1995年第1期 人口与经济 No. 1, 1995
(总第88期) POPULATION & ECONOMICS (Tot. No. 88)

关于一孩到二孩生育间隔模式变化规律的分析

陈建利

在应用孩次间隔递进比方法进行人口预测时，通过模拟分析我们发现出生数的估计对生育间隔模式的变化十分敏感。若给定准确的初始参数，则预测结果的准确性很大程度上依赖于对该模式的变化做出什么样的假设。因此，探讨模式变化的规律并据此寻求对其进行调控的思路及方法就成为预测过程中不可缺少的一环。本文采用数学函数模拟和统计分析的方法对1978—1992年全国29个省、市、自治区1孩到2孩生育间隔模式的变化进行了初步分析，并提出了几种不同的对模式进行调控和预测的思路。

一、生育间隔模式的概念

假设一批妇女共w人在t年内生了j孩，且这批妇女不受死亡的影响，我们来研究她们生育j+1孩的时间分布情况。记这批妇女在t+i年内生育j+1孩的妇女人数为W人，她们终身生育j+1孩的比例为q_0定义

$$v_i=w_i/w, i=0,1,2,\cdots,n \qquad (1.1)$$

(1.1)式中，n为最大间隔年数，V_i表示所研究的妇女队列在t+i年内生j+1孩的比例。显然，q和v_i有如下的关系

$$q=\sum_{i=0}^{n}w_i/w=\sum_{i=0}^{n}(w_i/w)=\sum_{i=0}^{n}v_i \qquad (1.2)$$

我们进一步定义

$$r_i=v_i/q,\ i=0,1,2,\cdots,n \qquad (1.3)$$

很显然

$$\sum_{i=0}^{n}r_i=1$$

向量$R=(r_0r_1\cdots r_n)$就是我们所定义的j孩到j+1孩的规范化间隔模式。以上是从队列的角度定义了j孩到j+1孩的生育间隔模式，在具体应用时，还需将它时期化。下面所分析的模式均为时期间隔模式。若给定规范化的j孩到j+1孩生育间隔模式，可采用不同的方法求得平均间隔年数，本文是按照下式计算的。

$$d=\sum_{i=0}^{n}ir_i \qquad (1.4)$$

二、数学函数模拟

从利用2‰抽样调查和38万抽样调查的资料所汇出的时期间隔模式来看，1孩到2孩的

生育间隔模式有着如下明显的特征——各个间隔年数的生育率都大于或等于零，从某个间隔年数起生育率从零开始平稳上升，并且在其后的某个间隔年数达到最高，然后开始持续下降，直到某个较高间隔年数降到零。从分布形状来讲，它属于正偏态的钟形曲线。另外，我们还发现各地区1孩到2孩生育间隔模式的时间序列均呈现出比较强的统计规律性。如果剔除随机因素的干扰，那么该模式是否服从或近似服从某种理想的概率分布？我们试图利用函数模拟的方法来验证上述猜测的正确与否，但选择理想的数学函数是一件非常复杂的事情，在很大程度上它取决于个人的经验和洞察力。一般来讲，理想的数学函数至少应满足以下三个条件：(1)具有良好的数学性质。(2)很好地拟合实际数据。(3)函数的形式和函数中的各参数要具有明确的人口学意义。鉴于上述情况，我们选取了广义伽玛分布密度函数。其定义如下：

$$f(x)=\begin{cases} k^{(x-x_0)^a}e^{-b(x-x_0)} & x>x_0 \\ o & x\leq x_0 \end{cases} \qquad (2.1)$$

f(x)是一个四参数函数，其中 x_0 是一个位置参数，它不影响函数曲线的形状；k 是一个水平参数，它保证 f(x)的规范性($\int_{-\infty}^{\infty}f(x)dx=1$)；a 和 b 是两个模式参数，它们影响函数 f(x)的形状。四个参数均大于零。若起始位置 x_0 给定，关键是估计 a 和 b 两个模式参数，求得 a 和 b 后，可由下式得到相应的 k 值。

$$k=1/\int_{x_0}^{\infty}(x-x_0)^a e^{-b(x-x_0)}dx \qquad (2.2)$$

模式的转变一般包括两个方面：曲线位置的变动和曲线形状的改变。生育模式的间隔年数的均值能较好地刻划曲线的位置，间隔年数的方差能较好地刻划曲线的形状。通过计算发现，密度函数 f(x)的均值与方差和 a、b 存在如下的关系：

$$e=\int_{-\infty}^{\infty}xf(x)dx=x_0+\frac{a+1}{b} \qquad (2.3)$$

$$d=\int_{-\infty}^{\infty}(x-e)^2f(x)dx=\frac{a+1}{b^2} \qquad (2.4)$$

从(2.3)、(2.4)式可以解得

$$a=\frac{(e-x_0)^2}{d}-1 \qquad (2.5)$$

$$b=\frac{e-x_0}{d} \qquad (2.6)$$

如果知道了一个生育模式的起点、间隔年数的均值和方差，我们就可以由(2.5)式、(2.6)式及(2.2)式计算出相应的 a、b、和 k。(2.2)式中，若估计出的参数 a 不是正整数，则不定积分无法用初等函数表示，只能用近似积分公式对其进行计算。这里，我们采用的是高斯——拉盖尔求积公式。

$$\int_0^{\infty}e^{-x}f(x)dx\approx\sum_{i=0}^{n}A_if(x_i) \qquad (2.7)$$

根据1988—1990年的1孩到2孩的平均生育间隔模式，计算出其均值与方差，我们求出了相应的函数 f(x)中各参数的估计值，见表2。模拟结果，除在个别点上，观察值与拟合值存在较大差异外，其余的均大致吻合。由此可认为上述所定义的广义伽玛分布密度函数能较好地对生育间隔模式进行模拟。

表1　高斯——拉盖尔求积公式的Xi和Ai的值(n=4)

X_i	A_i
0.2635603	0.5217556
1.4134031	0.3986668
3.5964258	0.0759424
7.0858100	0.0036118
12.6408008	0.0000234

资料来源：计算方法，西安交通大学出版社，1985年。

表2　　1孩到2孩的生育间隔模式 $f(x)=kx^ae^{-bx}$

地区	k	a	b	地区	k	a	b
北　京	0.3746085	2.0840	0.9343	河　南	0.2922465	0.7629	0.4763
天　津	0.2221559	1.0581	0.4874	湖　北	0.3993968	0.8938	0.6031
河　北	0.2543383	0.9225	0.4830	湖　南	0.3555373	0.7417	0.5268
山　西	0.2758342	1.2392	0.5928	广　东	0.3559725	1.3267	0.6898
内　蒙	0.2424945	2.1033	0.8173	广　西	0.6194428	1.2949	0.8663
辽　宁	0.0433710	2.0464	0.4546	海　南	0.4923003	1.0016	0.7021
吉　林	01316935	1.4279	0.4777	四　川	0.2297030	1.1705	0.5268
黑龙江	0.2337898	1.0772	0.5049	贵　州	0.4413850	1.7746	0.8902
上　海	0.4987870	0.7065	0.6313	云　南	0.5246345	1.1815	0.7738
江　苏	0.2629153	0.6055	0.4078	西　藏			
浙　江	0.0784453	1.4485	0.3913	陕　西	0.2527253	0.2957	0.3218
安　徽	0.4907961	1.5482	0.8568	甘　肃	0.3947473	0.6002	0.5240
福　建	0.7279068	1.6664	1.0343	青　海	0.4364090	1.2051	0.7179
江　西	0.5651398	1.1178	0..7832	宁　夏	0.4841162	0.7873	0.6403
山　东	0.2026922	0.5491	0.3327	新　疆	0.3073716	0.4367	0.4077

三、回归分析

假设从1孩到2孩的生育间隔模式的一般形式为 $R(t)=(r_0(t)\ r_1(t)\cdots r_n(t))$，0表示当年、n为最大间隔年数。通过观察与计算，我们发现R(t)中每个分量的变化趋向不很稳定，这主要是由随机因素的干扰造成的。因此，很难找到R(t)中每个分量的变化统计规律。1孩到2孩间隔模式的变化反映了计划生育政策执行的结果。更具体地讲，现行政策要求1孩到2孩的间隔为4年，$r_1(t)+r_2(t)+r_3(t)$ 的变化则体现了间隔政策执行的成败。定义 $s(t)=r_1(t)+r_2(t)+r_3(t)$。显然，s(t)弱化了随机因素的干扰，它是间隔政策执行结果的综合反映，其变化趋向要比R(t)中每个分量的变化趋向稳定的多。我们利用1978—1992年的1孩到2孩的生育模式从以下两个角度对s(t)的演变进行了分析：(1)s(t)随时间的变化统计规律，回归方程的形式为：$s(t)=a+b(t+1-t_0)$，t_0 为起始年份。(2)s(t)和1孩到2孩的平均间隔年数之间的关系，回归方程的形式为：$s(t)=a+b*d(t)$，d(t)为平均间隔年数。表3、表4给出了相应的估计值。

由表 3 可知，各地区的 s(t) 和时间变量 t 均存在较强的负相关关系，相关系数都大于 0.60，系数 a 和 b 的 T 检验值均通过了统计检验。从 s(t) 的历史演变过程中可看到，随着计划生育工作的不断加强和完善，s(t) 不断降低。系数 b 则从另一个侧面部分地反映了各地区间隔政策执行的力度。

从表 4 可看到，各地区的 s(t) 和平均间隔年数也存在较强的负相关关系，相关系数均大于 0.60，系数 a 和 b 的 T 检验值都通过了统计检验。一般来讲，s(t) 的增大必会造成平均间隔年数 d(t) 的减小；反之，s(t) 的减小肯定引起 d(t) 的增大。表 4 所给出的统计分析结果和上述主观判断是吻合的。

s(t) 随时间演变的统计规律及 s(t) 和平均间隔年数之间的关系可为模式的调控和预测提供重要的依据。

表 3　　$s(t)=a+b(t+1-t_0)$

地区	a	b	R－squared	t_0	地区	a	b	R－squared	t_0
北　京	—	—	—	—	河　南	0.7443 (20.60)	－0.0170 (－40.30)	0.65	1978
天　津	—	—	—	—	湖　北	0.8300 (20.04)	－0.0314 (－2.96)	0.69	1987
河　北	0.6598 (21.53)	－0.0098 (－2.72)	0.84	1978	湖　南	0.8991 (14.17)	－0.0373 (－3.98)	0.64	1982
山　西	0.7145 (30.78)	－0.0177 (－5.18)	0.75	1982	广　东	0.8773 (26.70)	－0.0215 (－5.95)	0.73	1978
内　蒙	0.8054 (17.27)	－0.0237 (－4.62)	0.62	1978	广　西	0.9253 (23.44)	－0.0312 (－4.83)	0.79	1983
辽　宁	0.6409 (10.83)	－0.0394 (－6.05)	0.74	1978	海　南	0.8204 (41.20)	－0.0127 (－4.38)	0.77	1978
吉　林	0.6763 (12.30)	－0.0261 (－4.38)	0.62	1978	四　川	0.5342 (37.51)	－0.0112 (－4.97)	0.87	1982
黑龙江	0.8067 (15.14)	－0.0270 (－4.61)	0.62	1978	贵　州	0.8802 (20.02)	－0.0245 (－4.10)	0.63	1981
上　海	—	—	—	—	云　南	0.9008 (25.94)	－0.0268 (－4.35)	0.73	1984
江　苏	0.6377 (20.71)	－0.0198 (－3.86)	0.73	1983	西　藏	—	—	—	—
浙　江	0.8737 (25.74)	－0.0505 (－13.52)	0.93	1978	陕　西	0.6849 (34.56)	－0.0063 (－2.14)	0.76	1982
安　徽	0.8641 (28.72)	－0.0182 (－6.23)	0.71	1978	甘　肃	—	—	—	—
福　建	0.8607 (56.21)	－0.0183 (－4.65)	0.84	1982	青　海	0.8479 (63.58)	－0.0179 (－6.76)	0.88	1978
江　西	0.8547 (125.71)	－0.0080 (－4.58)	0.84	1981	宁　夏	0.9133 (21.30)	－0.0161 (－3.36)	0.61	1978
山　东	0.6732 (16.95)	－0.0239 (－5.14)	0.71	1978	新　疆	0.7304 (31.02)	－0.0052 (－2.01)	0.62	1978

注：括号中的数字为 T 检验值

四、模式的调控及预测

前面的分析实际上给我们提供了对间隔模式进行调控和预测的思路。假设预测基年的模式为 $R=(r_0\ r_1 \cdots r_n)$，$s=r_1+r_2+r_3$，要预测的未来年份的模式为 $R(t)=(r_0(t)\ r_1(t)\ \cdots\ r_n(t))$，

表 4　　$s(t)=a+b*d(t)$

地区	a	b	R—squared	地区	a	b	R—squared
北　京	1.0848 (12.65)	−0.1455 (−6.74)	0.77	河　南	1.1811 (21.30)	−0.1567 (−10.63)	0.90
天　津	0.9463 (10.85)	−0.1066 (−5.27)	0.68	湖　北	1.0136 (16.32)	−0.0914 (−5.12)	0.70
河　北	1.0801 (16.56)	−0.1273 (−8.17)	0.84	湖　南	1.4153 (15.59)	−0.2225 (−8.16)	0.84
山　西	1.1383 (10.94)	−0.1469 (−5.09)	0.67	广　东	1.0406 (12.16)	−0.1436 (−5.64)	0.91
内　蒙	1.2771 (29.87)	−0.1882 (−15.75)	0.95	广　西	1.3802 (15.13)	−0.2149 (−6.54)	0.77
辽　宁	0.8917 (9.36)	−0.1007 (−6.20)	0.75	海　南	1.2276 (19.84)	−0.1533 (−7.57)	0.82
吉　林	1.0040 (16.01)	−0.1141 (−8.92)	0.86	四　川	0.8418 (10.33)	−0.0774 (−4.12)	0.68
黑龙江	1.0639 (33.79)	−0.1161 (−15.71)	0.95	贵　州	1.2926 (19.28)	−0.1809 (−8.44)	0.85
上　海	1.0272 (15.98)	−0.1375 (−9.22)	0.93	云　南	0.9472 (20.59)	−0.0558 (−3.75)	0.75
江　苏	1.2113 (6.86)	−0.1588 (−3.71)	0.60	西　藏	—	—	—
浙　江	1.2298 (13.70)	−0.1645 (−8.78)	0.86	陕　西	0.9497 (10.39)	−0.0838 (−3.47)	0.62
安　徽	1.3949 (15.76)	−0.2202 (−7.50)	0.81	甘　肃	1.1519 (15.08)	−0.1326 (−5.71)	0.71
福　建	1.3071 (56.57)	−0.1844 (−23.06)	0.98	青　海	1.3551 (10.80)	−0.2037 (−4.52)	0.61
江　西	1.2955 (15.21)	−0.1708 (−5.66)	0.78	宁　夏	1.4272 (10.01)	−0.2314 (−4.63)	0.62
山　东	1.0577 (13.32)	−0.1240 (−7.05)	0.79	新　疆	0.8680 (30.32)	−0.0474 (−6.13)	0.84

注：括号中的数字为 T 检验值

$s(t)=r_1(t)+r_2(t)+r_3(t)$。我们主要考虑了以下三种方法：(1)外推法。如果预测的年限较短时，可将 $s(t)=a+b(t+1-t_0)$ 进行简单的外推，求得相应的 $s(t)$ 值。这时，$R(t)$ 中的各分量可表示为如下的形式：

$$r_i(t)=\begin{cases} r_0, & i=0 \\ \alpha r_i, & 0<i<4 \\ \beta r_i, & i>=4 \end{cases} \tag{4.1}$$

其中，$\alpha=s(t)/s$，$\beta=(1-r_0-s(t))/(1-r_0-s)$。(2)参数控制法。如果给定未来某年份的平均生育间隔年数，则可通过控制方差由伽玛密度函数 $f(x)=kx^ae^{-bx}$ 计算出未来的生育模式；另外，也可利用 $s(t)$ 与平均间隔年数之间的经验关系〔$s(t)=a+b*d(t)$，$d(t)$ 代表平均间隔年数〕求得。(3)目标设置法。如果设定了未来某年份要达到的模式目标，则可以用线性插值或非线性插值的办法计算出各年度的生育模式。

（作者工作单位：国家计生委规统司）

参考文献：

1. 张二力、路磊：孩次——持续时间生育模型及其在人口预测中的应用，人口发展前景与对策科学讨论会论文，1993 年 2 月。
2. 路磊：试论人口预测中生育率参数的确定（中国人民大学硕士学位论文），1989 年 6 月。