Markov Chains for Investigating and Predicting Migration: A Case from Southwestern China

Bo Qin¹, Yiyu Wang², Haoming Xu³

¹School of Foreign Languages, University of Electronic Science and Technology of China, Chengdu, China
²Research Administration Office, Party School of Sichuan Provincial Committee of C.P.C for Provincial Authorities, Chengdu, China
³Department of Modern Science and Technology, Chengdu Administration Institute, Chengdu, China

*Corresponding author email: bobby.qin@qq.com

Abstract: In order to accurately predict the population's happiness, this paper conducted two demographic surveys on a new district of a city in western China, and carried out a dynamic analysis using related mathematical methods. This paper argues that the migration of migrants in the city will change the pattern of spatial distribution of human resources in the city and thus affect the social and economic development in all districts. The migration status of the population will change randomly with the passage of time, so it can be predicted and analyzed through the Markov process. The Markov process provides the local government and decision-making bureau a valid basis for the dynamic analysis of the mobility of migrants in the city as well as the ways for promoting happiness of local people's lives.

1. Background of this study

At present, under the macrocosmic background of the central government's in-depth implementation of the strategy for developing the western region, the Sichuan Provincial Party Committee and the Sichuan Provincial Government have explicitly mentioned that eight cities, including Panzhihua, should be built into mega-cities with a population of over 1 million and a GDP exceeding 100 billion. Accelerating development has brought important strategic opportunities and also brought a peak period for centralized construction of major projects for the accelerated development of the Huacheng New Area in the city. From a demographic point of view, the development and construction of Huacheng New District has provided a new space for the urban agglomeration in Panzhihua City; from a sociological point of view, the development of the New District has provided a new platform...
for perfecting the urban functions and provided a New window; from an economic point of view, the
construction of new district has provided a new impetus to the economic development of the city.

The implementation of Huacheng New District is a major decision made by Panzhihua Municipal
Party Committee and Municipal Government on the strategic and overall level. It is the key to
promoting the new-type urbanization in Panzhihua City and an important carrier for the construction
of megacities with a population of one million. This strategic objective involves many elements such
as urban planning, transportation construction, investment attraction and population aggregation. The
population factor is an important guarantee for the realization of many goals, especially for
mega-cities with a population of one million, and at the same time, an indicator of "happy city" The
main bearer, because the city's happiness in the final analysis is the public's happiness. Therefore,
mastering the population status and population development in Huacheng New District has important
decision-making value. In order to investigate the population status of Huacheng New District and
understand the population composition and source status of Huachen New District, it provides
valuable decision-making references and suggestions to government departments at all levels of the
province, city and county. Based on the principles and requirements of the prediction math model, two
sample surveys were organized respectively.

According to the plan, Huacheng New District starts from the Jinsha River to the north and reaches
the Kunming Expressway to the south. It runs from Panzhihua Avenue in the west to the natural hills
in the east, including Liusanpo, Bingsan District, Bingjiu District and Huashan-Basmas Dam Tong a
sand ditch a total of ribbon-made land. The new district's planning area is about 59.3 square
kilometers, and its planning population is about 350,000 people. The survey of this study will be
strictly controlled and conducted in the scope of the Huacheng New District.

2. Methodology and the Mathematical Modelling

Our two investigations carried out a comprehensive on-site confirmation in accordance with the
planning scope of Huacheng New District, and the scope of work was refined to buildings and roads.
The eastern part of the project included the Civil Construction and Community Park Road in
Bingcaogang Subdistrict, The streets of the bright, Jinfu community, Silver Town, Shaba Village;
Renhe District, including Dahe Zhong Road streets, South Road, bend the tree community, forward the
town of Dukou Village, Renhe Town Street, Shagou, Renhe Street, river community And Shagou
Village, Ganba Tong Village, Lixin Village and Zongfa Village of Zongfa Township. The above units
are partially or totally included in the red line of Huacheng New District. After the scope of the red
line was confirmed in the field, it took 3 months to investigate the population situation in Huacheng
New District. In the two surveys, the PIP system data, the "Sixth Nationwide population census" data
and the household survey collection were combined to take the community, village and residential
groups as a unit and the September 30, 2012 real-time data points to analyze the PIP system, "Sixth
Plan" data and the same code in the same area as the "Red Line" in Huacheng New District, the survey
of households in some areas that are not included in the PIP system or " (Habitat) did not leak group,
the group did not leak households, households do not leak people. In the second survey, based on the
Markov process, we established a population migration model to dynamically analyze the indicators of
population migration and reflect the dynamic evolution of the indicators in the sample survey to
provide information for the population flow and development in Huacheng New Area Provides a valid
prediction.
In addition, in order to find out the mentality and suggestions on the construction of Huacheng Urban District within Huacheng New District, our investigation team carried out the "Huacheng New District Demographic Mentality Survey" by means of questionnaire survey. In the eastern district, Renhe District selected representative sample points, each issued 300 questionnaires, a total of 545 valid questionnaires were recovered, of which 284 were registered population questionnaire, 261 migrant questionnaires, basically meet the household and migrant population Half of the data samples. In the second survey, a valid follow-up survey was conducted on the respondents from the same surveyed group to remove 364 valid questionnaires, which were not attended to and participated twice, but the second questionnaire was invalid. After the questionnaire is recovered, the professional survey software is used for input and data analysis. The first survey results form a series of situation analysis charts of "Huachen New District Demographic Mentality Survey". The second follow-up survey results are mainly used in the forecast part.

3. The Basic Situation of Population in Huacheng New District

3.1. Registered population and resident population

As of the end of September 2012, Huacheng New District has 83002 census registers (39,802 in the east and 43,200 in Renhe District), of which 26,006 are married women of child-bear age. The existing permanent population is 109,498 (52,691 people in East District, District 56,807 people), of which 30,287 were married women of childbearing age. According to the natural change of the population, there were 351 births with a birth rate of 3.2‰ between October 1, 2011 and September 30, 2012. The cumulative death toll was 173, with a population death rate of 1.6‰ and a natural population growth rate of 1.6‰. In terms of population, the number of 0-14-year-olds in Huacheng New District is 13977, the population of 15-64 is 81907, and the number of people over 65 is 13614.

3.2. The basic situation of floating population

In terms of the total population and composition of the floating population, the existing population of 31225 in Huacheng New Area is about 12.8% of the total floating population of the city. The floating population is mainly inflow population accounting for 92.5% of the total floating population in Huacheng New District, of which 25101 are from the city and 3722 are from the city, and 2402 are from the city, accounting for 7.5% of the total floating population of Huacheng New District, Among them, 1627 went to the outside of the city and 775 to other counties in the city. According to the reasons for the floating population, the number of migrant workers doing business and migrant work is 25628, accounting for 82.1% of the total. The number of floating population of 4407 is "fortunate and reliant", accounting for 14.1% of the total. "Other reasons" Of the floating population of 1190 people, accounting for 3.8% of the total. According to the sex and age composition of migrants, the male-female ratio of migrants is 116: 100, while that of men is slightly larger than that of women. The number of floating population aged 0-14 is 4992, accounting for about 16% of the total; the number of floating population aged 15-64 is 25048, accounting for about 80% of the total; and the number of floating population over 65 is 1185, accounting for about 4% of the total.

3.3. The Educational Level and Professional Titles of interviewees

There are 55,985 effective information about the education level of the population in Huacheng New District actually obtained in this survey. Among them, 34,643 have the education of "junior high
4. Population Sampling Survey in Huacheng New District

4.1 Positive Aspects

The Positive aspects are focused on five aspects as follows: The floating population has a high sense of happiness and belonging; People agree that attracting population is an important measure for the development of Huacheng New District; The floating population purchase intention is generally positive; The influx of young and middle-aged population will continue the demographic dividend; People have a high expectation for Huacheng new district construction.

4.2 Negative Aspects

The Negative aspects are focused on six aspects as follows: People do not know enough about Huacheng New District's development planning and positioning. People generally think that housing prices are high. The intention of settling migrants is not strong. Migrant population insured rate is not high. Housing is the biggest problem for the floating population. People's living standards are low.

5. Mathematical Analysis of the Second Sample Survey

In order to accurately predict the happiness index of Huacheng New District, we conducted two surveys on the surveyed population and conducted dynamic analysis using related mathematical methods. The flow of population in the city will change the pattern of spatial distribution of human resources in the city, thus affecting the social and economic development in all districts. The migration status of the population changes randomly over time, so it can be predicted and analyzed through the Markov process. Markov is a model describing a class of stochastic dynamic processes in which past historical states are irrelevant for predicting future states. The state at time \( t + 1 \) depends only on the state at time \( t \) and the transition probability, so Markov can describe the problem that changes according to the inherent laws and thus reflect the evolution of things. Markov process is the process of population migration in the city Dynamic analysis of the flow provides an effective basis.

5.1 Population Migration Model and Forecast Analysis

In the population migration model, the time and state are discrete Markov processes, which are Markov chains. Suppose the Markov chain is a stochastic process \( \{X_n, n \in \mathbb{N}\} \) is called the state of the space \( X_n \) the value of time \( n \) state. \( X_{n+1} \) satisfies the formula (1), then \( \{X_n, n \in \mathbb{N}\} \) is a Markov chain.

\[
P(X_{n+1} = x | X_0, X_1, X_2, \ldots, X_n) = P(X_{n+1} = x | X_n) \quad (1)
\]

Assume that the state probability of \( X_n = a_i \) is expressed as follows:
The array \( P = [p_1(n) p_2(n) \ldots p_N(n)]^T \), consisting of the state probabilities \( p_i(n) \), gives the probability distribution for all states of \( X_n \), with the sum of all states equal to 1 as follows:

\[
\sum_{i=1}^{N} p_i(n) = 1
\]

At time \( t \), under \( a_i \), the state transition matrix arriving at \( a_j \) at time \( t + 1 \) is

\[
p_{ij}(t,t+1) = P\{X_{t+1} = a_j | X_t = a_i\}
\]

According to the total probability formula, there is

\[
p_j(t+1) = \sum_{i=1}^{N} p[X_{t+1} = a_j, X_t = a_i] = \sum_{i=1}^{N} P[X_{t+1} = a_j | X_t = a_i]P[X_t = a_i]
\]

The Markov chain probability transfer matrix as,

\[
P(t,t+1) = P(t,t+1) = \begin{bmatrix}
P_{11}(t,t+1) & P_{12}(t,t+1) & \ldots & P_{1n}(t,t+1) \\
P_{21}(t,t+1) & P_{22}(t,t+1) & \ldots & P_{2n}(t,t+1) \\
\vdots & \vdots & \ddots & \vdots \\
P_{n1}(t,t+1) & P_{n2}(t,t+1) & \ldots & P_{nn}(t,t+1)
\end{bmatrix}
\]

In our prediction model, the prediction at time \( t \) depends only on the transformation at time \( t-1 \) and \( t \), then Markov’s transition probability \( P_{ij}(t,t+1) \) depends only on the difference between \( t + 1 \) and \( t \) and \( t \). Regardless of the value of \( t \) and \( t + 1 \) itself, the homogeneous Markov chain can be applied to calculate. When \( t + 1 - t = 1 \), \( P_{ij} \) denotes the transition probability of the state-wise \( a_i \) of the Markov chain after one-step transition to the state \( a_j \). Let the transpose of the one-step transfer matrix of homogeneous chains be:

\[
P^T(1) = \pi
\]

The homogeneous two-step transfer matrix can be expressed as

\[
P^T(2) = P^T(1) P^T(1) = \pi^2
\]

Homogeneous chain \( X_{n+1} \) in state \( a_i (i = 1, \ldots, N) \) state probability distribution can be listed as:

\[
P(n+1) = P^T(n) P(1) = \pi^n P(1)
\]

The law of the transfer between various states of the system is reflected in the state transition
probability matrix $P(t, t + 1)$. By analyzing the $P(t)$, we can predict the moment $t + 1$ (the development and transformation of the system in the future). Before predicting $t + 1$, we need to select the data in $t - m$ year, sum the array of probability matrix, and the state of the largest summation array is predicted.

5.2 The Trend of Happiness among Migrants Prediction model

The degree of happiness of floating population flows in the direction of happiness can be divided according to their degree of happiness as: good, just so so, not happy. Through this survey of the current status of the population in Panzhihua, combined with last year's survey data, we can draw the following table:

| The First Flow Direction | Very Happy | Good | So So | Not Happy |
|--------------------------|------------|------|------|-----------|
| Very Happy               | 0.76       | 0.23 | 0.01 | 0         |
| Good                     | 0.2        | 0.7  | 0.1  | 0         |
| So So                    | 0          | 0.3  | 0.6  | 0.1       |
| Not Happy                | 0          | 0    | 0.2  | 0.8       |

From the data in the table, we get the transfer matrix of the happiness flow in Panzhihua City:

$$
P(1) = \begin{bmatrix}
p_{11} & p_{12} & p_{13} & p_{14}
p_{21} & p_{22} & p_{23} & p_{24}
p_{31} & p_{32} & p_{33} & p_{34}
p_{41} & p_{42} & p_{43} & p_{44}
\end{bmatrix} = \begin{bmatrix}
0.76 & 0.23 & 0.01 & 0
0.2 & 0.7 & 0.1 & 0
0 & 0.3 & 0.6 & 0.1
0 & 0 & 0.2 & 0.8
\end{bmatrix}$$

(12)

Because Markov chains have ergodicity, the limit distribution $\Pi = (p_1, p_2, p_3, p_4)$ satisfies the following system of equations:

$$p_1 = 0.76 \times p_1 + 0.2 \times p_2,$$

$$p_2 = 0.23 \times p_1 + 0.7 \times p_2 + 0.3 \times p_3,$$

$$p_3 = 0.01 \times p_1 + 0.1 \times p_2 + 0.6 \times p_3 + 0.2 \times p_4,$$

$$p_4 = 0.1 \times p_3 + 0.8 \times p_4,$$

$$1 = p_1 + p_2 + p_3 + p_4.$$

By calculation, we can get the solution of this equation.
is $p_1 = 0.35, p_2 = 0.422, p_3 = 0.152, p_4 = 0.076$, so the distribution of limit

$\Pi = (0.35, 0.422, 0.152, 0.076)$ can show the proportions of happiness in 2014 are: happiness: 0.35, you can also: 0.422, the general: 0.152, not happy: 0.076.

5.3 The Forecasting Model for Trend of Sense of Belonging for Migrant Population forecasting model

The degree of belonging of floating population flows in the direction of mobility. According to the degree of population belonging, it can be classified as fully integrated, somewhat integrated, and not integrated. Through this survey of the current status of the population in Panzhihua, combined with last year's survey data, we can draw the following table:

| The Second Flow Direction | Fully integrated | Partly integrated | Not integrated |
|---------------------------|------------------|------------------|---------------|
| The First Flow Direction  |                  |                  |               |
| Fully integrated          | 0.76             | 0.23             | 0.01          |
| Partly integrated         | 0.2              | 0.7              | 0.1           |
| Not integrated            | 0                | 0.3              | 0.6           |

From the data in the table, we get the transfer matrix of Sense of belonging flow in Panzhihua City as:

$$P(1) = \begin{bmatrix} p_{11}(1) & p_{12}(1) & p_{13}(1) \\ p_{21}(1) & p_{22}(1) & p_{23}(1) \\ p_{31}(1) & p_{32}(1) & p_{33}(1) \end{bmatrix} = \begin{bmatrix} 0.88 & 0.14 & 0.03 \\ 0.33 & 0.67 & 0 \\ 0 & 0.4 & 0.6 \end{bmatrix}$$

According to Markov's calculation, there is a limit distribution $\Pi = (0.623, 0.323, 0.047)$. The distribution shows that the proportions of belonging degree in 2014 are: fully integrated: 0: 623, partially integrated: 0: 323, not integrated: 0: 047.

4.4 Prediction Model of Trend of Citizen Population Affecting the Development of Huacheng New District

Citizens' population development has a direct impact on the development of Huacheng New District. According to the influence of the population development on the development of Huacheng New District, it can be divided into "significant impact", "certain impact" and "no impact". Through this survey of the current status of the population in Panzhihua, combined with last year's survey data, we can draw the following table:

| The Second Flow Direction | significant impact | certain impact | no impact |
|---------------------------|--------------------|----------------|-----------|
| The First Flow Direction  |                    |                |           |
| significant impact        | 0.95               | 0.05           | 0         |
| certain impact            | 0.1                | 0.8            | 0.1       |
| no impact                 | 0                  | 0.28           | 0.72      |
From the data in the table, we get the transfer matrix of the influence of the population development of Panzhihua City on the development of Huacheng New District:

\[
P(1) = \begin{bmatrix}
p_{11}(1) & p_{12}(1) & p_{13}(1)
p_{21}(1) & p_{22}(1) & p_{23}(1)
p_{31}(1) & p_{32}(1) & p_{33}(1)
\end{bmatrix} = \begin{bmatrix}
0.95 & 0.05 & 0 \\
0.1 & 0.8 & 0.1 \\
0 & 0.28 & 0.72
\end{bmatrix}
\]  (15)

According to Markov's calculation, the limit distribution can be obtained (0: 595; 0: 298; 0: 107). This distribution can show that the proportions of the population development in Huacheng New District affected by the population development in 2014 are: 0: 595, have an impact: 0: 298, no effect: 0: 107.

References
[1] S Fine, Y Singer, N Tishby: The Hierarchical Hidden Markov Model: Analysis and Applications, Kluwer Academic Publishers, 1998, 32 (1):41-62
[2] SJ Lycett, CJ Norton: A demographic model for Palaeolithic technological evolution: The case of East Asia and the Movius Line, Quaternary International, 2010, 211 (1–2) :55-65
[3] Q Chen, MR Elliott, R Little: Bayesian penalized spline model-based inference for finite population proportion in unequal probability sampling., Survey Methodology, 2017, 36 (1) :23-34
[4] QIN Bo, XU Haoming: The Application of Machine Learning in the Weight Assignment of Party and Government Cadres' Daily Assessment Model, Study on Party and Government, 2015, (3): 28-31
[5] M Cui, Z Chen: The exact solution of nonlinear age-structured population model, Nonlinear Analysis Real World Applications, 2007, 8 (4) :1096-1112
[6] J He, K Wang: The survival analysis for a single-species population model in a polluted environment, Applied Mathematical Modelling, 2007, 31 (10) :2227-2238
[7] ML Burton, Mathematical Anthropology, Annual Review of Anthropology, 1973 , 2 (1) :189-199