

# 模拟茧丝纤度曲线的统计特征<sup>\*</sup>

白 伦 谢 佳 李林甫<sup>\*\*</sup> 周 韶 王建民

(苏州大学 苏州, 215006)

**摘 要** 对于茧丝纤度曲线的阶梯式自回归模型,讨论了用该模型生成的茧丝纤度统计特征及自回归模型次数选择等问题。通过分析模型的茧丝纤度平均值误差,给出纤度平均值误差修正方法。此外给出根据粒内偏差选择自回归次数时的试验误差实验结果。

**关键词:** 茧丝纤度曲线,自回归次数,模拟实验,误差修正。

**中图法分类号:** TS142.22

前报(白伦等,1997)中,我们给出了一种以茧丝回归直线为准线的阶梯式自回归模型,用以描述茧丝纤度曲线这种具有有限长度的非平稳时间序列,并给出该模型中自回归系数的最小二乘估计以及残差方差的最优值。按这个自回归模型开发出茧丝纤度曲线模拟生成程序并通过模拟试验讨论了该生成方法的实用性

在对用这种模拟方法生成的茧丝纤度曲线的特征进行统计分析时,我们注意到茧丝纤度曲线的平均值总有偏小的倾向。此外前报中给出的模拟结果可以看到,当阶梯式自回归模型的次数不同时,生成茧丝纤度的粒内偏差值渐次发生变化。为了得到与原料茧庄口特征相符的模拟茧丝纤度曲线,在本文中我们讨论用阶梯模型生成的茧丝纤度特征数及自回归模型次数选择等有关问题。通过分析模型的茧丝纤度平均值误差,给出误差修正方法。此外为了根据粒内偏差的变化选择合适的自回归次数并使选择的结果有较高的可信赖度,我们调查了模拟生成的茧丝纤度的偏差的波动状况,以给出简便的选定自回归次数的方法。

在本研究中,我们继续利用前文中由山崎等(1983)提供的中国、日本、巴拉圭、保加利亚原料茧纤度基础调查数据。并对山崎教授提供这些数据表示谢意。

根据前报的讨论,从调查到的茧丝纤度样本数据,可以推定原料庄口的茧丝纤度主要特征数,如茧丝纤度平均值 $\bar{f}$ 、纤度偏差 $\hat{\epsilon}$ 、粒内偏差 $\hat{\epsilon}_n$ 、粒间偏差 $\hat{\epsilon}_s$ 、初始纤度平均值 $\hat{f}_0$ 与偏差 $\hat{\epsilon}_0$ 、茧丝长 $L$ 的平均值 $\bar{L}$ 与偏差 $\hat{\epsilon}_L$ 等。对样本茧丝纤度曲线可以作出回归直线

\* 本课题为三项费用研究项目

本文于 1998 年 4 月 6 日收到

\*\* 丝绸工程 95 级硕士研究生

白 伦,男 1948 年出生,教授

$$\hat{x}_i = \bar{x} + U(i - \frac{L}{2}) \quad (1)$$

并推定出回归直线斜率平均值  $\hat{U}$  及偏差  $\hat{Q}_U$  以回归直线为基准, 将茧丝纤度序列  $\{x_i\}$  变换为在回归直线上下波动的准纤度序列  $\{z_i\}$ , 即

$$z_i = x_i - \hat{x}_i \quad (i = 0, 1, \dots, L) \quad (2)$$

则准茧丝纤度序列利用以下的阶梯式自回归模型进行预测

在给出茧丝的准初始纤度  $z_0$  及茧丝长  $L$  时, 对于一定值  $k$ , 当  $i \leq k$  时,

$$z_i = \sum_{r=1}^i a_r z_{i-r} + X \quad (i = 1, 2, \dots, k) \quad (3)$$

当  $i > k$  时,

$$z_i = \sum_{r=1}^k a_r z_{i-r} + X \quad (i = k+1, k+2, \dots, L) \quad (4)$$

式中,  $X$  为白色噪音。自回归系数  $\{a_i\}$  按最小二乘估计在使误差  $X$  的方差  $\sigma^2$  最小的条件下求出。由此预测得到的准茧丝纤度序列与回归直线上相应的值相加而获得茧丝纤度序列。

## 1 生成茧丝纤度的平均值

考虑由上述阶梯式自回归模型生成的准茧丝纤度序列的平均值的期望值 因  $X$  的期望值为 0, 故有

$$\begin{aligned} E\left(\sum_{i=1}^L z_i\right) &= E\left[\sum_{i=1}^k \sum_{r=1}^i a_r z_{i-r} + \sum_{i=k+1}^L \sum_{r=1}^k a_r z_{i-r}\right] \\ &= E\left[\sum_{r=1}^k a_r \sum_{i=0}^{L-r} z_i\right] \\ &= E\left[\sum_{r=1}^k a_r \sum_{i=0}^L z_i - \sum_{r=1}^k a_r \sum_{i=L-r+1}^L z_i\right] \end{aligned}$$

记茧丝纤度平均值为

$$\bar{x} = \frac{1}{L} \sum_{i=0}^L z_i$$

即由上式得到

$$E(\bar{x}) - \frac{1}{L} E(z_0) = \sum_{r=1}^k a_r E(\bar{x}) - \frac{1}{L} \sum_{r=1}^k a_r E\left(\sum_{i=L-r+1}^L z_i\right)$$

若此式中最后的和式内的  $z_i$  均用其平均值  $\bar{x}$  来代替时, 近似地有

$$E(\bar{x}) = \frac{E(z_0)}{L - \sum_{r=1}^k a_r} \quad (5)$$

式中  $E(z_0)$  为准茧丝纤度的初始均值 因准茧丝纤度序列是以回归直线为基准变换而得到的, 故其平均值应该为 0 因此, (5) 式给出了茧丝纤度平均值的偏移量平均值。

为了了解利用这个偏移量对茧丝纤度平均值进行修正的效果, 我们对各个不同原料茧庄口, 按不同次数的自回归模型模拟生成的茧丝, 计算出其平均茧丝纤度。再根据 (5) 式修正后与原庄口的纤度平均值进行比较。表 1 给出中国秋茧等共 6 个庄口的原料茧在自回归次数分别为 2, 5, 8 时所生成的茧丝平均纤度及其修正的值 在这里, 各生成茧丝平均纤度均由自回归模型生成的 500 根茧丝求得。

表 1 茧丝平均值的偏差修正及模拟试验结果

自回归次数 $k$		中 国 秋 茧	巴拉圭 原料茧	保加利亚 原料茧	日 本 春 茧	日 本 初秋茧	日 本 晚秋茧
$k = 2$	修正量	0.08280	0.08063	0.10600	0.04045	0.07631	0.06456
	修正前	2.235	2.353	2.273	3.030	2.491	2.636
	修正后	2.318	2.434	2.379	3.071	2.567	2.701
$k = 5$	修正量	0.04413	0.04836	0.05502	0.02266	0.03197	0.02454
	修正前	2.267	2.383	2.296	3.070	2.514	2.627
	修正后	2.311	2.431	2.351	3.092	2.546	2.652
$k = 8$	修正量	0.02784	0.02877	0.03730	0.01669	0.01943	0.01549
	修正前	2.260	2.402	2.323	3.055	2.510	2.666
	修正后	2.288	2.431	2.360	3.072	2.529	2.681
原庄口值		2.323	2.451	2.399	3.113	2.570	2.704

注: 实验值为 500 根生成茧丝的平均纤度。

从表 1 可以看到,在生成的茧丝纤度不加修正时,平均纤度确实存在系统的偏移。如果不加修正,直接利用这样生成的茧丝进行模拟缫丝时,模拟生成的统计特性将受到平均纤度偏移误差的影响而使精度下降。表中的结果表明,通过利用(5)式进行修正,尽管各庄口的纤度特性的差异颇大,但在用不同次数的自回归模型而获得的平均纤度都与原庄口茧丝平均纤度更为一致了。虽然与原庄口平均值相比,修正后的平均值仍稍有偏小的倾向,但茧丝平均纤度模拟值的系统偏移状况得到了明显的改善。

2 茧丝粒间偏差

因为用阶梯式自回归模型生成茧丝纤度曲线时,以回归直线方程(1)式作为基准,故生成的茧丝纤度曲线的中心应该与回归直线的中心一致。事实上,因生成茧丝纤度曲线时,是先作出回归直线后再生成准茧丝纤度值的,故由于生成过程中须加入随机误差项等原因,二者的中心位置间存在一定的误差是不可避免的。但在大量的茧丝纤度曲线生成中,这些误差对于总的平均状态影响是不大。这一点,我们可以通过模拟曲线的生成实验来确认。由(1)式可知,回归直线的中心即是各茧丝纤度的平均纤度,而该值是依靠生成具有庄口粒间偏差的正态分布的方法获得的。因此生成的茧丝纤度曲线平均值之间的波动应与庄口茧丝粒间偏差一致。此外我们还注意到,因为回归直线的生成,并不受自回归次数大小的影响,故生成茧丝纤度的粒间偏差的大小与自回归模型的次数之间可以视为是无关系的。

表 2 给出供试的 6 个原料茧庄口的粒间偏差的试验结果。这些结果由模拟生成的 500 根茧丝的纤度值而计算得到。由表中可以看到由模拟试验生成的粒间偏差与各原料茧庄口的相应值相比,仍有一定程度偏小倾向,但几乎所有的模拟试验值对原庄口值的误差都在 10% 的范围内。并且模拟的粒间偏差随自回归模型的次数并无明显的规则的变化。因此可以认为,在实用上模拟茧丝生成方法获得的粒间偏差基本是符合要求的。

表 2 茧丝粒间偏差模拟试验结果

自回归次数 $k$		2	3	4	5	6	7	8
中国秋茧	实验值	0. 3262	0. 3472	0. 3378	0. 3505	0. 3524	0. 3524	0. 3586
	粒间偏差 原庄口值	0. 3527						
巴拉圭原料茧	实验值	0. 2524	0. 2613	0. 2688	0. 2601	0. 2763	0. 2755	0. 2641
	粒间偏差 原庄口值	0. 2783						
保加利亚原料茧	实验值	0. 3729	0. 4100	0. 3844	0. 3966	0. 3921	0. 3936	0. 3960
	粒间偏差 原庄口值	0. 4063						
日本春茧	实验值	0. 3374	0. 3276	0. 3381	0. 3219	0. 3353	0. 3352	0. 3426
	粒间偏差 原庄口值	0. 3454						
日本初秋茧	实验值	0. 3256	0. 3480	0. 3393	0. 3385	0. 3558	0. 3553	0. 3367
	粒间偏差 原庄口值	0. 3645						
日本晚秋茧	实验值	0. 3684	0. 3474	0. 3645	0. 3534	0. 3707	0. 3771	0. 3928
	粒间偏差 原庄口值	0. 3974						

注: 实验值为 500根生成茧丝的粒间偏差

3 粒内偏差与自回归次数的选择

生成的茧丝纤度曲线的粒内偏差主要由曲线的形状特征所影响。而由生成模型 (3) (4) 式可以看出,生成茧丝纤度曲线的形状特征主要由自回归系数  $\{a_i\}$  及随机误差  $X$ 所决定。当设定不同的自回归次数  $k$ 时,相应的自回归系数及随机误差的方差都发生变化,这就导至生成的茧丝纤度曲线的形状特征随之变化,从而使茧丝的粒内偏差发生变化。表 3为各供试庄口的粒内偏差及随机误差  $X$ 的均方差随自回归次数  $k$ 而变化的模拟试验结果。表中的实验值均为由生成 500根茧丝而计算得到。

表 3 误差均方差与茧丝粒内偏差模拟试验结果

自回归次数 $k$		2	3	4	5	6	7	8
中 国 秋 茧	误差均方差	0. 1220	0. 1184	0. 1161	0. 1135	0. 1120	0. 1097	0. 1081
	粒内 实 验 值	0. 4654	0. 5050	0. 5216	0. 5373	0. 5502	0. 5742	0. 5829
	偏差 原庄口值	0. 55						
巴拉圭原料茧	误差均方差	0. 1219	0. 1198	0. 1178	0. 1171	0. 1138	0. 1126	0. 1104
	粒内 实 验 值	0. 4431	0. 4781	0. 4980	0. 5048	0. 5184	0. 5371	0. 5403
	偏差 原庄口值	0. 5366						
保加利亚原料茧	误差均方差	0. 1322	0. 1286	0. 1239	0. 1205	0. 1185	0. 1170	0. 1154
	粒内 实 验 值	0. 5611	0. 5826	0. 5967	0. 6060	0. 6245	0. 6356	0. 6438
	偏差 原庄口值	0. 6648						
日 本 春 茧	误差均方差	0. 1657	0. 1633	0. 1618	0. 1597	0. 1582	0. 1576	0. 1572
	粒内 实 验 值	0. 6438	0. 6330	0. 6351	0. 6530	0. 6561	0. 6736	0. 7531
	偏差 原庄口值	0. 6637						
日 本 初秋茧	误差均方差	0. 1347	0. 1301	0. 1283	0. 1250	0. 1229	0. 1214	0. 1203
	粒内 实 验 值	0. 4758	0. 5138	0. 5556	0. 5745	0. 5994	0. 5997	0. 6016
	偏差 原庄口值	0. 5594						
日 本 晚秋茧	误差均方差	0. 1158	0. 1133	0. 1108	0. 1059	0. 1046	0. 1034	0. 1022
	粒内 实 验 值	0. 5505	0. 5741	0. 5917	0. 6261	0. 6475	0. 6875	0. 6977
	偏差 原庄口值	0. 5982						

注: 粒内偏差实验值为 500根生成茧丝的结果

由表 3 中各庄口的结果可以看到,各庄口的误差均方差均随着自回归次数的增大而渐次减少。这表明对于原庄口的调查数据来说,自回归模型的次数大一些时,有利于减少自回归模型对于原庄口数据的描述误差。但从表中粒内偏差的实验值一栏中可以看到,随着自回归次数的渐次增大,粒内偏差明显地呈递增的趋势。当自回归次数取得比较小时,生成的粒内偏差低于原庄口的粒内偏差,而随着自回归次数的增大,生成的粒内偏差逐渐达到原庄口值,而后又越过原庄口值而递增。这时,我们可以找到一个使粒内偏差最接近原庄口值的次数  $k$  值。为了直观地了解  $k$  在比较极端的情况下生成的蚕丝纤度曲线的形态,图 1 给出中国秋茧当自回归次数分别为  $k=2$  与  $k=8$  时模拟生成蚕丝纤度曲线的各一组结果。比较图 1 的这两组曲线可以看到,在蚕丝纤度曲线的模拟生成中,当  $k$  比较小(图中  $k=2$ )时,生成的曲线倾向于比较平坦。这时,尽管与原庄口曲线嵌配的相对差异(由误差方差表示)比较大,即生成曲线时必须加上较大的随机误差成分,但因曲线总体比较平坦故粒内偏差最终呈较低的值。而当  $k$  比较大(图中  $k=8$ )时,模拟生成的蚕丝曲线具有与原庄口曲线嵌配得较好的弯曲度,这使得嵌配误差(误差方差)比较小;尽管如此,这样的嵌配误差与较强的曲线弯曲度的叠加作用可能产生进一步强化弯曲效果,从而使得粒内纤度偏差增大以至超越原庄口值的限度。因此, $k$  过大时,反而会带来很高的模拟粒内偏差。通过表 3 的调查结果及以上的讨论得知,显然我们可以根据生成蚕丝粒内偏差与原庄口值的接近程度来选择自回归次数  $k$  值;二者最接近时的  $k$  值是最优的。这样选择  $k$  值的方法,电子计算机上,是非常方便地就能完成的。

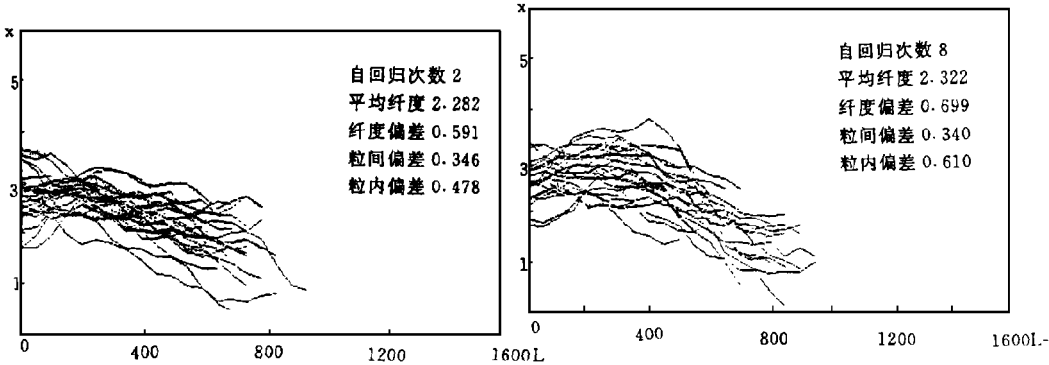


图 1 自回归次数  $k$  不同时的蚕丝纤度曲线

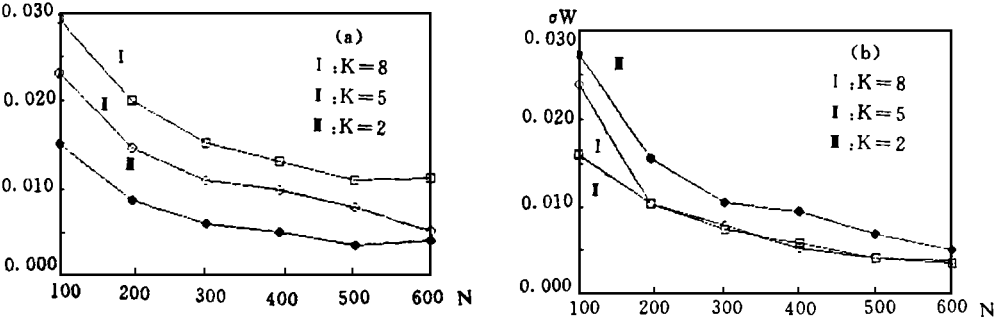


图 2 粒内偏差的均方差与蚕丝样本数

我们曾试图用非参数估计的  $AIC$  方法来选择最优的  $k$  值,但未能得到良好的结果。这种状况在  $AIC$  方法中因参数退化不良等原因(坂元等,1982)是可能出现的。目前我们正在继续探讨产生本课题中这种现象的原因。

## 4 选择自回归次数的样本数

按照前述用电子计算机来选择次数  $k$  值时,为了保证能高效率地选择到可靠的  $k$  值,还需要对生成蚕丝纤度曲线的样本数进行确定。生成的蚕丝纤度曲线根数太少时,由其计算出的粒内偏差的波动性较大,这使得由比较粒内偏差而获得的  $k$  值可靠性下降。另一方面,虽然电子计算机可以很快地生成蚕丝纤度曲线,但样本数取得太大时,显然是非能率的,也是不经济的。图 2 给出中国秋茧(a)及日本春茧(b)在生成蚕丝纤度曲线的样本数变化时,粒内纤度偏差的波动变化状况。图中横轴为生成的蚕丝纤度曲线样本数,纵轴为相应粒内偏差的均方差。对于一定的样本数量,计算粒内偏差的均方差时反复进行试验次数为 50 次。对于每一种原料茧分别就自回归次数  $k$  等于 2, 5, 8 三种情况进行了模拟试验并得到如图所示的 3 根曲线。从图中可以看到,两种原料茧的粒内偏差的均方差都随着样本数增大而减少,并在样本数达到 500 以上时,粒内偏差的均方差的减少渐趋平缓。即样本数再继续增加时,粒内偏差的均方差减少不再有明显的效果。这种情况对于  $k$  取各不同值都基本上是一致的。由此可以知道,我们在选择次数  $k$  时,样本数量达到 500 时,一般就可以获得稳定的结果。

## 5 蚕丝初始纤度与茧丝长

在模拟生成蚕丝纤度曲线时,茧丝的初始纤度及茧丝长都是独立地用正态分布随机数而获得的(白伦等,1997),故模拟的蚕丝初始纤度及茧丝长的性质,主要由随机数生成方法所决定。反复的模拟试验结果说明,本方法中,我们可以获得与原料茧庄口十分一致的蚕丝初始纤度与茧丝长模拟结果。试验数据在此不再罗列。

## 6 蚕丝纤度曲线的终点纤度

蚕丝纤度曲线的终点纤度是指自然落绪时的纤度,在这里我们把按照检查丝长调查茧丝纤度时最后一个纤度丝的纤度称为终点纤度。茧丝的终点纤度在本方法中是不能在生成过程中直接控制的,其性质主要由自回归递推方法所制约。表 4 给出对供试的 6 个原料茧庄口终点纤度平均值及偏差的模拟试验结果与原庄口值的对照。由表中的结果可以看到,无论是终点纤度平均值还是均方差,模拟试验结果与原庄口值相比都偏大。在利用本方法生成茧丝纤度曲线用以分析生成纤度偏差时,茧丝纤度偏差,粒间偏差与粒内偏差的影响是主要的,而终点纤度的误差影响是不大的。但用以分析生丝的匀度变化时,终点纤度统计特性的误差所产生的影响则是不应该忽视。这种情况下,须要考虑修正或者调整方法等措施。

## 7 结语

以上我们对于茧丝纤度曲线的阶梯式自回归模型,讨论了用该模型生成的茧丝纤度统计特征及自回归模型次数选择等问题.通过分析模型的茧丝纤度平均值误差,给出纤度平均值误差修正方法.此外根据模拟试验调查了粒内偏差模拟值的均方差随样本茧丝数而变化的情况,得知为了选择合适的自回归次数  $k$  而调查粒内偏差时,模拟样本数达到 500 根,一般就可以获得可靠的结果.最后,通过模拟调查还了解到,用本方法生成的茧丝纤度曲线的终点纤度平均值及均方差,与原庄口相应值相比都有偏大的倾向.

## 参考文献

- 1 白 伦, 謝 佳 等 (1997): シミュレーション  
実験による繭絲纖度曲線の生成, 製絲絹研  
究会誌, 第 6 卷, pp. 8-15
- 2 坂元慶行等 (1982): 情報量統計学, 共立出版株  
式会社, pp. 236
- 3 嶋崎昭典等 (1983): 異質原料繭の合併理論に  
関する研究, 昭和 57 年度科学研究費補助金  
研究成果報告書

## STATISTICAL CHARACTERS OF THE SIZE CRUVE OF COCOON FILAMENT BY COMPUTER SIMULATION

*Bai Lun, Xie Jia, Li Linpu, Zhou Zhao, Wang Jianmin*

( Suzhou University, Suzhou, 215006)

### Abstract

In this paper, we discussed statistical characters of the size curve of cocoon filament generated with the stage-autoregressive model by computer simulation and the problem of choosing rank of auto-regressive model. We gave a method to amend the error of average of cocoon-filaments size by simulation, and gave dates for investigation of the test error choosed rank of auto-regressive model through the size deviation between cocoon filaments.

**Key Words** Curve of cocoon filament, rank of auto-regressive model, computer simulation, Revision of averages error.