

洪水频率分析的非参数变换回归方法

董洁, 谢悦波

(河海大学水资源环境学院, 江苏 南京 210098)

摘要:提出适用于小样本的非参数变换核估计方法, 并将其用于洪水频率分析. 结果显示, 非参数变换核估计法不仅具有事先不需要假定总体线型的优点, 而且估计的结果具有较高的精度.

关键词:非参数变换; 核估计; 窗宽; 洪水频率

中图分类号:TV122

文献标识码:A

文章编号:1006-7647(2004)03-0019-02

在各种水利工程规划设计中, 需要确定满足工程要求的设计值, 计算方法主要有参数法和非参数法. 我国目前的计算规范中主要采用参数法. 但由于非参数法具有不需要模型假设这一基本优点, 加之计算机技术的快速发展, 使得非参数法越来越受到重视. 本文采用变换估计方法进行洪水频率分析, 从而避开了窗宽选择的难题, 达到提高小样本估计精度的目的.

1 非参数变换法

核估计是常用的非参数估计. 小样本时估计性能差的原因是估计偏差和方差对窗宽的矛盾比较突出. 例如, 非参数密度估计的偏差与窗宽成正比, 而方差与窗宽平方成反比. 核估计的偏差还与待估计函数的形状有关. 如果待估函数的形状比较陡峭, 则在陡峭部分核估计不能得到无偏估计, 产生较大偏差. 本文提出一种新的核估计——变换核估计. 它将原始数据样本变换成估计精度较高的新样本, 对新样本进行函数估计后, 再反变回去, 就得到原函数较好的估计结果. 变换估计克服了非参数法针对性差的缺点, 避开了窗宽选择的难题, 提高了小样本的估计精度, 特别适用于洪水频率分析.

1.1 分布函数的非参数变换估计

目前, 在洪水频率分析的非参数方法中, 一般采用核估计方法先由样本估计出总体的密度函数, 再对其积分得出设计值. 由于密度估计不是无偏估计, 对其积分得到的分布函数会将偏差积累起来. 密度和分布函数是随机变量的两种不同属性, 虽然两者的关系是微分、积分关系, 但两者的估计方法不一

样. 如果要从密度估计得到分布函数估计, 密度估计应该是无偏的, 方差大点没关系, 因为积分会消除方差的影响; 如果要从分布函数估计得到密度估计, 分布函数应该连续光滑, 允许有点偏差, 因为数值积分用邻近几个点估计一个点的微分值, 对非光滑点较敏感. 因此, 本文介绍一种直接对经验分布函数拟合的非参数回归的变换方法.

1.1.1 样本经验分布函数

传统的分布函数非参数估计是经验分布函数, 经验分布函数有较好的极限性, 当 n 趋于无穷大时, $F_n(x)$ 几乎处处收敛. 但小样本时性能不好, 方差较大, 表现为阶梯形函数. 可以利用经验分布函数的无偏性, 然后加以平滑, 就得到偏差和方差都较小的估计. 经验分布函数估计是常用的分布函数估计, 设 $x_1, x_2, \dots, x_n$ 是随机样本, 有分布函数 $F(x)$, 则经验分布函数为

$$F_n(x) = \frac{i}{n+1} \quad (i = 1, 2, \dots, n) \quad (1)$$

1.1.2 样本经验分布函数的变换回归

记 $y_i = F_n(x_i)$, 则 (x_i, y_i) 构成随机变量, 可以进行非参数回归计算. 但这里的 y_i 不是通常回归分析中事先给出的样本, 是分布函数在 x_i 点的无偏估计. 将原样本 (x_i, y_i) 变成新样本 (v_i, u_i) , 其中 v_i 是均值为 1 的随机变量, u_i 是均匀分布的随机变量.

记 $\hat{f}(x)$ 为 x 的密度核估计, 则 $\hat{f}(x) = \frac{1}{nh} \cdot$

$\sum_{i=1}^n k\left(\frac{x-x_i}{n}\right)$, 其中 $\sum k\left(\frac{x-x_i}{n}\right)$ 是核函数, h 是窗宽. 迭代算法如下:

$$y_i = \frac{i-1+\alpha}{n} \quad (2)$$

$$\hat{F}_l(x) = \frac{\sum_{i=1}^n y_i k\left(\frac{x-x_i}{h}\right)}{\sum_{i=1}^n k\left(\frac{x-x_i}{h}\right)} \quad (3)$$

$$u_{l,i} = \hat{F}_l(x_i) \quad (4)$$

$$u = \hat{F}_l(x) \quad (5)$$

$$v_{l,i} = y_i - u_{l,i} + 1 \quad (6)$$

$$\hat{F}_{l+1}(u) = \frac{\sum_{i=1}^n v_{l,i} k\left(\frac{u-u_{l,i}}{h}\right)}{\sum_{i=1}^n k(u-u_{l,i}) + u-1} \quad (7)$$

$$\hat{F}_{l+1}(x) = \hat{F}_{l+1}[\hat{F}_l^{-1}(u)] \quad (8)$$

变换回归的偏差很小,基本上是无偏回归,因样本经验分布函数也是无偏估计,所以最后得到的分布函数估计是无偏估计。

1.2 回归的偏差分析

核回归直接从条件均值出发构造统计量,以期得到最小均方差估计,但小样本时回归偏差较大。核回归本质上是平滑运算,若待回归的函数起伏较大,例如存在峰值,平滑后可能削掉峰。小样本时邻近点的真实值相差较大,平滑产生的偏差大,回归精度不如参数回归法。而变换回归在小样本时精度超过参数法。

设样本为 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, 满足模型:

$$y_i = m(x_i) + \varepsilon_i, \quad (i = 1, 2, \dots, n) \quad (9)$$

其中 $m(x_i)$ 满足 Lipschitz 条件,其核估计为^[1]

$$\hat{m}(x) = \frac{\frac{1}{nh} \sum_{i=1}^n y_i k\left(\frac{x-x_i}{h}\right)}{\hat{f}(x)} = \frac{\alpha(x)}{\hat{f}(x)} \quad (10)$$

分别考虑上式分子分母的偏差 $bias^{[2]}$, 则

$$bias_x(\alpha) = \frac{1}{2} h^2 [m(\cdot)f(\cdot)]^{(2)}(x) k_2 + o(h^2) \quad (11)$$

$$bias_x(\hat{f}) = \frac{1}{2} h^2 f''(x) k_2 + o(h) \quad (12)$$

其中

$$[m(\cdot)f(\cdot)]^{(2)}(x) = m''(x)f(x) + 2m'(x)f'(x) + m(x)f''(x)$$

为了减少分子和分母的偏差,将原样本 (x_i, y_i) 变换成新样本 (v_i, u_i) ,通过变换, $[m(\cdot)f(\cdot)]^{(2)}(x)$ 和 $f''(x)$ 除端点外将为零,在端点 $(0, 1)$ 处,虽然 $f''(x)$ 不为零,但分子和分母的数值是差不多相同的,具有同样的偏差,因而偏差互相抵消,能恢复核

回归的“端点效应”。

1.3 核函数和窗宽的选择

1.3.1 核函数的选择

常用的核函数有 Epanechnikov 核函数、矩形核函数、正态核函数等,他们都是有界的函数。从渐近效率看, Epanechnikov 核函数有较好的渐近效率,但定义域有限,小样本时落入定义域的样本个数很少。正态核函数渐近效率低,但定义域是全数轴,对小样本有利,所以本文选正态核函数。

1.3.2 窗宽的选择

窗宽的选择是一个较复杂的问题,因为回归的方差除了与条件方差、核函数、样本有关外,还与窗宽有关。条件方差由样本决定,核函数对方差影响不大,减小方差的有效途径是加大窗宽。因为回归估计当迭代收敛时是无偏的,偏差与窗宽无关,所以进行变换核回归时,迭代开始可取较小的窗宽,使初始估计的偏差较小,收敛快些,而迭代收敛时,可用较大窗宽,以减少回归的方差^[3]。

在实际回归问题中,条件方差往往不是常数,随 x 有较大变化,有时某些样本是异常点,远离群体,最佳窗宽应是可变窗宽,随方差的变化而变化。

2 估计实例

基本数据选自文献[4]中某水文站年径流资料,原文假设总体服从 P-III 型分布,采用参数适线法对经验频率曲线进行拟合,最后得到一条拟合曲线(图1),通过内插即可求出给定频率下的设计值。本文不需任何假设条件,用非参数变换法对样本进行拟合,只需4次迭代就能很好地拟合,该方法具有很好的稳健性。

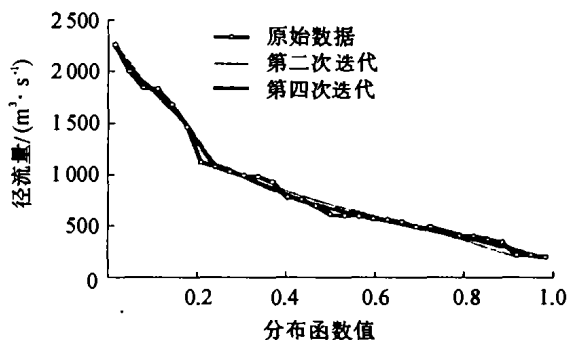


图1 洪水频率分析拟合曲线

3 结 语

非参数变换核估计是一个迭代过程,包括样本变换、新样本函数估计和函数反变换。迭代过程中新样本逐步逼近最佳样本,最后经过函数反变换得到原样本的较好估计。

(下转第44页)

从上部掏砂口进行掏砂作业,只要盾构在纵向的布置合理,则掏砂操作距离就可以很容易缩短到需要的距离。

盾构的结构分析可按构筑期和掏砂施工期分别进行,构筑期可按成层地基 Rankine 主动土压力计算理论进行计算分析:

$$E_a = \frac{1}{2} \rho g H^2 K_a - 2cH \sqrt{K_a} + \frac{2c^2}{\rho g}$$

$$K_a = \tan^2 \left(45^\circ - \frac{\varphi}{2} \right)$$

式中: c 为土的粘聚力; φ 为土的内摩擦角; ρ 为土的密度,地下水位以下用有效密度; g 为重力加速度; H 为盾构的侧壁高度, mm。

侧压力考虑土压力、水压力、基底压力,取单位墙长计算。盾构的平面位置布置如图 1 中虚线所示,截面尺寸见图 2。在掏砂施工期,上部掏砂口的混凝土块已部分或全部撤去,一定距离内的砂已从掏砂口中掏去,基底压力已无法传递给盾构侧壁,筏底压力作用在主动破坏楔体之外,侧壁压力也就大大减少,所以按构筑期设计已能满足施工要求。

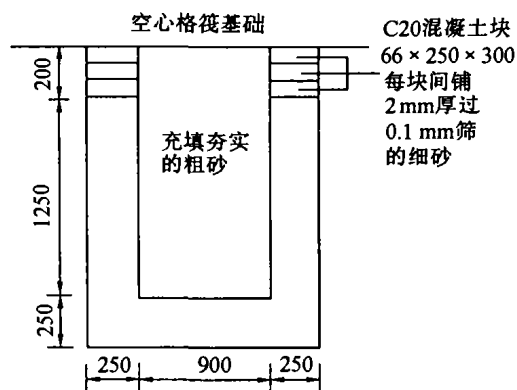


图 2 盾构设置(单位:mm)

施工期间应抽取地下水,降低地下水位到盾构底部以下,掏砂施工不宜过快,以防基础一边突沉。应使建筑物各部位能按既定的计划下沉或抬升,掏砂纠偏施工结束后应将盾构用砂土填实。本工程建成投入使用后,经观测,空心格筏基础已能使筒仓沉降满足设计要求,盾构设施暂未使用。

4 结 论

a. 空心格筏基础是介于箱基与筏基之间的一种新型基础形式,具有较大的刚度和整体性,能有效地控制复杂地基上建(构)筑物的不均匀沉降,在节约建材、方便施工、缩短工期和降低造价等方面具有较大的优越性。

b. 预埋式盾构纠偏技术的应用是一种新的思路,也是控制多层建(构)筑物不均匀沉降的一条行之有效的途径。空心格筏基础下设砂垫层时,一旦建

筑物出现较大的不均匀沉降,可以采用掏砂纠偏的方法进行补救。在建筑物的纠偏工程中采用掏砂法往往是较经济,在盾构的掩护下进行掏砂,操作距离可大为缩短,可控性好。

c. 本文工作仅是结合实际工程对空心格筏基础和预埋式盾构纠偏技术作些探索性的研究,应用中有关技术问题待进一步研究,如空心格筏基础的沉降随其面积和刚度的变化规律等。

参考文献:

- [1] JGJ6—99,高层建筑箱形与筏形基础技术规范[S].
- [2] 沈杰.地基基础设计手册[M].上海:上海科学技术出版社,1985.
- [3] GB50007—2002,建筑地基基础设计规范[S].
- [4] 华南理工大学,东南大学,浙江大学,等.地基及基础[M].北京:中国建筑工业出版社,1995.

(收稿日期:2003-11-03 编辑:傅伟群)

(上接第 20 页)

变换核估计的优点是窗宽和估计精度与待估计函数关系不大,因为各种分布样本都变换成接近最佳分布的样本进行估计。某些不好估计的密度函数,如具有较陡的起始部分和长尾,通过变换估计能显著改善其精度。

变换估计的缺点是计算量较大,较高的估计精度是靠较多的计算量换来的。但变换估计主要用于小样本,总的计算量实际并不大。

变换核估计用于洪水频率分析,是一种简单有效的方法。因为我国洪水实测资料年限一般较短,仅有几十年的观测资料系列,属于小样本。用这仅有的几十年资料推求工程上所需要的百年、千年、甚至万年一遇的设计洪水精度明显偏低。另外,现有的洪水频率计算方法,一般假设洪水系列来自 P-Ⅲ型总体,这与实际不一定符合,因而由此推求的设计值也有很大偏差。

参考文献:

- [1] 陈希孺.非参数统计教程[M].上海:华东师范大学出版社,1986.
- [2] Muller H G, Stadtmuller U. Variable bandwidth kernel estimators of regression curves[J]. Ann Statist, 1987, 15: 182 ~ 202.
- [3] Derroye F M, Gyorf L. Nonparametric density estimation[M]. New York: Wiley Inc, 1985. 244 ~ 252.
- [4] 王俊德.水文统计[M].南京:河海大学出版社,1992.

(收稿日期:2003-09-25 编辑:张志琴)