第24卷 增刊2 工 程 数 学 学 报 Vol. 24 Supp.2
2007年12月 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Dec. 2007

文章编号:1005-3085(2007)08-0039-08

基于 Leslie 模型的中国人口预测及蒙特卡罗仿真

陈 鹏, 张成龙, 高斯蒙

指导教师: 梅长林

(西安交通大学, 西安 710049)

编者按: 本文在 Leslie 模型的基础上加上迁移项, 其参数的发展模型规范、合理. 其最大的特色是加入蒙特卡罗仿真, 进一步验证其合理性。

摘 要: 本文主要预测了中国人口增长情况及增长过程中相关人口指数的变化. 考虑到城镇(市、镇)和农村在迁移、生育、死亡等方面的差异, 本文对城镇和农村分别讨论. 建立了长期(2001-2050)人口变化的带有迁移项的 Leslie 模型。进一步利用蒙特卡罗方法对人口动态变化过程进行了计算机仿真, 其结果与模型求解结果吻合较好。在总和生育率为 1.8 的情况下的主要预测结果为: 总人口在 2038 年左右达到 14.9 亿的峰值, 到 2050 年回落至接近 14 亿; 人口平均寿命升高, 老龄化比例持续升高, 2020 年达到近 20%, 2050 年达到 35%; 人口抚养比总体呈升高趋势, 到 2050 接近 70%。

关键词: 人口预测; Leslie 模型; 人口迁移; 蒙特卡罗仿真

分类号: AMS(2000) 62J10 **中图分类号:** O212 **文献标识码:** A

1 引言

人口问题一直是我国关注的焦点之一, 对我国人口各项指标的预测有着深远的意义。我国的人口特点除了人口基数大, 人口出生率持续偏高外, 近年来有呈现出如老龄化进程加速、出生人口性别比持续升高及乡村人口城镇化等新的特点。影响人口变化的因素众多, 如出生、死亡、迁移、性别比、人口素质、社会环境、生育政策等等, 同时各影响因素之间存在着相互影响关系, 而且也随时间发生变化, 导致预测模型尤其是长期预测模型难度较大。至今有关人口预测的模型主要有人口发展方程、Leslie 模型、Logistic 模型、回归分析等。

长期预测需综合考虑各影响因素。通过分析, 本文以人口的生育、死亡、迁移作为主要影响因素, 并通过 Leslie 模型将三种影响因素联系起来, 确定扩展的 Leslie 模型。同时, 本文分别采用 Logistic 模型, 图形分析, 函数拟合, 概率密度分布等对以上三种因素做出随时间变化的动态方程, 即参数发展方程, 然后用差分方法求解。为了保证上述模型预测的准确性以及稳定性, 我们采用蒙特卡罗计算机仿真的方法对上述模型进行检验。

2 模型及求解

2.1 基本假设

1) 我国看成一个封闭系统, 即没有人口的迁入和迁出;
2) 人口增长只与人口基数、生育、死亡和迁移有关;
3) 国内人口迁移为农村向城市的单向迁入, 且关于年龄的分布不随时间变化;
4) 我国城市化水平上限为现代发达国家最高城市化水平;
5) 90 岁以上人口统视为一个年龄段群体。

2.2 模型建立[1,2]

长期的人口预测必须考虑人口的出生、死亡、迁移等主要因素，为此我们以已知的 Leslie 模型为基础。通过建立带有迁移项的 Leslie 模型来解决相关问题。

取一岁为一个年龄段，一年为一个时段。设人口按年龄分组为0到 $n-1$ 岁及大于等于 n 岁者(n 岁以上视为同一年龄段)共 $n+1$ 个年龄段(这里 n 为90)。

设 $p_r^i(t)$、$d_r^i(t)$、$b_r^i(t)$、$h_r^i(t)$、$k_r^i(t)$、$v_r(t)$、$f_r(t)$ 分别表示 t 到 $t+1$ 年第 r 个年龄段总人口、人口死亡率、人口出生率、女性生育模式、女性性别比、净迁移人口、人口迁移率(迁移人口比总人口)，$R(t)$ 为第 t 年城市化水平(城市人口比总人口)。为了分别考察城镇、农村人口的发展，以上各参数上标 i 为1时代表城镇，为2是代表农村，以下各参数上标同此。

添加迁移项的 Leslie 模型建立如下

$$\begin{cases} P^i(t+1) = H^i(t)P^i(t) + \beta(t)B^i(t)P^i(t) + (-1)^{i+1}V(t), \\ P^i(0) = (p_0^i(0), p_1^i(0), \cdots, p_n^i(0))^T, \end{cases} \tag{1}$$

其中

$$H^i(t) = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1-d_0^i(t) & 0 & 0 & \cdots & 0 \\ 0 & 1-d_1^i(t) & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & 0 \\ 0 & 0 & 0 & 1-d_{n-1}^i(t) & 1-d_n^i(t) \end{pmatrix},$$

$$B^i(t) = \begin{pmatrix} 0 & \cdots & 0 & b_{r_1}^i & \cdots & b_{r_m}^i & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix},$$

$$\begin{cases} p^i(t) = (p_0^i(t), p_1^i(t), \cdots, p_n^i(t))^T, \\ V(t) = (v_0(t), v_1(t), \cdots, v_n(t))^T, \\ v_r(t) = (R(t+1) - R(t))^* \sum_{r=0}^n (p_r^1(t) + p_r^2(t)) f_r(t),\ r = (0, 1, \cdots, n), \\ b_r^i(t) = (1-d_0^i(t))k_r^i(t)h_r^i(t),\ r = r_1, \cdots, r_m, \text{其中 } [r_1, r_m] \text{ 为生育区间.} \end{cases} \tag{2}$$

如果可以确定时刻 t 各年龄段人口总数 $p_r^i(t)$，则可以确定总人口，老龄人口，人口抚养比等各类人口指标，由上述模型可知，除了人口初始值外还需要确定以下三项参数

1) 迁移项：$R(t)$、$f_r(t)$； 2) 死亡项：$d_r^i(t)$； 3) 生育项：$h_r^i(t)$、$k_r^i(t)$、$\beta(t)$。

为了分别考察城镇、农村人口的发展，以上各参数上标 i 为1时代表城镇，为2是代表农村。

2.3 模型中各参数的确定

2.3.1 迁移项分析

根据中国的发展现状，我们仅需要考虑从农村向城市迁入人口的变化情况。城市化水平随时间的的变化是表征人口迁移的重要指标；同时，为了对城乡劳动力人口和老龄人口作出准确

预测，迁移人口随年龄的分布也是确定人口迁移的重要因素。因此，本文分别从城市化水平和迁移人口随年龄的分布两方面对人口迁移作出预测。

1) 城市化水平的 Logistic 预测模型[3]

根据发达国家的城市化经历，一个国家或地区的城市化过程大致呈一条拉平的“s”形曲线。因为 Logistic 函数具有典型的“s”形曲线特征，同时含有“环境容纳量”和“内禀增长率”等较深刻的生物学意义，比较符合自然界和人类社会发展规律，因此本文选用 Logistic 模型来预测我国的城市化进程。模型的方程建立如下

$$\frac{\mathrm{d}R(t)}{\mathrm{d}t} = kR(t)(R_{\max} - R(t)), \quad (3)$$

其中 $R(t)$ 为城市化水平，即城市人口占总人口的比例，$R_{\max} = 0.85$(以发达国家现阶段城市化水平为参考，取值 0.85)为城市化水平制约项(环境容纳量)，k 为城市化水平增长率(内禀增长率)。

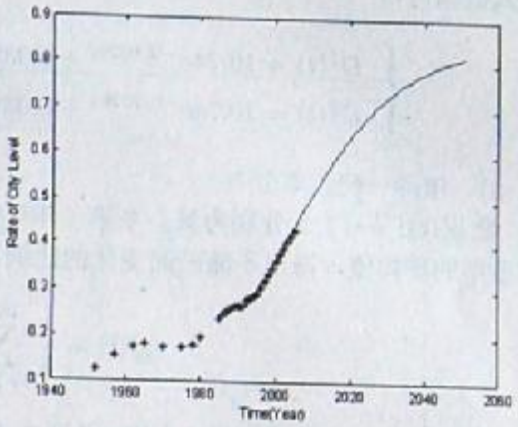

图 1: 我国城市化水平至 2050 年的预测

求解(3)得 $R(t) = \frac{0.85}{(1+0.85ce^{-0.85kt})}$，利用国家统计局对我国三十年来城市化水平的统计数据[4]，通过最小二乘法确定 $k = 0.080$, $c = 2.24$。预测结果如图 1

2) 年龄－迁移率分布

由基本假设 3)，迁移率随年龄的分布不随时间变化，因此只需要确定某年迁移率随年龄的分布即可。取上海市 2000 年总体迁移人口中各年龄段人口所占比例为代表数据，借助 Matlab 拟合工具箱对数据做三项移动平均法平滑处理，根据方差分析经多次尝试后选取混合正态分布函数做最终拟合函数，归一化后如下

$$f_r(t) = 0.0305e^{-\frac{(r-25.08)^2}{83.41}} + 0.0118e^{-\frac{(r-37.67)^2}{669.77}}, \quad (4)$$

拟合显著性判断参数为 SSE: 3.22、R-square: 0.9791。

2.3.2 死亡项分析

对人口死亡率的预测分为两个方面，一方面是总人口死亡情况随时间的变化关系，另一方面是不同年龄段人口死亡情况随时间的变化关系。

1) 城乡总人口死亡率[5]

建国以来，我国人口死亡率整体呈现下降趋势；改革开放以来，死亡率逐渐平缓，且呈现周期性的波动。因此，本文将总人口死亡率分解为趋势项和周期项两部分。总人口死亡函数建立为

$$D(t) = C(t) + T(t),$$

其中 $D(t)$ 为总人口死亡率，$C(t)$ 为趋势项，$T(t)$ 为周期项。由建国来总人口死亡率数据[4]，对趋势项进行负指数函数拟合得 $C(t) = 10.74e^{-0.1026t} + 634$，拟合显著性判断参数 SSE: 1.648，R-square: 0.9893；对周期项进行三角函数拟合得 $T(t) = 0.039\sin(0.47t) + 0.029\cos(0.47t) + 0.062$，拟合显著性判断参数 SSE: 0.010、R-square: 0.6246。

设 $D_1(t)$, $D_2(t)$, 分别为城镇人口死亡率和乡村人口死亡率，本文用总人口死亡率加上20年来[7]已知的两地区人口死亡率与总人口死亡率之差的平均值来代替两地区人口死亡率。即

$$D^i(t) = D(t) + \frac{1}{20}\sum_{t=1}^{20}(D^i(t) - D(t)), \quad i = 1, 2,$$

代入数据可得

$$\begin{cases} D^1(t) = 10.74\mathrm{e}^{-0.1026t} + 0.039\sin(0.47t) + 0.029\cos(0.47t) + 5.454, \\ D^2(t) = 10.74\mathrm{e}^{-0.1026t} + 0.039\sin(0.47t) + 0.029\cos(0.47t) + 6.765. \end{cases} \tag{5}$$

2) 年龄—死亡率分布

设 $g_r^i(t)$, $i = 1, 2$ 分别为城、乡第 t 年年龄为 r 的的死亡人口占城、乡总死亡人口的比例，取五年的平均值，得出不随时间变化的比例，表达式如下

$$g_r^i(t) = \frac{1}{5}\sum_{t=2001}^{2005} g_r^i(t), \quad i = 1, 2,$$

同时，该比例=该年龄段死亡人口÷总死亡人口，即

$$g_r^i(t) = \frac{d_r^i(t)p_r^i(t)}{\sum\limits_{t=2001}^{2005} d_r^i(t)p_r^i(t)}, \quad i = 1, 2,$$

综合以上两式，得年龄为 r 的人在第 t 年的死亡率=该年龄段死亡人口÷该年龄段总人口，即

$$d_r^i(t) = \frac{g_r^i(r)D^i(t)P^i(t)}{p_r^i(t)}, \quad i = 1, 2. \tag{6}$$

2.3.3 生育项分析

对人口出生率的预测不仅要考虑育龄女性人口比例，也要考虑生育模式，即育龄女性的生育率随年龄的概率密度分布。

1) 育龄女性人口比例分析

对2001年至2005年五年内的育龄女性人口比例数据及其平均值绘图2分析可得：

1) 五年的育龄女性人口比例围绕平均值上下波动，但波动范围较小，波动随机性较强，波动规律不明显；

2) 育龄女性人口比例随年龄的分布虽然随着时间的变化而发生扰动，但扰动比较微弱，规律不显著。

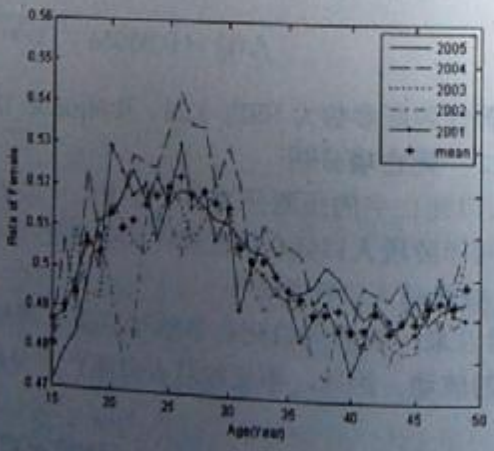

图 2: 育龄女性人口比例随年龄的分布关系

因此本文用五年内育龄女性人口比例的平均值作为对育龄女性人口比例的预测值。

2) 生育模式的确定

与1)做同样分析，此处对生育模式(育龄妇女生育率)可取五年内的平均值作为预测值。鉴于生育模式图形(如图3)具有显著的 Γ 分布[6]的形状，本文构造 Γ 分布对生育模式进行预测。

设

$$h_r = \frac{(r-\tau_1)^{\alpha-1}e^{-\frac{r-\tau_1}{\theta}}}{\theta^\alpha\Gamma(\alpha)}, \qquad r > \tau_1,$$

(此处 h_r 为方程中 $h_r^i(t)$，非时间变量)

其中 τ_1 为15，$\theta = 2$，$\alpha = \frac{n}{2}$，$r_c = \tau_1 + n - 2$，其中 r_c 为最大概率值所对应的点，此时的分布化为 χ^2 分布，由矩估计法可得

$$n = E(r) = \frac{1}{N}\sum_{i=1}^{N}(r_i - 15)$$

因为40岁以后的密度很小，几乎为0，这里我们忽略40岁以后的生育率影响可得，

$$n = E(r) = \frac{1}{25}\sum_{i=1}^{25}(r_i - 15) = 13,$$

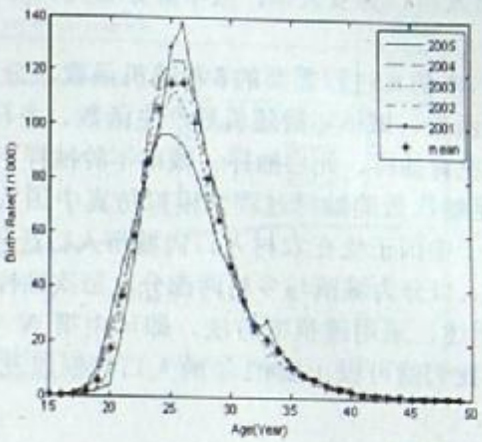

图 3: 2001-2005年五年中市育龄女性生育率与年龄的关系图

进而可得 $r_c = 26$，由图3中平均值分布可以看出，最大概率值所对应的点为27，验证了该取法的正确性。最后可得

$$h_r = \frac{(r-15)^{5.5}e^{-\frac{r-15}{2}}}{2^{6.5}\Gamma(6.5)}, \quad r > 15, \tag{7}$$

该分布的决定因子 τ_1 在此处有特殊的含义，即最小生育年龄，不同的最小生育年龄可以带来不同的总和生育率，因此国家可以通过法律来规定 τ_1 取值，从而可以有效实施计划生育。

2.4 模型求解

由各已知参数求解模型，根据求解结果可得到以下重要指标

1) 城市、农村及全国总人口

$$P(t) = \sum_{r=0}^{n}(p_r(t)). \tag{8}$$

2) 老龄化指数：未来某年 r_0 岁以上的老人占总人口的比例

$$\sigma(t) = \frac{\sum_{r=r_0}^{n}(p_r(t))}{P(t)}. \tag{9}$$

3) 人口扶养比：单位劳动力所能供养的非劳动力人口

$$\begin{cases} \rho(t) = \frac{P(t)-L(t)}{L(t)}, \\ L(t) = \sum_{r=l_1}^{l_2} p_r(t), \end{cases} \tag{10}$$

计算的具体结果见模型检验

3 蒙特卡罗方法[6]与计算机仿真对模型的检验

用蒙特卡罗方法仿真人口问题，根据现有的生育、生存、死亡、男女比例等统计资料，我们找出相应的概率分布，称其为概率分布函数。对每个人口样本，用一组随机数来模拟其发展，这个人口个体的集合就相当一组人口样本。分析这组样本，便可找出某些问题的解答。

3.1 蒙特卡罗人口仿真过程

首先，本仿真过程需要已知起始以及终止年份、生育率、时变死亡率、第一年的人口分布(包括总人口，男女人口，按年龄分布人口)、时变迁移率。这些条件在先前的论文中均已求出。

其次，本仿真过程需要的5种随机函数。分别是：出生概率随机数产生函数、死亡概率随机数产生函数、城镇年龄随机数产生函数、乡村年龄随机数产生函数、性别随机数产生函数以分别完成生育抽样、死亡抽样、城市年龄抽样、乡村年龄抽样以及性别抽样。

通过足够次数的抽样过程来模拟仿真中国人口的变化过程，每次抽样完成后进行人口结构调整。由于中国正处在农村人口向城市人口迁移的阶段，所以对人口迁移的考虑是必须的，这里把中国人口分为城镇与乡村两部分，每次抽样过后，还要进行因迁移而造成的人口调整。

综上所述，采用递推的方法，即可由第 N 年的人口数据信息推出 $N+1$ 年的人口数据信息。这样我们就可以由2001年的人口数据以及各种参数变化率的分布函数，推出往后每年的人口数据。

3.2 流程图

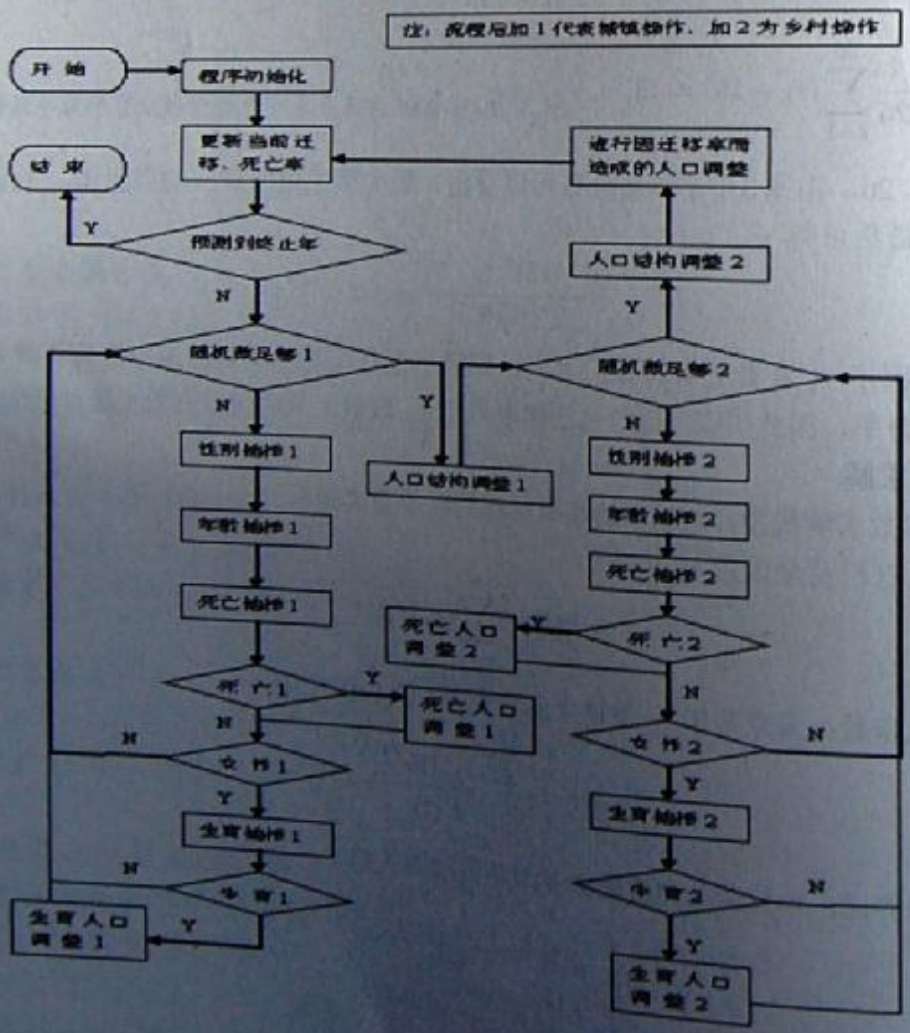

图4: 城镇、农村人口发展流程图

3.3 检验模型简析

由于此题是对中国人口进行仿真模拟，采用一般的模拟仿真方法时间效率不太理想，也不能够完全体现出统计学中随机的特点。这里采用了蒙特卡罗方法，不仅仅在时间效率上得到大大提高，更为重要的是，它可以通过改变其中一些参数的分布函数，来研究此参数对中国人口发展的影响。

3.4 模型及其检验结果

根据人口发展方程模型求解并与仿真结果作对比，对比图形如5，6，7

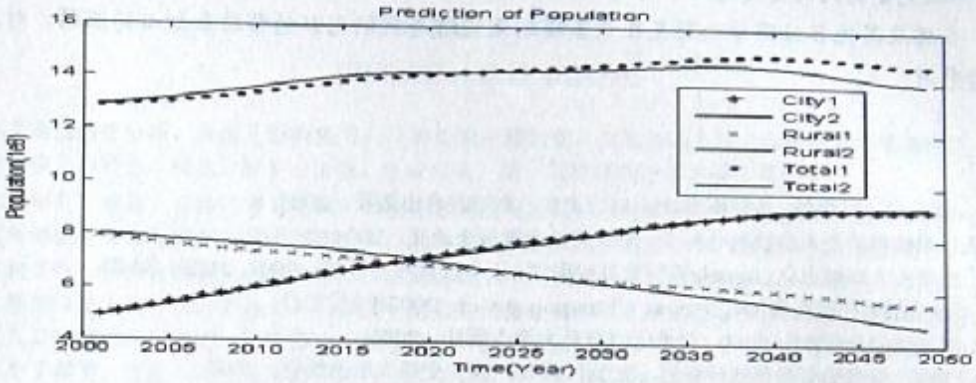

图5: 城镇、乡村和总人口预测及仿真结果

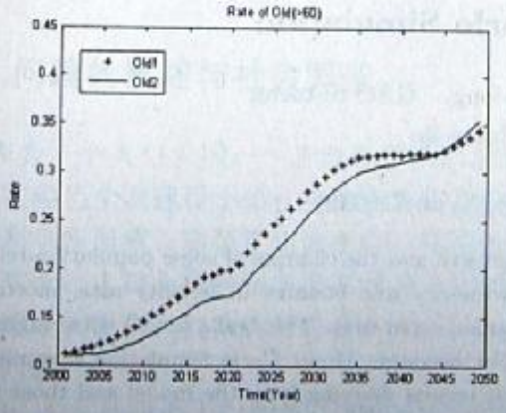

图6: 老龄化人口(超过60岁)比例预测与仿真结果

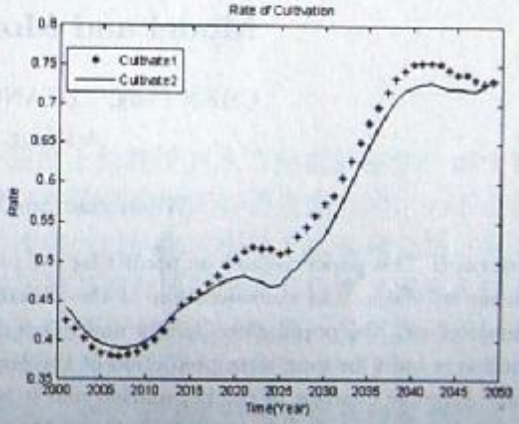

图7: 人口扶养比预测与仿真结果

4 模型评价

4.1 模型的优点与不足

本文在模型的选择、建立、求解和检验过程中有以下优点：

1) 本文充分考虑实际国情，对城、乡人口分别预测，结果涵盖了各主要人口指标。

2) 对人口的长期预测中，综合考虑了人口发展最主要的影响因素，并根据各影响因素的时变特征做了较为实际的预测。

3) 本文借助蒙特卡罗随机仿真的方法对本文的主体模型进行了动态仿真和检验，检验结果良好；通过仿真，还可以灵活调整各种参数，从而对各种参数在人口发展中所起作用得到验

证。鉴于该问题的复杂性，本文模型也省略和简化了一些非重要影响参数，如人口的受教育水平，人口所处的自然和社会环境等；同时本文把一些难以确定但影响不大的参数(如育龄女性数量比例，死亡率在各年龄阶段的分布等)设为时间的不变量，难免产生一定的误差。

4.2 模型的推广

对人口的预测加入人口素质，政府政策，社会经济环境等影响因素，更加具体细致的预测未来人口的发展。本文所用到的函数拟合、Leslie 模型、Logistic 模型、蒙特卡罗随机仿真等方法在很多领域内都有广泛的应用。

致谢：本次竞赛得到西安交通大学的支助及各位指导老师的支持，在此表示我们由衷的感谢。教练梅长林老师、乔璐及其他各位同学和朋友在竞赛培训期间给与我们无私的帮助和细心的照顾，对此表达我们最真诚的感谢。

参考文献：

[1] 姜启源，谢金星，叶俊编. 数学模型-3版[M]. 北京：高等教育出版社，2003, 8
[2] 陈强. 人口系统模型及人口状况分析[J]. 南京理工大学硕士论文，2004, 3: 5-6
[3] 李百岁. 内蒙古人口城市化Logistic模型及其应用[J]. 干旱区自然与环境，2007, 21(2): 32-35
[4] 国家人口发展战略研究报告[OL]. www. Chinapop.gov.cn, 2007年9月25日
[5] 潘红宇. 时间序列分析[M]. 北京：对外经济贸易大学出版社，2005
[6] 施雨，李耀武. 概率论与数理统计应用-2版[M]. 西安：西安交通大学出版社，2005
[7] 杨耀臣. 蒙特卡罗方法与人口仿真学[M]. 合肥：中国科学技术大学出版社，1997, 7

Prediction of Chinese Population Based on the Leslie Model and Monte Carlo Simulation

CHEN Peng, ZHANG Cheng-long, GAO Si-meng
Advisor: MEI Chang-lin

(Xi'an Jiaotong University, Xi'an 710049)

Abstract: This paper focuses on predicting the population growth and the changes of some population-related indexes in China. In consideration of the differences between city and country in fertility rate, mortality, migration etc., the populations in city and in country are studied separately. The Leslie model with migration function is built for long term prediction of the population. Furthermore, Monte Carlo Simulation is applied to study dynamic changes of the population and it is found that results deriving from the model and those from the simulation are very close. With the total fertility rate 1.8, our prediction is as follows: the total population in China will peak to 1.49 billion around 2038, and will drop to 1.4 billion in 2050; the Chinese life expectancy will increase; the percent of aged people will reach 20% in 2020 and 35% in 2050; the demographic dependency ratio will reach nearly 70%.

Keywords: population prediction; Leslie model; population migration; Monte Carlo simulation