离散人口模型随机预测误差的估计

·沈 建 法·

（一）引言

人口研究表明，一定区域的人口可以用偏微分方程描述。目前广泛用来预测的是其一年的离散形式：

$$\begin{cases} x_0(t+1)=[1-\mu_{00}(t)]\beta(t)\sum\limits_{i=r_1}^{r_2} k_i(t)h_i(t)x_i(t) \\ x_i(t+1)=[1-\mu_{i-1}(t)]x_{i-1}(t)+f_{i-1}(t),\ i=1,\ 2,\ \cdots,\ \omega \\ x_i(o)=x_{io},\ i=0,\ 1,\ \cdots,\ \omega \end{cases} \qquad (1)$$

这里$x_i(t)$、$k_i(t)$、$h_i(t)$、$\mu_i(t)$、$f_i(t)$分别表示预测期t年龄为i岁的人口数、女性比例、生育概率、死亡概率、人口净迁移量；$\beta(t)$为预测期t总和生育率；$\mu_{00}(t)$为预测期t的婴儿死亡率；x_{io}为基期第i岁人口；r_1、r_2分别为妇女生育年龄下限、上限；ω为人能活到的最高年龄。

人口的动态基本上决定于总和生育率$\beta(t)$，生育概率$h_i(t)$，死亡率$\mu_i(t)$及人口净迁移量$f_i(t)$。由于这些因素的随机性，将导致实际人口发生偏离预测结果的波动。本文把各年龄人口的生育和死亡看成一个随机过程，用契比雪夫不等式估计了生育频率和死亡频率的随机波动范围，由此得到了总人口数相对预测误差的估计式，并用湖州市实际人口资料进行计算。

（二）随机预测误差的估计

用模型（1）进行人口预测时使用的是根据历史人口资料估计得到的生育概率和死亡概率，但实际人口的动态是由生育频率和死亡频率所确定的。由于各种随机因素的影响，生育频率和死亡频率将分别在生育概率和死亡概率上波动，这样就会使预测结果偏离实际人口状况，导致预测误差。使用控制理论中的卡尔漫（Kalman）滤波器可以对预测结果进行校正，但仍无法彻底消除误差。如果把各年龄人口的生育和死亡看成一个随机过程，则我们就可以借助概率论的理论对预测误差作出恰当的估计，以正确评价人口预测的可靠性。

首先讨论生育频率的随机波动所引起的出生人数的随机误差。

设育龄妇女按龄人数为n_i，$n_i=k_ix_i$，$i=r_1\cdots r_2$；按龄生育概率为h_i。则i岁的某个妇女在一年当中是否生育服从0—1分布：

$$Y_{ij}=\begin{cases}0\text{未生育} & i=r_1\cdots r_2\\ 1\text{生育} & j=1,\ 2\cdots n_i\end{cases}$$

Y_{ij}表示i岁的育龄妇女中的第j个妇女在一年中生育的情况（为方便起见，这里设一个妇女一年最多只生育一个孩子）。

Y_{ij}的数学期望$EY_{ij}=h_i$，方差$DY_{ij}=h_i(1-h_i)$。

若i岁妇女实际生育频率为Q_i，则

$$Q_i=\frac{1}{n_i}\sum_{j=1}^{n_i}Y_{ij},\quad EQ_i=h_i,\quad DQ_i=\frac{1}{n_i}(1-h_i)h_i$$

据契比雪夫不等式有：

$$P\{\ |Q_i - h_i| < \varepsilon_i\ \} \geqslant 1 - \frac{h_i(1-h_i)}{n_i \varepsilon_i^2}$$

可得保证度为a（%）的生育频率的随机波动范围：

$$|\varepsilon_i| \leqslant \sqrt{\frac{h_i(1-h_i)}{n_i(1-a)}}$$

$$\text{总出生数相对误差} = \sum_{i=r_1}^{r_2} \beta n_i |\varepsilon_i| \leqslant \sum_{i=r_1}^{r_2} \beta \sqrt{\frac{k_i x_i h_i(1-h_i)}{(1-a)}} \qquad (2)$$

同理可得，死亡频率的随机波动范围：

$$|\varepsilon'_i| \leqslant \sqrt{\frac{\mu_i(1-\mu_i)}{x_i(1-a)}}$$

$$\text{总死亡数误差} \leqslant \sum_{i=0}^{\omega} \sqrt{\frac{x_i \mu_i (1-\mu_i)}{(1-a)}} \qquad (3)$$

由（2）、（3）可得总人数相对误差E的估计：

$$E \leqslant \frac{k}{\sqrt{1-a}} \cdot \frac{1}{\sqrt{M}} \qquad (4)$$

其中M为总人口，a为保证度，K取决于具体人口的生育概率h_i、死亡概率μ_i、女性比例k_i、按龄人口比例g_i（$g_i = X_i/M$）。

$$K = \sum_{i=r_1}^{r_2} \beta \sqrt{k_i g_i h_i(1-h_i)} + \sum_{i=0}^{m} \sqrt{g_i \mu_i (1-\mu_i)}$$

（4）式表明当用离散人口模型（1）进行人口预测时，总人口数的相对预测误差同总人口数的平方根成反比，误差的估计范围还同保证度有关，保证度要求越高，则相对误差越大，反之则越小。

表1是在不同保证度下，用湖州市实际人口资料计算得到的人口误差（当时湖州市总人口约为94万）。表2表示不同人口规模时总人口预测误差的变化情况，其中E为总人口为M时的相对误差。

表1　总人口相对预测误差

保证度%	99	80	75
相对误差‰	11.2	2.5	2.2

表2　预测误差与总人口的关系

总人口	相对误差
100M	0.1E
10M	0.32E
M	E
0.1M	3.16E
0.01M	10E

（三）结论

用离散人口模型（1）进行人口预测时，关于人口预测误差可以有以下二点结论：

1.总出生人数、总死亡人数的绝对预测误差可分别由（2）、（3）式估计，它们分别同总人口的平方根成正比。

2.总人口数的相对预测误差E同总人口的平方根成反比，并由下式估计（保证度为a）：

$$E \leqslant \frac{K}{\sqrt{1-a}} \cdot \frac{1}{\sqrt{M}}$$

参考文献：宋健、于景元、李广元，“人口发展过程的预测”，中国科学，1980（9），P、920—932

（作者系华东师大研究生）