第7卷 第2期　　水 科 学 进 展　　Vol.7, No.2
1996年6月　　ADVANCES IN WATER SCIENCE　　Jun., 1996

模糊水文统计分析研究*

黄伟军　丁　晶

（四川联合大学水利系　成都 610065）

摘　要　水文统计分析的观测数据和调查数据往往不够精确，为此，提出在模糊样本条件下进行水文统计分析的方法体系，用模糊数进行模拟，并以范例作了简要说明。

关键词　水文统计　模糊数　模糊样本　模糊随机过程　模糊贝叶斯分析

分类号　P333.6

1 引　　言

水文统计分析结果的可靠性往往取决于所拥有的信息，经典统计学常假设采样数据是精确的，分析者也可准确地理解所有的试验结果。然而事实非并如此，在采样数据中，除了经典的统计变化（随机性）外，还涉及到模糊性这一非概率的不确定性。这种信息的模糊性主要来源于试验观测、模型解和调查结果等途径的欠精确性，如洪水频率分析中样本的欠精确性就特别明显。因此，在进行统计分析时应当把随机性和模糊性这两类不确定性耦合起来，以进一步改善水文分析与计算的成果。

Kwakernaak[1]最先提出模糊随机变量的概念，即以模糊数而非实数表示的随机变量，同时他还讨论了模糊随机变量的期望、条件期望以及独立模糊随机变量的基本性质。Kruse[2]和Miyakoshi 等[3]研究了模糊随机变量的强大数定律。而 Stein 等[4]则根据 Nahmias[5]的模糊变量的概念对凸模糊随机变量作了探讨。模糊随机变量这一概念具有重要的价值，能将系统中由欠精确测量或量化上固有含糊性而不确知的元素与因随机变化而不完全知道的元素区别开来。在传统的应用中，前一种不确知（定）性通常表示为清晰量，用敏感性分析研究，或者用随机变化的量表示。但这两种表示都不很恰当。相反，模糊集对此提供了十分令人信服的表示方式，它与随机理论结合，即能完成对更复杂的不确定性的分析。同时，分析的结果最终仍以模糊集表示，就可以显式地考虑不确定性因素对结果的影响。模糊水文统计分析正是基于这样一条思路。

2 基本框架

2.1 模糊数

模糊数是实数集 R 上的特殊模糊集，同时也是实数和区间数概念的推广，最早见诸文献

本文于1994年11月7日收到，1995年10月16日收到修改稿。

*博士点基金资助项目。

[6]。Negoita 等[7]则首先将模糊数看成是一个区间数族，即含参数的区间数。设 x^* 是一模糊数，其隶属函数为 $\mu_{x^*}(x)$，则其区间数族可用 α-截集族表示，即

$$C(x^*)_\alpha = \{x \in R; \mu_{x^*}(x) \geqslant \alpha\} \qquad \forall \alpha \in [0,1] \tag{1}$$

其中（1）$\alpha=1$ 时 $C(x^*)_\alpha$ 非空；（2）$C(x^*)_\alpha$ 是 R 上的有界闭区间 $[C_L(x^*)_\alpha, C_U(x^*)_\alpha]$，$L$，$U$ 分别表示区间的下、上界。

显然，x^*、$(C(x^*)_\alpha)_{\alpha\in[0,1]}$和 $\mu_{x^*}(x)$ 这三个概念是等价的。近年来，模糊数常用以表示欠精确量（如 [8]），对观测数据的误差进行有效描述。因为经典的误差理论中用正态分布来表示误差分布，我们即以与它形状相似又满足正则性要求的指数型函数来刻画水文采样数据的模糊性。

模糊数的概念可以推广到模糊向量的情形。

2.2 模糊样本

通常给定一个水文样本（x_1，x_2，…，x_n）相应的隶属函数时，就可以得到一个以模糊向量（x_1^*，x_2^*，…，x_n^*）表示的模糊样本 X^*，其联合隶属函数为

$$\mu_{X^*}(X) = f(\mu_{x_1^*}(x_1), \mu_{x_2^*}(x_2), \cdots, \mu_{x_n^*}(x_n); x_1, x_2, \cdots, x_n) \tag{2}$$

若无进一步的假定，$f(\cdot)$ 非唯一，主要与扩张原理有关。不同的扩张原理对应着不同的隶属函数合成规则。目前常用的有最小规则

$$\mu_{X^*}(X) = \min(\mu_{x_1^*}(x_1), \mu_{x_2^*}(x_2), \cdots, \mu_{x_n^*}(x_n)) \tag{3}$$

和乘积规则

$$\mu_{X^*}(X) = \prod_{i=1}^{n} \mu_{x_i^*}(x_i) \tag{4}$$

扩张原理与 α-截集具有相容性[9]。这种相容性确保模糊数的 α-截集表示形式可以很方便地应用于模糊环境中的统计分析之中，因为一般情况下，直接确定统计推断结果比确定其 α-截集族要困难得多。

2.3 模糊参数估计

在模糊样本条件下的统计分析途径可以分为两类：

（1）基于模糊集理论的分解定理与表现定理将模糊数转化为经典集合中的区间数，然后进行集值统计，求估计量。例如，给定隶属度 α，相应的截集为 $C(X^*)_\alpha=[C_L(X^*)_\alpha, C_U(X^*)_\alpha]$，即可以把一个模糊随机变量分解成两个经典随机变量（含参数 α），利用其联合概率分布的概念进行多元统计分析。

（2）按经典统计方法求出估计量，然后运用模糊集理论的扩张原理，求估计量的隶属函数。此时，估计量仍以模糊数表示。显然，模糊估计量可用其 α-截集形式表示。本文采用这种途径。

若 $X \sim f(x|\theta)$（$\theta \in \Theta$ 为参数）是随机变量 X 的一个经典概率模型，（x_1，x_2，…，x_n）为 X 的已知精确样本值，θ 的点估计量为 $t(x_1, x_2, \cdots, x_n)$。当考虑成模糊样本时，θ 对应的模糊估计量可通过 Θ 上的模糊子集 θ^* 的隶属函数 $\xi_{\theta^*}(\theta)$ 定义

$$\xi_{\theta^*}(\theta) = \sup\{\mu_{X^*}(X): \theta = t(x_1, x_2, \cdots, x_n)\} \tag{5}$$

由此可见，无论对于隶属函数合成的哪种规则而言，$\xi_{\theta^*}(\theta)$ 的推求是一列以 x_i（$i=1$，2，…，n）为变量的非线性规划模型的求解过程。只有当 $\theta=t(x_1, x_2, \cdots, x_n)$ 是线性关系

时，才有可能求得解析解。

对于高阶参数 C_v、C_s 等的模糊估计也可以类似的方式求得，只不过在求解上要困难一些，一般只能得到数值解。

另外，对于水文统计中的非参数问题，如洪水频率分析中的经验频率分布，也可用模糊统计分析进行处理，这样既可以考虑模糊样本的不确定性，同时又可以考虑经验频率本身的不确定性。

2.4　模糊随机过程

若在模糊随机变量中考虑时间因素，即对 X^* 加入 $t\in T$ 变量后就变成 $X^*(t)$，形成模糊随机过程。如天气预报中的晴、雨、阴、云等时间序列，水文中的径流过程、预报序列等。对于 $X^*(t)$ 而言，同样可以用 α-截集形式表示为 $X^*(t)=[C_L(X^*(t))_\alpha, C_U(X^*(t))_\alpha]$ $(t\in T)$。以 $AR(1)$ 模型为例，其基本参数均值、方差和一阶自回归系数均可按模糊样本条件下的参数估计方法求得，仍为模糊数。

3　模糊贝叶斯分析

同经典统计学相比，贝叶斯统计学具有一些很重要的特点，如它的学习性质、信息综合能力等，因而在水文水资源领域取得了比较广泛的应用[10]。模糊随机环境中的贝叶斯分析取决于对模糊随机变量的定义方式。Zadeh 把模糊事件的概率看成是其隶属函数的期望，这种意义上的模糊贝叶斯分析[11]尽管具有一定的理论和实用意义（特别在离散状态空间情形），但对大多数水文统计问题没有太大的实用性。另外，按乘积规则建立的模糊贝叶斯分析框架计算十分繁复，有时甚至无法求得某些极值问题，也难以在水文统计中获得理想的应用效果。为此我们主要探讨最小规则下的模糊贝叶斯分析。

3.1　模糊似然函数

在最小规则 $\mu_{X^*}(X)=\min\limits_{i=1,2,\cdots,n}\mu_{x_i^*}(x_i)$ 下可以证明

$$C(X^*)_\alpha=C(x_1^*)_\alpha\times C(x_2^*)_\alpha\times\cdots\times C(x_n^*)_\alpha \tag{6}$$

即模糊向量 $X^*=(x_1^*, x_2^*, \cdots, x_n^*)$ 的 α-截集是其各分量 α-截集的欧氏积。因此，若随机变量 X 的精确样本 $(x_1, x_2, \cdots, x_n)$ 的似然函数 $L(X|\theta)=\prod\limits_{i=1}^{n}f(x_i|\theta)$，考虑 X 的模糊性，称其相应的模糊样本 $(x_1^*, x_2^*, \cdots, x_n^*)$ 的似然函数（它是一个泛函）为模糊似然函数，其最小规则下的 α-截集表示为

$$C_L(L(X^*|\theta))_\alpha=\min_{X\in C(X^*)_\alpha}L(X|\theta)=\prod_{i=1}^{n}(f_i)_\alpha^L(x_i|\theta) \tag{7}$$

$$C_U(L(X^*|\theta))_\alpha=\max_{X\in C(X^*)_\alpha}L(X|\theta)=\prod_{i=1}^{n}(f_i)_\alpha^U(x_i|\theta) \tag{8}$$

其中 $(f_i)_\alpha^L(x_i|\theta)=\min\limits_{x_i\in C(x_i^*)_\alpha}f(x_i|\theta)$，$(f_i)_\alpha^U(x_i|\theta)=\max\limits_{x_i\in C(x_i^*)_\alpha}f(x_i|\theta)$

3.2　模糊贝叶斯公式

假设 $f_0(\theta)$ 为一般意义下 θ 的先验分布，根据贝叶斯定理，在模糊似然函数条件下所要

推断的模糊参数 θ^* 的非标准化的模糊后验分布为

$$C_L(g(\theta^*|X^*))_\alpha = f_0(\theta)\prod_{i=1}^{n}(f_i)_\alpha^L(x_i|\theta) \tag{9}$$

$$C_U(g(\theta^*|X^*))_\alpha = f_0(\theta)\prod_{i=1}^{n}(f_i)_\alpha^U(x_i|\theta) \tag{10}$$

很明显，这里把（5）式中求 n 维极值问题转化为密度函数 f $(x_i|\theta)$ 在 X_i 一维空间中的局部极值问题，使计算量大大减小。另外，可以证明，当增加新的样本点 x^*_{n+1} 等以后，非标准化的模糊后验分布只需在原有基础上乘以 $(f_{n+1})_\alpha^L$ $(x_{n+1}|\theta)$ 和 $(f_{n+1})_\alpha^U$ $(x_{n+1}|\theta)$，即依然保持了经典贝叶斯分析中的学习性质，这也是它所具有的理论上的重要特点之一。

经典的贝叶斯估计通常采用 θ 的期望值（相应于二次损失函数），同样可以推广到模糊贝叶斯分析中。改变 α，使之遍历所要求的值，如 0.0，0.1，…，0.9，1.0，则可获得 θ 的一个完整的模糊贝叶斯估计。

3.3 P-Ⅲ型分布 C_s 的模糊贝叶斯估计

基于汲取专家经验和地区信息的思想，笔者提出一种分阶段、分层次估计 P-Ⅲ型分布参数的地区贝叶斯方法[①]：(1) 按传统方法（如概率权重矩法）求均值，也可综合各种方法的估计值；(2) 通过地区综合途径求 C_v 值，即分别赋予单站估计值和地区加权平均值以不同的权重进行综合；(3) 运用贝叶斯方法推求 C_s 值，即假定地区内各站点 C_s 值服从某一分布（如贝塔分布）作为 C_s 的先验分布，然后结合单站样本信息，通过贝叶斯公式推断出单站 C_s 的后验分布并以其期望值作为 C_s 的估计值。统计试验研究表明，地区贝叶斯方法具有优良的统计性能，特别是有效性方面，要比概率权重矩法、优化适线法等传统方法优越得多。

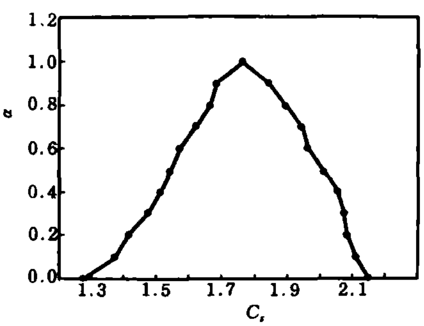

图1 太平驿洪峰系列 C_s 的模糊贝叶斯估计

Fig. 1. Fuzzy Bayesian estimation of C_s of Taipingyi Station

在模糊样本条件下，如果暂不考虑 $\bar{x}$ 和 C_v 估计的模糊性，我们运用岷江上游太平驿的洪水资料（假定观测误差不超过10%，调查洪水误差不超过20%）对 C_s 的模糊性进行了分析计算，见图1。从中可以看出，资料的欠精确性对 C_s 估计的影响很大，这在传统的估计方法中是反映不出来的。

4 结　语

模糊随机现象是水文中客观存在的不确定性表现形式，本文提出了将水文统计从精确环境推广到模糊环境的一个初步的框架，特别是对基于最小规则的模糊贝叶斯方法作了较为详细的分析。模糊水文统计结果可以提供较传统方法更多的信息，而后者仅是前者的特例。最后值得一提的是，本文提出的方法并不是唯一的，可以根据其它的定义方式构建不同的体系。

---

① 黄伟军．灰色模糊贝叶斯理论及其在水文水资源中的应用．四川联合大学博士学位论文．1995

参 考 文 献

1 Kwakernaak H. Fuzzy random variables Ⅰ，Ⅱ. Information Science. 1978，15：1～29，1979，17：253～278

2 Kruse R. The strong law of large numbers for fuzzy random variables. Information Science. 1982，28：233～241

3 Miyakoshi M and Shimbo M. A strong law of large numbers for fuzzy random rariables. Fuzzy Sets and Systems. 1984，12：133～142

4 Stein W，and Talati K. Convex fuzzy random variables. Fuzzy Sets and Systems. 1981.6：271～283

5 Nahmias S. Fuzzy variables. Fuzzy Sets and Systems. 1978，1：97～110

6 Chang S S L. and Zadeh L A. On fuzzy mapping and control. IEEE Trans. systems，Man，Cybernet. 1972. 2（1）：30～34

7 Negoita C V，and Ralescu D A. Application of Fuzzy Sets to System Analysis. Wiley，New York，1975. 1～15

8 Viertl R. Statistical inference for fuzzy data in environmetrics. Environmetrics. 1990，1：37～42

9 Nguyen H T. A note on the extension principle for fuzzy sets. J. Math. Anal. Appl，1978，64（2）：369～380

10 黄伟军、丁晶. 水文水资源系统贝叶斯分析现状与前景. 水科学进展. 1994，5（3）：242～247

11 Piasecki K. On the Bayes formula for fuzzy probability measures. Fuzzy Sets and Systems. 1986，18：183～185

Research on Fuzzy Hydrologic Statistical Analysis

Huang Weijun and Ding Jing

(Dept. of Hydraulic Eng.，Sichuan Union Uni.，Chengdu，610065)

Abstract：Hydrologic observed data and investigated (or historical) data are always imprecise，and may be modeled by fuzzy numbers，which forms fuzzy sample (s). This paper studies the methodology of fuzzy hydrologic statistical analysis based on the concepts of fuzzy number and fuzzy sample，and describes the fuzzy Bayesian analysis approach in detail. The methodology is briefly interpreted by examples.

Key words：hydrologic statistics；fuzzy number；fuzzy sample；fuzzy stochastic process；fuzzy Bayesian analysis.