

模糊均生函数模型及其应用

魏凤英 曹鸿兴

(中国气象科学研究院,北京 100081)

提 要

将均生函数(MGF)概念推广到模糊集中,定义了模糊均生函数(FMGF),给出了相应的建模方案及实施步骤。淮河、长江流域夏季降水的计算实例表明,这一方案具有较高的长期预报能力。

关键词: 模糊均生函数 数学模型 降水预报

引 言

为了建立具有多步预测能力的数学模型,作者拓广了数理统计中算术平均值的概念,定义了时间序列的均值生成函数(mean generating function,缩写为MGF),简称均生函数,提出了视均生函数MGF为原序列生成的、具有周期性的基函数的新构想。在此基础上相继给出了几种适于不同类型序列的建模方案^[1-2]。这些建模方案的最大特点是,可以外推多步,并对极值有较好的拟合和预测,因此具有较强的实用价值。为了使这一建模方法更加完善,应用范围更加广泛,本文将均生函数概念推广到模糊集中,推导出不同类型序列的隶属度,并定义了模糊均生函数,给出了相应的建模方案及实施步骤。作为应用实例,文中给出了在江淮流域夏季降水预报中的使用情况及其与用均生函数建模方案的比较。

1 均生函数

设一时间序列

$$X(t) = \{X(1), X(2), \dots, X(N)\} \quad (1)$$

式中 N 为样本量。定义式(1)的均生函数

$$\bar{X}_l(i) = \frac{1}{n_l} \sum_{j=0}^{n_l-1} X(i+jl) \quad i=1, \dots, l \quad (2)$$
$$1 \leq l \leq M$$

其中 $n_l = \text{INT}(N/l)$, $M = \text{INT}(N/2)$ 或 $\text{INT}(N/3)$, INT 表示取整。对 $\bar{X}_l(i)$ 作周期性延拓

$$f_l(t) = \bar{X}_l(t - \text{INT}(\frac{t-1}{l})) \quad (3)$$

即可得到式(2)的外延序列。视 $f_l(t)$ 为基函数,便可以讨论时间序列 $X(t)$ 的建模问题了。

2 模糊均生函数

把均生函数的概念推广到模糊集中。设论域 $U = \{u_i | i=1, 2, \dots, N\}$, 在 U 上构造模糊子集 A

$$A = \mu_1/u_1 + \mu_2/u_2 + \dots + \mu_N/u_N \quad (4)$$

为叙述方便,记起报时刻为 t_N 。那么,要对未来时刻 $t_{N+1}, t_{N+2}, \dots, t_{N+q}$ 作出预报。从预报经验及物理意义上考虑,愈靠近起报时刻的观测值包含对预报有用的信息愈多,对预报愈有价值。在决定性预报或马尔柯夫过程中甚至认为,系统所处的状态与时刻 t_N 以前所处的状态无关,只有 t_N 时刻的观测值才对预报有用。也就是说,若序列具有马尔柯夫性,则可以把隶属函数定义为

$$A_M = 0/u_1 + 0/u_2 + \dots + 1/u_N \quad (5)$$

A_M 表示具有马尔柯夫性序列的隶属函数。

若从统计学观点出发,则把样品 $X(1)$,

$X(2), \dots, X(N)$ 等概率对待,用 A_s 表示统计学意义下的隶属函数,即有

$$A_s = 1/u_1 + 1/u_2 + \dots + 1/u_N \quad (6)$$

也就是隶属度均取1,即 $\mu(t) \equiv 1$ 。这是普通均生函数的做法,将每一样品视为等同的。

在实际问题中,我们既不忍心丢舍过多以往的信息,又想使近期观测值对预报发挥较大作用。为此,我们设计了随 t_N 的远近以指数形式下降的隶属度,即

$$\mu_A(t_i) = \begin{cases} e^{-\beta(u_N - t_i)} & t_i < t_N \\ 1 & t_i \geq t_N \end{cases} \quad (7)$$

当 $\beta=0$ 时式(7)蜕化为式(6)。若等间隔采样,令 $\Delta t=1, t_i=i\Delta t, t_N=N\Delta t$, 则(7)可写为

$$\mu_A(i) = \begin{cases} e^{-\beta(N-i)} & i < N \\ 1 & i \geq N \end{cases} \quad (8)$$

式中 β 按对过去观测值重视程度事先给定。显然, $0 \leq \mu_A(i) \leq 1$ 。

若序列(1)具有周期性,则可令隶属度为

$$\mu_A(i) = \begin{cases} r \sin \frac{2\pi}{l}(N-i) & i < N \\ 1 & i \geq N \end{cases} \quad (9)$$

其中 l 为周期长度。 r 为由经验或试算确定的常数。显然,当 $r \leq 1$ 时有 $0 \leq \mu_A(i) \leq 1$ 。

若既考虑观测值随起报时刻远近效用逐渐下降又体现周期性,则令隶属度为

$$\mu_A(i) = \begin{cases} r e^{-\beta(N-i)} \sin \frac{2\pi}{l}(N-i) & i < N \\ 1 & i \geq N \end{cases} \quad (10)$$

基于 μ_A 构造模糊向量

$$a = (\mu_A(1), \mu_A(2), \dots, \mu_A(N))$$

在论域 U 上构造另一个模糊子集 B , 它的隶属函数取为

$$\mu_B(i) = X(i)/X_{\max}$$

式中 $X_{\max} = \max X(i)$, 显然 $0 \leq \mu_B(i) \leq 1$ 。构造模糊向量

$$b = (\mu_B(1), \mu_B(2), \dots, \mu_B(N))$$

用代数加和乘定义模糊向量的内积^[3]

$$a \cdot b = \frac{1}{N} \sum_{i=1}^N \mu_A(i) \mu_B(i) \quad (11)$$

根据式(11)定义模糊均生函数(FMGF)

$$X_l(i) = \frac{C}{n_i} \sum_{j=0}^{n_i-1} \mu_A(i+j) \mu_B(i+j) \quad (12)$$

式中 C 为给定的常数,使得 FMGF 与序列 $X(t)$ 的量级相同。方便地可取 $C = X_{\max}$, 式(12)就变为

$$\bar{X}_l(i) = \frac{1}{n_i} \sum_{j=0}^{n_i-1} \mu_A(i+j) X(i+j) \quad i = 1, 2, \dots, l \quad (13)$$

同样用式(3)作周期性延拓,得到外延序列。

下面是一个具体构造模糊均生函数外延序列的例子。资料取长江中下游1951—1990年6—8月降水量,这里 $N=40, M=20$ 。据式(13)和式(3)可以得到模糊均生函数外延矩阵 $F = (f_{ij})_{40 \times 20}$ 。图1为周期长度 l 为2、3、4、5年的模糊均生函数外延序列变化曲线。

3. 建模方案及实施步骤

设一原始序列

$$X^{(0)}(t) = \{X^{(0)}(1), X^{(0)}(2), \dots, X^{(0)}(N)\}$$

建立其预测模型的方案及实施步骤如下:

3.1 对 $X^{(0)}(t)$ 作一、二阶差分运算,得到

$$X^{(1)}(t) = \{\Delta X^{(1)}(1), \Delta X^{(1)}(2), \dots, \Delta X^{(1)}(N)\}$$

$$X^{(2)}(t) =$$

$$\{\Delta X^{(2)}(1), \Delta X^{(2)}(2), \dots, \Delta X^{(2)}(N)\}$$

这一过程实际起到高通滤波作用。

3.2 对序列 $X^{(0)}(t), X^{(1)}(t)$ 和 $X^{(2)}(t)$ 分别用式(13)求出模糊均生函数并作周期性延拓。这里 μ_A 选用式(8)计算,通过试算取 $\beta=0.01$ 。得到三组模糊均生函数延拓序列 $f_i^{(0)}(t), f_i^{(1)}(t), f_i^{(2)}(t), i=1, 2, \dots, M, M = \text{INT}(N/2)$ 。

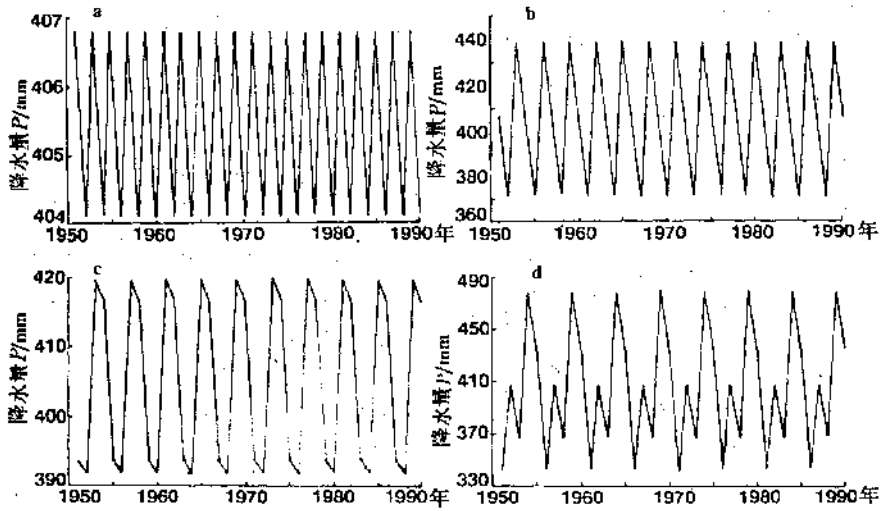


图1 模糊均生函数外延序列

a. $l=2$ b. $l=3$ c. $l=4$ d. $l=5$

3.3 构造一组累加延拓序列,即

$$f_l^{(3)}(t) = X^{(0)}(1) + \sum_{i=1}^{t-1} f_l^{(1)}(i+1) \quad (14)$$

$t = 2, 3, \dots, N$
 $l = 1, 2, \dots, M$

其中 $f_l^{(3)}(1) = X^{(0)}(1)$

3.4 引用双评分准则对上述生成的 $l \times M$ 个延拓序列逐一筛选。双评分准则兼顾数量评分和趋势评分。记 S_1 为检验预报效果的数量评分, S_2 为趋势评分, 则双评分准则定义为:

$$CSC = S_1 + S_2 \quad (15)$$

其中定义数量评分为

$$S_1 = (N - K) \left(1 - \frac{Q_K}{Q_Y}\right) \quad (16)$$

这里 K 为统计模型中因子个数, Q_K 为残差平方和

$$Q_K = \sum_{t=1}^N (X(t) - \hat{X}(t))^2$$

$\hat{X}(t)$ 为 $X(t)$ 的估计值。令

$$Q_Y = \sum_{t=1}^N (X(t) - \bar{X})^2$$

\bar{X} 为 $X(t)$ 的平均值。

趋势评分取最小判别信息统计量

$$S_2 = 2I = 2 \left[\sum_{i=1}^G \sum_{j=1}^G n_{ij} \ln n_{ij} + N \ln N - \left(\sum_{i=1}^G n_i \ln n_i + \sum_{j=1}^G n_j \ln n_j \right) \right] \quad (17)$$

其中 G 为趋势类别数, n_{ij} 为 i 类事件与 j 类估计事件列联表中的个数。

假定 S_1 与 S_2 独立, 当样本量足够大时可求得 CSC 为渐近 χ^2 分布^[2,4], 故可用 χ^2 检验来对预报因子进行初步筛选。当

$$CSC(X) > \chi_{\alpha}^2$$

时, 因子 X 入选。式中 $\nu = k + (G-1)(G-1)$, α 取 0.05, 0.01 或 0.001。

3.5 由(4)步粗选得的因子, 再进行精选, 显然用所有可能子集回归是最好的, 但计算量很大, 因此这里采用一个简便方法。按 CSC 值由大到小, 将粗选出的模糊均生函数采用前向筛选逐一进入回归方程。当 CSC 出现极大值时, 停止筛选。假设引入 K_0 个模糊均生函数时 CSC 值为极大, 那么预报方程为

$$\hat{X}(t) = a_0 + \sum_{i=1}^{K_0} a_i f_i(t) \quad (18)$$

4 江淮流域汛期降水预测模型

我们分别用长江、淮河两流域 1951—

1990年6—8月降水量建立基于模糊均生函数的预测模型,对1991年降水量作了试报,同时与用一般均生函数建模的结果进行了比较。为了进一步验证模糊均生函数的预测能力,又用1951—1986年资料建模,试报1987—1991年降水量。

4.1 淮河流域降水模型

淮河流域选用新浦、清江、徐州、蚌埠、阜阳、南阳、信阳、东台、邳县9站平均6—8月降水量。 $N=40$,周期最大长度取 $M=INT[40/3]=13$ 。表1为淮河流域预测模型的回归系数及引入的模糊均生函数。

表1

| 序号 | 系数 | FMGF |
|-------|---------|-------------------|
| a_0 | -1312.7 | |
| a_1 | 1.4804 | $f_{12}^{(3)}(t)$ |
| a_2 | 1.0518 | $f_{11}^{(3)}(t)$ |
| a_3 | 3.4347 | $f_9^{(3)}(t)$ |
| a_4 | 0.1426 | $f_{12}^{(1)}(t)$ |
| a_5 | -2.3063 | $f_9^{(0)}(t)$ |
| a_6 | -0.3187 | $f_7^{(1)}(t)$ |
| a_7 | 0.4737 | $f_7^{(0)}(t)$ |

表中 $f_{12}^{(3)}(t)$ 表示周期长度为12年的模糊均生函数累加延拓序列, $f_9^{(0)}(t)$ 为原序列周期长度为9年的模糊均生函数外延序列, $f_7^{(1)}(t)$ 为一阶差分序列周期长度为7年的模糊均生函数外延序列,余类推。

图2为所建模型的拟合和实况值。从图中

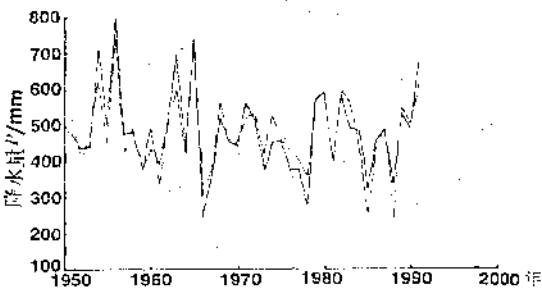


图2 淮河流域1951—1991年6—8月降水量变化
实线为实况,虚线为拟合,1991年为试报

可以看出,模型将历史的旱涝趋势非常好的拟合出来了,尤其淮河流域明显大水年,像1954、1956、1963、1965、1980年等都拟合得相当满意。拟合均方根误差为45.6mm。

1991年淮河流域6—8月降水量为674mm,是自1965年以来降水量最大的一年,比多年平均高出40%,模糊均生函数模型预报值为615mm,效果令人满意。

用相同的资料建立一般均生函数模型,模型由5个均生函数组成,拟合均方根误差为73.06mm。1991年试报值为531mm,比模糊均生函数模型的误差要大。

表2是用1951—1986年资料建立模糊均生函数模型试报1987—1991年的结果及与实况的比较。从表中看出,5年中只有1989年的预报趋势与实际多水的趋势相反,其余4年预报效果尚好,像1988年的大旱、1991年的大水趋势均能预报出来。

表2

| 年份 | 预报值/mm | 实况值/mm |
|------|--------|--------|
| 1987 | 448 | 489 |
| 1988 | 213 | 334 |
| 1989 | 447 | 535 |
| 1990 | 469 | 491 |
| 1991 | 568 | 674 |

4.2 长江中下游降水模型

长江中下游降水量资料是选取南京、合肥、上海、杭州、安庆、屯溪、九江、汉口、钟祥、岳阳、宜昌、常德、宁波、衢县、贵溪、南昌、长沙17站6—8月平均降水量。

表3是预测模型的回归系数及所对应的模糊均生函数外延序列。

模型对长江中下游夏季降水有非常理想的拟合(见图3)。尤其像1954、1969、1980及1983年的大水都能很准确地拟合出来。拟合均方根误差为65.7mm。

外推一步作出1991年降水量的预报值为575mm,实际观测值为552mm,仅差23mm。

而用相同资料建立的一般均生函数模型1991年的试报值为437mm,与实况误差较大。

表3

| 序号 | 系数 | FMGF |
|----------|---------|-------------------|
| a_0 | -1380.4 | |
| a_1 | 1.5519 | $f_{13}^{(3)}(t)$ |
| a_2 | -0.9285 | $f_{12}^{(2)}(t)$ |
| a_3 | -0.1036 | $f_{13}^{(1)}(t)$ |
| a_4 | 2.3666 | $f_7^{(2)}(t)$ |
| a_5 | -2.1092 | $f_{10}^{(2)}(t)$ |
| a_6 | 2.9361 | $f_{10}^{(3)}(t)$ |
| a_7 | 0.3952 | $f_3^{(2)}(t)$ |
| a_8 | 0.2994 | $f_6^{(1)}(t)$ |
| a_9 | 0.9211 | $f_{11}^{(2)}(t)$ |
| a_{10} | -1.8396 | $f_7^{(3)}(t)$ |

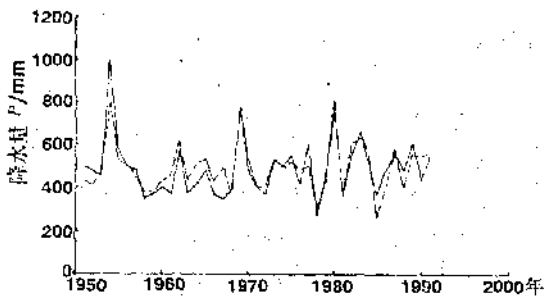


图3 长江中下游1951—1991年6—8月降水量变化曲线

1991年的虚线值为预报

表4是用1951—1986年资料建立模糊均生函数模型试报1987—1991年的结果与实况对照。

表4

| 年份 | 预报值/mm | 实况值/mm |
|------|--------|--------|
| 1987 | 540 | 558 |
| 1988 | 304 | 480 |
| 1989 | 490 | 606 |
| 1990 | 611 | 433 |
| 1991 | 526 | 552 |

其中1987和1991年预报与实况最接近,1989年效果尚可,1988和1990年预报与实况误差较大。

参考文献

- 曹鸿兴、魏凤英,基于均值生成函数的时间序列分析,数值计算与计算机应用,1991,12(12):82—89.
- 魏凤英、曹鸿兴,长期预测的数学模型及其应用,北京:气象出版社,1990:49—82.
- 汪培庄,模糊集合论及其应用,上海科学技术出版社,1983:86—97.
- Kullback, S., Information Theory and Statistics. Dover Publications, Inc., New York, 113—119.

A Fuzzy Mean Generating Function(FMGF)Model and Its Application

Wei Fengying Cao Hongxing

(Chinese Academy of Meteorological Sciences)

Abstract

Generalizing mean generating function(MGF) into fuzzy set, fuzzy mean generating function(FMGF) is defined and a modelling scheme and its calculating procedure are suggested. Experiments of summer precipitation along the reaches of the Huaihe and the Changjiang Rivers result in that the scheme is suitable for longterm prediction with high accuracy.

Key Words: fuzzy mean generating function modelling scheme precipitation prediction