中国人口中长期预测及未来人口政策建议

摘　要

本文通过对以往数据的分析与建模，对2020年后的中国人口中长期(80年)变化进行了预测，得到了详细且高精度的结果，并给出了预测区间的估计。

在预测过程中，我们首先基于人口平衡方程和Leslie矩阵建立了贝叶斯分层模型，并通过马尔科夫链蒙特卡洛(MCMC)方法进行了带扰动项的逐次迭代，得到了总和生育率(TFR),预期寿命(LE)的中长期估计，并以此作为数据，建立队列要素模型对中国总人口和年龄结构变化的预测。结果表明，**中国人口将在2025年左右达到峰值水平(14.6亿人)，此后将逐年下降，直到2100年减少至12亿人左右。**由于中国生育政策的影响，我们对育龄妇女建立了孩次递进模型，分析了一孩妇女在不同二孩概率下，中国人口长期变化的趋势，结果显示，**当二孩出生率低迷时，中国人口将在未来80年内锐减近半数**，而二孩出生率高时，结果较为接近贝叶斯分层模型得到的结果。基于以上的模型和预测数据，我们对政府部门有如下的三点建议:

第一，指定福利政策鼓励生育，解决低生育率陷阱。

第二，减缓劳动力的老化与加速建设创新型发展。

第三，提供养老福利和全面社会保障，解决结构失衡问题。

关键字　Leslie矩阵　贝叶斯分层模型　MCMC方法　队列要素模型　孩次递进模型

Abstrat

In this paper, we analyze and model previous data to forecast the medium- and long-term (80year) changes of China's population after 2020, and obtain detailed and high-precision results and give estimates of the forecast intervals.

In the forecasting process, we first build a Bayesian hierarchical model based on the population equilibrium equation and Leslie matrix, and perform iterations with perturbation terms by Markov chain Monte Carlo (MCMC) method to obtain the medium- and longterm estimates of total fertility rate (TFR), life expectancy (LE), and use them as data to build a cohort element model for the change of total population and age structure in China. The results show that China’s population will change in the next 20 years. The results suggest that China’s population will reach its peak level (1.46 billion) around 2025 and will decline annually thereafter until it decreases to about 1.2 billion in 2100. Due to the impact of China’s fertility policy, we developed a child progression model for women of childbearing age and analyzed the trends of long-term changes in China’s population for one-child women with different probabilities of having two children. The results show that when the birth rate of two children is low, China’s population will plummet by nearly half in the next 80 years, while when the birth rate of two children is high, the results are closer to those obtained from the Bayesian hierarchical model. Based on the above model and predicted data, we have three suggestions for government depart ments:

First, designate welfare policies to encourage childbearing and address the low fertility trap.

Second, slow down the aging of the labor force with accelerating the construction of innovative development.

Third, provide old-age benefits and comprehensive social security to address structural imbalances.

Key words : Leslie matrix, Bayesian hierarchical model, MCMC method, Cohort component model Child progression model

一、 问题提出

随着第七次普查数据的公布，数据中揭示了我国人口发展呈现出一些新的特征。例如人口总量持续增长，但人口增速有所放缓；劳动年龄人口下降，人口老龄化逐步加速等特征。关于我国人口问题历年来已有大量的研究并积累了丰富的数据资料，2007 年全国大学生数学建模竞赛中的 A 题就探讨过我国人口的增长趋势。而近十年来，不断开放的生育政策和逐年降低的生育率值得我们对这些新的动向和趋势进行关注。

根据题中所给要求以及我们浏览的资料与文献，我们将问题分为以下几个部分：

第一部分：搜索并整理相关数据资料，结合今年第七次人口普查的数据，考虑人口变化的新趋势，对多种影响人口增长和总量计算因素进行综合考虑，建立相应的人口增长模型。

第二部分：依据建立的模型对我国中长期的人口变化做出预测，我们假设此处的中长期人口变化是指 2021 年至 2100 年的人口总量、年龄结构以及性别比例的变化。

第三部分：根据第二部分的结果，对不同参数情况下的预测进行比较和分析，得出我们关于政府部门在以后的规划和发展中的建议。

二、 问题分析

结合题目背景，需要考虑到中国实际情况和人口增长的特点出发，首先需要查阅资料和整理数据，根据人口变化特点建立人口平衡方程，再分别研究生育率、死亡率、预期寿命等预测模型。利用人口演变特点进行序列外推，从而计算得到人口总数的预测值，同时得到年龄结构、性别比例等预测结果。另外，考虑到中国于 2016 年实施推广二胎生育政策，2021 年鼓励三胎生育，政策对人口增长也具有不可忽略的影响，于是还需要建立基于孩次递增模型研究政策对中国人口总数的影响进行预测估计。

三、 模型假设与符号说明

1、 假设不考虑人口的迁入和迁出情况。

2、 假设在长期内中国社会环境相对稳定，没有出现大型的战乱或大型灾难。

3、 假设数据来源准确可信度高，能够真实反映整体的实际情况。

4、 假设生育率与预期寿命相互独立。

四、 符号说明

表 1 符号说明

符号	含义
P	总人口
B	出生人口
E	迁出人口
D	死亡人口
I	迁入人口
c	某指定国家
t	某t时刻
$\ell_{c,t+1}$	预期寿命增加量
$F_{c,x,t}$	c国r年龄组在t时刻的生育率

五、 模型建立

(一) 基本定义

人口预测是根据一个国家或一个地区的人口现状，还有以往人口发展的规律性，对影响人口发展的各种因素的假设等因素，对未来人口发展状况进行的推算。人口预测是通过直接影响人口变化的出生、死亡、迁移三大要素来进行的。出生、死亡、迁移三要素的变化，是社会经济等因素影响作用的结果。因此，进行人口预测必须把定量分析与定性分析紧密结合起来。人口预测实质上是以现有社会经济发展对人口发展的影响分析为依据，结合其人口的现状及其人口现象间的稳定比例关系，按照未来社会经济的前景，及其对未来人口发展过程的影响，对未来人口的发展趋势及构成变动做出恰当的估计，它体现了在一定条件下，以数量的形式反映人们对未来人口发展及其状况的认识水平。

1. 出生率(Birth Rate)

出生率是指一定时期内（通常为一年）出生活婴数与同期平均育龄妇女人数之比，通常用千分数表示。是总出生数与相应人口中育龄妇女人数之间的比例，亦称育龄妇女生育率。

2. 死亡率(Death Rate)

死亡率是指用来衡量一部分种群中，一定规模的种群大小、每单位时间的死亡数目（整体或归因于指定因素），是在种群层面上研究的问题。人类死亡率通常以每年每一千人为单位来表示；例如：在死亡率为 9.5‰的 10 万人口中，表示这一人口中每年死去 950 人。

3. 总生育率(Total Fertility Rate)

总生育率是指该国家或地区的妇女在育龄期间，每个妇女平均的生育子女数。

4. 预期寿命(Life Expectancy)

预期寿命是指假若当前的分年龄死亡率保持不变，同一时期出生的人预期能继续生存的平均年数。它以当前分年龄死亡率为基础计算，但实际上，死亡率是不断变化的，因此，平均预期寿命是一个假定的指标。

5. 男女比例(Male/Female Ratio)

男女比例，即性别比，是在种群层面上研究的问题，是指族群中男性对女性的比率。男女比例主要受男女出生比和男女死亡率的影响，男女出生比正常范围在 103- 107，也就是说出生 100 个女孩的相应的会有 103 到 107 个男孩出生，但是在现实社会中，女性死亡率低于男性，所以男性与女性人数大致相等，社会维持在一个稳定状态。但目前我国男女出生比超过 110，这不仅将导致男女比例失调，还会对人口的预测产生影响，所以在人口预测时必须将男女比例问题考虑进去。

(二) 人口平衡方程

引起人口总量和结构变化的因素包括生育、死亡和迁移,即人口平衡方程：

$$P_{(t+1)}=P_{(t)}+B_{(t,t+1)}-D_{(t,t+1)}+I_{(t,t+1)}-E_{(t,t+1)} \tag{1}$$

$P_{(t+1)}$为时刻$t+1$的总人口，P_t为时刻t总人口，$B_{(t,t+1)}$为时刻t至$t+1$的出生人口，$D_{(t,t+1)}$为时刻t至$t+1$的死亡人口，$I_{(t,t+1)}$为时刻t至$t+1$迁入人口，$E_{(t,t+1)}$为时刻t至$t+1$迁出人口。$B_{(t,t+1)}$和$D_{(t,t+1)}$是人口自身变化引起的，称为人口的自然变动。$I_{(t,t+1)}$和$E_{(t,t+1)}$是人口外部变动引起的，也称为人口的机械变动。研究序列$P_{(t+1)}$，就需要对$B_{(t,t+1)}$,$D_{(t,t+1)}$,$I_{(t,t+1)}$和$E_{(t,t+1)}$进行深入的研究。由于中国人口净流出/流入值相对于总人口的比重十分小，而疫情当下的移民和出国人数骤减，所以我们预计在短中期之内，人口的流动都不会成为影响人口总量的重要因素。

(三) 队列要素人口预测模型

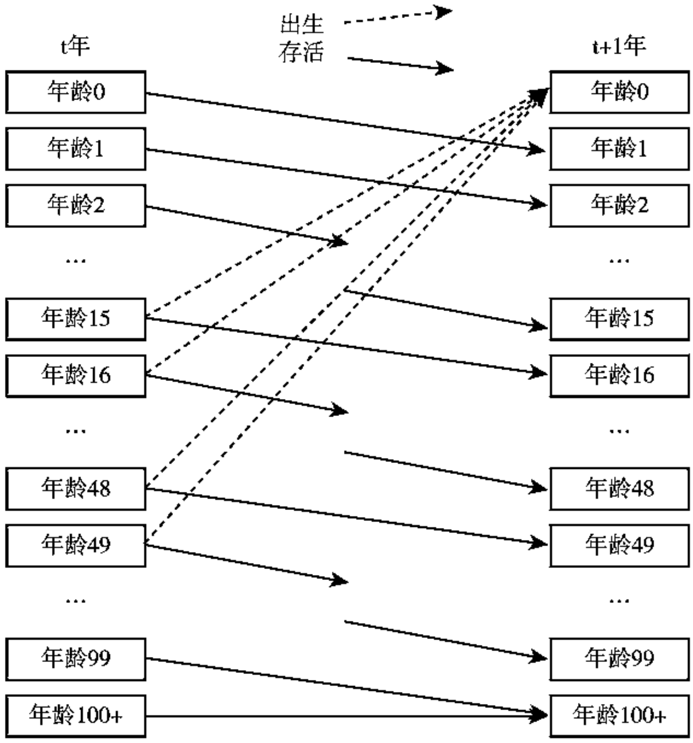

图 1 人口年龄结构推算示意图

目前使用较为广泛，方法比较先进准确的人口预测方法的核心是队列要素预测或 Leslie 矩阵法。我们考虑一个性别(例如女性)，并将人口分为N个k岁年龄组;第x个年龄组的年龄从$k(x-1)$岁到$(kx-1)$岁。对人口的预测通常使用k年作为一个周期（本题中我们选择的$k=5$）。第t个时间段的开始时刻为t。

将$n_{x,t}$定义为时刻t时在第x个年龄组中女性的个数，$S_{x,t}$定义为第t个时期内第x个年龄组存活率，$B_{x,t}$定义为在t时刻的第x个年龄组中女性生出女性后代的数量，再除以$n_{x,t}$，$m_{x,t}$定义为处于第x个年龄组人口在第t时刻净迁移人口。对于最高年龄组N，我们假设存活的人在下一个时间点仍在同一年龄组。

故人口平衡方程可以转换为以下恒等方程组：

$$n_{1,t+1}=\sum_{x=1}^{N}B_{x,t}n_{x,t}+m_{1,t},n_{x+1,t+1}=S_{x,t}n_{x,t}+m_{x+1,t}$$

其中$x=1,...,N-2$且$n_{N,t+1}=S_{N-1,t}n_{N-1,t}+S_{N,t}n_{N,t}+m_{N,t}$.

我们定义周期为$t,\boldsymbol{P}_t$的预测矩阵如下：

$$
\boldsymbol{P}_t = \begin{bmatrix} B_{1,t} & B_{2,t} & \cdots & B_{N-1,t} & B_{N,t} \\ S_{1,t} & 0 & \cdots & 0 & 0 \\ 0 & S_{2,t} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & S_{N-1,t} & S_{N,t} \end{bmatrix}
$$

然后将模型改写为矩阵形式为

$$
\boldsymbol{n}_{t+1} = \boldsymbol{P}_t \boldsymbol{n}_t + \boldsymbol{m}_t
$$

其中$\boldsymbol{n}_t = (n_{1,t}, \ldots, n_{N,t})^T$，$\boldsymbol{m}_t = (m_{1,t}, \ldots, m_{N,t})^T$。这个模型又被称为Leslie 模型。根据这个矩阵，按周期递归计算可以得到总人口的预测。将其中的人口数据用男/女性人口数进行替换，可以直接用于预测男/女性。

(四)　人口预测模型

根据文献可知各个国家的生育率具有一定共性，生育率表现得显著变化具有三个阶段，包括生育率下降开始阶段，下降最大幅度阶段和缓慢下降阶段， 在许多国家都观察到这种生育率下降速度增加到减少的现象，并根据目前的联合国人口研究结果，将其作为总生育率的函数，用来对预测生育率五年下降速度进行建模。

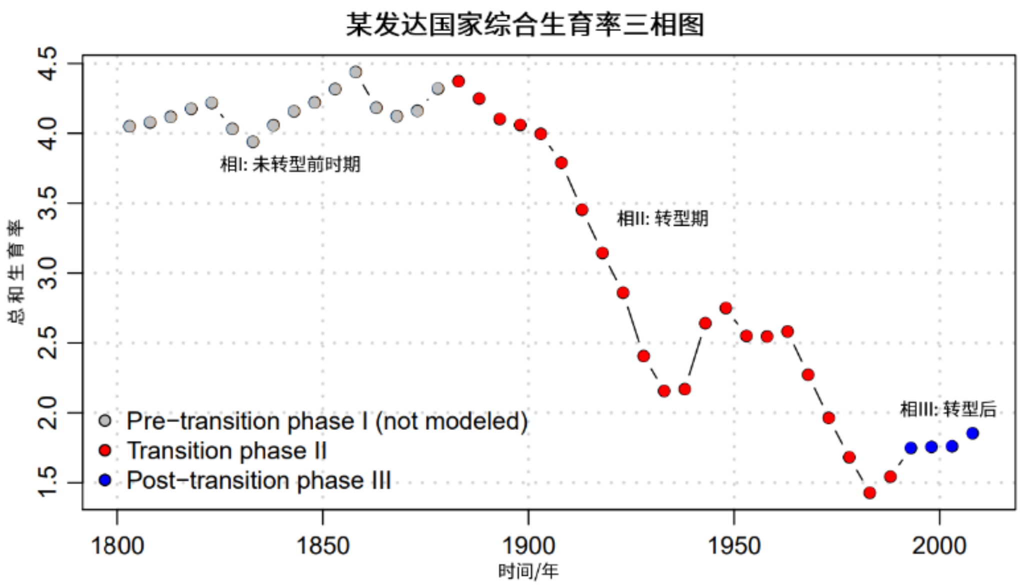

图 2 生育率三相图

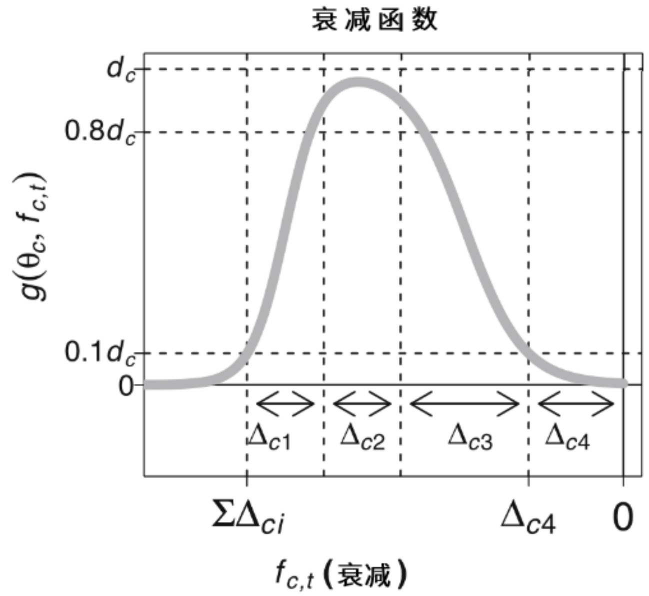

图 3 衰减函数

为了模拟和预测生育率的变化，将生育率下降分解为系统下降部分和随机扰动项，定义一个衰减函数模拟生育率下降得趋势。参考联合国预测预期寿命的双重 logistic 法，

$$g(\theta_c, f_{c,t}) = \frac{-d_c}{1+\exp\left(-\frac{2\ln(9)}{\Delta_{c1}}\left(f_{c,t}-\sum_i \Delta_{ci}+0.5\Delta_{c1}\right)\right)} + \frac{d_c}{1+\exp\left(-\frac{2\ln(9)}{\Delta_{c3}}\left(f_{c,t}-\Delta_{c4}-0.5\Delta_{c3}\right)\right)}$$

参数含义也可由上图知，其中参数d_c表示为国家 c 最有可能在五年内的衰减量，实际上的衰减比d_c略小，衰减速度取决于生育下降速度$\Delta_{c1},\Delta_{c2},\Delta_{c3},\Delta_{c4}$。

同理，假设国家c在时期t的五年内预期寿命的增加量为$\ell_{c,t+1}$，则可以通过双重 logistic 函数进行计算：

$$\ell_{c,t+1} = \ell_{c,t} + g(\ell_{c,t} \mid \boldsymbol{\theta}^c)$$

其中 5 年内的预期寿命增长量$\ell_{c,t+1}$为：

$$g(\ell_{c,t} \mid \boldsymbol{\theta}^c) = \frac{k^c}{1+\exp\left(-\frac{A_1}{\Delta_2^c}\left(\ell_{ct}-\Delta_1^c-A_2\Delta_2^c\right)\right)} + \frac{z^c-k^c}{1+\exp\left(-\frac{A_1}{\Delta_4^c}\left(\ell_{ct}-\sum_{i=1}^{3}\Delta_i^c-A_2\Delta_4^c\right)\right)}$$

在上式中，$\theta^c=(\Delta_1^c,\Delta_2^c,\Delta_3^c,\Delta_4^c,k^c,z^c)$关于国家c 的 logistic 函数的 6 个参数，

A1和A2为常数。

不同国家的参数也具有差异，根据联合国人口预测研究中给出的五种预定模式选择适合中国的参数进行计算。

衡量总体生育率水平最常用的方法是c国在t时刻的总生育率(TFR)，定义为$f_{c,t}=k\sum_{x=1}^{N-1}F_{c,x,t}$，其中$F_{c,x,t}$是$c$国r年龄组在$t$时刻的生育率。TFR是在$t$时刻位于某年龄组的一名妇女能够生育的孩子的平均数量。

(五) 贝叶斯人口预测模型

联合国目前的预测方法有些主观，因为计算过程中的双重logistic参数是由分析师从少量预先确定的可能性中选择的，而不是从数据中估计的。它也有一定的局限性，不能涵盖全有可能性。

为了解决这些问题，使用一种贝叶斯概率预测方法。这包括建立贝叶斯层次模型来预测总生育率（TFR）和预期寿命，再将每一个模型进行后验预测分布从而产生大量可能的未来轨迹，然后对数据进行队列要素预测，以提供所需的后向预测分布。

1. 贝叶斯生育率预测模型

如上文所述，将生育率变化为三个阶段，结合目前的中国国情考虑，中国正在处于生育率由高到低的过渡中期，所以本文主要围绕第二阶段和第三阶段进行建模。在预测生育率的第二阶段中，对双重logistic函数进行修正，添加异方差误差项；其次，考虑到单个国家的数据点较少，于是模型中加入世界各国的数据得到更多的数据点。每个国家的预期生育率下降量取决于过去数据以及选取特定部分国家的历史数据经验，从而得到预测结果更加稳定的层次模型。

带漂移的随机游动的TFR模型如下：

$$f_{c,t+1}=f_{c,t}-r(f_{c,t}\mid\boldsymbol{\delta}^c)+a_{c,t},$$

其中五年生育率的减少量为：

$$r(f_{c,t}\mid\delta^c)=\frac{-d^c}{1+\exp\left(-\frac{2\ln(9)}{\nabla_1^c}\left(f_{c,t}-\sum_{i=2}^{4}\nabla_i^c+0.5\nabla_1^c\right)\right)}+\frac{d^c}{1+\exp\left(-\frac{2\ln(9)}{\nabla_3^c}(f_{c,t}-\nabla_4^c-0.5\nabla_3^c)\right)}$$

其中

$$\boldsymbol{\delta}^c=(\nabla_1^c,\nabla_2^c,\Delta_3^c,\nabla_4^c,d^c)$$

代表各个国家的向量参数，另外$a_{c,t}$随机扰动项的方差与时间和生育水

平相关：

$$a_{c,t} \rightarrow N\left(0, \sigma(t, f_{c,t})^2\right)$$

利用马尔可夫链蒙特卡罗估计得到贝叶斯分层模型。

根据中国历年人口统计数据可知，中国已连续两年 TFR 低于 2，可认为中国进入了生育率变化的第三阶段。根据一阶自回归模型性质可知，模型的长期平均值等于第三阶段所有国家的近似生育水平。

$$f_{c,t+1} - \mu = \rho(f_{c,t} - \mu) + b_{c,t}$$

其中$b_{c,t} \xrightarrow{\text{iid}} N(0, \sigma_b^2)$，参数ρ和$\sigma_b$可以根据极大似然估计得到，结果为$\hat{\rho} = 0.89$，$\hat{\sigma}_b = 0.10$。

2. 贝叶斯预期寿命率预测模型

我们对女性的预期寿命与第二阶段的总生育率类似，使用双重 Logisitic 函数预测寿命预期的增加值，同时增加异方差误差项，并且参数是连续变化的，而不是局限于联合国设定的五个预测模型。

$$\ell_{c,t+1} = \ell_{c,t} + g(\ell_{c,t} \mid \boldsymbol{\theta}^c) + e_{c,t}$$

$$g(\ell_{c,t} \mid \boldsymbol{\theta}^c) = \frac{k^c}{1 + \exp\left(-\frac{A_1}{\Delta_2^c}(\ell_{ct} - \Delta_1^c - A_2\Delta_2^c)\right)} + \frac{z^c - k^c}{1 + \exp\left(-\frac{A_1}{\Delta_4^c}\left(\ell_{ct} - \sum_{i=1}^{3}\Delta_i^c - A_2\Delta_4^c\right)\right)}$$

其中$e_{c,t} \xrightarrow{\text{ind}} N\left(0, \omega(\ell_{c,t})^2\right)$，其中$\omega(\ell_{c,t})$是一个代表标准差的平滑函数，取决于预期寿命。

$$\begin{aligned} &\Delta_i^c \xrightarrow{\text{iid}} \mathrm{TN}_{[0,100]}(\Delta_i, \sigma_{\Delta_i}^2), \quad i = 1, \ldots, 4 \\ &k^c \xrightarrow{\text{iid}} \mathrm{TN}_{[0,10]}(k, \sigma_k^2) \\ &z^c \xrightarrow{\text{iid}} \mathrm{TN}_{[0,1.15]}(z, \sigma_z^2) \end{aligned}$$

其中$\mathrm{TN}_{[a,b]}(\mu, \sigma^2)$表示在a到b之间的截断正态分布。

基于历史数据可知，几乎每个国家的预期寿命是渐近线性增长，男性预期寿命与女性的预期寿命是高度相关的，且总是稍微低于女性预期寿命，因此，可以共同预测女性与男性预期寿命，首先先预测女性的预期寿命，然后再预测两者之间的差距。可以利用误差项服从 t 分布的回归模型预测男女预期寿命的差距：

$$G_{c,t+1} = \min\left\{(G_{c,t+1}^*)_+, 18\right\}$$

其中

$$G^*_{c,t+1}=\begin{cases}\beta_0+\beta_1\ell_{c,1950-1955}+\beta_2G_{c,t}+\beta_3\ell_{c,t}+\beta_4(\ell_{c,t}-75)_++\epsilon_{c,t}, & \text{if } \ell_{c,t}\le M\\ \gamma_1G_{c,t}+\epsilon_{c,t}, & \text{if } \ell_{c,t}>M\end{cases}$$

$$\epsilon_{c,t}\xrightarrow{\text{iid}}t(\mu=0,\sigma^2=0.0665,\nu=2),$$

其中，$M=86.2$，根据已知各国差距的最大值，设定男女预期寿命差距的最大值限制在 18 年。

3. 贝叶斯总人口预测模型

根据后向预测分布模型了从 2020 年到 2100 每五年期总生育率的多条轨迹，以及男女预期寿命的联合轨迹，然后再根据年龄别进行分组，根据队列要素预测模型对未来人口数量进行预测。

(六) 孩次递进模型

所以根据上述假设，我们改进了我们的模型，使其更加符合我国的政策与实际情况。

孩次递进生育模型与队列要素预测模型在生命表的建立，平均寿命预测邓方面有很多相似之处，可以沿用前个模型中的数据进行推演，但两者最大的不同是在于孩次递进模型对育龄妇女和孩次结构的记录和推演，而在迭代过程中，最重要的部分也是育龄妇女年龄层次的构成以及孩次递进比的计算。

孩次递进模型可以分为以下几个部分：

第一部分：育龄妇女分年龄、分孩次递进；

第二部分：未递进妇女存活人数和新增初育 0 孩妇女计算；

第三部分：育龄妇女孩次结构更新和迭代。

其中定义如下表所示：

表 2 孩次递进模型符号说明

符号	含义
i	年龄
j	孩次
$women[i,j]$	i 岁 j 孩育龄妇女
$newBornWowen[i+1,j+1]$	i 岁妇女生育 j 孩子女
$wLx[i]$	i 岁至 $i+1$ 岁妇女存活年数
$proht[i,j+1]$	i 岁 $j+1$ 孩育龄妇女递进比

$NonBornWomen[i+1,j]$	为i岁且未发生递进生育的j孩妇女
$female[i]$	i岁妇女人数

各部分的模型表示如下：

育龄妇女年龄别、孩次别递进生育子女数：

$$newBornWomen[i+1,j]=$$

$$women[i,j]*(wLx[i+1]/wLx[i])*(1-proht[i,j+1])$$

育龄妇女年龄别、孩次别没有递进生育：

$$NonBornWomen[i+1,j]=women[i,j]\times(wLx[i+1]/wLx[i])\times(1-proht[i,j+1])$$

$$NonBornWomen[15,0]=female[14]\times(wLx[i+1]/wLx[i]);$$

这里我们将国际标准 15 岁作为育龄妇女的最低年龄。

年龄别、孩次别妇女构成生育更新：

$$women[i,j]=NonBornWomen[i,j]+newBornWomen[i,j]$$

经过一个年度的递进生育后，当前妇女的分年龄孩次结构可以分为两种情况。第一种情况，发生递进生育的育龄妇女的年龄和孩次都发生变化；第二种情况，未发生递进生育的妇女只是年龄增长，孩次没有发生变化。记录育龄妇女的年龄、孩次构成，目的是进行下一次递进生育和状态更新，这样既有低龄无孩妇女源源不断地进入递进生育，也有超出育龄期妇女不断退出。

1. 递进生育率与递进生育模式

通过基础数据计算各孩次的年龄别递进生育率和递进生育模式，目的是设定分孩次总和递进生育率参数时，根据新的参数计算分年龄、分孩次递进比。 分孩次递进生育率与递进生育模式计算方法简单总结如下。

(1) 对于 $0\to1$ 孩年龄别递进生育率：

$$aspfr(i,1)=h(i,1)\times\prod_{k=15}^{i-1}[1-h(k-1,1)]$$

其中，i、k为年龄，$aspfr(i,1)$中为 $0\to1$ 孩 i岁 1 孩递进生育率，1 代表 0→1 孩递进；h为i岁 1 孩递进比，规定$h(14,1)=0$，那么 $1-h(14,1)=1$，也就是说，认为 14 岁没有初育行为；$\prod_{k=15}^{i-1}[1-h(k-1,1)]$ 为 14 岁开始到 $i-1$ 岁连续未递进比例的乘积。

$0\to1$ 孩总和递进生育率为：

$$TPFR_1=\sum_{15}^{49}aspfr(i,1)$$

(2) $1\rightarrow 2$ 孩年龄别递进生育率：

对于不同生育间隔的 $1\rightarrow 2$ 孩年龄别递进生育率：

$$subaspfr(i,p,2)=aspfr(i-p,1)\times h(i,2)\times \prod_{k=i-p+1}^{i-1}[1-h(k,2)]$$

其中，i,k 为年龄且$i,k\geqslant 16$，p 为 1→2 孩递进的时间间隔，取值$1\sim 34$；

1→2 孩年龄别递进生育率为该年龄 1→2 孩各个生育间隔递进生育率的合计：

$$aspfr(i,2)=\sum_{p=1}^{i-15}subaspfr(i,p,2)$$

孩总和递进生育率：

$$TPFR_2=\sum_{15}^{49}aspfr(i,2)$$

总和递进生育率：

$$TPFR=TPFR_1+TPFR_2$$

递进生育模式：

$$PG1(x)=ASPFR1(x)/TPFR_1$$
$$PG2(x)=ASPFR2(x)/TPFR_2$$

经过上述算法得到孩次别递进生育模式为孩次递进预测参数奠定基础。

2. 分孩次年龄别递进比参数

根据给定的分孩次递进生育模式和分孩次总和递进生育率得到年龄别孩 次递进生育率。然后，根据年龄别孩次递进生育率推算分孩次年龄别递进比参数，运算的过程是第二部分的逆运算，具体算法如下。

(1) 年龄别孩次递进生育率

根据现有育龄妇女孩次结构和递进生育数据得到：

$$PG1(x)=ASPFR1(x)/TPFR_1$$
$$PG2(x)=ASPFR2(x)/TPFR_2$$

假定 $TPFR`[j]$为已知或设定参数 $TPFR`[j]$，可以推算递进生育模式 不变且假定参数条件下，年龄别各孩次的递进生育率：

$$aspfr`[i,j]=pg[i,j]\times TPFR`[j]$$

当然，也可以推算生育模式转变条件下的年龄别孩次递进生育率，但需要注意的前提条件是递进生育模式的转变要与递进生育模式转变的实际规律相符，这样推算出来的参数才有意义，但有些人为设定的参数有可能是不符合实际规律的，即有可能与实际规律相左，计算的结果可能会出现负值的不符合实际人口含义的结果，这里是要有人口学常识和知识，并需要特别注意的。

3. 从年龄别孩次递进生育率推算年龄别孩次递进比

从年龄别孩次递进生育率推算年龄别孩次递进比是一个从孩次递进比计 算孩次递进生育率的逆运算。由于是一个递推的逆运算，之 所以可以连续递推，原因是假定已知分孩次年龄别递进率且起始年龄孩次递进 生育率与起始年龄孩次递进比相等。 一孩递进的情况比较简单，$0 \to 1$ 年龄别孩次递进比的推算方法如下。由于

$$aspfr`(i,1) = h(i,1) \times \prod_{k=15}^{i-1} (1 - h(k-1,1))$$

其中，$aspfr`[i,1]$ 已知，而且 $h[15,1] = aspfr`[15,1]$，当 i>15 时， 递推各年龄 1 孩年龄别递进比的方法如下。

$$h[i,1] = \frac{aspfr`[i,1]}{\prod_{k=16}^{i-1}(1 - h[k-1,1])}$$

$1 \to 2$ 年龄别孩次递进比的推算方法如下：

已知 $aspfr`[i,1]$ 和 $aspfr`[i,2]$，根据

$$subaspfr(i,p,2) = aspfr(i-p,1) \times h(i,2) \times \prod_{k=i-p+1}^{i-1}[1 - h(k,2)];$$

$$aspfr(i,2) = \sum_{p=1}^{i-15} subaspfr(i,p,2);$$

即：

$$aspfr(i,2) = \sum_{p=1}^{i-15}\left\{aspfr(i-p,1) \times h(i,2) \times \prod_{k=i-p+1}^{i-1}[1 - h(k,2)]\right\};$$

进而推算：

$$proht[16,2] = aspfr`[16,2] / aspfr`[15,1];$$

16 岁及以上 1 孩间隔递进：

$$ht[16,2] = aspfr`[16,2] / aspfr`[15,2];$$

$$duration[i,p,2] = aspfr[i-p,1] \times \prod_{k=i-p+1}^{i-1}(1 - ht[k,2]) \qquad p = [1,34], i = [16,49];$$

$$h(i,2) = \frac{aspfr(i,2)}{\sum_{p=1}^{i-15} duration[i,p,2]}$$

同理 2→3 孩的递进生育率为：

$$h(i,3) = \frac{aspfr(i,3)}{\sum_{p=1}^{i-15} duration[i,p,3]}$$

根据不同假定的孩次总和递进生育率和年龄别孩次递进模式推算新的年龄别孩次递进比参数，根据 2010 年全国人口普查育龄妇女的曾生孩次结构和

2020 年度分年龄分孩次生育数据计算 0→1 孩、1→2 孩和 2→3 孩年龄别孩次递进比和总和递进生育率为基础。2010 年 0→1 孩、1→2 孩和 2→3 孩总和递进生育率分别为 0.9733, 0.3582, 0.0427 设定新的 0→1 孩和 1→2 孩总和递进生育率分别为 0.95,0.4 和 0.0427，如果递进生育模式不变，那么，可以推算出与此孩次递进生育率相对应的年龄别孩次递进比并用于未来的人口预测。

六、 模型求解与结果分析

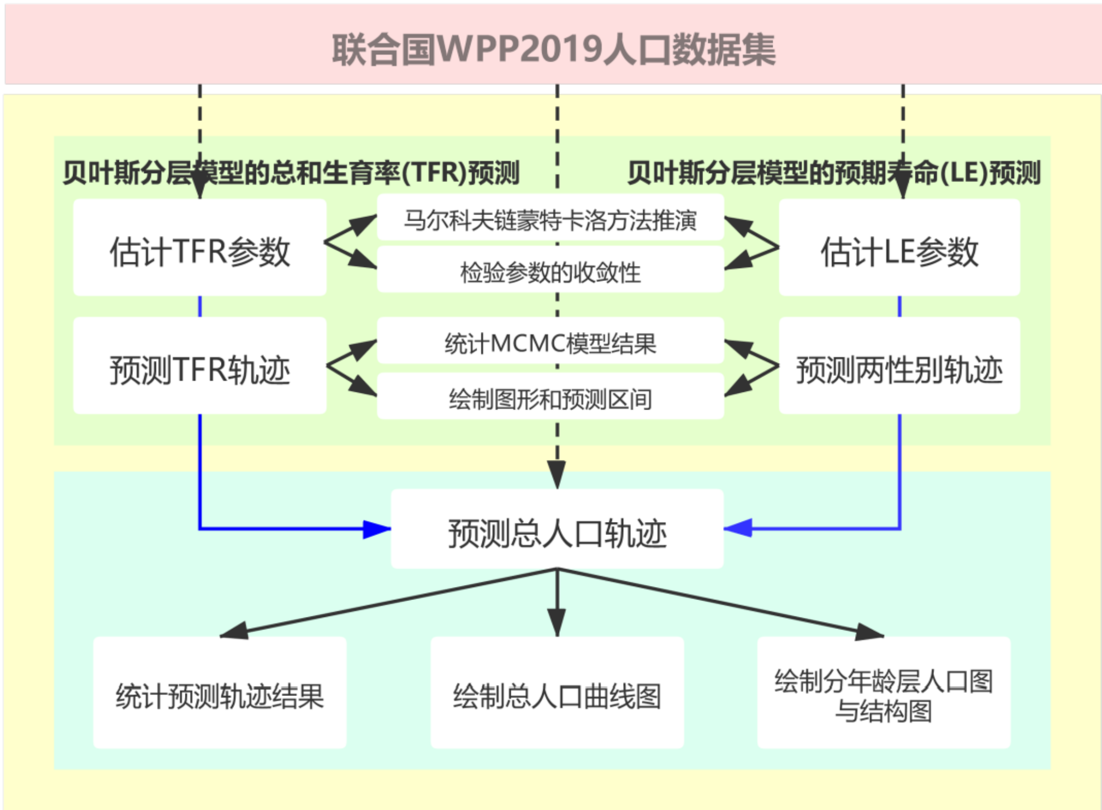

图 4 建模过程流程图

（一） TFR 预测

训练模型给定参数为 24 次轨迹预测以及每次的 1000 次迭代。得到的结果如下图所示，图中灰色线为每次轨迹预测所得到的预测值，而红色曲线代表轨迹预测的中位数，两种不同点数的虚线分别代表了 80%和 95%预测区间：

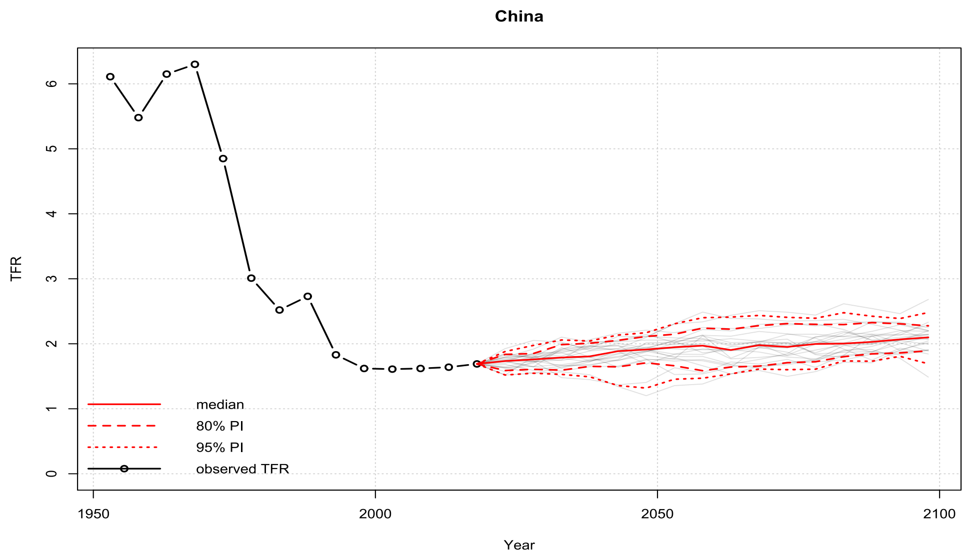

图 5 中国总生育率预测结果

由图中的总和生育率预测可以看出，从 2020 至 2100 年的总和生育率基本保持平稳，有微弱的上涨趋势，在 2 左右波动。其 AR(1)模型的预测参数为:

表 3 参数结果

μ	ϕ	σ
1.9	0.886	0.102

该 AR(1)模型满足弱平稳性，参数收敛，认为其性质较好。可以作为总人口数预计的准备数据。

（二） 预期寿命预测

与 TFR 类似，我们训练模型给定参数为 24 次轨迹预测以及每次的 1000 次迭代，并区分男性和女性的预期寿命进行迭代。带到的结果如下图所示，其中灰色线为每次预测所得到的预测值，粉红实线代表女性预期寿命，绿色实线代表男性预期寿命：

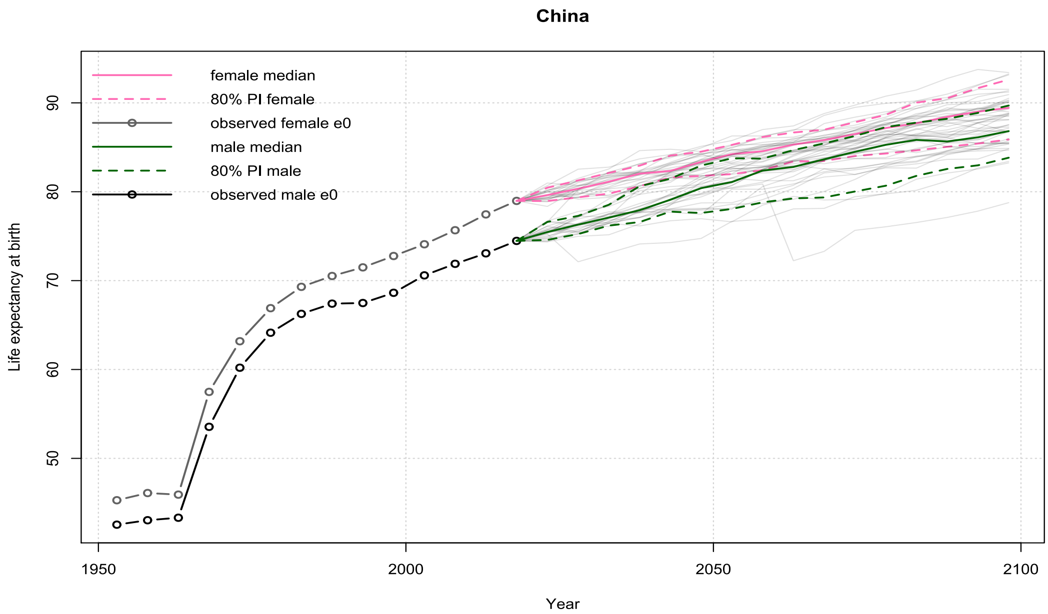

图 6 预计寿命预测结果

从图中可以看出两大趋势，第一点为中国人均寿命在不断地稳步上升，2100 年可以达到 87 岁左右，第二点为女性预期寿命要比男性高 5 岁左右，但差距在不断缩小。这对我们接下来进行的人口总量迭代计算提供了预期寿命的数据。

（三） 人口总量预测

在人口总量预计中，我们分别对男性和女性进行了代际迭代，首先得到了人口总量的预测值，其中灰色线为每次迭代地预测值，红色实线为预计的中位数，而虚线分别代表了 80%和 95%预测区间：

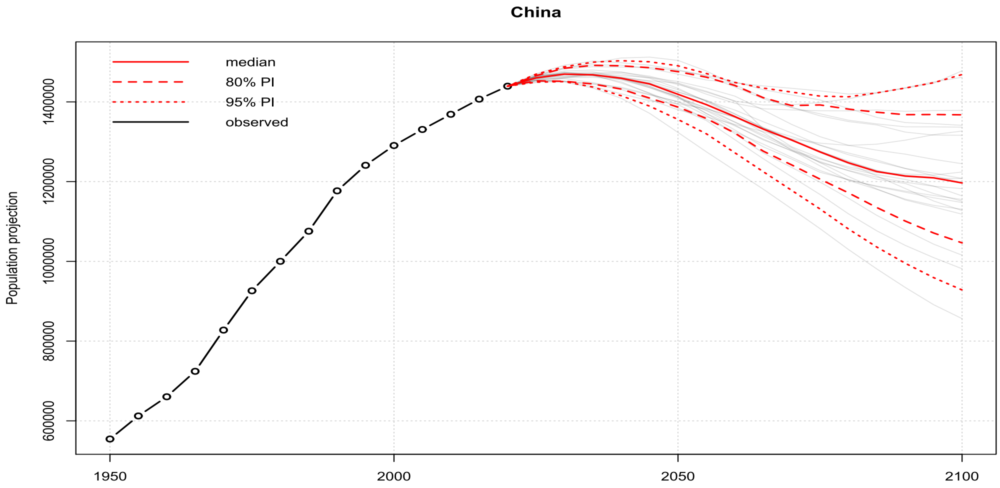

图 7 人口总量预测结果

从图中我们可以明显地看出，在 2025 年左右，中国人口达到峰值，约为 14.7 亿人，而之后的几十年里，人口总数则不断下降，直到 2100 达到低于 12 亿人的水平。

而针对年龄层的分布，我们对 Leslie 矩阵进行迭代后得到的结果如下图所示，我们选取了 9 个代表性的年龄层对不同年龄阶段的人口数目推演进行分析，红色实线为预测结果：

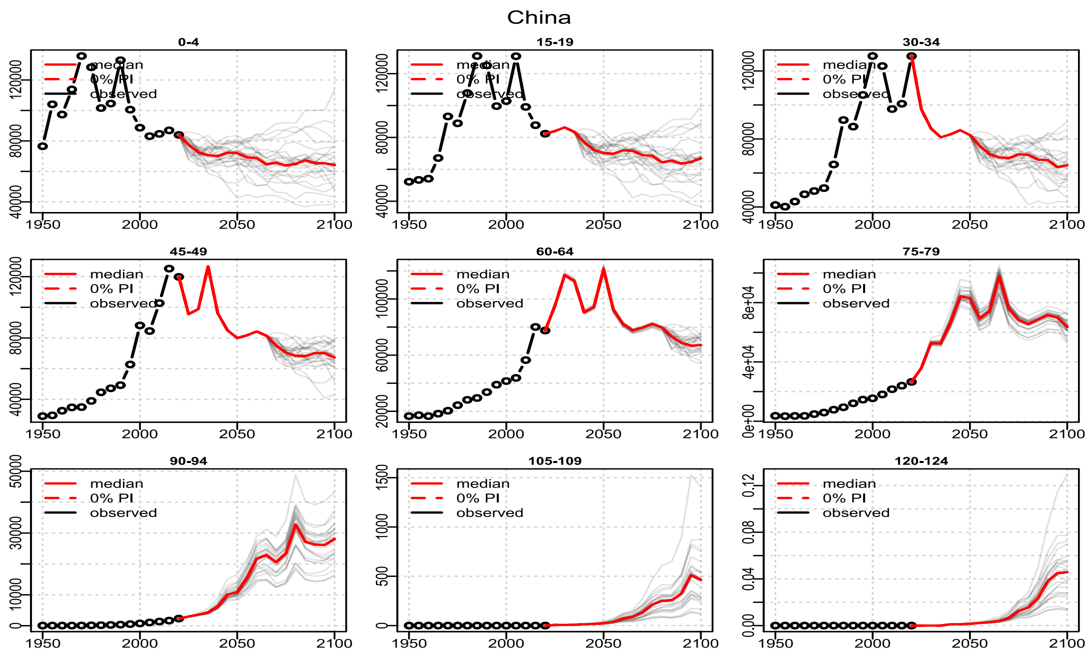

图 8 年龄层人口数目预测结果

从图中我们可以看出，0-20 岁的幼儿和青少年的人口在接下来的 80 年间将不断减少，而 30-64 岁的劳动力在经过当前青年衰老后的增长后逐年波动下降，而高于 65 岁的老年人口则在快速上升后进入平台期。这一现象代表了中国可能将进入严重老龄化社会和少子化社会的状态，老年人口的比例超过了总人口的 30%。通过下面的年龄结构金字塔图可能更好的看出年龄层之间在 80 年内

的变迁，四张图中的年份分别为 2020，2050，2080 和 2100 年：

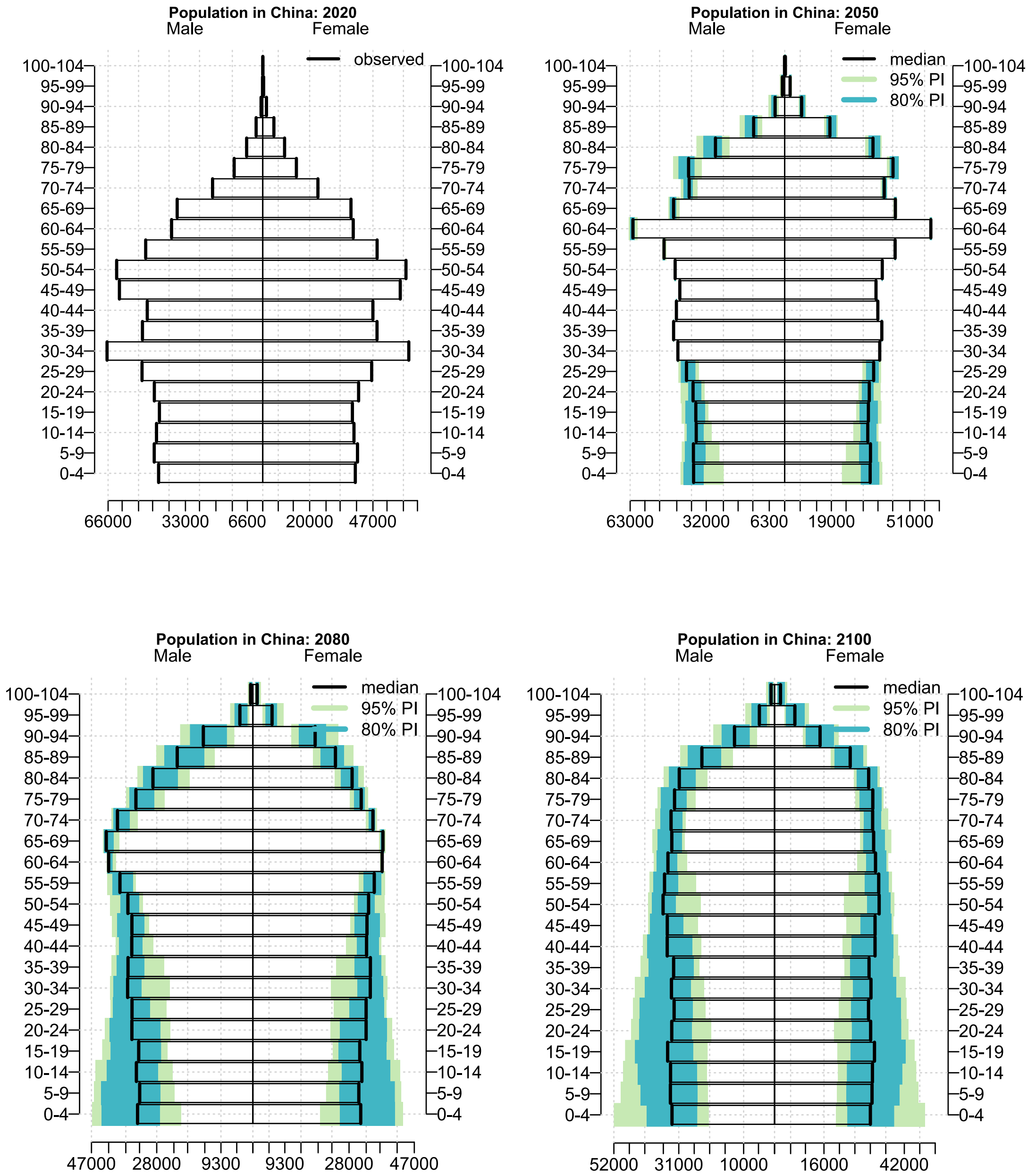

图 9 2020，2050，2080 和 2100 年年龄结构预测结果

从这四张图可以看出，未来的人口趋势将倾向于各年龄层人口的均匀分布，劳动力人口比例持续下降。

（四） 孩次递进模型

考虑国家出台的生育政策对人口长期趋势的影响，我们对多个情境下的孩次模型进行了迭代计算。为简化情境条件，更凸显不同情境下生育水平对人口

的影响，平均预期寿命、出生性别比等参数都设定为相同的区间，对于 1 孩以及 3 孩以上同样采用相同参数，只对 2 孩生育水平设定 3 种情境，分别为：

二孩生育率保持 2010 年水平：35.82%的一孩妇女会选择生二孩；

二孩生育率保持在 0.6；

二孩生育率保持在 0.8；

完整的情景预测参数如下表所示：

表 4 情景预测参数表

<table>
<tr><th colspan="3">指标</th><th>2010 年</th><th>2020 年</th><th>2050 年</th><th>2100 年</th></tr>
<tr><td rowspan="6">递进生育率</td><td colspan="2">0-1</td><td>0.9733</td><td>0.9733</td><td>0.9733</td><td>0.9733</td></tr>
<tr><td rowspan="3">1-2</td><td>情境一</td><td>0.3582</td><td>0.3582</td><td>0.3582</td><td>0.3582</td></tr>
<tr><td>情境二</td><td>0.6</td><td>0.6</td><td>0.6</td><td>0.6</td></tr>
<tr><td>情境三</td><td>0.8</td><td>0.8</td><td>0.8</td><td>0.8</td></tr>
<tr><td colspan="2">2-3</td><td>0.0427</td><td>0.0427</td><td>0.0427</td><td>0.0427</td></tr>
<tr><td colspan="2">3-4+</td><td>0.0058</td><td>0.0058</td><td>0.0058</td><td>0.0058</td></tr>
<tr><td rowspan="2">平均预期寿命</td><td rowspan="2">值</td><td>男</td><td>75.77</td><td>76.67</td><td>79.79</td><td>85</td></tr>
<tr><td>女</td><td>80.60</td><td>81.37</td><td>84.23</td><td>89</td></tr>
<tr><td>出生性别比</td><td colspan="2">均值</td><td>117.96</td><td>107</td><td>107</td><td>107</td></tr>
</table>

得到的结果为：

表 5 预测结果

<table>
<tr><th>指标</th><th>二孩递进生育率</th><th>2010 年</th><th>2020 年</th><th>2050 年</th><th>2100 年</th></tr>
<tr><td rowspan="3">总人口</td><td>0.3582</td><td>13.33</td><td>13.71</td><td>12.2</td><td>6.57</td></tr>
<tr><td>0.6</td><td>13.33</td><td>13.90</td><td>13.05</td><td>8.73</td></tr>
<tr><td>0.8</td><td>13.33</td><td>14.07</td><td>13.86</td><td>11.02</td></tr>
</table>

从该表中我们可以发现，在其他参数不变的情况下，如果只改变二孩递进生育率，在生育率低迷时中国人口的长期趋势将会呈现大幅下跌甚至人口减半的趋势，而在大力推广二孩政策的情境下，人口在 2020 年代达到峰值后，将会逐渐下降值 11 亿左右。值得注意的是，如果二孩递进率过低，社会老年化的问题会持续加重，在二孩递进生育率为 0.35 左右时，2100 年的老年人口（65 岁及以上）将接近总人口的 40%，这对社会或个人压力会极度。在这个模型中，因为只考虑二孩递进生育率的改变而忽略了其他参数的变化，数据与第一个模型得到的结果可能出现出入，但总体的趋势均是下降的。

七、 模型的改进和推广

在运用贝叶斯分层模型对生育率、预期寿命等进行估计时，使用 Mcmc 算法耗费时间长，计算过程较为复杂，造成计算效率低，可以对模型算法进行改进优化，提高算法的效率，精简过程。

本文建立的模型没有考虑人口的迁入迁出影响，对未来预测会造成一定的误差，该模型还可以在模型的完整性上进行优化，将迁入迁出作为未来人口数

量预测的因素，则预测结果将会更加准确，使得模型具有较高的精确度和普遍适用性。

八、 政策与建议

中国人口发展的历史决定人口发展的未来。对比人口发展情景模拟结果可以看到中国人口发展的内在规律。人口政策特别是生育政策是未来中国人口发展符合人口发展战略要求，进入良性循环轨道的辅助工具。面对中国人口未来三个方面的严峻挑战，人口政策更需要与时俱进。

第一，指定福利政策鼓励生育，解决低生育率陷阱。对照中国人口长期发展目标，即总和生育率稳定在 1.8 左右的目标，目前的生育政策调整是无法实现的。同时，当前和未来促进生育率持续下降的因素不断强化，比如，育龄妇女受教育程度不断提高，城镇人口比例持续增加和子女养育成本居高不下，其结果是生育率长期走低的风险远远大于稳定回升的可能，跨越低生育率陷阱的任务非常艰巨。

第二，减缓劳动力的老化与加速建设创新型发展。人口快速老化的前提是劳动力快速老化，劳动力快速老化的结果是严重阻碍产业或经济的转型升级。众所周知，科学技术发展核心竞争是创新，经济的快速发展动力也是创新。对绝大多数历史经验丰富但知识、技能陈旧的高龄劳动力的教育、培训是一个非常艰难的系统工程，如何面对持续、快速老化的劳动年龄人口，将是未来创新经济、创新社会和创新国家的严峻挑战。

第三，提供养老福利和全面社会保障，解决结构失衡问题。人口快速老化和老年人口比例长期居高不下，老年人口总量和比例持续增加所形成的人口结构失衡问题凸显，其结果是一方面加大养老金支付系统的压力和在业人口负担，同时长寿因素也是养老系统安全运行的不确定性风险；另一方面家庭养老支持人均照料负担前所未有，如何破解人口快速老化对养老保障和社会支持系统的持续增压，也将是全社会面临的研究课题和不可回避的突出矛盾。

总之，未来中国人口政策将是稳定、协调中国人口总量和结构问题，解决近期和长期的矛盾冲突的重大战略举措之一，也就是解决所谓的时期和队列或代内和代际人口问题重要策略之一。人口问题的解决不能寄希望于通过短期举措就能够解决长期历史积累的问题。因此，不断根据人口形势变化，前瞻性政策调整和干预才是缓解主要矛盾的必要条件。

参考文献

[1] 王广州．中国人口预测方法及未来人口政策[J].财经智库,2018,3(03):112-138+144.

[2]Adrian E. Raftery et al. Bayesian Probabilistic Projections of Life Expectancy for All Countries[J]. Demography, 2013, 50(3) : 777-801

[3]Leontine Alkema et al. Probabilistic Projections of the Total Fertility Rate for All Countries[J]. Demography, 2011, 48(3) : 815-839.

[4]Adrian E. Raftery and Leontine Alkema and Patrick Gerland. Bayesian Population Projections for the United Nations[J]. Statistical Science, 2014, 29(1) : 58-68.

[5]Lamnisos D and Giannakou K and Siligari T. Demographic forecasting of population projection in Greece: A Bayesian probabilistic study[J]. European Journal of Public Health, 2019, 29(Supplement4)

[6]张丽萍,王广州.中国育龄人群二孩生育意愿与生育计划研究[J].人口与经济,2015(06):43-51.

[7]乔晓春. “单独二孩”政策下新增人口测算方法及监测系统构建[J].人口与发展,2014,20(01):2-12.

[8]Francesco C. Billari and Rebecca Graziani and Eugenio Melilli. Stochastic Population Forecasting Based on Combinations of Expert Evaluations Within the Bayesian Paradigm[J]. Demography, 2014, 51(5) : 1933-1954.

部分代码:

```
library(bayesPop)

setwd("D:/MESS/XS")

tfr.dir  <-  "D:/MESS/XS/tfr"
tfr.pred  <-  get.tfr.prediction(tfr.dir)
tfr.trajectories.plot(tfr.pred,  "China",  nr.traj=24,half.child.variant  =  F)
summary(tfr.pred)
```

```
e0.dir="D:/MESS/XS/e0"
e0.pred=get.e0.prediction(e0.dir)
e0.trajectories.plot(e0.pred,"China",nr.traj = 24,pi=80,both.sexes
 = T)
summary(e0.pred)

country="China"
pop.dir="D:/MESS/XS/pop"
pop.pred=get.pop.prediction(pop.dir)
pop.trajectories.plot(pop.pred,country,sex = "both",nr.traj=24,age=
c(1, 4, 7, 10, 13, 16, 19, 22, 25),pi=0,dev.ncol = 3)
pop.trajectories.plot(pop.pred,country,sex = "both",nr.traj = 24,
sum.over.ages = T)
pop.pyramid(pop.pred,country,year=c(2020))
pop.pyramid(pop.pred,country,year=c(2050))
pop.pyramid(pop.pred,country,year=c(2080))
pop.pyramid(pop.pred,country,year=c(2100))
summary(pop.pred)
```