

# 初婚模型——对中国人口数据适用性的探讨

黄荣清

**摘要** 由于婚姻在人口再生产中所起的重要作用,人口学家历来重视婚姻研究。从七十年代以来,以美国人口学家寇尔为首,在定量研究方面取得了不少成就。本文较为详细地介绍了人口学家在人口初婚模型建立方面所作的一些重要工作和初婚模型间的联系,并把这些模型用于中国女性人口的调查数据上来探讨它们的适用性,最后还探讨了在实际中使用这些模型时的一些注意点。

**作者** 黄荣清,男,1946年生,首都经济贸易大学人口经济研究所研究员。(北京市100026)

婚姻是人类两性结合的一种社会形式。人类的婚姻始终受社会因素影响。由于不同社会在法律制度、社会习俗、生活方式、思想观念方面的不同,这些社会因素对婚姻过程,包括婚姻开始时间(初婚起始年龄),婚姻速度(年龄别初婚概率),婚姻的完成(终身已婚率)作用的结果,形成在不同时期,不同地区,不同人口集团在年龄别婚率上的差别来。但尽管存在着不同人口集团的婚率差别,作为一个共同的从未婚到已婚的过程,是否存在着一个反映这一相同过程共同规律?如果这种规律存在,那末记述这种变化过程的方法是什么?找出这种反映从未婚到已婚变化过程的方法,就是所谓的人口初婚模型的研究。以下,我们将介绍人口学家关于初婚模型的一些重要研究成果以及用于中国人口的情况。

## 一、初婚模型

### 1、寇尔的初婚模型

提起初婚模型,首先要提人口学家寇尔(Coale)的研究。寇尔最初的工作是通过对一些国家和地区在不同时期已婚率的比较,以瑞典女性1865~1869年已婚数据为基础,制作了以年龄为自变量的初婚进度表,以此为标准来配制不同人口的已婚率。<sup>(1)</sup>后来,他与麦克内尔(Mc Neil)(1972)合作,把表格数据改成初婚概率(风险函数)形式,这就是:

$$r_s(x) = 0.174 \exp[-4.411 \exp(-309x)] \quad (1)$$

而其它人口的已婚率可以通过 $r_s(x)$ 调整已婚率的最高水平,初婚开始发生的年龄和尺度参数求得。<sup>(2)</sup>

后来,他们又把初婚变动写成更一般的形式,这一般形式的密度函数为:

$$f(x) = \frac{\lambda}{\Gamma(a/\lambda)} \exp\{-\alpha(\chi-\theta) - \exp[-\lambda(\chi-\theta)]\} \quad (2)$$

注:把概率分布函数 $F(x)$ 与 $x$ 岁已婚率相对应时, $x$ 必须减去初始结婚年龄 $\chi_0$ ,即与 $x - \chi_0$ 相对应。

对(2)式积分,可得其分布函数:

$$F(x) = \frac{1}{\Gamma(a/\lambda)} \int_{\exp(-\lambda(x-\theta))}^{\infty} y^{\frac{a}{\lambda}-1} e^{-y} dy \quad (3.1)$$

(3.1)式常常写成以下形式:

$$F(x) = 1 - I\{\exp[-\lambda(x-\theta)], \frac{a}{\lambda}\} \quad (3.2)$$

这里  $I(x, p) = \frac{1}{\Gamma(p)} \int_0^x y^{p-1} e^{-y} dy$ , 即常用的不完全  $\Gamma$  函数的形式。由式(3)可知,寇尔等最后提出的初婚模型是有 3 个不同参数的模型。

寇尔等还根据一般形式与标准形式的关系,通过对参数关系的限制,把上述三参数削减为 2 个,其分布函数为:

$$F(x) = 1 - I\left\{\exp\left[\frac{0.288}{K}(x-x_0) - \frac{K}{0.288} \frac{r'}{r}\right], 0.604\right\} \quad (4)$$

(4)式右边括号中  $r, r'$  分别为  $\Gamma$  函数和  $\Gamma$  函数的导数在 0.604 的值。

### 2. 赫内斯和逻辑斯蒂对数模型

初婚密度函数类似于逻辑斯蒂函数的变化。对每一个出生同期群来说,其初婚频率,在开始进入初婚年龄时,是随年龄增加而增加的,到一定年龄后初婚频率达到最大值,以后逐渐下降。其累积频率最后趋向于一个定值。所以初婚密度函数具有:

$$f(x) = \frac{dF(x)}{dx} = q(x)F(x)(1-F(x)) \quad (5)$$

的形式。这里方程中  $\frac{dF}{dx} = F \cdot (1-F)$  的  $F$  是典型的逻辑斯蒂(Logistic)函数,而  $q(x)$  为刻划初婚随年龄变化对逻辑斯蒂函数修正的函数。

赫内斯(Hernes, 1972)认为,  $q(x) = Ab^x (0 < b < 1)$ , 解方程式(5)可得:

$$F(x) = \left[1 + \frac{1-F(0)}{F(0)} \cdot \frac{\exp(A/\ln b)}{\exp(Ab^x/\ln b)}\right]^{-1} \quad (6)$$

这就是赫内斯初婚模型。<sup>[3]</sup>

坎里(Keeley, 1979)和苏申等(Sorensen, 1986)认为,  $q(x) = a/x$ 。把  $q(x)$  代入(5)解出  $F(x)$ , 得到  $F(x)$  为如下形式:

$$F(x) = \frac{\exp(\beta + a \ln x)}{1 + \exp(\beta + a \ln x)} = \frac{\exp(\beta)x^a}{1 + \exp(\beta)x^a} \quad (7)$$

(7)式称为逻辑斯蒂对数(log-logistic)模型。<sup>[4][5]</sup>

### 3. 极大值极值分布模型和伽马分布模型

在寇尔的初婚模型(2)中,若令  $\lambda = \alpha$ , 则有:

$$f(x) = \alpha \exp\{-\alpha(x-\theta) - \exp[-\alpha(x-\theta)]\} \quad (8)$$

对(8)式积分可得:

$$F(x) = \exp\{-\exp[-\alpha(x-\theta)]\} \quad (9)$$

这就是极大值的极值分布模型。在工业产品的寿命研究中,当一个工业制品有许多零件(这些零件的寿命具有独立的有相同的指数分布式伽马分布等指数形式)组成,则这个制品的零件寿命的最大值就是该制品的寿命分布,其分布函数的形式即是(9)的形式。<sup>[6]</sup>

在寇尔初婚的一般模型中,初婚分布通过不完全伽马函数积分形式来表示的,到某一个年龄点  $x$  的已婚率,可通过这个积分求得,其下限是和  $x$  有关的一个值,而上限为无穷大,这种表示形式和习惯的分布函数表示形式有所不同。如果我们保持被积函数(密度函数)为伽马分布密度不变,而积分上下限按通常的形式,即下限为初婚年龄的起点,上限为某一个年龄点  $x$ , 用这样的积分值是否可以用来表示初婚模型呢? 即用如下形式:

$$F(x) = \frac{K^p}{\Gamma(p)} \int_0^x y^{p-1} e^{-ky} dy \quad (10)$$

式(10)即是伽马分布函数。在人口生育率研究中,它常被用来表示生育模型。

#### 4. 对数正态分布模型

作者(黄荣清,1990)曾提出用对数正态分布作为初婚模型。对数正态分布的形式为

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^x \frac{1}{y} \exp\left[-\frac{(y-\theta)^2}{2\sigma^2}\right] dy \quad (11)$$

在此模型中共有 2 个参数,它们分别为  $\theta$  和  $\sigma$ 。(7)

#### 5. 初婚间接估计模型。

上述模型中,都是把已婚率看作是年龄的函数。即年龄为自变量,已婚率为因变量的概率函数模型。如果模型中参数确定以后,则可以根据年龄的变化直接估计已婚率的值。布拉斯(W. Brass)的初婚模型却不同,它是以某一个确定的人口已婚率来估计另一个人口的已婚率,所以称之为间接估计模型。这个模型类似于人口死亡研究中的 logit 体系。他的估计方法是这样的:设有标准的年龄别已婚率为  $F_s(x)$ ,待估计的年龄别已婚率为  $F(x)$ ,  $F_s(x)$  和  $F(x)$  有关系

$$\ln\left(\frac{1-F(x)}{F(x)}\right) = \alpha + \beta \ln\left(\frac{1-F_s(x)}{F_s(x)}\right) \quad (12)$$

如果参数  $\alpha$  和  $\beta$  确定后,则可以根据式(12)确定  $F(x)$  的值。

其实,如果婚率变动若服从对数逻辑斯蒂模型,则由(7)式,我们就能推出两个人口的已婚率有(12)式的关系。事实上,由于

$$F(x) = \frac{\exp(\beta)x^a}{1 + \exp(\beta)x^a} \quad (13)$$

则

$$\frac{F(x)}{1-F(x)} = \exp(\beta)x^a$$

两边取对数并把左边写成 logit 体系的形式,则有

$$\ln\left(\frac{1-F(x)}{F(x)}\right) = b_0 + b_1 \ln x \quad (14)$$

$$(b_0 = -\beta, b_1 = -a)$$

(8)式表示,已婚率  $F(x)$  的 logit 形式与  $\ln x$  成线性关系。则由线性关系的传递性质,任意两个人口已婚率  $F(x)$  和  $F_s(x)$  的 logit 形式必有线性关系,即式(12)的关系。

有趣的是,当假定已婚率服从极大值极值分布时,可以推导出已婚率  $F(x)$  的两重对数,  $\ln\ln\left(\frac{1}{F(x)}\right)$  和年龄  $x$  成线性关系,从而推导出任意两个人口的已婚率的双重对数也有线性关系,有兴趣的读者不妨推导一下。

## 二、模型用于中国人口数据的结果

建立初婚模型的目的是要从理论上把握人口从未婚到已婚的变化过程。同时,模型也可作为表示这种变化的记述方法。检验一个模型是否成功往往要看它和实际的人口过程是否符合。下面我们来看一下上述模型用于中国人口数据的结果。

本文使用的数据是根据国家计生委在 1988 年实施的全国千分之二妇女生育节育调查结果中人口初婚率数据和 1990 年人口普查中女性人口的已婚率。设在 1988 年调查时  $Z$  岁人口在 14 岁、15 岁…… $Z-0.5$  岁(由于该调查时点是在年中,所以 1988 年调查时  $Z$  岁人口仅过了半年,因此估计为  $Z+0.5$ )的已婚率为  $M(15)$ , … $M(Z+0.5)$ ,到 1990 年人口普查时,1988 年  $Z$  岁的人口年龄为  $Z+2$  岁,其已婚率为  $M(Z+2.5)$ 。这样,对每一个 1988 年  $Z$  岁人口的同期群,由两次调查的结果,可知其年龄点已婚率数据为  $n=Z-13$  个,它们分别为  $M(x)$  ( $x=15, \dots, Z+0.5, Z+2.5$ )。

从 1988 年调查得到的已婚率数据,由于是抽样调查得到的,所以不可避免存在着抽样误差。把人口普查和抽样调查数据联用,不但扩大了可利用的信息量,并且普查结果同时也是对抽样调查结果准确性的一种检

验。对于人口普查,由于调查对象是全人口,如果没有调查错误,应该是没有误差的。但每个人的婚姻,与谁结婚,在什么年龄结婚本身就有很大偶然性,在不同社会条件和不同情况下,他(她)的结婚对象和结婚年龄可能就会不一样,所以说婚姻事件是个随机事件,则已婚率也是有随机变异的。从模型设计的目的来看,是要寻找一个不受随机干扰影响而能反映已婚进程的特征的函数;从数据的图形上看,是要寻找一条曲线;一方面,它是光滑的,另一方面,它在各年龄上的值与调查数据非常接近。而上面提到的各种初婚模型就是我们所希望寻找的函数。在各种模型中,判断哪一个模型更好一些的重要标准之一就是看该模型函数的值与调查值接近一些。

每个模型由该模型的参数来确定。因此,寻找参数的过程就是确定模型的过程。设某个模型中含有  $K$  个参数,  $A = (A_1, A_2, \dots, A_K)$ , 在本文提到的 6 个初婚模型中,  $K$  或者等于 2, 或者等于 3。此模型在各年龄  $x$  的值为  $F(x; A) (x=15, \dots, Z+0.5, Z+2.5)$ , 则寻找模型参数可归纳为以下过程:

$$\begin{cases} \text{Min} \sum_x [F(x; A) - M(x)]^2 & A \in R \subseteq E^n \\ R = \{A/g_j(A) \geq 0, j=1, 2, \dots, l\} \end{cases} \quad (15)$$

这里  $R$  为参数  $A = (A_1, A_2, \dots, A_K)$  的可行域, 而  $g_j(A) \geq 0$  表示可行条件。这是因为在每个初婚模型中, 其参数并不是任意的, 例如在寇尔初婚模型(3.2)中, 其参数  $\alpha, \lambda, \theta > 0, \alpha - \lambda > 0$  即必须附加 4 个条件( $l=4$ )不等式才成立。

确定式(12)中  $A$  实际上是求解非线性规划问题。本文利用可变误差多面体算法,<sup>[8]</sup>对在 1988 年调查时年龄为 25 岁到 60 岁共 36 个同期队列数据, 确定其在各种初婚型下的参数, 并设平均残差

$$\Delta = \sum_x [F(x; A) - M(x)]^2 / n$$

来比较各种模型。所得结果如表 1 所示(见下页)。

由表 1 可知, 这 36 个队列年龄别已婚率的调查值与模型拟合值的平均残差, 在寇尔的一般性初婚模型, 逻辑斯蒂对数模型, 极大值极值分布以及对数正态分布模型中都较小, 都在千分之五以下, 其中以寇尔模型为最小, 接近千分之四; 而赫内斯模型与伽马分布模型较大。由于实际的已婚率数据主要来自抽样调查, 对于已婚率这样一个比例指标, 平均误差若能控制到千分之五以下可以说是一个很好的结果了, 换个角度说, 上述前面的 4 个模型, 可以认为对中国人口的数据是适用的。

若仔细观察各个队列的平均残差, 我们发现上述结论下得似乎嫌早。因为, 从各个队列的平均残差来看, 以寇尔模型为例, 在 35 岁以上的队列, 其平均残差仅千分之三左右(一些平均残差大的队列, 大都是数据质量较差的), 而在 35 岁以下的一些队列中, 其平均残差在千分之七左右, 即是说, 年轻队列的平均残差要大大高于年长的队列。一般来说, 队列的年龄越小, 事件的发生(结婚)与报告(调查时间)相隔的时间越短, 报告发生的错误可能性则变小, 数据质量应该提高。相应于质量较高的数据模型误差反而增大, 这似乎可证明这些模型对年轻队列不太适用。

对实际数据的观察有助于我们对这一现象的解释。图 1 是 25, 30, 35, 40 岁四个队列的年龄别初婚率。从图形上看, 30, 35 岁队列的初婚率曲线非常出奇。一般来说, 初婚率的年龄分布曲线应该是单峰的, 但 35 岁队列的图形是双峰的; 而 30 岁队列的峰特别地高而窄, 按照通常的经验, 峰值一般在 0.15 左右, 而该队列的峰值在 23 岁却走到了 0.2 以上, 这就是说, 该队列几乎有百分之二十人都在同一年龄结婚。

上述两个队列出现的特殊图形, 当然不能排斥在其中包含有调查误差。但若联系我国社会情况, 就可知道这特殊情况的出现决不是偶然的。1988 年 35 岁的队列, 在 1973 年为 20 岁, 按照比她们年龄大的队列婚姻时间表来说, 20 岁以后婚率还应增大, 婚率峰值应在 21—22 岁, 但由于 70 年代以后大规模实施计划生育后, 许多地区政策上规定 23 岁为女性初婚起始年龄, 这样, 正常的结婚进程被打乱了, 许多原拟在 20 岁至 22 岁结婚人推迟到政策允许的年龄上才结婚, 于是出现了 21 岁, 22 岁初婚率下降, 23 岁初婚率再次提高, 因而形成了两个峰值。

表 1

各种初婚模型的平均残差(单位:10<sup>-3</sup>)

队列年龄	寇尔	逻辑斯蒂 对数	对数正态	极大值 极值	伽马	赫内斯	平均
25	9.1	9.8	9.6	10.1	10.1	9.1	9.6
26	5.4	6.3	6.0	6.8	6.0	6.1	6.1
27	3.5	4.5	4.4	5.2	4.0	4.9	4.4
28	4.5	5.3	5.5	6.2	5.5	4.7	5.3
29	5.0	5.2	5.6	6.6	6.5	4.0	5.5
30	7.0	7.0	7.1	8.3	7.5	5.5	7.1
31	7.1	7.6	7.7	8.5	7.8	6.6	7.6
32	7.6	8.9	8.6	9.1	7.5	8.4	8.4
33	6.5	8.3	8.0	8.0	6.7	8.9	7.7
34	4.6	7.4	7.2	6.1	7.4	9.4	7.0
35	3.4	5.7	5.2	3.9	8.4	8.9	5.9
36	3.6	6.1	5.2	4.1	6.9	9.3	5.9
37	2.3	3.9	2.9	2.3	8.7	8.5	4.8
38	5.6	3.1	3.8	5.7	12.1	6.2	6.1
39	2.0	3.9	3.6	3.1	2.5	5.2	3.4
40	2.1	2.9	3.2	2.1	5.4	6.5	3.7
41	3.7	3.5	4.1	3.8	8.7	7.6	5.2
42	3.1	5.4	4.6	3.5	4.4	8.9	5.0
43	2.6	2.8	2.3	2.7	6.9	8.1	4.2
44	3.9	2.1	2.8	4.1	7.4	6.3	4.4
45	2.3	3.3	3.1	2.2	2.2	5.9	3.2
46	5.3	3.2	4.5	5.5	9.0	5.8	5.6
47	1.6	4.0	3.4	2.0	3.7	7.8	3.8
48	2.2	4.3	3.4	2.2	4.4	8.6	4.2
49	4.9	2.5	3.4	5.1	10.1	7.8	5.6
50	2.8	1.8	2.2	2.9	7.5	7.8	4.2
51	1.9	3.8	3.3	1.8	3.0	7.7	3.6
52	4.8	2.8	3.9	4.9	9.8	8.1	5.7
53	3.2	3.7	4.3	3.4	8.8	9.6	5.5
54	2.8	5.8	5.2	2.8	5.1	10.7	5.4
55	2.8	2.7	3.1	3.0	8.5	9.7	5.0
56	5.7	3.0	4.3	5.9	10.4	8.3	6.3
57	2.2	2.4	2.8	2.4	7.0	8.5	4.2
58	4.9	2.8	4.0	5.0	10.0	7.6	5.7
59	2.3	3.2	3.4	2.4	5.0	6.5	3.8
60	2.1	1.5	2.0	2.2	3.7	4.5	2.7
25—60	4.0	4.5	4.5	4.6	6.9	7.4	4.6
15—41	4.9	5.8	5.7	5.9	7.2	7.0	5.2
42—60	3.2	3.2	3.5	3.4	6.7	7.8	4.0

注 1: 平均残差 =  $\sqrt{\sum_x [F(x) - M(x)]^2 / n}$

这里:  $M(x)$  为到  $x$  岁年龄已婚率的调查值

$F(x)$  为到  $x$  岁年龄已婚率的模型值

$n$  为年龄组数。

注 2: 这里初婚起始年龄为 14 岁。

资料来源: 由千分之二妇女生育节育调查资料计算

1988 年 30 岁队列的婚姻又是另外一种情况。该队列在 23 岁以前, 是在严格的晚婚政策下进行的, 但传

统上婚烟习俗并未多大改变,不过是在政策限制下原来拟在 23 岁以前结婚的人推迟到 23 岁结婚,而 1981 年开始贯彻实施新婚姻法,新婚姻法把 20 岁规定为婚姻起始年龄,这样,大批 23 岁以前的人进入已婚队伍,对于 1988 年 30 岁队列的人来说,原来是因为把 23 岁作为初婚起始年龄,所以 24 岁,25 岁结婚亦不算晚,但现在大批的比 23 岁年龄小的人结婚,使她们感到了社会压力,于是她们也更主动地进入婚姻角色,即后面队列早婚者把她们拉入了已婚队伍,这“推”和“拉”两种力量,使该队列在 23 岁严重堆积,造成了在现有人口各队列中从未有过的初婚率高值。

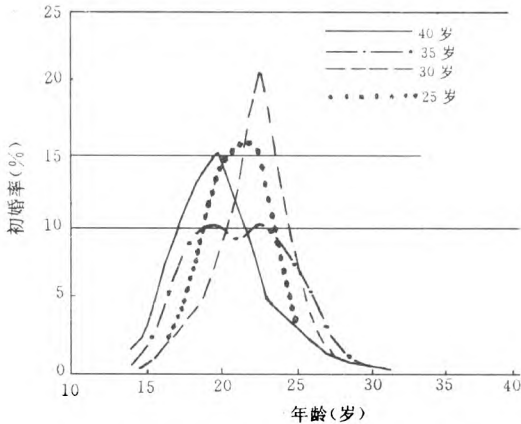


图 1 25 岁、30 岁、35 岁、40 岁队列的年龄别初婚率

资料来源:根据 1988 年中国千分之二人口生育节育抽样调查结果作出。

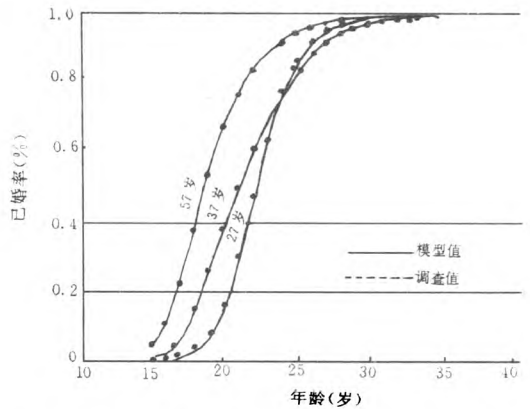


图 2 27 岁、37 岁、57 岁队列各年龄已婚率的调查值与模型值比较

注:27 岁、37 岁、57 岁的寇尔模型参数分别为:

$$27 \text{ 岁: } \theta = 9.6797 \quad \lambda = 0.3234 \quad \alpha = 0.6405$$

$$37 \text{ 岁: } \theta = 5.9753 \quad \lambda = 0.3340 \quad \alpha = 0.3384$$

$$57 \text{ 岁: } \theta = 3.9084 \quad \lambda = 0.4226 \quad \alpha = 0.4236$$

人们以什么年龄作为最佳初婚年龄,这种观念是在社会习俗影响下形成的,当然,时代在发生变化,社会习俗也在发生变化,但这种变化是渐进的,缓慢的。当代大多数国家,都把婚姻看作是个人行为,除了从生理健康上加以一些限制外,一般不直接干预。人口学家在设计婚姻数理模式时,实际上是把婚姻进程看作恒定的或者变动较小的社会习俗力量对人们婚姻行为作用的结果,所以它的婚率变化是稳定的,均匀的。我国从 70 年代开展计划生育以后,把实施晚婚看作是控制人口增长的一种重要手段,对人们的婚姻进行了行政干预,而这种干预又是突然而来的,且时强时弱,因时而异,因地而异,因人而异。当人们还未根据新的社会条件调整好自己的婚姻行为时,新的社会作用力又出现,在这种场合下,人们的婚姻进程自然就成为无序的,当然,这种普通的婚姻模式也就不能准确记述婚姻变化的进程了。

年轻的一代正在按新婚姻法的规定调整自己的婚姻进程。从图 2 我们可以看到,27 岁的队列除了在 20 岁以前调查值与实际值有较大差别后,在 20 岁以后是相当接近的。这一迹象表明:只要我们的政策保持稳定,上述的一些已婚模式还是能适用于我国人口的。25 岁队列的平均残差较大,可能和她们相当部分尚未结婚,而模型值和实际值在初婚起始年龄上往往差别较大的缘故。

### 三、应用初婚模型的几点体会

社会现象,从根本上说是由社会自身运动所决定的。描述社会现象的数学模型,总是假定在一定的社会条件下才成立,如果社会条件发生了变动,该模型就可能发生改变甚至不成立。从这个意义上说,在稳定的社会条件,均质的人口(即包含在该人口中的每个人有共同的社会性,如宗教信仰,社会习俗,收入、教育等等)中,模型使用的效果往往好一些,反之则效果差一些。我们有时常常诧异一些国外常用的人口模型用到中国人口数据上效果就不太理想,实际上就是由于模型的效果取决于它依赖的社会条件的缘故。这样,我们在判断我国人口数据的可靠性时,首先应该联系我国的社会实际,探讨这些数据是否有合理的一面。以上述初婚数据为例,从模型角度来看,30—35 岁的数据很不正常,因为它的初婚曲线出现了双峰以及峰值太大;而联系

到社会背景来考察,这异常的数据正反映了我国当时的现实,所以这些数据是基本合理的,若简单地以模型值来对这些队列初婚数据修匀、调整,则调整后的数据反而会与现实发生的婚姻情况大相径庭,数据修正的结果反而弄巧成拙了。

当然,我们这样说并不是否定模型在人口分析中的作用。如上所述,在社会变化相对稳定,人口的社会性质相对一致的社会条件下,初婚模型能很好地反映人口的初婚进程。如我们看到的,我国 35 岁以上的妇女的初婚数据和模型拟合值就能较好地一致。由于实际调查得到的初婚数据存在抽样误差和非抽样误差,利用模型就能对这些数据进行调整和修匀。事实上,在 1988 年的抽样调查结果中,我们可以发现许多问题。以 39 岁这一队列为例,它在 35 岁时的已婚率(1988 年调查结果)就已经高于它在 41 岁的已婚率(1990 年普查结果),对同一队列,已婚率随年龄增加只会增大,所以可认为该队列的初婚调查结果存在着一些问题。类似的情况在 42 岁,45 岁,47 岁,51 岁,52 岁,54 岁等队列都存在,对这些队列数据的修正,正是模型可以发挥作用的地方。当然,建立初婚模型的意义还不仅在数据修正方面。

另外,从各种初婚模型使用上说,应注意以下几点:

1、从模型精度上看,寇尔的一般已婚率模型要大于极大值极值分布模型。这是可以想像的。因为极大值极值分布是寇尔模型在  $\lambda = \alpha$  时的特例,寇尔模型比极大值分布多一个参数,即多一个自由度,模型在拟合时要更灵活一些,所以精度也会高一些。但如果比较这两个模型精度的差别可以发现它们是极小的,仅是在万分位上。由于极大值分布模型的参数较少,参数减少不但意味着计算量的减小,而且使模型参数的意义更加明确。所以,在大部分场合,用极大值极值分布来替代寇尔的一般初婚模型是适用的。

2、在赫内斯模型中, $F(0)$ 是当作常数看待的,常常被预先规定为一个很小的值(本文在拟合时规定  $F(0) = 10^{-5}$ ),但  $F(0)$ 本身也可认为在不同队列中是变化的,即也可以把  $F(0)$ 当作一个参数来对待。如果把  $F(0)$ 当作一个参数,则拟合的精度自然比表 1 中精度有所提高,与此同时,也增加了计算的难度。

3、在对逻辑斯蒂对数模型和极大值极值分布的参数估计时,通过适当的形式变换,也可化成线性规划的形式来求解,这样可使解法变得非常简单。但经过变换后的形式求出的解往往不是与实际已婚率距离最小的最优解。而在实际求解非线性规划问题时,初值的选择往往是很关键的。在实际中求解这两个模型参数时,可以先把它们转变为线性规划问题来求解,把解出的参数作为非线性规划解的初值,然后再求出相应的非线性规划的解。

4、利用逻辑斯蒂对数模型和极大值分布的性质,可以推得任意两个同期群的年龄别已婚率的函数有线性关系。这一结果大大丰富了寇尔最早提出的初婚模型的内容,寇尔的方法是建立标准的已婚率曲线和其它已婚率发生联系,而在这里,一个数据质量较好的已婚率数据,都可以作为标准数据来使用,这将为实际应用带来许多方便。

#### 参考文献:

- 1 Coale. A. J. Age pattern of marriage Population Studies, 1971, Vol 25, No. 2.
- 2 Coale. A. J and D. R. McNeil. The distribution by age of the frequency of first marriage in a female cohort. Journal American Stastical Association, 1972, Vol. 67, No. 4.
- 3 Hernes, G. The process of entry into first marriage. American Sociological Review, 1972 Vol. 37, No. 1.
- 4 Keeley. M. C, An analysis of the age pattern of first marriage. International Economic Review, 1979 Vol. 20, No. 2.
- 5 Sorensen A. and A. B. Sorensen. An event history analysis of the process of entry into first marriage. Current Perspectives on Aging and Life Cycle, 1986. Vol. 2.
- 6 大谷宪司. 关于包括结婚到妊娠过程的几个时间分布. 人口问题研究(日本), 1990, 第 45 卷, No. 4.
- 7 黄荣清、元昕. 中国妇女的初婚与初育模型. 中国人口科学, 1990, 第 4 期.
- 8 刘德贵等. Fortran 算法汇编. 国防工业出版社, 1983.