

中国随机人口预测*

李 南 胡华清

【提要】 本文应用 LTC 方法首次完成了中国随机人口预测,并全面介绍了这一方法。给出了从 1990 年到 2050 年的包括人口数、老龄比和负担比变化的期望和置信区间等主要结果,显示了这种预测的多种用途,如评价确定性方案预测的可能性等,弥补了传统的“高一低”方案法所无法回答发生某个状态概率的缺点,为科学决策服务。

【作者】 李 南 西安交通大学管理学院,教授、博士生导师;胡华清 中国民航科学技术研究中心规划财务研究室,博士。

1. 引言

由 Lee, Tuljapurkar (1991) 及 Carter (Lee and Carter, 1990) 提出的随机人口预测 LTC 方法,由随机死亡与生育预测及随机人口预测两部分构成。其中,随机死亡与生育预测克服了经典随机人口模型 (Pollard, 1975) 无法处理按龄生育与死亡率的随机性时变的困难;而李南与 Tuljapurkar (1995) 通过将时间序列模型扩展为时间——区域序列模型解决了中国及许多发展中国家的死亡时间序列数据的缺乏问题,成功地进行了中国随机死亡率预测;李南与申卯兴又完成了中国随机生育率预测 (李南、申卯兴, 1996);这就完成了 LTC 方法的第一部分并提供了进行中国随机人口预测的基础。实际上,此时的随机人口预测已可通过经典仿真方法进行,而 LTC 方法在此部分的核心,在于提供了解析的近似计算方法,从而解决了经典仿真方法中结果的不可重现问题。本文使用 LTC 方法进行中国随机人口预测,一方面通过给出在未来某时刻发生某种人口状态的概率,来定量地评价未来某时刻实现作为基本国策的体现人口控制目标的可能性,为科学决策服务;另一方面,也介绍了 LTC 方法。

2. 中国随机死亡预测主要结果

记 0 时刻为 1990 年年中,中国随机死亡预测主要结果 (李南、Tuljapurkar, 1995) 为:

$$m_x = e^{(am_x + bm_x \cdot k_t)} \quad (1)$$

$$k_t = k_{t-1} - a + em_t \quad (2)$$

其中:

m_x 为 t 时刻 x 岁人口死亡率;

am_x 和 bm_x 为模型值,具体值见表 1; $k_0 = -3.0388$, $a = 0.371$

$$em_t \sim N(0, \sigma_m^2); E(em_t, em_s) = 0, t \neq s; \sigma_m = 0.476 \quad (3)$$

* 本文受国家自然科学基金资助。

表 1 中国生育与死亡预测参数

x	am _x	bm _x	af _x	bf _x	x	am _x	bm _x	af _x	bf _x
0	-3.2792	0.0683	0.0000	0.0000	45	-5.3160	0.0458	0.0038	-0.0134
1	-5.3986	0.2131	0.0000	0.0000	50	-4.8539	0.0402	0.0000	0.0000
5	-6.6260	0.1888	0.0000	0.0000	55	-4.3877	0.0319	0.0000	0.0000
10	-7.2312	0.0980	0.0000	0.0000	60	-3.8680	0.0289	0.0000	0.0000
15	-6.9351	0.0155	0.0213	-0.0442	65	-3.4222	0.0162	0.0000	0.0000
20	-6.5935	0.0175	0.2046	-0.1441	70	-2.9167	0.0143	0.0000	0.0000
25	-6.5300	0.0387	0.2111	-0.2505	75	-2.4737	0.0123	0.0000	0.0000
30	-6.3499	0.0422	0.1143	-0.3135	80	-2.0082	0.0072	0.0000	0.0000
35	-6.0715	0.0582	0.0650	-0.2631	85	-1.6173	0.0072	0.0000	0.0000
40	-5.7300	0.0556	0.0288	-0.1265					

在上述预测下, 中国人口的期望寿命的期望值将以递减的速度上升, 在 2050 年达到 76.7 岁, 而其 95% 的置信区间将逐渐扩大, 到 2050 年为 71.9 岁至 81.4 岁。

3. 中国生育预测主要结果

中国随机生育预测主要结果 (李南、申卯兴, 1996) 为:

$$f_{xt} = af_x + bf_x \cdot f_t \quad (4)$$

$$f_t = c_0 + c_1 \cdot f_{t-1} + ef_t; \quad (5)$$

其中:

f_{xt} 为 t 时刻中国 x 岁女性人口生育率;

$f_0 = 0.175$, $c_0 = 0.02546$, $c_1 = 0.8826$

$$ef_t \sim N(0, \sigma_f^2); E(ef_t, ef_s) = 0, t \neq s; \sigma_f = 0.0489 \quad (6)$$

ef_t 和 em_t 独立, af_x 和 bf_x 的值见表 1。

在上述预测下, 中国人口的总和生育率在 2050 年达到 2.0165, 而其 95% 的置信区间将逐渐扩大, 到 2050 年为 1.31 至 3.1703。置信区间的这种不对称性是由于当 f_t 的状态在偏离期望非常远时使某些年龄的按龄生育率为负造成的, 这是由经典时间序列模型中的随机扰动为正态分布的假定引起的; 这不仅是 LTC 方法, 也是一般地将经典时间序列模型用于逻辑上有界的变量时必然遇到和需要解决的问题, 虽然它不对通常关心的期望附近的状态产生显著影响。

4. LTC 方法简介

记 t 时刻 Leslie 阵为 X_t , 则预测按龄人口向量 N_t 需计算 X_t 的连乘积 M_t :

$$M_t = \prod_{i=1}^t X_i = \prod_{i=1}^t (b_i + Z_i); E(Z_i) = 0 \quad (7)$$

其中 b_i 描述生育与死亡预测中确定变化, 而 Z_i 描述其中的随机扰动影响。

按 LTC 方法 (Lee, Tuljapurkar, 1991), 当 Z_i 各元素的方差不大时, M_t 可对随机扰动二阶近似为:

$$M_t \approx a(t, 1) + S_{1t} + S_{2t} \quad (8)$$

其中:

$$a(t_2, t_1) = \begin{cases} b_{t_2} b_{t_2-1} \cdots b_{t_1}, & t_2 > t_1 \\ b_{t_1}, & t_2 = t_1 \\ I, & t_2 < t_1 \end{cases} \quad (9)$$

$$S_{it} = \sum_{i=1}^t a(t, i+1) Z_i a(i-1, 1) \quad (10)$$

$$S_{2t} = \sum_{i=1}^{t-1} \sum_{j=1}^{t-i} a(t, i+j+1) Z_{i+j} a(i+j-1, i+1) Z_i a(i-1, 1) \quad (11)$$

此时已可计算 N_t 及其比值 (如总人口、老龄人口、负担比等), 定义选择向量 u 和 w , 则:

$$(u, N_t) = [u, a(t, 1) + S_{1t} + S_{2t}] \cdot n_0 \quad (12)$$

$$\begin{aligned} E(u, N_t) &= [u, a(t, 1) \cdot n_0] + E(u, S_{1t} \cdot n_0) \\ &= (u, n_0) + [u, E(S_{1t} \cdot n_0)] \end{aligned} \quad (13)$$

$$Var(u, N_t) = (u \otimes u) \cdot E(S_{1t} \otimes S_{1t}) \cdot (n_0 \otimes n_0) \quad (14)$$

⊗ 表示张量积或 Kronecker 积 (须田信英, 1979)。而对于比值

$$D_t = \frac{(w, N_t)}{(u, N_t)}, \quad d_t = \frac{(w, n_0)}{(u, n_0)} \quad (15)$$

$$E(D_t) = d_t - \frac{[E(U_1 W_1)/u + E(W_2) - d_t E(U_2) + d_t^2 E(U_1^2)/u]}{u} \quad (16)$$

$$Var(D_t) = \frac{d_t^2 E(U_1^2) + E(W_2^2) - 2d_t E(U_1 W_1)}{u^2} \quad (17)$$

其中:

$$E(W_2) = (w, ES_{2t} n_0) \quad (18)$$

$$E(U_2) = (u, ES_{2t} n_0) \quad (19)$$

$$\begin{aligned} E(U_1 W_1) &= E(u, S_{1t} n_0) (w, S_{1t} n_0) \\ &= (u \otimes w) ES_{1t} \otimes S_{1t} (n_0 \otimes n_0) \end{aligned} \quad (20)$$

$$\begin{aligned} E(U_1^2) &= E(u, S_{1t} n_0)^2 \\ &= (u \otimes u) ES_{1t} \otimes S_{1t} (n_0 \otimes n_0) \end{aligned} \quad (21)$$

$$E(W_2^2) = (w \otimes w) ES_{2t} \otimes S_{2t} (n_0 \otimes n_0) \quad (22)$$

注意到其中方差是正态随机变量的平方或积的期望, 因而在 t 时刻上述变量与比值的分布均可近似地视为正态分布, 从而以期望为中心的某种程度的置信区间均可按正态分布计算。

以上就是有关 LTC 方法的主要结果, 剩下的只是根据生育与死亡预测的结果算出 $E(S_{2t})$ 及 $E(S_{1t} \otimes S_{1t})$, 即由已知数据来实现。法。

5. LTC 方法的实现

基于简单和平滑的原因, 随机死亡和生育预测是在简缩生命表和 5 岁组按龄生育率基础上进行的, 因而人口预测将以 5 年的间隔进行。这样, 实现方法的主要问题在于给出 5 年 5 岁组按龄留存率和生育率的表达式。

由于 $S_{it} (i=1, 2)$ 仅包含正态随机变量的 i 阶项, 所以上述表达式中仅包含正态随机变量的一阶项。

5.1 分年龄组留存率

记 t 时刻第 x 年龄组留存到下一时刻和年龄组的留存率为 p_x , 其期望为 $E(p_x) = \hat{p}_x$

其中 x 按顺序对应的年龄段为 $0 \sim 4$ 岁、 $5 \sim 9$ 岁, \dots 。

5.1.1 分年龄组死亡率 m_x

记 t 时刻第 x 年龄组死亡率为 m_x , 其期望为 $E(m_x) = \hat{m}_x$, 其中 x 按顺序对应的年龄段为 0 岁、 $1 \sim 4$ 岁、 $5 \sim 9$ 岁, \dots 。则由 (1), (2):

$$\begin{aligned} m_x &= e^{[am_x + bm_x \cdot (k_0 - 5a \cdot t + \sum_{i=1}^t am_i)]} \\ &\approx e^{[am_x + bm_x \cdot (k_0 - 5a \cdot t)]} [1 + bm_x (\sum_{i=1}^t am_i)] \\ &= \hat{m}_x (1 + bm_x \eta_t) \\ &= \hat{m}_x + q_x \eta_t \end{aligned} \quad (23)$$

其中, η_t 为零均值正态变量。

5.1.2 分年龄组起始年龄人数 l_x

记 t 时刻生命表中第 x 年龄组起始年龄人为 l_x , 其期望为 $E(l_x) = \hat{l}_x$, 其中 x 按顺序对应的年龄段为 0 岁、 $1 \sim 4$ 岁、 $5 \sim 9$ 岁, \dots 。则由 (蒋正华, 1984):

$$\begin{aligned} m_x &= \frac{l_{xt} - l_{(x+1)t}}{a_x l_{xt} + (1 - a_x) l_{(x+1)t}} \\ &= \frac{l_{xt} - l_{(x+1)t}}{a_x l_{xt} + b_x l_{(x+1)t}} \end{aligned} \quad (24)$$

$$a_0 = 0.3; a_1 = 2; a_x = 2.5, x \geq 3 \quad (25)$$

可得:

$$l_{(x+1)t} = [(1 + a_x m_{xt}) / (1 + b_x m_{xt})] l_{xt}, l_{0t} = \hat{l}_{0t} = 1 \quad (26)$$

于是有:

$$l_{(x+1)t} = \hat{l}_{(x+1)t} + \tau_{(x+1)t} \eta_t \quad (27)$$

$$\tau_{(x+1)t} = \hat{l}_{(x+1)t} \tau_{xt} + [a_x q_{xt} \hat{l}_{xt} / -b_x q_{xt} \hat{l}_{(x+1)t}] / (1 + b_x \hat{m}_{xt}), \tau_{0t} = 0 \quad (28)$$

5.1.3 分年龄组平均人数 L_x

记 t 时刻生命表中第 x 年龄组平均人数为 L_x , 其期望为 $E(L_x) = \hat{L}_x$, 其中 x 按顺序对应的年龄段为 $0 \sim 4$ 岁、 $5 \sim 9$ 岁, \dots , 则由:

$$L_{0t} = a_0 + (b_0 + a_1) l_{1t} + b_1 l_{2t} \quad (29)$$

$$\begin{aligned} L_{xt} &= a_x l_{xt} + b_x l_{(x+1)t} \\ &= \hat{L}_{xt} + \gamma_{xt} \eta_t, \quad x \geq 1 \end{aligned} \quad (30)$$

其中:

$$\gamma_{0t} = (b_0 + a_1) \tau_{1t} + b_1 \tau_{2t} \quad (31)$$

$$\gamma_{xt} = a_x \tau_{(x+1)t} + b_x \tau_{(x+2)t}, \quad x \geq 1 \quad (32)$$

5.1.4 分年龄组留存率 p_x

记 t 时刻第 x 年龄组留存到下一时刻和年龄组的留存率为 p_x , 其期望为 $E(p_x) = \hat{p}_x$, 其中 x 按顺序对应的年龄段为 $0 \sim 4$ 岁、 $5 \sim 9$ 岁, \dots 。则由:

$$p_x = L_{(x+1)t} / L_{xt} \quad (33)$$

$$\text{有: } p_{xt} = \hat{p}_{xt} + v_{xt} \eta_t \quad (34)$$

$$\text{其中: } v_{xt} = (\gamma_{(x+1)t} - \hat{p}_{xt}\gamma_{xt}) / \hat{L}_{xt} \quad (35)$$

5.2 分年龄组生育率

记 t 时刻女性第 x 年龄组单年生育为 f_{xt} , 其期望为 $E(f_{xt}) = \hat{f}_{xt}$, 其中 x 按顺序对应的年龄段为 $0 \sim 4$ 岁、 $5 \sim 9$ 岁, ……。则由 (4), (5):

$$\begin{aligned} f_{xt} &= af_x + bf_x(c_0 + c_1 f_t) \\ &= af_x + bf_x(c_0 + c_1 \hat{f}_t) + bf_x \xi_t \\ &= \hat{f}_{xt} + bf_x \xi_t \end{aligned} \quad (36)$$

其中:

$$\hat{f}_{xt} = c_1^{st} f_0 + c_0 \sum_{i=1}^{st} c_1^{st-i} \quad (37)$$

$$\xi_t = ef_t + \sum_{i=1}^{st-1} c_1^{st-i} ef_i \quad (38)$$

可见 ξ_t 是零均值正态分布变量且与 η_t 独立。

记 t 时刻第 x 年龄组 5 年生育率为 F_{xt} , 其期望为 $E(F_{xt}) = \hat{F}_{xt}$, 其中 x 按顺序对应的年龄段为 $0 \sim 4$ 岁、 $5 \sim 9$ 岁, ……。则由:

$$F_{xt} = s \cdot L_{1t} [f_{xt} + p_{xt} f_{(x+1)t}] / 2 \quad (39)$$

$$\text{有: } F_{xt} = \hat{F}_{xt} + \delta_{xt} \xi_t + \mu_{xt} \eta_t \quad (40)$$

(39) 式中 s 为基准年女性人口与总人口之比, 这是因为死亡预测是对不分性别的人口进行的, 因而是 LTC 方法在应用中的有待改进之处。

5.3 构成 Leslie 阵 X_t

$$\begin{aligned} X_t &= B_t + C_t \eta_t + D_t \xi_t \\ &= B_t + Z_t \end{aligned} \quad (41)$$

其中, B_t 首行对应元素为 \hat{F}_{xt} , 下次对角线上对应元素为 \hat{p}_{xt} , 其他元素为零; C_t 首先对应元素为 μ_{xt} , 下次对角线上对应元素为 v_{xt} , 其他元素为零; D_t 首行对应元素为 δ_{xt} , 其他元素为零。

5.4 计算 $E(S_{2t})$ 及 $E(S_{1t} \otimes S_{1t})$

有了 Z_t 表达式 (41) 并注意到 ξ_t 与 η_t 独立后, 即可按 (10)、(11) 计算 $E(S_{2t})$ 及 $E(S_{1t} \otimes S_{1t})$

$$\begin{aligned} E(S_{2t}) &= \sum_{i=1}^t \sum_{j=1}^{t-i} a(t, i+j+1) C_{i+j} a(i+j-1, i+1) C_i a(i-1, 1) E[\eta_{i+j} \eta_i] \\ &\quad + \sum_{i=1}^t \sum_{j=1}^{t-i} a(t, i+j+1) D_{i+j} a(i+j-1, i+1) D_i a(i-1, 1) E[\xi_{i+j} \xi_i] \end{aligned} \quad (42)$$

$$\begin{aligned} E(S_{1t} \otimes S_{1t}) &= \sum_{i=1}^t \sum_{j=1}^t [a(t, i+1) C_i a(i-1, t)] \otimes [a(t, j+1) C_j a(j-1, 1)] E[\eta_i \eta_j] \\ &\quad + \sum_{i=1}^t \sum_{j=1}^t [a(t, i+1) D_i a(i-1, t)] \otimes [a(t, j+1) D_j a(j-1, 1)] E[\xi_i \xi_j] \end{aligned} \quad (43)$$

$$E(\eta_{i+j} \eta_i) = 5i\sigma_m^2 \quad (44)$$

$$E(\xi_{i+j} \xi_i) = \sigma_f^2 \left(\sum_{m=1}^{st} \pi_{i+j-m} \pi_{i-m} \right) \quad (45)$$

最后, 由于预测按 5 年步长进行, 而生育与死亡率应取为 5 年间隔中点的值, 因而应将 (2) 与 (5) 中 k_0 与 f_0 改为按模型倒推 2.5 年, 即:

6. 中国随机人口预测主要结果

这里，给出从1990年到2050年的总人口、老龄比（65岁及以上为老年）、负担比（15岁到64岁为工作年龄）及少年和老年负担比作为主要结果。其中的期望变化在正态分布的意义下也就是最可能的变化。在此意义下，中国人口总数将在2045年达到最大值15.9亿然后开始下降。其他细节见图1~5及表2。

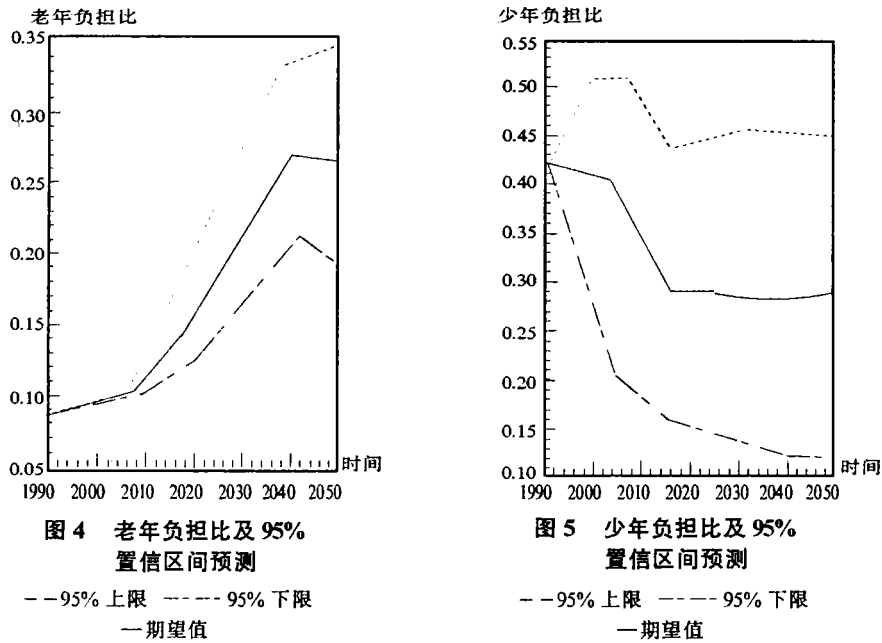
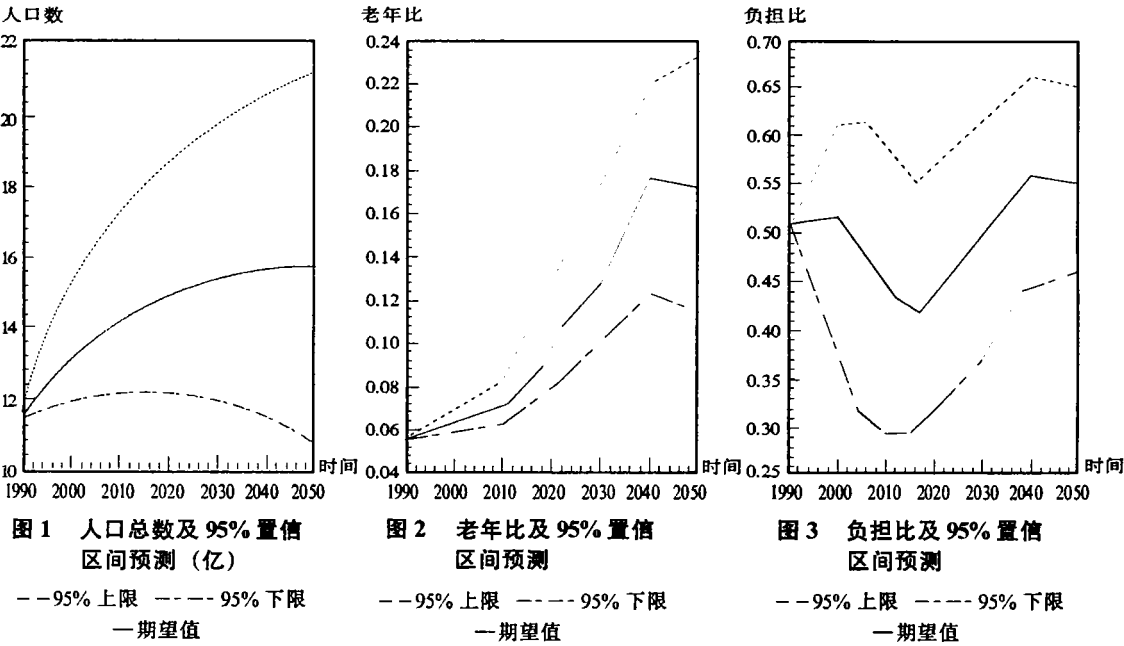


表 2 中国人口总数的期望与 95% 置信区间预测

亿

年份	区间下限	期望	区间上限	年份	区间下限	期望	区间上限
1995	11.6472	12.1059	12.5646	2025	12.1965	15.3493	18.5021
2000	11.9142	12.8324	13.7506	2030	11.9954	15.6227	19.2499
2005	12.0566	13.4188	14.7810	2035	11.7275	15.7883	19.8492
2010	12.1700	13.9394	15.7089	2040	11.4118	15.8783	20.3448
2015	12.2750	14.4601	16.6452	2045	11.0455	15.9069	20.7682
2020	12.2955	14.9513	17.6071	2050	10.6425	15.8873	21.1321

图中显示了变量的期望与 95% 置信区间的随时间变化的情况, 其中负担比的 95% 置信区间在 2025 年后有随时间轻微减小现象, 这是 LTC 方法中忽略了随机变量的高阶项引起的。

表 3 对联合国中国人口方案预测的评价

年份	亿方案(亿)	高方案(亿)	处于高低方案 间的可能性(%)
1995	12.3196	12.4416	11.65
2000	12.9076	13.2710	26.28
2005	13.2516	13.9540	37.42
2010	13.5561	14.5617	42.13
2015	13.8443	15.1943	45.42
2020	14.0722	15.8389	48.60
2025	14.1707	16.4223	51.59

从预测结果不仅可得到变量在给定时刻的期望及其某种程度的置信区间, 而且可得变量在给定时刻的分布, 这就可对各种确定性人口方案预测的实现可能性进行定量的评价。例如, 联合国 (United Nations, 1992) 的中国人口“高低方案”预测和按随机预测的对处于高低方案间的可能性的评价见表 3。其中短期的可能性较小是由与其 1990 年基数与随机预测 (11.30 亿, 国务院人口普查办公室, 1993) 的不同, 为 11.54 亿。

此外, 观察可能性的大小还应注意到方案的覆盖

范围, 范围较小时可能性就较小; 但是, 如此范围包含了期望而且关于期望是对称的, 那么它的可能性就是同大小范围中最大的, 方案预测在 2005 年之后就接近这种情况。

参 考 文 献

1. 国务院人口普查办公室等编:《中国 1990 年人口普查资料》, 第二册, 中国统计出版社, 1993 年。
2. 蒋正华:《人口分析与规划》, p53 ~ 57, 陕西科学技术出版社, 1984。
3. Lee, R.D, S. Tuljapurkar, 1991, "Stochastic Population Forecasts for the US: Beyond High, Medium and Low", presented at the 1991 meetings of the Population Association of America in Washington D.C.
4. Lee, R.D. and L. Carter, 1990, "Modeling and Forecasting US Mortality", 1990 Annual meetings of the Population Association of America (Toronto).
5. 李南、S. Tuljapurkar:《基于时间-区域序列的随机死亡率预测及对中国数据的应用》,《人口研究》1995(4), p58-63。
6. 李南、申卯兴:《基于随机模型的中国生育率预测》,《预测》1996 (6)。
7. Pollard, J.H. 1975, "Mathematical Models for the Growth of Human Population", University Press, Cambridge. p112-134.
8. United Nations, 1993, "World Population Prospects, The 1992 Revision" United Nations Publication, Sales No. E.93. X III. 7 p418-419.
9. 须田信英等著, 曹长修译:《自动控制中的矩阵理论》, 科学出版社, p155~157, 1979。

(本文责任编辑: 朱 萍)