비모수 퍼지회귀모형

최승회[1) 김해경[2) 성나영[3)

요 약

본 연구에서는 크리스프자료(crisp data)인 독립변수와 퍼지자료(fuzzy data)인 종속변수 사이의 관계가 특정한 함수로 표현되지 않는 비모수 퍼지회귀모형을 분석하기 위하여 퍼지수 순위와 퍼지순위변환방법을 소개하고, 모의실험을 통하여 퍼지순위변환방법의 효율성을 조사한다.

주요용어 : 비모수 퍼지회귀모형, 퍼지수 순위, 퍼지순위변환방법.

1. 서론

자연과학, 공학, 그리고 사회과학등 여러 분야에서 독립변수와 종속변수에 대한 관계를 설명하기 위하여 회귀모형을 이용한다. 회귀모형을 분석하는 회귀분석에 있어서, 종속변수의 실제값과 관찰값과의 차는 측정오차로 설명된다. 그러나, Tanaka와 Ujima, 그리고 Asai(1982)에 의하여 시작된 퍼지회귀모형은 종속변수의 실제값과 관찰값의 차를 모형의 부정확성, 즉 퍼지성으로부터 설명된다. Tanaka et al.(1982)와 Savic과 Pedrycz(1991) 그리고 Chang과 Ayyub(2001)등에 의하여 활발히 연구된 모수적 퍼지회귀모형은 독립변수에 크게 의존하고, 독립변수와 종속변수의 관계가 미지인 경우가 존재함으로 이를 보완하기 위한 새로운 퍼지회귀모형을 Cheng과 Lee(1999)가 제시하였다. 크리스프자료인 독립변수(x_i)와 퍼지자료인 종속변수(Y_i) 사이의 관계가 특정한 함수로 표현되지 않는 비모수 퍼지회귀모형은

$$Y_i = m(x_i) + \varepsilon_i, \quad i = 1, \cdots, n$$

와 같다. 여기서 ε_i는 퍼지측정오차이다.

Cheng과 Lee는 비모수 퍼지모형을 분석하기 위하여, 비모수 회귀분석에서 사용되는 kernel 방법과 k-NN 방법을 퍼지모형에 적용하였다. 그러나, 평균 개념을 이용한 kernel 방법과 k-NN 방법은 이상치에 민감할 수 있다.

본 연구에서는 비모수 퍼지회귀모형을 분석하기 위하여 퍼지수 순위와 퍼지순위변환방법을 소개하고, 모의실험을 통하여 퍼지순위변환방법의 효율성을 조사한다.

2. 퍼지순위변환방법

퍼지수(fuzzy number)를 정의하기 위하여 함수 $L(x)$과 $R(x)$은 구간 $[0, \infty)$에서 감소하고

1) 한국항공대학교 교양학부 부교수, 경기도 고양시 덕양구 화전동 200-1
2) 연세대학교 수학과 교수, 서울시 서대문구 신촌동 120-749
3) 연세대학교 수학과 석사과정, 서울시 서대문구 신촌동 120-749

$L(-x)=L(x), R(-x)=R(x)$ $L(0)=R(0)=1$ 그리고 $L(1)=R(1)=0$을 만족한다고 하자. $L-R$퍼지수 A의 멤버쉽함수는

$$\mu_A(x)=\begin{cases} L\left(\dfrac{m-x}{e_A^l}\right) & \text{for} x\le m \\ R\left(\dfrac{x-m}{e_A^r}\right) & \text{for} x\ge m \end{cases}$$

이고, $A=(e_A^l, m, e_A^r)_{LR}$와 같이 표현하자. 여기서 e_A^l과 e_A^r은 양수이다. 특히, 함수 L과 R의 차수가 일차이면 삼각형 퍼지수라 하고 $A=(e_A^l, m, e_A^r)$와 같이 표현하고 한다. $L-R$퍼지수 A의 α-레벨집합 $A_\alpha=\{x|\mu_A(x)\ge\alpha\}$에 대하여 $\underline{A_\alpha}\equiv infA_\alpha$와 $\overline{A_\alpha}\equiv supA_\alpha$라 하자. $L-R$퍼지수의 집합 F_{LR}에 속하는 두 $L-R$퍼지수 A와 B는 $e_A^l\le e_B^l$일 때 $e_A^r\le e_B^r$을 만족한다고 하자. 집합 F_{LR}에 속하는 퍼지수 A에 대한 순위(rank)는 $\underline{A_\alpha}$에 대한 크기에 따라 정의할 수 있다. 그리고 $\underline{A_\alpha}$의 값이 동일할 경우는 $\overline{A_\alpha}$의 크기에 따라 정하고, 동일한 순위에 대하여서는 순위의 평균값을 순위로 정한다.

비모수 퍼지회귀모형에서 독립변수 $x=x_i^*$에서 종속변수 Y_i^*에 대한 예측 퍼지수 $\widehat{Y_i^*}$을 구하기 위하여 다음과 같은 퍼지순위변환방법을 이용하자.

1. 독립변수 x_i의 순위 $R(x_i)$와 종속변수 Y_i의 α-cut $\mu_{Y_i}^{-1}(\alpha)$의 순위 $R\left(\mu_{Y_i}^{-1}(\alpha)\right)$에 대한 선형관계식을 결정.
2. x_i^*의 순위 $R(x_i^*)$을 결정.
3. $R(x_i^*)$을 이용하여 $R\left(\mu_{Y_i^*}^{-1}(\alpha)\right)$을 결정.
4. 주어진 선형관계식을 이용하여 $\mu_{Y_i^*}^{-1}(\alpha)$을 결정.
5. 예측 퍼지수 $\widehat{Y_i^*}$을 결정.

3. 퍼지순위변환방법의 효율성

앞에서 추정된 퍼지수가 주어진 퍼지회귀모형을 얼마나 잘 설명하고 있는지를 조사하기 위하여 추정된 퍼지수의 정도(precision)를 설명하는 두 가지 측도(measure)를 생각하여 보자. 두 $L-R$퍼지수 $A=(e_A^l, m_A, e_A^r)_{LR}$와 $B=(e_B^l, m_B, e_B^r)_{LR}$의 차

$$(A-B)_{LR}\equiv(e_A^l-e_B^l, m_A-m_B, e_A^r-e_B^r)$$

에 대한 거리와 면적을 이용하여 정의하여 보자. 실제 퍼지수의 집합 $\{Y_1, \cdots, Y_n\}$에 대한 추정 퍼지수의 집합 $\{\widehat{Y}_1, \cdots, \widehat{Y}_n\}$에 대한 효율성을 다음과 같이 정의하자.

$$1.\ \text{SDE}=\sum_{i=1}^{n} D^p(Y_i, \widehat{Y}_i)$$

$$2.\ \text{SAE}=\sum_{i=1}^{n} M(Y_i, \widehat{Y}_i)$$

여기서, $D^p(A,B)=|e_A^l-e_B^l|^p+|m_A-m_B|^p+|e_A^r-e_B^r|^p$ $M(A,B)$는 퍼지수 $(A-B)_{LR}$의 멤버

십합수 $\mu_{(A-B)_{LR}}(x)$와 x축에 의해서 둘러싸인 영역의 면적을 표시한다. Cheng과 Lee(1999)가 제시한 kernel 방법이나 k-NN 방법과 퍼지순위변환방법의 효율성을 비교하기 위하여 다음과 같은 모형을 생각하자.

$$1.\ f(x_i) = \frac{x_i^2}{5} + 2e^{\frac{x_i}{10}}, \quad i = 1.\cdots.10.$$

$$2.\ f(x_i) = \sin(x_i) + 2e^{\frac{x_i}{10}},\ i = 1, \cdots, 10.$$

독립변수 x_i는 균등분포 $U(0,10)$에서 생성하고, 종속변수 $Y_i = (m_i, a_i)$의 중심(m_i)과 스프레드(a_i)는 각 각 $m_i = f(x_i) + \eta_i$와 $a_i = x_i + \zeta_i$에 의하여 생성된다. 여기서 x_i는 균등분포 $U(0,5)$을 확률오차항 η_i와 ζ_i는 각 각 균등분포 $U(-0.5, 0.5)$를 따른다고 하자. 이와 같은 환경에 따라 실험한 결과, 모형1에서는 퍼지순위변환방법이 kernel 방법이나 k-NN 방법보다 더 효율적이고 모형2에서는 kernel 방법이 퍼지순위변환방법보다 효율적이었다.

참고문헌

[1] Y. O. Chang and B. M. Ayyub(2001), Fuzzy regression methods-a comparative assessment, Fuzzy Sets and Systems, Vol.119, 187-203.

[2] C. B. Cheng and E. S. Lee(1999), Nonparametric fuzzy regression, An International Journal Computers and Mathematics with Applications, Vol.38, 239-251.

[3] C. H. Cheng(1998), A new approach for ranking fuzzy numbers by distance method, Fuzzy Sets and Systems, Vol.95, 307-317.

[4] D. Savic and W. Pedrycz,(1991), Evaluation of fuzzy regression models, Fuzzy Sets and Systems Vol.39, 51-63.

[5] H. Tanaka, S. Uejima and K. Asai, Linear regression analysis with fuzzy model, IEEE, Systems, 903-907.

[6] K. J. Kim and H. R. Chen(1997), A comparison of fuzzy and nonparametric linear regression, Computers Ops Res, Vol.6, 505-519.

[7] R. L. Iman and W. J. Conver(1979), The use of the rank transform in regression., Technometrics, Vol.21, 499-509.