

# 四类预测人口方法的对比及 Logistic 人口生长模型的改进

王沛林

(北京理工大学 数学与统计学院, 北京 100081)

**摘 要:** 分析当前中国的人口形势, 基于现阶段人口政策, 利用 1970—2014 年的中国总人口数据, 建立了马尔萨斯人口模型、多项式拟合模型、ARIMA 模型和 Logistic 人口生长模型对我国未来的总人口进行预测, 并将这 4 种预测方法进行对比, 得出 Logistic 人口生长模型的拟合优度和预测效果更优的结论, 而后又针对 Logistic 模型存在的缺陷分别从方程和参数的角度提出了 2 种改进办法, 使模型的精确度进一步提高. 最后预测 2030 年我国的总人口数将在 14.2 亿~14.6 亿波动.

**关键词:** 人口预测; Logistic 人口生长模型; 人口数据

**中图分类号:** O213.9

**文献标志码:** A

**文章编号:** 1674-2494(2022)03-0087-12

在当今科技经济飞速发展的社会中, 人口问题一直是世界各个国家和各个领域所关注的焦点问题, 它是一个社会发展的最基础也是最关键的问题. 为了维持社会生活的基本平衡与稳定, 把人口数量保持在一个比较理想化的状态, 就要对国家和地区的人口总数进行合理预测, 以便制定未来人口政策, 使社会有序发展. 在早期的研究中, 进行人口定量预测的基础方法就是依据国家统计局的人口结构和规模的资料和数据, 以当前人口为出发点, 并且对未来人口的变化趋势作出合乎常理的假设, 运用科学的数学模型进行拟合演算, 进而预测出未来人口总数、性别比例、结构组成等一些重要人口要素.

社会的进步需要人口稳定可持续发展, 基于中国人口数量现状和上述已知方法, 本文通过比较马尔萨斯人口模型、多项式拟合模型、ARIMA 模型和 Logistic 人口生长模型的预测效果, 着眼于未来人口总数的估计分析, 关注人口发展的总体趋势, 并将现有 Logistic 人口生长模型从微分方程和模型参数拟合优化角度进行改进, 得到了更加准确的预测方程, 以便更好地预测今后一段时间中国的总人口数, 具有一定的预见性.

## 1 中国人口发展预测模型的建立

### 1.1 数据的选取

拟合数据的选取十分重要. 经过资料调查, 我们知道从 20 世纪 70 年代, 在全国范围内开始大力实施计划生育政策<sup>[1]</sup>, 所以在预测全国总人口时, 我们选取 1970 年为基础年份, 1970 年的总人口数量为最初人数. 进而考虑到数据选取的原则, 数据选择得越多, 即样本越多, 能准确预测拟合人口数量的可能性越大, 结果就越精确, 误差就越小, 所以选择的年份不能少于 25 年, 否则将很难发现人口数量变动的趋势. 再有, 我国自 2016 年 1 月 1 日起, 正式开始实行“全面二孩”政策<sup>[2]</sup>, 这项政策的实施在短期内不会对我国总人口数的变化趋势产生较大的影响. 因此, 在本文中先不考虑近几年二胎方针的影响.

收稿日期: 2021-11-26

作者简介: 王沛林(1997-), 女, 河北保定人, 硕士研究生, 主要研究方向为统计学.

最后通过查阅中国国家统计局网站,获得 1970—2014 年每年中国人口总数,以这 45 年的数据作为拟合数据,保证了一定的准确性.

## 1.2 模型假设

- 1) 现有的基本国策“计划生育”保持不变.
- 2) 不考虑战争等不可预见因素.
- 3) 由于环境等因素,人口不能无限制地增长.

## 1.3 马尔萨斯人口模型及其应用

### 1.3.1 模型的理论背景

人口学家马尔萨斯在生物增长定律中提到:人口的变化率与人口总数成正比<sup>[3]</sup>,且这个比率不随时间变化,为常量.由此,我们设定人口总数为  $p(t)$ ,人口变化率为  $l$ ,若已经知道在  $t_0$  时刻的人口数量值  $p_0$ ,则就可以得到马尔萨斯人口模型:

$$\begin{cases} \frac{dp(t)}{dt} = lp(t), \\ p(t_0) = p_0. \end{cases} \quad (1)$$

这就是非常著名的分析人口数量问题的基础微分方程模型,通过求解方程,我们可以得到它的另外一种表达形式:

$$p(t) = p_0 e^{l(t-t_0)}. \quad (2)$$

其中,  $t_0$  表示最开始的时间(即初始年份),  $p_0$  表示  $t_0$  时刻的人口总数,  $t$  表示一般意义上的时间,  $p(t)$  表示  $t$  时刻的人口总数.

### 1.3.2 模型的建立与实现

将 1970—2014 年每一年的总人口数作为基础数据,代入模型中.为了进行数据拟合,首先要考虑模型的求解实现问题.

参数估计:  $l, p_0$  可以用已知数据,运用线性最小二乘法进行估计.(2)式两边取对数,得到:

$$\ln p(t) = \ln p_0 + l(t-t_0). \quad (3)$$

作对数变换后,再以 1970—2014 年的人口数据拟合(3)式,经计算得到:  $l=0.0112, p_0=88\,521.43$ .

最后,得到马尔萨斯人口模型:

$$p(t) = 88\,521.43 e^{0.0112(t-t_0)}. \quad (4)$$

其中  $t_0=1970$ .

得到模型的表达式以后,我们进而对 2015—2019 年的人口进行预测.

### 1.3.3 模型的误差检验分析

评价一个模型的好坏以及拟合效果,首先可以通过实际值和拟合值作出的图像来直观地对比分析.

如图 1 所示,在马尔萨斯人口模型预测图中,这 45 年来实际人口和预测人口的总体增长趋势大体是一致的,但是到后期预测人口的增长率要明显大于实际人口的增长率.

通过计算误差来进一步分析模型的优

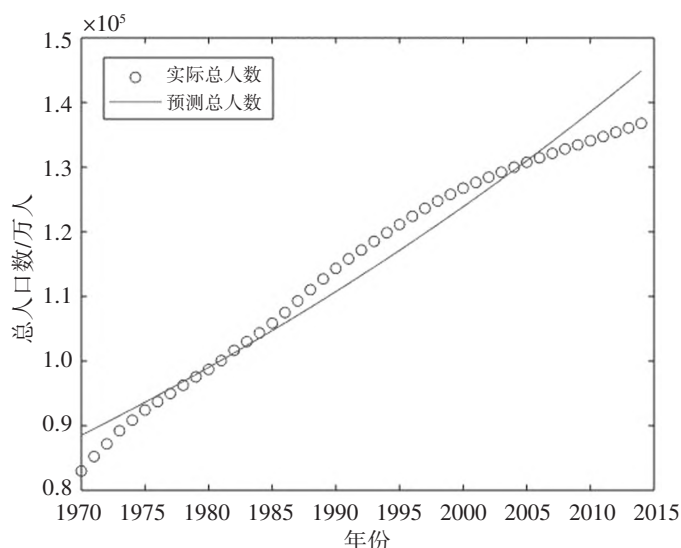


图 1 马尔萨斯人口模型预测 1970—2014 年总人口

劣.在统计学中,我们惯用  $y_i$  来代表第  $i$  组数据的真实值,  $\hat{y}_i$  来代表第  $i$  组数据的预测值,用  $n$  来表示样本数据的组数,所以有误差指标为:相对误差( $RE$ ):  $\frac{|y_i - \hat{y}_i|}{y_i}$ , 残差平方和( $resnorm$ ):  $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ , 标准误差( $RMSE$ ):  $\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$ ,  $R_{new}: 1 - (\frac{resnorm}{\sum y_i^2})^{\frac{1}{2}}$ .

相对误差越小,模型可信度越高;残差平方和越小,模型拟合效果越好;标准误差越小,模型精密程度越高; $R_{new}$  为非线性回归方程的拟合优度指标,其值越靠近 1,模型拟合程度越高<sup>[4]</sup>.

知悉这几个误差指标后,用 Matlab 计算出由马尔萨斯人口模型拟合的这 45 年数据的相对误差最大值为 0.066 6,残差平方和为 4.941 7,标准误差为 0.331 4,  $R_{new}$  为 0.971 3.

在物理学中,一般相对误差不超过 5%算是比较精确<sup>[5]</sup>,但此模型最大相对误差 0.066 6 要高于 0.05,并且  $R_{new}$  的值离 1 还有一定距离,由此看来马尔萨斯人口模型的精确度不是很高.

### 1.3.4 模型的预测效果及总体评价

将前文利用马尔萨斯人口模型预测出的 2015—2019 年中国总人口数与国家统计局网站上的真实数据进行对比,以此来判断预测效果.

由表 1 可以计算出,这 5 年的相对误差最大值为 0.094 6,远大于 0.05,误差较大,预测效果并不是很理想.

进一步绘制 2015—2030 年的人口预测图,从图 2 中可以看出预测人口的年增长率过高,人口数量呈几何状增长,不太符合实际.

从马尔萨斯人口模型的拟合效果和预测结果来看,计算出来的误差较大,模型并不是很精确,其预测的人口数量呈指数增长,增长率过高,故该模型的现代利用价值有限.

## 1.4 多项式拟合模型及其应用

### 1.4.1 模型的理论背景

从图 1 中我们可以观察到,人口数量的变化过程线上各点斜率都不太相同,这就表明各个阶段的人口增长速率是不一样的,这时用单纯的一条直线来描绘人口的发展显然是不合适的,所以要用更贴近实际的曲线来拟合人口数量的变动,这就产生了多项式拟合模型.

### 1.4.2 模型的建立与实现

为了保证多项式拟合模型的合理性和准确性,并且兼顾误差最小的原则,还要使模型具有可解释性,故本文选取了三次多项式对 1970—2014 年的总人口数据进行拟合.

将时间  $t$  作为解释变量,中国总人口数量  $p(t)$  作为预测变量,建立多项式确定它们之间的关系.得出如下结果:

$$p(t) = -0.000\ 035\ 76t^3 + 0.212\ 3t^2 - 420t + 276\ 900. \quad (5)$$

表 1 中国 2015—2019 年实际总人口数与预测总人口数 万人

年份	实际总人口数	预测总人口数
2015	137 462	146 530
2016	138 271	148 180
2017	139 008	149 850
2018	139 538	151 540
2019	140 005	153 250

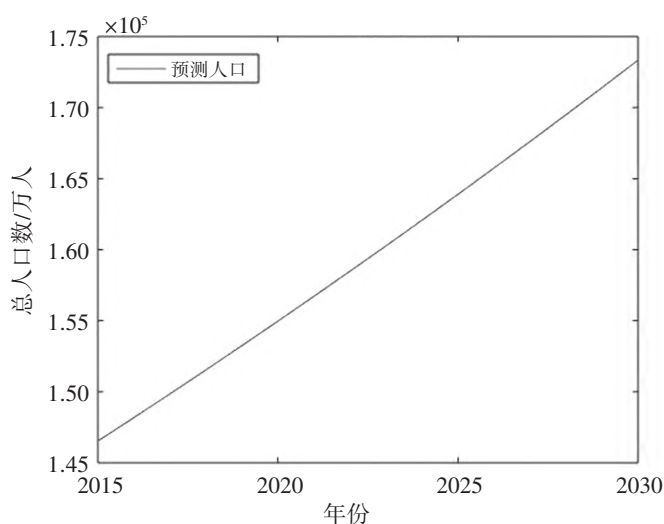


图 2 马尔萨斯人口模型预测 2015—2030 年总人口

### 1.4.3 模型的误差检验分析

通过图 3 拟合图像可以看出,在 1970—2014 年中,拟合的曲线很贴近实际的数据,三次多项式拟合模型的短期预测效果很好,没有太大的误差。

三次多项式的相对误差最大值为 0.013 3,残差平方和为 0.122 3,标准误差为 0.052 1, $R_{new}$  为 0.995 5.由误差可以看出,多项式拟合模型的精确度较马尔萨斯人口模型有很大程度的提高。

### 1.4.4 模型的预测效果及总体评价

利用三次多项式拟合模型的表达式,预测出 2015—2019 年中国人口总数分别为 136 090 万人、136 194 万人、136 218 万人、

136 161 万人、136 021 万人,将这 5 年来的预测人口同实际人口相比较,我们得出最大的相对误差值为 0.028 5,没有超过 0.05,且要远低于马尔萨斯人口模型的误差值,预测效果比较好。

从三次多项式拟合模型的曲线贴合效果和部分预测结果来看,各种误差都比较小,模型比较合理,但是 2015—2019 年预测出来的人口数量总体呈现先增长后下降的趋势,人口数量下降在我国现阶段几乎是不可能的,所以笔者认为多项式拟合模型只适合短时段 1~3 年内的人口预测,长期预测则会出现大的偏差,效果不会很好。

## 1.5 ARIMA 模型及其应用

### 1.5.1 模型的理论背景

由前面的多项式拟合模型可以看出,人口数量是随着时间不断变化的.但是人口与时间的关系又不能简单地用多项式来准确的描述,为了寻找更合理的人口预测模型,我们想到了经济学中的 ARIMA 模型。

预测一种事物的变化时,用它的过去推断它的未来,即用时间序列的往日数据表示事物随时间变化的种种规律,并将两者之间的模式应用到未来,进而对未来的数据作出预测. ARIMA 模型就是一种常用的时间序列模型。

ARIMA( $p, d, q$ ) 模型含有 3 个参数,这里面  $p$  表示自回归(AR)阶数,  $q$  表示移动平均(MA)阶数,  $d$  表示模型的差分阶数<sup>[6]</sup>.该模型有 3 种基本类别<sup>[7]</sup>: AR( $p$ ) 模型、MA( $q$ ) 模型、ARIMA( $p, d, q$ ) 模型。

### 1.5.2 模型的建立与实现

利用 1970—2014 年的中国总人口数据,建立 ARIMA( $p, d, q$ ) 模型.模型的建立步骤如下:

1) 载入数据,进行一次单位根检验,判定数据序列的平稳性.本文得到  $p$  值为 0.99,大于检验的临界值 0.05,所以该序列不平稳。

2) 对不平稳序列进行差分处理,差分的次数为  $d$ ,使其成为平稳序列.经过了三次差分操作后,一次单位根检验的  $p$  值小于了 0.05,序列化为平稳序列,进而确定  $d=3$ 。

3) 对已经平稳的序列进行白噪声检验,判断其是否为纯粹的随机序列,由于对白噪声进行下一步的处理和预测没有任何价值,故使用 Ljung-Box 方法<sup>[8]</sup>进行检测,  $p$  值为 0.008 808,小于标准值 0.05<sup>[8]</sup>,所以三次差分后的平稳序列不是随机序列,可以展开下一步工作。

4) 确定模型的阶数.画出序列的自相关图(ACF)和偏自相关图(PACF),见图 4.据观察可以发现

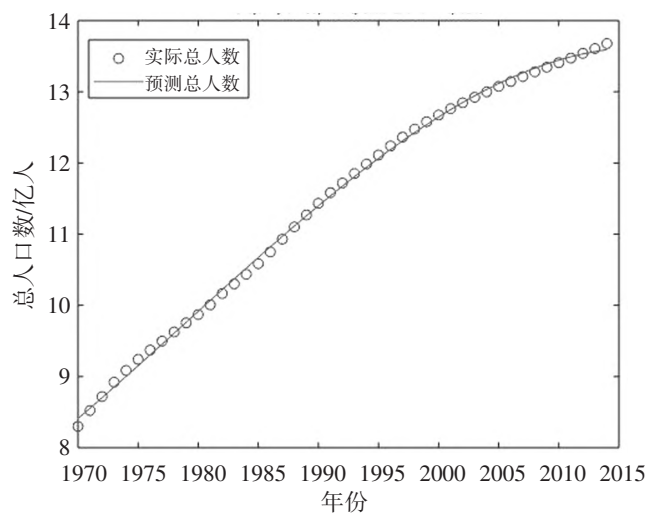


图 3 三次多项式拟合模型预测 1970—2014 年总人口



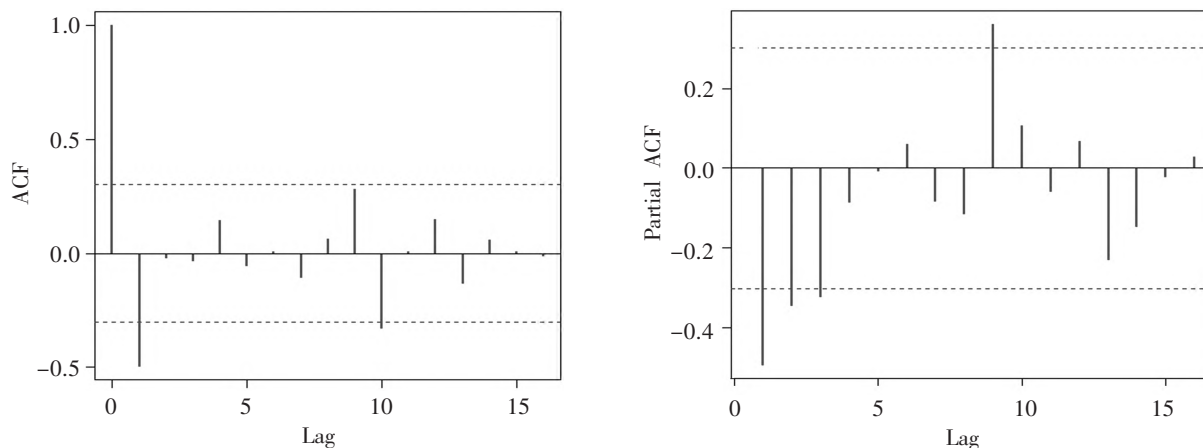


图 4 自相关图与偏自相关图

ACF 1 阶截尾, PACF 具有拖尾性. 综上, 初定模型为 ARIMA(0,3,1). 接下来用 R 语言软件中的自动定阶功能(auto.arima)来对上面的定阶结果进行准确性检验, 结果也显示为 ARIMA(0,3,1) 模型.

5) 确定模型的系数. 得出该模型为  $x_t = \varepsilon_t + 0.769\varepsilon_{t-1}$ .

6) 模型的显著性检验. 使用 Ljung-Box 方法<sup>[8]</sup>检验残差:  $p$  值为 0.287 8, 大于 0.05, 说明残差是随机序列, 白噪声检验通过. 参数的显著性检验: 用 R 软件测出的系数 -0.769 0 除以它的标准误差 0.122 7, 商的绝对值为 1.96, 大于  $T$  检验统计量在 5% 水平的临界值, 断定系数显著不等于零<sup>[9]</sup>, 检验通过.

7) 用 ARIMA(0,3,1) 模型预测 2015—2019 年的中国人口总数.

### 1.5.3 模型的误差检验分析

从图 5 可以看出, ARIMA 模型作出的预测数据与实际数据比较吻合, 相差不大, 效果良好.

相对误差最大值为 0.003 3, 残差平方和为 0.006 3, 标准误差为 0.011 8,  $R_{new}$  为 0.999 0. 从误差值来看, 模型的拟合水平非常高, 较前面 2 种模型而言, 有着更高的可信度.

### 1.5.4 模型的预测效果及总体评价

利用 ARIMA 模型预测 2015—2019 年的中国人口总数, 将预测值与实际值作比较, 得出相对误差的最大值为 0.002 8, 远小于前 2 种模型的误差值, 预测效果有了质的飞跃.

进而利用此模型预测出了 2015—2030 年中国的总人口数, 并绘成图 6. 看出未来 10 年我国人口将呈缓慢增长态势, 到 2030 年不会破 15 亿大关.

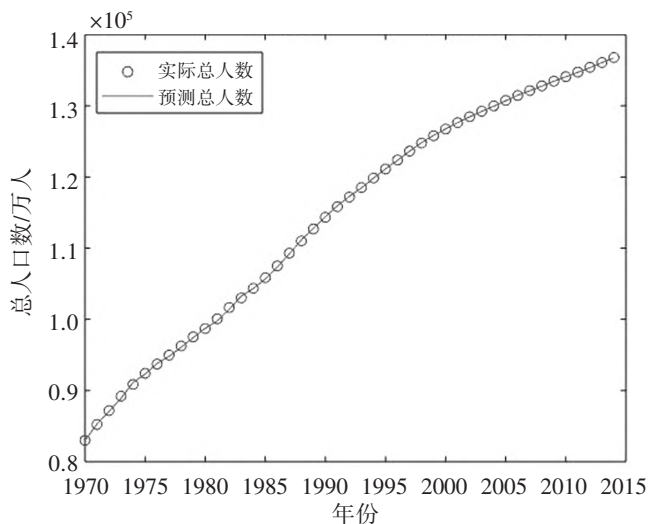


图 5 ARIMA 模型预测 1970—2014 年总人口

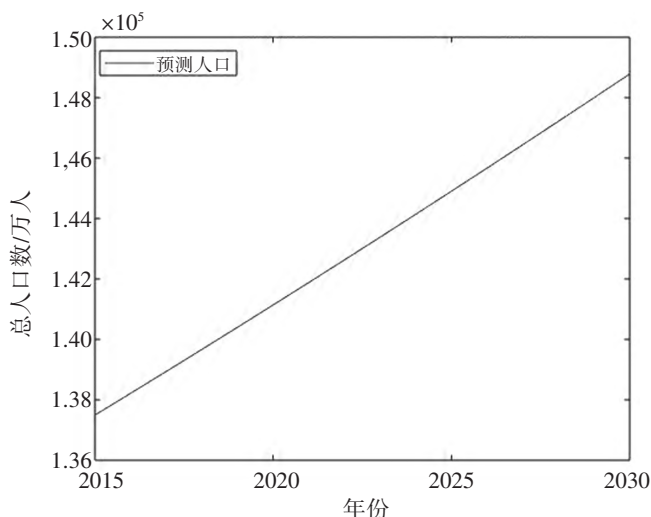


图 6 ARIMA 模型预测 2015—2030 年总人口

根据综上所述各种论证来看, ARIMA 模型的预测效果远优于前 2 种模型, 给出了较为合理的人口随时间变化的趋势, 精确度比较高, 有较大的参考价值.

## 1.6 Logistic 人口生长模型及其应用

### 1.6.1 模型的理论背景

在前面讨论过的马尔萨斯人口模型中有致命的缺点——人口变化率  $l$  为常量. 它只假设出了人类发展最理想的情况, 并得出人口数量将呈指数上涨的极端结论. 认识到马尔萨斯人口模型的不足之后, 有学者在其基础上进行了改进, 形成了 Logistic 人口生长模型. 模型的中心观点为: 人口数量并不能无穷尽地上涨, 它会受到各种因素的约束, 且随着人口的不断增加, 这种约束力会不断变强, 最终人数会到达一个极值.

这种约束力作用在人口变化率  $l$  上, 它将不再是一个常量, 人口数量越多,  $l$  就会变得越小. 如果将  $l$  表示为关于人口总数  $p(t)$  的函数  $l(p)$ , 那么它是减函数. 则有方程:

$$\begin{cases} \frac{dp(t)}{dt} = l(p)p, \\ p(t_0) = p_0. \end{cases} \quad (6)$$

假设  $l$  与  $p(t)$  是最简单的线性关系, 有:

$$l(p) = l - sp, \quad l > 0, s > 0. \quad (7)$$

这里的  $l$  与上面的意义有所不同, 它被称作内禀增长率<sup>[10]</sup>, 即人口数几乎为零并且不受条件约束时的增长率. 现在需要弄明白系数  $s$  的内涵, 我们引入当前社会环境可以容纳的最大人口数量  $p_m$ , 当人口总数  $p$  达到  $p_m$  时, 人口不再增长, 即  $l(p_m) = 0$ , 代入方程(7)得  $s = \frac{l}{p_m}$ , 接着方程(7)变为:

$$l(p) = l \left( 1 - \frac{p}{p_m} \right). \quad (8)$$

把此式再代入方程(6), 就得到 Logistic 人口生长模型:

$$\begin{cases} \frac{dp(t)}{dt} = l \left( 1 - \frac{p}{p_m} \right) p, \\ p(t_0) = p_0. \end{cases} \quad (9)$$

这个方程可用分离变量法解出得:

$$p(t) = \frac{p_m}{1 + \left( \frac{p_m}{p_0} - 1 \right) e^{-l(t-t_0)}}. \quad (10)$$

### 1.6.2 模型的建立与实现

从上述的理论基础中可知, 如果想获取方程(10)的解, 就要估计出  $p_m$  和  $l$  这 2 个参数的值, 求解方法有 2 种.

#### 1.6.2.1 三点等间距法

利用昔日的人口数据, 粗略计算出  $l$  和  $p_m$  的值. 再采用  $t_0, t_1, t_2$  这 3 个年份的人口总数量  $p_0, p_1, p_2$ , 且满足条件  $t_2 - t_1 = t_1 - t_0 = a$ , 代入方程式(10)进行计算, 可得出:

$$l = \frac{1}{a} \ln \frac{\frac{1}{p_0} - \frac{1}{p_1}}{\frac{1}{p_1} - \frac{1}{p_2}}, \quad p_m = \frac{1 - e^{-la}}{\frac{1}{p_1} - \frac{1}{p_0} e^{-la}}. \quad (11)$$

取 1970 年作为开端年份  $t_0$ , 取 1992 年为中间年份  $t_1$ , 取 2014 年为最末年份  $t_2$ , 此时  $a=22$ , 将相应数值代入公式(11), 获取  $l$  和  $p_m$  的具体值; 将计算出的 2 个参数值代入方程(10)中, 得到模

型:  $p(t) = \frac{150\ 210}{1 + 0.809\ 9 e^{-0.048(t-1970)}}$ , 模拟图见图 7.

#### 1.6.2.2 *lsqcurvefit* 最小二乘法

在 Matlab 软件中有一个 *lsqcurvefit* 命令, 依据最小二乘原理:  $\min \sum_{i=1}^n (y_i - \hat{y}_i)^2$ , 即使模拟数据和实际数据之间的残差平方和最小, 也可拟合出最优化的非线性曲线.

选取  $t_0=1970$ ,  $p_0=82\ 992$  代入式 (10), 然后将式 (10) 作为未知的需要拟合的非线性曲线,  $l$  和  $p_m$  为要求解的参数. 给出两参数的初值解, 1970—2014 年的总人口为实际数据, 用 *lsqcurvefit* 函数进行拟合,  $p_m$  为 153 147 万人,  $l$  为 0.045 4, 有模型:  $p(t) = \frac{153\ 147}{1 + 0.845\ 3 e^{-0.045\ 4(t-1970)}}$ , 最后画出对比图, 见图 8.

#### 1.6.3 模型的误差检验分析

从图 7、图 8 中可以剖析出, 三点等间距法和最小二乘法的模拟结果都不错.

三点等间距法: 相对误差最大值为 0.018 2, 残差平方和为 0.305 0, 标准误差为 0.082 3,  $R_{new}$  为 0.992 9. 最小二乘法: 相对误差最大值为 0.013 5, 残差平方和为 0.247 7, 标准误差为 0.074 2,  $R_{new}$  为 0.993 6. 很明显, 最小二乘法还是要略优于三点等间距法, 这可能是因为 *lsqcurvefit* 函数能够找到全局最优解, 而三点等间距法只是拿出 3 年的数据来求解方程 (10), 有局限性.

#### 1.6.4 模型的预测效果及总体评价

综上, 我们选用最优的 *lsqcurvefit* 最小二乘法预测出了 2015—2019 年的总人口数, 与实际人口比, 最大相对误差值为 0.003 9. 进而本研究用 Logistic 人口生长模型预测出了 2015—2030 年的总人口, 见图 9.

从图 9 可以发现, 在未来 10 年我国人口增长有逐渐放缓的趋势, 总体较平稳.

总体来说, Logistic 人口生长模型比马尔萨斯人口模型考虑的影响人口变动的因素更多更充分, 预测结果也更加确切、合理. 但是, 从误差的角度来看, 该模型还不如 ARIMA 模型精确, 有改进优化的空间.

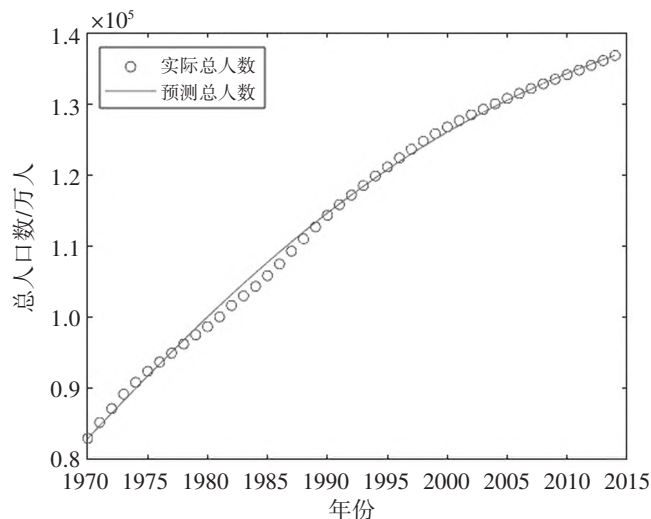


图 7 Logistic 模型三点等间距法预测 1970—2014 年总人口

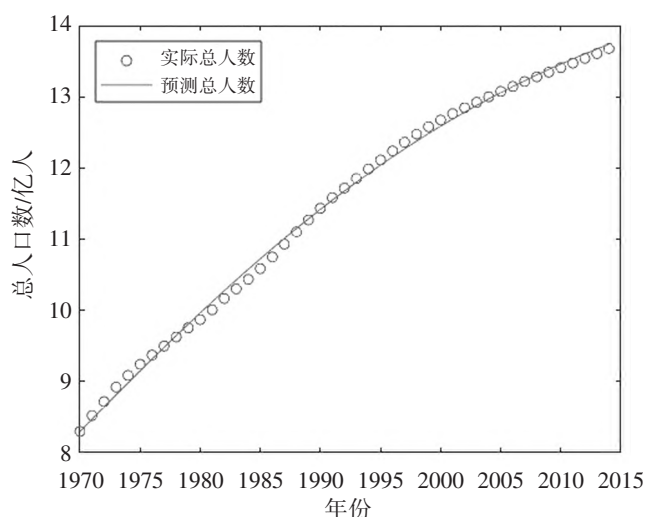


图 8 Logistic 模型最小二乘法预测 1970—2014 年总人口

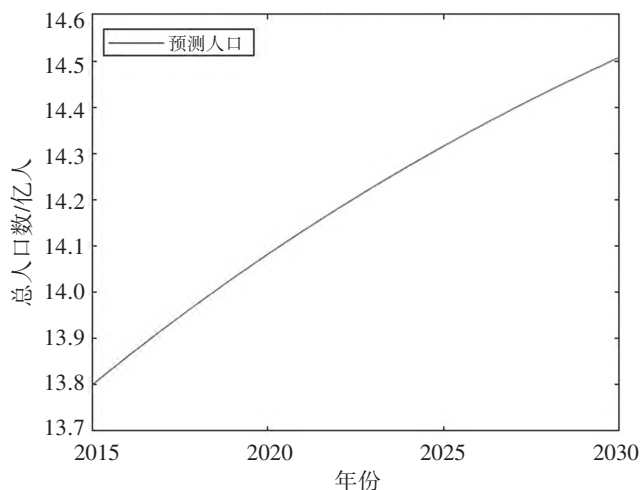


图 9 Logistic 人口生长模型预测 2015—2030 年总人口

## 2 4 种预测方法的比较分析

通过表 2 我们清楚地认识到,无论是从误差最小化角度,还是从模型的拟合效果和预测结果方面,Logistic 人口生长模型和 ARIMA 模型都要优于其他两种模型,并且 ARIMA 模型的精确性还要比 Logistic 人口生长模型突出,但是 ARIMA 模型的局限

表 2 4 种预测模型的比较分析

模型类别	相对误差	残差平方和	标准误差	$R_{new}$
马尔萨斯人口模型	0.066 6	4.941 7	0.331 4	0.971 3
多项式拟合模型	0.013 3	0.122 3	0.052 1	0.995 5
ARIMA 模型	0.003 3	0.006 3	0.011 8	0.999 0
Logistic 人口生长模型	0.013 5	0.247 7	0.074 2	0.993 6

性在于它只适合短时间内的预测<sup>[1]</sup>,而 Logistic 人口生长模型作为中长期人口预测的一种典型方法,它的精确度还有待提高.为了更好地进行人口中长期预测,下文将探讨 Logistic 人口生长模型的改良方法.

## 3 Logistic 人口生长模型的改进

通过分析看出,Logistic 人口生长模型在预测人口时,其自身的合理性和精确度还有进一步提升的空间,为此我们从 2 种不同角度探索对此模型修正的方法.

### 3.1 方程角度改进及其应用

#### 3.1.1 改进的方法依据

从我国的当前人口形势出发,由于计划生育政策已实行了多年,我国过去人口猛烈增长的势头已经得到了很好的控制,久而久之又出现了生育率低、人口老龄化的现象.如今我国的人口增长率已经非常低,如方程(7)的假设——人口变化率  $l$  和人口总数  $p(t)$  是最简单的线性关系,已经不再适用于当下.现在我们从  $l$  与  $p(t)$  的关系入手,对  $l(p)$  函数式进行改进.

作出假设:人口上涨已经快到了负荷的程度,上涨率慢慢变小最终将趋向零.

根据假设, $l(p)$  理应表达为  $e^{-p}$  的函数,设:

$$l(p)=l(p_0) e^{-kp}. \quad (12)$$

其中  $l(p_0)$  是初始年份为  $t_0$ 、总人数为  $p_0$  时的人口增长率, $b$ 、 $k$  为待求系数.

于是,修正以后的模型为:

$$\begin{cases} \frac{dp(t)}{dt}=l(p_0) e^{-kp}p, \\ p(t_0)=p_0. \end{cases} \quad (13)$$

#### 3.1.2 模型的建立与实现

依旧使用 1970—2014 年的中国人口总数作为研究数据,这样上下文形成对照,方便比较拟合效果.先来求解方程(12)中的参数值  $b$ 、 $k$ .

首先,对方程(12)的两边作对数变换得到:

$$\ln l(p)=\ln l(p_0)+b-kp. \quad (14)$$

接着用  $\frac{p(1971)-p(1970)}{p(1970)}=0.026\ 954$  作为初始的人口增长率  $l(p_0)$ ,再算出 1971 年到 2014 年逐年的人口增长率作为  $l(p)$ .然后,将最后计算出的  $\ln l(p)$  和  $\ln l(p_0)$ 、1970—2014 年中国总人口数量都代入式(14)中,利用最小二乘法解出系数  $b$ 、 $k$  的值分别为 2.337 6 与 0.287 7.

最后,得到从方程角度进行修改后的模型:



$$\begin{cases} \frac{dp(t)}{dt} = 0.026954 e^{2.3376 - 0.2877p} p, \\ p(t_0) = p_0. \end{cases} \quad (15)$$

其中起始年份  $t_0$  为 1970 年.求出了方程(15)的数值解,并绘制拟合图像,见图 10.

### 3.1.3 模型的误差检验分析

根据图 10 可看出,从方程角度改进后的模型绘制的拟合曲线与现实中的数据比较贴近,在开始的 5~6 年和末尾的 5~6 年预测效果尤其好,在中间年份或多或少出现了偏差.依据程序运行结果,相对误差最大值是 0.026 3,残差平方和是 0.796 3,标准误差是 0.134 5,  $R_{new}$  是 0.988 4.从误差值的大小来看,此方法修正后的 Logistic 人口生长模型并没有较大改善.

### 3.1.4 模型的预测效果及总体评价

与前文一样,我们用改良的模型预测出了 2015—2019 年的中国人口总数,与实际人口比较,这 5 年的最大相对误差值为 0.004 0,这与没有修正之前的最大相对误差值 0.003 9 差距非常小,几乎没有改善.之后,用改进后的模型预测了 2015—2030 年的中国总人口,见图 11.

将图 11 与图 9 对比发现,改进前预测 2015—2030 年的中国总人口大约是从 13.8 亿上涨到 14.5 亿,改进后预测这 16 年的人口数大约是从 13.8 亿上涨到 14.7 亿多,首尾年都相差不大,说明我国未来人口的变化有很大可能按照这种趋势发展,并在 2030 年达到 14.6 亿左右的人口数.

无论是从误差指标的角度来看,还是从预测效果来看,修正了主方程的 Logistic 人口生长模型并没有很大程度上提高模型的效率和精确性,与没修改之前的预测大致相同,借鉴意义有限.

## 3.2 参数角度的模拟仿真改进及其应用

### 3.2.1 改进的方法依据

本文又从另一个角度——使参数值更加精确化,来对模型的修正问题进行探究.

在 Logistic 人口生长模型中,存在关系函数  $l(p) = l - sp$ ,  $l > 0, s > 0$ , 此式中含有 2 个待求参数  $l$  和  $s$ . 我们知道,这 2 个参数主要是通过专家预估<sup>[10]</sup>或者统计得来,而没有从函数式本身入手去求其值.所以,修正模型中  $l(p)$  的决定方式,根据 1970—2014 年逐年的总人口增长率与每年人口总数的关系,运用最小二乘线性拟合的方法来明确参数  $l, s$  的切实值,即通过真实数据的模拟仿真来确定参数值,

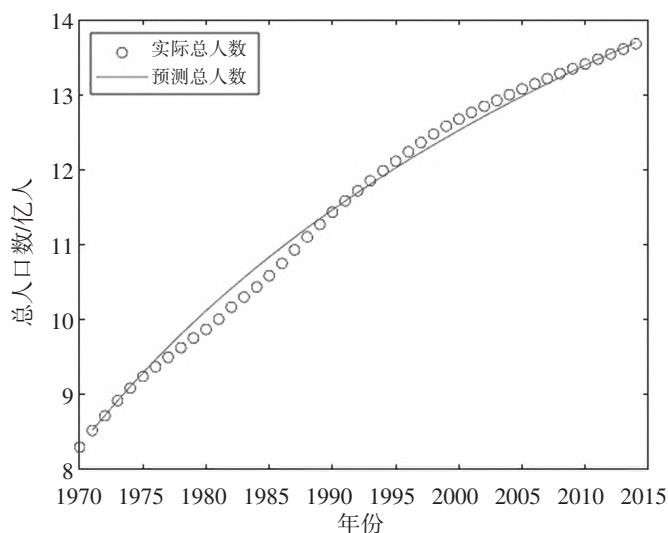


图 10 方程角度改进的 Logistic 人口生长模型  
预测 1970—2014 年总人口

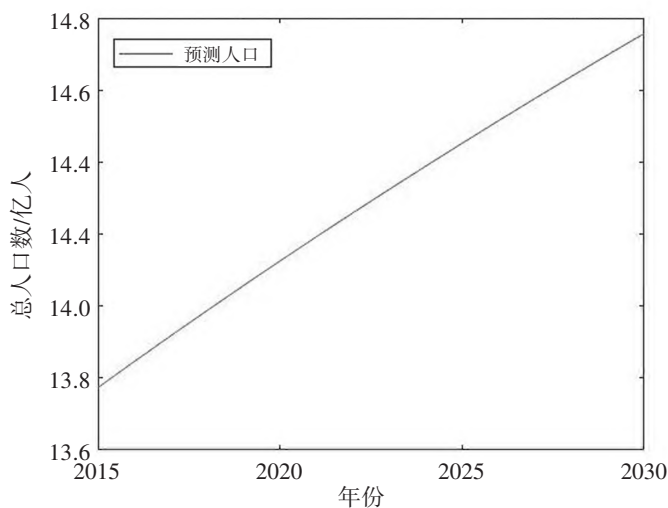


图 11 方程角度改进的 Logistic 人口生长模型  
预测 2015—2030 年总人口

而不是无端的估计统计,这样预测水平可能会有所提高.

### 3.2.2 模型的建立与实现

应用国家统计局的数据计算出 1970—2014 年中国总人口的逐年变化率,并制成散点图.

由图 12 可以观察到,总人口的增长率自 1987 年开始,几乎呈直线型下降,因此可以采用直线拟合的方式来得出函数  $l(p)=l-sp, l>0, s>0$  的具体表达式.

于是,1987—2014 年逐年的总人口增长率作为  $l(p)$ , 1987—2014 年每年的总人口数为  $p$ , 运用线性最小二乘法来拟合它们之间的关系式  $l(p)=l-sp$ , 最终得到具体表达式为:

$$l(p)=0.0664-0.0046p. \quad (16)$$

将 1987 年当作初始年份  $t_0$ , 1987 年所对应的总人口数为  $p_0$ , 2 个数值与式 (16) 一起代入方程 (6), 得到方程:

$$\begin{cases} \frac{dp(t)}{dt}=(0.0664-0.0046p)p, \\ p(1987)=10.93. \end{cases} \quad (17)$$

而后再求解微分方程 (17) 得到了预测总人口数  $p(t)$  的解析式:

$$p(t)=\frac{332}{7.3751e^{-0.0664(t-1987)}+23}. \quad (18)$$

实际总人口与预测总人口相比较, 见图 13.

### 3.2.3 模型的误差检验分析

查看拟合图 13, 我们发现改进后的模型作出的预测数据与真实数据很接近, 总体的增长趋势也相当一致.

通过计算, 得到了该模型下相对误差最大值是 0.0069, 残差平方和是 0.0860, 标准误差是 0.0554,  $R_{new}$  是 0.9956. 改进后模型计算出的误差指标值要比原模型计算出的误差值小得多, 拟合优度  $R_{new}$  也更加接近 1, 结果良好.

### 3.2.4 模型的预测效果及总体评价

同样, 我们用改良后的模型预测出了 2015—2019 年的中国总人口数, 分别为 137 480 万人、137 901 万人、138 298 万人、138 671 万人、139 023 万人, 预测值与实际 2015—2019 年总人口数的最大相对误差值是 0.0070, 与之前未改进时的预测效果相近. 接着我们又预测了 2015—

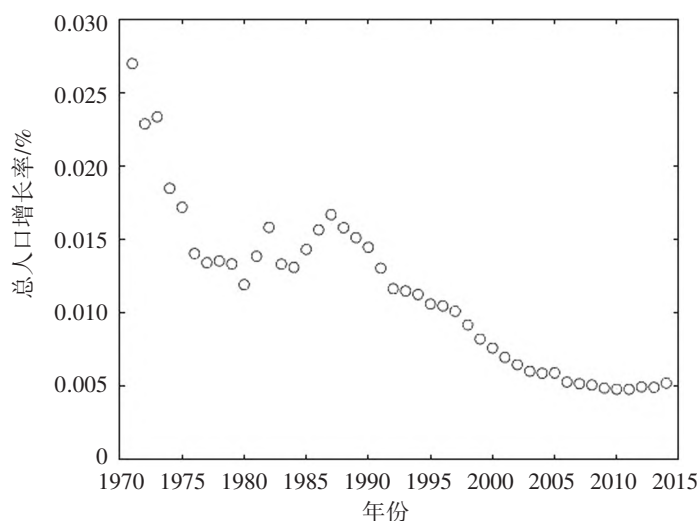


图 12 1970—2014 年中国总人口的增长率变化

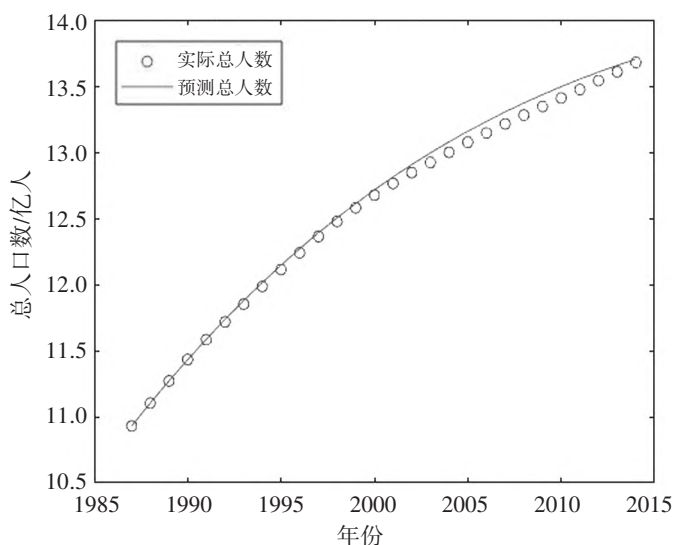


图 13 参数角度改进的 Logistic 人口生长模型  
预测 1987—2014 年总人口

2030 年的中国总人口数,见图14.

可以看出,按此种模型进行推测,我国的人口增长率减小的趋势将加快,人口增长将放慢.

利用真实数据对未知函数进行模拟仿真,进而求解出更加有实际意义的参数值,这种改进方法经过实践证明具有不错的效果,使模型的拟合优度有了较大提高,从误差值来看也比原始模型更具有说服力,精确性大幅度提高.

### 3.3 模型的改进效果剖析

无论是从模型存在的偏差还是从模型的拟合优度指标来看,从参数角度进行的

模拟仿真改进远远要比从方程入手对模型的改进效果理想得多.改进后的模型,参数值被赋予了更加具体的内涵而且也精确了许多,同时,与原有的 Logistic 人口生长模型相比,从参数角度改进后的模型精确度提高了,预测水平也升高了.

本文通过比较马尔萨斯人口模型、多项式拟合模型、ARIMA 模型、Logistic 人口生长模型这 4 种人口预测方法的优劣,认为 ARIMA 模型和 Logistic 人口生长模型的预测效果比较理想,并用比较合理的 Logistic 人口生长模型对未来的人口进行了预测,进一步针对 Logistic 人口生长模型存在的不足提出了 2 种改进办法,使模型的精确度进一步提高.最后预测,到 2030 年,我国的总人口将在 14.2 亿~14.6 亿波动.

### 参考文献:

- [1]王雪超.马克思主义人口理论视角下中国人口生育政策演变研究(1949—2018 年)[D].呼和浩特:内蒙古工业大学,2019.
- [2]全面二孩政策[J].中国经济周刊,2019(18):203.
- [3]沈小凤.几类人口预测模型的改进及实证分析[D].北京:中央民族大学,2012.
- [4]陈斯斯,徐爽,许丹,等.基于作者储备规模及作者生产率分布的中印、中日医学学科成熟度及科研实力分析[J].中华医学图书情报杂志,2017,26(7):21-26.
- [5]刘斌,周炜明.大学物理实验中误差教学衔接问题研究[J].广西物理,2015,36(1):46-49.
- [6]赵子铭.基于 ARIMA 模型的中国人口预测[J].赤峰学院学报(自然科学版),2019,35(9):10-12.
- [7]利小玲.浅析处理经济数据的几种方法[J].金融经济,2008(16):75-76.
- [8]Jafar Nouri Hajwal Hashem. 时间序列模型的应用研究[D].武汉:华中师范大学,2014.
- [9]蔺平爱,董晨昱,薛志诚.实证资产定价研究中的  $t$  统计量比较与应用[J].山西大学学报(自然科学版),2021,44(1):13-19.
- [10]顾忠盈,张孝义,耿济国.温度对药材甲实验种群生长的影响[J].南京农业大学学报,1993(1):33-37.
- [11]李姝敏,张勋尘.基于 ARMA 模型的兰州市人口短期预测研究[J].中国市场,2015(39):54-55.

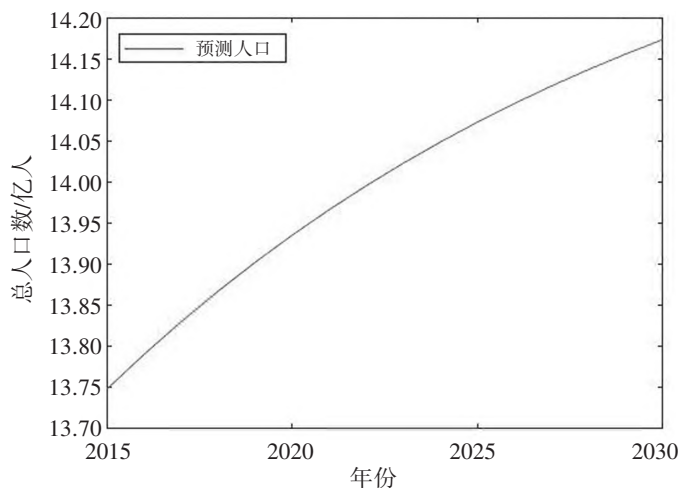


图 14 参数角度改进的 Logistic 人口生长模型  
预测 2015—2030 年总人口

## Comparison of Four Kinds of Population Prediction Methods and Improvement of Logistic Population Growth Model

Wang Peilin

(School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China)

**Abstract:** Based on the current population policy and the data of China's total population from 1970 to 2014, this paper analyzes the current population situation in China, establishes Malthus population model, polynomial fitting model, ARIMA model and Logistic population growth model to predict the future total population of our country, and compares these four prediction methods. The conclusion that Logistic population growth model is better in goodness of fit and prediction effect is obtained. Then two improved methods are proposed from the perspectives of equations and parameters to further improve the accuracy of the model according to the defects of Logistic model. Finally, it is concluded that the total population of China will fluctuate between 1.42 billion and 1.46 billion by 2030.

**Key words:** population prediction; Logistic population growth model; population data

(责任编辑 陈 静)

---

(上接第 86 页)

[3]郭 芳,刘志勇.函数的中心对称性及其在定积分中的应用[J].保定师范专科学校学报,2007(2):19-20.

[4]人民教育出版社中学数学室.几何:第二册[M].北京:人民教育出版社,1998.

## The Application of Central Symmetry Around the Point $(h, k)$ on the Definite Integral Operation

Guo Fang, Tian Shuhuan, Wang Shuyan

(College of Date Science and Software Engineering, Baoding University, Baoding, Hebei 071000, China)

**Abstract:** In this paper, the author studies the application of central symmetry on the definite integral operation. The necessary and sufficient condition of central symmetry around the Point  $(h, k)$  about the graph of function is given. Furthermore, the operational formula of the definite integral is given when the kind of function is integrand.

**Key words:** central symmetry; definite integral; the graph of function

(责任编辑 陈 静)