

文章编号:1005-3085(2007)08-0079-04

关于中国人口预测的数学模型

唐 云

(清华大学数学科学系, 北京 100084)

摘 要: 本文对于2007年全国大学生数学建模竞赛A题介绍了命题的背景和解答要点, 对同学的解答中一些优点和问题进行评述。最后还展望进一步研究的前景。

关键词: 数学模型; 中国人口; 预测

分类号: AMS(2000) 91D20

中图分类号: O29

文献标识码: A

1 问题背景

人口问题是一个永恒的话题。从马尔萨斯的《人口原理》起, 人们对于人口规律的研究已有二百多年的历史。关于人口问题的一般数学模型也已写进一些教材中, 如姜启源等人编写的《数学模型》^[1]中就介绍了人口增长的指数增长模型、阻滞增长模型、偏微分方程模型和 Leslie 迭代矩阵模型。中国是个人口大国, 几十年来, 专家们对中国人口发展和预测的数学模型研究积累了丰富的成果, 还研制出一些专用于人口预测的软件。

人口问题是复杂的, 涉及到社会的方方面面。目前, 对近期的预测已达到相当的精度, 但对人口发展做出较确切的中长期预测还是难以实现的。这些都为人口问题的数学模型研究创造了条件, 也为人口发展的中长期预测留下了研究的空间。

2007年初发布的《国家人口发展战略研究报告》^[2]指出, “人口众多始终是我国考虑一切问题的出发点和落脚点”。该报告分析了当前中国人口增长的特点, 对今后几十年中国人口的发展进行了预测, 并提出了人口总量控制的战略目标。但该报告没有公布数学模型。这样, 如何通过数学建模的方法对中国人口做出中长期的预测, 便成了我们感兴趣的一个课题。除了一般的规律外, 中国人口的增长还有着自身的特点和发展趋势, 这正是我们建立中国人口预测模型的出发点。在本赛题中也提到“老龄化进程加速、出生人口性别比持续升高, 以及乡村人口城镇化等因素, 这些都影响着中国人口的增长”。在已有模型研究成果的基础上进一步发展, 建立一个符合中国人口增长特点的数学模型, 使得我们能通过数学的方法来对中国人口的发展趋势做出较长期的预测, 并加强我们的国情意识, 这就是本命题的初衷。

2 解答要点

2.1 模型类型

首先要考虑的是建立什么类型的数学模型, 是确定性还是随机性? 由于人口增长的机理还是比较清楚的, 而且本题要求做出较长期的预测, 因此可以选择确定性系统的数学模型, 建立一个描述随时间演化过程的微分方程或差分方程。上述一般性的人口模型, 从指数增长模型到 Leslie 迭代矩阵, 也都属于这种类型。

建立微分或差分方程模型的关键在于如何确定可能起着重要作用的状态变量和参数, 并弄清它们之间的关系。这就需要对相关的数据进行深入分析。只有基于对实际数据的分析, 才可

能提出合理的假设、从而建立起反映各状态变量之间关系的模型；也只有在对实际数据分析的基础上，才能确定模型中的各项参数，并对所解出的结果进行检验，最后做出预测。

2.2 数据分析

应抓住人口增长中能反映规律性的数据进行分析。我们知道，对于一个地区或一类人群而言，人口的自然增长率=出生率-死亡率+迁移率，这里的迁移率符号当迁入时为正，迁出时负。而：

$$\text{出生率} = \text{育龄妇女的生育率} \times \text{育龄妇女在总人口中所占的比例}。$$

因此，必须对上述相关数据进行分析。

通过从《年鉴》^[3]摘录的“附录2”中的数据进行分析表明：

1) 生育率及出生人口的男女性别比在农村和城镇有着较大差别，农村育龄妇女的生育率及初生儿性别比明显高于城镇。而女孩偏少又将进一步影响出生率。因此有必要将出生率和男女性别比在农村与城镇人口中分别处理；

2) 大量农村人口持续不断转化为城镇人口对于中国人口的增长将造成很大影响，因此必须把城镇化作为农村人口向城镇迁移的主要因素考虑在模型中去；

3) 从中国人口年龄分布的数据分析中可以看出，近年来人口出生呈下降趋势。而死亡率的数据分析则告诉我们，从50岁起，人口的死亡率才逐步上升，在这之前一直是比较低的。这表明在相当一个时期人口进入老龄化的“速率”将高于出生率，致使老龄人口的比重不断加大。另一方面，老龄化进程的加速反过来又对人口的总体出生率和死亡率造成影响，所以必须把老龄化这个年龄的因素考虑进去。

2.3 模型的建立

由于所考虑的人口模型是一个动态的确定性模型，在建模时就应当首先考虑到如何设置状态变量。据上分析，应当把城镇人口与农村人口，及男女性别区分开来；另一方面，注意到育龄妇女的生育率是决定人口增长的主要因素，以及老龄化在人口死亡率中所起的主要作用，所以还应当把人口的年龄分布考虑在模型中。

较多的同学用到了 Leslie 人口迭代矩阵的思想，其特点是引进了年龄结构。这是合适的。但是 Leslie 矩阵原来的模型^[1]比较粗糙，比如，在讨论到人口的出生率时没有将育龄妇女的生育率及出生人口的性别比按(城镇与农村)不同区域分离开来，当然也没有这些区域之间的迁移过程，因此，在用该模型思想时必须把这些因素考虑进去。这样便可以建立起一个差分方程组初值问题的模型。

还有不少同学用到偏微分方程模型^[2]，这是在宋健等^[4]的人口模型中采用过的。该模型含有年龄结构，所以也是合适的。当然也需要将人口的出生率及性别比分城镇和农村两个“区域”讨论，并引进“区域”之间的迁移过程。所得到的偏微分方程的初边值问题有一般的数值解法。许多学生取一年作为步长对该类方程离散化，其过程与上述矩阵迭代方法基本一致。

不管用什么方法，这次发表的优秀论文均考虑到中国人口在城镇和乡村不同区域的差异和迁移过程，这都是成功的。

本期有一篇《中国人口增长的常微分方程模型》，是我学生本科毕业期间做的。虽不很令人满意，考虑到还没有见到同类模型，发表出来供参考。

2.4 确定参数

本赛题的另一个重点是对参数的估计。模型中可能出现的参数(包括初值条件)很多，如人口密度、出生率、(总和)生育率、生育模式、性别比、死亡率、存活率、迁移率、城镇化率、平均年龄、平均寿命及老龄化指数等。这些参数都要求根据《年鉴》^[3]或别的文献中的实际数据进行分析来确定。

从实际数据出发,基于一定的机理,便可以通过一定的拟合方式来确定这类参数。所采用的方法很多,如通常的曲线拟合法、最小二乘法、回归分析法、BP神经网络算法、灰色系统GM(1,1)算法和支持向量机(SVM)等方法。一些重要参数的拟合函数还可以在文献中查到,如与生育率相关的正态分布和 Γ 分布。这些方法散见于同学们的论文中,在此次发表的优秀论文中也有提及,限于篇幅,本文就不一一介绍了。

这里说一下关于城镇化参数的估计。在“报告”^[2]中估计今后20年内人口城镇化水平是年均增长1%。这里把市和镇放在一起,称为“城镇”,而乡单列,称为“乡村”或“农村”。但《年鉴》^[3]中的人口数据都是按市、镇和乡(或县)分别给出的,每年数据不一样。本题的“城镇化”也可以这样理解。这就需要对“市”和“镇”的每个要用到数据进行加工,即按“附录2”最右边给出的每年人口的抽样调查数据进行加权换算。有许多同学没有这样做。当然,一些同学把城、镇、乡分成三个不同区域来讨论它们之间的迁移关系,这也是可以的。

此外,在本题的数据说明中曾指出“个别数据有异常,原文如此,可酌情处理”。实际上,“附录2”中关于2003年育龄妇女的生育率数据应该是百分比,但按《年鉴》^[3]原是以千分比计,与实际相差10倍。许多同学注意到这一点,把它们作为异常数据剔除,这是合适的。

2.5 模型的求解和预测

用适当的数值方法进行迭代,即可求得数学模型的解,从而得到今后几十年的预测结果。许多同学都能够根据自己建立的数学模型计算出预测结果,绘出图像。

模型的预测分为中短期预测和长期预测,而“中短期”和“长期”的概念与区分是要同学们自己定的。一般把中短期定为少则5年,多则一、二十年,而长期至多是到五、六十年,这都是合适的。在《报告》^[2]中预测到2050年。可以把自己的结果与《报告》等文献中的相应结果进行对照分析。如出现较大差异,则应找出原因,予以改进,或提出自己的看法。

有些同学在完成所要求的中短期及长期预测的同时,还就不同的控制措施(生育政策)对未来人口发展的影响做了模拟和分析;还有些同学对老年化或性别比的变化进行预测,这些做法都是值得肯定的。

3 问题与展望

3.1 发现的问题

从阅卷中发现在同学论文中有下面一些问题:

1) 关于模型的选择:前面说过,本模型宜考虑用具有确定性的微分或差分方程。有的同学采用灰色系统模型,或一些统计模型,这些对于确定参数及人口的近期预测是比较有效的,但难以对人口中长期的变化做出预测。

2) 有许多同学没有考虑到农村人口城镇化。事实上,城镇化是本题模型与一般人口模型的一个重要区别。因为育龄妇女生育率及出生人口性别比在城镇与乡村的差异导致人口发展的不均衡,而持续的城镇化将对中国人口的变化产生重要影响,在预测模型中是不可缺少的。

3) 年龄结构:特别是对于中国人口增长问题,诸如生育率(包括晚婚晚育)、老龄化及死亡率之类的问题都与年龄结构密切相关。有些同学这方面注意不够。

4) 学风问题:预测的结果本应是根据自己的模型计算出来的,但发现有些同学的模型与所得的数值结果明显不一致。有的对所得结果优缺点的评价也不够实事求是。最后应列出自己参考过文献,包括论文、著作和数据,都要注明出处,如果是网上的,则应列出网址,但发现仍有些同学没有这样做。这些都是学风问题,应引起注意。

3.2 展望

关于人口增长的数学建模问题可以从深度和广度两个方面做进一步的研究。深度是指模型本身的数学与数值方面；广度则是指与该模型之外相关领域的联系。

我们看到，人口增长有迟滞效应。在《报告》^[2]中提到“由于20世纪80年代至90年代第三次出生人口高峰的影响”，导致在2005-2020年出生人口数量会“出现一个小高峰”，这就是迟滞效应。如果在模型中适当引进迟滞项，就可预测到这种“小高峰”现象。由此可以建立起一个关于人口增长的迟滞微分方程模型，当然，此时的初值应当是一个近几十年来人口的变化函数。此外，人口在发展过程中还会受到各种随机因素的干扰，由此可以考虑建立一个关于人口变化的随机微分方程模型。这些都涉及到更多的数学知识。而以人口增长预测为动因，还可以发展出更加有效的计算技术、方法和工具。

另一方面，人口发展是个社会问题。《报告》^[2]就是把它作为当代中国社会整体中的一个部分来研究的。从人口的发展与其他领域的关系，比如从人口与经济、人口与资源环境、人口与文化教育等不同方面的研究中，都可以引申出很多很有意义的数学建模课题。

参考文献：

- [1] 姜启源, 谢金星, 叶俊. 数学模型[M]. 北京: 高等教育出版社, 2003
- [2] 国家人口发展战略研究课题组. 国家人口发展战略研究报告[J]. 人口研究, 2007, 31(1): 1-10
- [3] 国家统计局人口统计就业司. 中国人口统计年鉴[M]. 北京: 中国统计出版社, 1995-2006
- [4] 宋健, 于景元. 人口控制论[M]. 北京: 科学出版社, 1985

On Mathematical Models for Prediction of China Population

TANG Yun

(Department of Mathematical Sciences, Tsinghua University, Beijing 100084)

Abstract: In this article the background and solving idea for the contest problem A of CUMCM-2007 are introduced, and comments on solution papers of the participants are proposed. Finally, prospects for the further investigation are shown.

Keywords: mathematical modeling; China population; prediction