municipality is the same, and a summation of the municipalities to the Kanton level yields to the correlation structure of the Kanton.

It can be shown that if the marginal totals of a multiway table are known and a sample from the population, which generated these marginals, is given, IPF gives a constrained maximum entropy estimate of the true proportions in the complete population multiway table (Ireland and Kullback, 1968).

To start the iteration process, we need an initial multiway table, generated e. g. from the PUS. The IPF algorithm estimates zero cells for all cells that are zero in the sample. Since we have only samples of the population available to construct the initial distribution, not all of the zero cells in the multiway table might be zero in the population. Therefore it is useful to allocate all zero cells which are not zero due to logical constraints with a small value, e.g. 0.001.

4. First Results for the Kanton Zürich

In this chapter, the generation of approximately one million agents is presented. The data sources used for that run are the census data from 2000 (not published yet) from Switzerland for the generation of the synthetic population and the traffic microcensus data from 1974-2000 for the mapping of the activities and journeys to the agents. Due to the available categories in the census data, we use only two categories here: the age of the persons and the sex of the persons.

With the notation from above for the IPF runs, the first task is to build the multiway table (1,1) from the given marginal distributions (1,0) and (0,1) for the Kanton Zürich.

$$(1,0); (0,1) \Rightarrow (1,1) \qquad (4.1)$$

(1,0) is given by Table 3, (0,1) is given by Table 4. (1,1) is obtained by running the basic IPF procedure once using the 5% sample values for the initial (1,1) multiway table. Table 5 shows the result for the Kanton Zürich.

Table 5 Result of the IPF Procedure for Age x Sex for the Kanton Zürich

Age (Categories)	1 -Male	2 - Female
1	6745	4848
2	29882	21246
3	37286	27376
4	36268	27957
5	35773	28911
6	42132	35469
7	49533	43745
8	57257	53510
9	55523	55888
10	46195	50310
11	40248	46714
12	38771	47360
13	34865	44810
14	26265	35499
15	22260	31152
16	19171	27610
17	15390	22738
18	9858	14907
19	6289	9653
20	2737	4247
21	548	855
22	40	63
23	1	1

For the labels see Table 3 & 4

To disaggregate this data to the municipality level, a second IPF run is made. The desired result is the (1,1,1) multiway table. The third category is the number of the municipality in the Kanton Zürich. With that notation, (1,1,0) is the result from the first IPF run, shown in Table 5. (1,0,1) are the marginal age distributions for each municipality and the (0,1,1) table is the marginal sex distribution for each municipality. Parts of that data is shown in Appendix A. Therefore the following IPF run has to be executed:

$$(1,0,1);\ (0,1,1);\ (1,1,0) \quad \Rightarrow \quad (1,1,1) \qquad (4.2)$$

This yields in a multiway table for age x sex x number of the municipality which can be separated in a set of multiway tables age x sex for each municipality. Then the spatial disaggregation is done. To produce the population, one can easily choose the available category combinations with the obtained frequencies.

Another way to produce the population is done as follows: Let {5,1,1} be a certain choice of the category combinations. Here it means that the agents have age category 5, i.e. the agent is between 15 and 19 years old, that the agents have sex category 1, i.e. the agent is a male and that the agent live in the municipality number 1, which is Aeugst am Albis. Refer to the Appendix B for all municipality labels from Kanton Zürich. The number of agents with these particular properties is given from the IPF run and in that case the result is 12.

One can assume, that for example the age in one category is equally distributed, and therefore one can choose a random number between one and five, to get the age of the agent with a one year resolution instead of a five year resolution.

We have extracted activities and journeys from the traffic microcensus data from the Kanton Zürich according to the categories available in the synthetic population. These activities can be mapped randomly to the agents. The final result is a xml file with entries for each individual agent, its soziodemographics and its activities as well as its journeys. In Appendix C a small part of this file with a few agents, their activities and plans, is presented.

5. Discussion and further work

In the present work, the generation of a synthetic population for the Kanton Zürich is presented. The first result include only two sociodemographic variables, the age and the sex of the people, but the method presented to generate the population is more general in terms of the possible sociodemographic variables which can be calculated. We currently try to include all variables mentioned in the introducing chapter.

With the approach of high dimensional IPF it is possible to include different types of data in terms of the dimension of cross-classified data available from censuses.

The present paper is only a first approach in the Generating of Synthetic Populations. The next step is to generate a set of households according to the estimated multiway table (Beckman et al., 1996).

Acknowledgement - This work is part of a project "Large scale multi-agent simulation of travel behaviour and traffic flow", which is sponsored by the ETH Zürich.

6. References

Smith L., Beckman R., Anson D., Nagel K. and Williams M. (1995) TRANSIMS: Transportation SIMulation System. Los Alamos National Laboratory Unclassified Report, LA-UR-95-1664, Los Alamos, NM 8744.

Axhausen, K. W. and Gärling T. (1991) Activity-based Approaches to Travel Analysis: Conceptual Frameworks, Models and Research Problems. U.S. Department of Commerce, National Technical Information Service, TSU Ref:628.

Gärling T., Kwan M. and Colledge R. G. (1994) Computational-process modeling of household activity scheduling. Transportation Research, B 28, 355-364.

Papacostas C. S. and Prevedouros P. D. (1993) Transportation Engineering and Planning, (2ed Edition). Prentice Hall, Englewood Cliffs, NJ.

Anderson, E.B. (1997) Introduction to the Statistical Analysis of Categorical Data, Springer, Berlin

Deming, W. E. and Stephan, F. F. (1940) On a least squares adjustment of a sampled frequency table when the expected marginal tables are known. Annals of Mathematical Statistics 11, 427-444.

Voellmy, A., M. Vrtic, B. Raney, K.W. Axhausen, K. Nagel (2001) *Status of a TRANSIMS implementation for Switzerland*, Networks and Spatial Economics

Beckman, R. J., K.A. Baggerly und M. D. McKay, (1996) *Creating Synthetic Baseline Populations*, Transportation Research-A, **30 (**6), 415-429

Ireland C. T. and Kullback S. (1968) Contingency tables with given marginals. Biometrica 55, 179-188

Bundesamt für Statistik (BFS) / Bundesamt für Raumentwicklung (ARE), (2001), Mikrozensus Verkehrsverhalten 2000, Bundesamt für Statistik, Bern, 2001.

Bundesamt für Statistik (BFS) / Bundesamt für Raumentwicklung (ARE), (1995), Mikrozensus Verkehrsverhalten 1994, Bundesamt für Statistik, Bern, 1995.

Bundesamt für Statistik (BFS) / Bundesamt für Raumentwicklung (ARE), (1990), Mikrozensus Verkehrsverhalten 1989, Bundesamt für Statistik, Bern, 1990.

Bundesamt für Statistik (BFS), (2001), Public Use Samples 1970, 1980, 1990, Bundesamt für Statistik, Bern, 2001.

Appendix A: Parts of the 1D Marginal Distributions for the Municipalities

The inner bracket gives the categories, followed by the frequency of occurrence.

A 1: The (1,0,1) Marginal Tables for age category 1

{{1, 1}, 17}, {{1, 2}, 93}, {{1, 3}, 51}, {{1, 4}, 28}, {{1, 5}, 29},
{{1, 6}, 5}, {{1, 7}, 12}, {{1, 8}, 8}, {{1, 9}, 50}, {{1, 10}, 47},
{{1, 11}, 18}, {{1, 12}, 7}, {{1, 13}, 30}, {{1, 14}, 30}, {{1, 15}, 6},
{{1, 16}, 3}, {{1, 17}, 1}, {{1, 18}, 5},
{{1, 19}, 15}, {{1, 20}, 9}, {{1, 21}, 28}, {{1, 22}, 11}, {{1, 23}, 14},
{{1, 24}, 17}, {{1, 25}, 22}, {{1, 26}, 9}, {{1, 27}, 16}, {{1, 28}, 15},
{{1, 29}, 17}, {{1, 30}, 10}, {{1, 31}, 12}, {{1, 32}, 11}, {{1, 33}, 5},
{{1, 34}, 6}, {{1, 35}, 5}, {{1, 36}, 5},
{{1, 37}, 4}, {{1, 38}, 4}, {{1, 39}, 32}, {{1, 40}, 70}, {{1, 41}, 127},
{{1, 42}, 86}, {{1, 43}, 18}, {{1, 44}, 63}, {{1, 45}, 20}, {{1, 46}, 26},
{{1, 47}, 21}, {{1, 48}, 29}, {{1, 49}, 9}, {{1, 50}, 158}, {{1, 51}, 13},
{{1, 52}, 34}, {{1, 53}, 11}, {{1, 54}, 106}, {{1, 55}, 36}, {{1, 56}, 23},
{{1, 57}, 101}, {{1, 58}, 4}, {{1, 59}, 15}, {{1, 60}, 36}, {{1, 61}, 5},
{{1, 62}, 16}, {{1, 63}, 42}, {{1, 64}, 34}, {{1, 65}, 24}, {{1, 66}, 60},
{{1, 67}, 2}, {{1, 68}, 20}, {{1, 69}, 37},
{{1, 70}, 90}, {{1, 71}, 22}, {{1, 72}, 46}, {{1, 73}, 10}, {{1, 74}, 18},
{{1, 75}, 4}, {{1, 76}, 151}, {{1, 77}, 50}, {{1, 78}, 5}, {{1, 79}, 12},
{{1, 80}, 18}, {{1, 81}, 20}, {{1, 82}, 9}, {{1, 83}, 41}, {{1, 84}, 47},
{{1, 85}, 63}, {{1, 86}, 27}, {{1, 87}, 100}, {{1, 88}, 34}, {{1, 89}, 91},
{{1, 90}, 99}, {{1, 91}, 12}, {{1, 92}, 67}, {{1, 93}, 185}, {{1, 94}, 127},
{{1, 95}, 27}, {{1, 96}, 176}, {{1, 97}, 13}, {{1, 98}, 54}, {{1, 99}, 56},
{{1, 100}, 42}, {{1, 101}, 95}, {{1, 102}, 51}, {{1, 103}, 25},
{{1, 104}, 142}, {{1, 105}, 200}, {{1, 106}, 39}, {{1, 107}, 63},
{{1, 108}, 70}, {{1, 109}, 70}, {{1, 110}, 81}, {{1, 111}, 99},
{{1, 112}, 47}, {{1, 113}, 100}, {{1, 114}, 57}, {{1, 115}, 34},
{{1, 116}, 104}, {{1, 117}, 38}, {{1, 118}, 55}, {{1, 119}, 39},
{{1, 120}, 124}, {{1, 121}, 7}, {{1, 122}, 35}, {{1, 123}, 81},
{{1, 124}, 32}, {{1, 125}, 4}, {{1, 126}, 19},
{{1, 127}, 17}, {{1, 128}, 9}, {{1, 129}, 170}, {{1, 130}, 74},
{{1, 131}, 61}, {{1, 132}, 50}, {{1, 133}, 90}, {{1, 134}, 18},
{{1, 135}, 65}, {{1, 136}, 288}, {{1, 137}, 154}, {{1, 138}, 39},
{{1, 139}, 7}, {{1, 140}, 6}, {{1, 141}, 20}, {{1, 142}, 16}, {{1, 143}, 5},
{{1, 144}, 12}, {{1, 145}, 31}, {{1, 146}, 6}, {{1, 147}, 35},
{{1, 148}, 5}, {{1, 149}, 24}, {{1, 150}, 3}, {{1, 151}, 41},
{{1, 152}, 29}, {{1, 153}, 19}, {{1, 154}, 4},
{{1, 155}, 78}, {{1, 156}, 42}, {{1, 157}, 39}, {{1, 158}, 847},
{{1, 159}, 35}, {{1, 160}, 9}, {{1, 161}, 61}, {{1, 162}, 210},
{{1, 163}, 32}, {{1, 164}, 52}, {{1, 165}, 14}, {{1, 166}, 141},
{{1, 167}, 24}, {{1, 168}, 12}, {{1, 169}, 90}, {{1, 170}, 38},
{{1, 171}, 3191}...

A2: The (0,1,1) Marginal Tables for municipality 1 and 2

{{1, 1}, 771}, {{1, 2}, 254200/50},
{{1, 3}, 1957}, {{1, 4}, 1519},
{{1, 5}, 1474}, {{1, 6}, 445},
{{1, 7}, 737}, {{1, 8}, 6875/25},
{{1, 9}, 1875}, {{1, 10}, 105700/50},
{{1, 11}, 26750/25}, {{1, 12}, 8600/25},
{{1, 13}, 1349}, {{1, 14}, 1861},
{{1, 15}, 7525/25},
{{1, 16}, 8700/25}, {{1, 17}, 300},
{{1, 18}, 9025/25}, {{1, 19}, 19200/25},
{{1, 20}, 7275/25}, {{1, 21}, 71900/50},
{{1, 22}, 592}, {{1, 23}, 15425/25},
{{1, 24}, 821}, {{1, 25}, 790},
{{1, 26}, 5325/25}, {{1, 27}, 892},
{{1, 28}, 695}, {{1, 29}, 44100/50},
{{1, 30}, 524}, {{1, 31}, 642},
{{1, 32}, 862}, {{1, 33}, 326},
{{1, 34}, 509}, {{1, 35}, 4140/20},
{{1, 36}, 21400/50}, {{1, 37}, 143},
{{1, 38}, 8925/25}, {{1, 39}, 1576},
{{1, 40}, 187100/50}, {{1, 41}, 342100/50},
{{1, 42}, 153850/50}, {{1, 43}, 1462},
{{1, 44}, 3944}, {{1, 45}, 53300/50},
{{1, 46}, 1751}, {{1, 47}, 806},
{{1, 48}, 29200/25}, {{1, 49}, 9825/25},
{{1, 50}, 8767}, {{1, 51}, 14850/25},
{{1, 52}, 113850/50}, {{1, 53}, 492},
{{1, 54}, 612000/100},
{{1, 55}, 1698},
{{1, 56}, 1099}, {{1, 57}, 296650/50},
{{1, 58}, 273}, {{1, 59}, 673},
{{1, 60}, 1715}, {{1, 61}, 7100/25},
{{1, 62}, 530}, {{1, 63}, 2136},
{{1, 64}, 42375/25}, {{1, 65}, 905},
{{1, 66}, 123050/50}, {{1, 67}, 5720/20},
{{1, 68}, 1218},
{{1, 69}, 1859}, {{1, 70}, 192300/50},
{{1, 71}, 27175/25}, {{1, 72}, 64500/25},
{{1, 73}, 15900/25}, {{1, 74}, 927},
{{1, 75}, 6075/25}, {{1, 76}, 7791},
{{1, 77}, 2770}, {{1, 78}, 7900/25},
{{1, 79}, 13550/25}, {{1, 80}, 871},
{{1, 81}, 1426}, {{1, 82}, 498},
{{1, 83}, 104450/50}, {{1, 84}, 2687},
{{1, 85}, 153000/50}, {{1, 86}, 978},
{{1, 87}, 4328}, {{1, 88}, 75600/50},
{{1, 89}, 229350/50}, {{1, 90}, 266500/50},
{{1, 91}, 16900/25}, {{1, 92}, 205200/50},
{{1, 93}, 443700/50}, {{1, 94}, 391850/50},
{{1, 95}, 950}, {{1, 96}, 427100/50},
{{1, 97}, 442}, {{1, 98}, 3395},
{{1, 99}, 162950/50}, {{1, 100}, 109400/50},
{{1, 101}, 249750/50},
{{1, 102}, 114350/50},
{{1, 103}, 899},
{{1, 104}, 7645},
{{1, 105}, 473350/50},
{{1, 106}, 110650/50},
{{1, 107}, 133800/50},
{{1, 108}, 177350/50},
{{1, 109}, 287950/50},
{{1, 110}, 197400/50},
{{1, 111}, 271150/50},
{{1, 112}, 2169},
{{1, 113}, 279600/50},
{{1, 114}, 121900/50},
{{1, 115}, 2198},
{{1, 116}, 267350/50},
{{1, 117}, 106700/50},
{{1, 118}, 116950/50},
{{1, 119}, 1481},
{{1, 120}, 357400/50}, {{1, 121}, 5150/25},
{{1, 122}, 2095},
{{1, 123}, 236450/50},
{{1, 124}, 2007}, {{1, 125}, 3560/20},
{{1, 126}, 1411},
{{1, 127}, 914}, {{1, 128}, 10975/25},
{{1, 129}, 549350/50},
{{1, 130}, 179600/50},
{{1, 131}, 3180},
{{1, 132}, 64675/25}, {{1, 133}, 213100/50},
{{1, 134}, 79800/50}, {{1, 135}, 52750/25},
{{1, 136}, 704050/50},
{{1, 137}, 354600/50},
{{1, 138}, 151950/50},
{{1, 139}, 310}, {{1, 140}, 25250/50},
{{1, 141}, 22775/25}, {{1, 142}, 11150/25},
{{1, 143}, 6700/25}, {{1, 144}, 668},
{{1, 145}, 1812}, {{1, 146}, 9475/25},
{{1, 147}, 74500/50}, {{1, 148}, 13950/25},
{{1, 149}, 1159}, {{1, 150}, 5375/25},
{{1, 151}, 2245},
{{1, 152}, 63900/50},
{{1, 153}, 1027},
{{1, 154}, 320},
{{1, 155}, 163950/50},
{{1, 156}, 101500/50},
{{1, 157}, 1993},
{{1, 158}, 4378800/100},
{{1, 159}, 113200/50}, {{1, 160}, 12025/25},
{{1, 161}, 137850/50},
{{1, 162}, 1063000/100},
{{1, 163}, 2283},
{{1, 164}, 2903},
{{1, 165}, 27500/25}, {{1, 166}, 335900/50},
{{1, 167}, 1766},
{{1, 168}, 1399},
{{1, 169}, 233750/50}, {{1, 170}, 97400/50},

{{1, 171}, 175828}

Appendix B: The Municipality Numbers and Labels for the Kanton Zürich

1	0001	Aeugst am Albis
2	0002	Affoltern am Albis
3	0003	Bonstetten
4	0004	Hausen am Albis
5	0005	Hedingen
6	0006	Kappel am Albis
7	0007	Knonau
8	0008	Maschwanden
9	0009	Mettmenstetten
10	0010	Obfelden
11	0011	Ottenbach
12	0012	Rifferswil
13	0013	Stallikon
14	0014	Wettswil am Albis
15	0021	Adlikon
16	0022	Benken (ZH)
17	0023	Berg am Irchel
18	0024	Buch am Irchel
19	0025	Dachsen
20	0026	Dorf
21	0027	Feuerthalen
22	0028	Flaach
23	0029	Flurlingen
24	0030	Andelfingen
25	0031	Henggart
26	0032	Humlikon
27	0033	Kleinandelfingen
28	0034	Laufen-Uhwiesen
29	0035	Marthalen
30	0036	Oberstammheim
31	0037	Ossingen
32	0038	Rheinau
33	0039	Thalheim an der Thur
34	0040	Trüllikon
35	0041	Truttikon
36	0042	Unterstammheim
37	0043	Volken
38	0044	Waltalingen
39	0051	Bachenbülach
40	0052	Bassersdorf
41	0053	Bülach
42	0054	Dietlikon
43	0055	Eglisau
44	0056	Embrach
45	0057	Freienstein-Teufen
46	0058	Glattfelden
47	0059	Hochfelden
48	0060	Höri
49	0061	Hüntwangen
50	0062	Kloten
51	0063	Lufingen
52	0064	Nürensdorf
53	0065	Oberembrach
54	0066	Opfikon
55	0067	Rafz
56	0068	Rorbas
57	0069	Wallisellen
58	0070	Wasterkingen
59	0071	Wil (ZH)
60	0072	Winkel
61	0081	Bachs
62	0082	Boppelsen
63	0083	Buchs (ZH)
64	0084	Dällikon
65	0085	Dänikon
66	0086	Dielsdorf
67	0087	Hüttikon
68	0088	Neerach
69	0089	Niederglatt
70	0090	Niederhasli
71	0091	Niederweningen
72	0092	Oberglatt
73	0093	Oberweningen
74	0094	Otelfingen
75	0095	Regensberg
76	0096	Regensdorf
77	0097	Rümlang
78	0098	Schleinikon
79	0099	Schöfflisdorf
80	0100	Stadel
81	0101	Steinmaur
82	0102	Weiach
83	0111	Bäretswil
84	0112	Bubikon
85	0113	Dürnten
86	0114	Fischenthal
87	0115	Gossau (ZH)
88	0116	Grüningen
89	0117	Hinwil
90	0118	Rüti (ZH)
91	0119	Seegräben
92	0120	Wald (ZH)
93	0121	Wetzikon (ZH)
94	0131	Adliswil
95	0132	Hirzel
96	0133	Horgen
97	0134	Hütten
98	0135	Kilchberg (ZH)

99 0136 Langnau am Albis
100 0137 Oberrieden
101 0138 Richterswil
102 0139 Rüschlikon
103 0140 Schönenberg (ZH)
104 0141 Thalwil
105 0142 Wädenswil
106 0151 Erlenbach (ZH)
107 0152 Herrliberg
108 0153 Hombrechtikon
109 0154 Küsnacht (ZH)
110 0155 Männedorf
111 0156 Meilen
112 0157 Oetwil am See
113 0158 Stäfa
114 0159 Uetikon am See
115 0160 Zumikon
116 0161 Zollikon
117 0171 Bauma
118 0172 Fehraltorf
119 0173 Hittnau
120 0174 Illnau-Effretikon
121 0175 Kyburg
122 0176 Lindau
123 0177 Pfäffikon
124 0178 Russikon
125 0179 Sternenberg
126 0180 Weisslingen
127 0181 Wila
128 0182 Wildberg
129 0191 Dübendorf
130 0192 Egg
131 0193 Fällanden
132 0194 Greifensee
133 0195 Maur
134 0196 Mönchaltorf
135 0197 Schwerzenbach
136 0198 Uster
137 0199 Volketswil
138 0200 Wangen-Brüttisellen
139 0211 Altikon
140 0212 Bertschikon
141 0213 Brütten
142 0214 Dägerlen
143 0215 Dättlikon
144 0216 Dinhard
145 0217 Elgg
146 0218 Ellikon an der Thur
147 0219 Elsau
148 0220 Hagenbuch
149 0221 Hettlingen
150 0222 Hofstetten bei Elgg
151 0223 Neftenbach
152 0224 Pfungen
153 0225 Rickenbach (ZH)
154 0226 Schlatt (ZH)
155 0227 Seuzach
156 0228 Turbenthal
157 0229 Wiesendangen
158 0230 Winterthur
159 0231 Zell (ZH)
160 0241 Aesch bei Birmensdorf
161 0242 Birmensdorf (ZH)
162 0243 Dietikon
163 0244 Geroldswil
164 0245 Oberengstringen
165 0246 Oetwil an der Limmat
166 0247 Schlieren
167 0248 Uitikon
168 0249 Unterengstringen
169 0250 Urdorf
170 0251 Weiningen (ZH)
171 0261 Zürich

Appendix C: Parts of the Generated Population in XML Format


```
<?xml version="1.0" encoding="de_CH" ?>

<plans>
.
.
.
	<person id="19941413401001" >
		<plan>
			<act type="hx" x100="679590" y100="242200" />
			<leg mode="walk" dep_time="08:00" trav_time="00:10" />
			<act type="w" x100="679590" y100="242200" />
```

```
                <leg mode="walk" dep_time="12:00" trav_time="00:10" />
                <act type="w" x100="679590" y100="242200" />
                <leg mode="walk" dep_time="13:25" trav_time="00:10" />
                <act type="w" x100="679590" y100="242200" />
                <leg mode="walk" dep_time="15:30" trav_time="00:10" />
                <act type="w" x100="679590" y100="242200" />
                <leg mode="bike" dep_time="18:10" trav_time="00:04" />
                <act type="l" x100="679590" y100="242200" />
                <leg mode="bike" dep_time="20:00" trav_time="00:10" />
                <act type="hx" x100="679590" y100="242200" />
        </plan>
    </person>
    <person id="200096712001" >
        <plan>
                <act type="hx" x100="614240" y100="211610" />
                <leg mode="n.a." dep_time="08:05" trav_time="00:05" />
                <act type="w" x100="614240" y100="211610" />
                <leg mode="n.a." dep_time="12:00" trav_time="00:10" />
                <act type="w" x100="614240" y100="211610" />
                <leg mode="n.a." dep_time="13:30" trav_time="00:05" />
                <act type="w" x100="614240" y100="211610" />
                <leg mode="n.a." dep_time="15:35" trav_time="00:05" />
                <act type="w" x100="614240" y100="211610" />
                <leg mode="car" dep_time="17:15" trav_time="00:05" />
                <act type="s" x100="679590" y100="242200" />
                <leg mode="car" dep_time="19:25" trav_time="00:05" />
                <act type="hx" x100="614240" y100="211610" />
        </plan>
    </person>
    <person id="200085770004" >
        <plan>
                <act type="hx" x100="689870" y100="120250" />
                <leg mode="car" dep_time="08:45" trav_time="00:05" />
                <act type="w" x100="689870" y100="120250" />
                <leg mode="walk" dep_time="11:50" trav_time="00:25" />
                <act type="w" x100="689870" y100="120250" />
                <leg mode="car" dep_time="14:00" trav_time="00:02" />
                <act type="l" x100="689870" y100="120250" />
                <leg mode="car" dep_time="16:00" trav_time="00:05" />
                <act type="hx" x100="689870" y100="120250" />
        </plan>
    </person>
.
.
.
    <person id="200085770005" >
        <plan>
                <act type="hx" x100="689870" y100="120250" />
                <leg mode="car" dep_time="08:45" trav_time="00:05" />
                <act type="w" x100="689870" y100="120250" />
                <leg mode="walk" dep_time="11:50" trav_time="00:25" />
                <act type="w" x100="689870" y100="120250" />
                <leg mode="car" dep_time="14:00" trav_time="00:02" />
                <act type="l" x100="689870" y100="120250" />
                <leg mode="car" dep_time="16:00" trav_time="00:05" />
                <act type="hx" x100="689870" y100="120250" />
        </plan>
    </person>
```

```
    <person id="1994503701013" >
        <plan>
            <act type="hx" x100="607900" y100="210260" />
            <leg mode="walk" dep_time="11:15" trav_time="00:06" />
            <act type="l" x100="607900" y100="210260" />
            <leg mode="walk" dep_time="16:00" trav_time="00:06" />
            <act type="hx" x100="607900" y100="210260" />
        </plan>
    </person>
.
.
.
</plans>
```