수리통계학 교육에서 상호정보의 활용에 대한 연구

장대홍[1)]

요 약

상호정보를 이용하면 두 확률변수 사이의 종속의 정도를 평가할 수 있는 측도를 제시할 수 있고 두 변수 사이의 상관관계를 나타내는 표본상관계수의 단점을 보완한 일반화상관계수를 정의할 수 있다.

주요용어 : 상호정보, 독립성, 표본상관계수

1. 서론

상호정보(mutual information)는 다음과 같이 정의된다(Hutter와 Zaffalon(2005) 참조).

1. (연속형)

$$I_{XY} = \int\int f(x,y) \ln \frac{f(x,y)}{f_1(x) f_2(y)} dx dy$$

2. (이산형)

$$I_{XY} = \sum_x \sum_y f(x,y) \ln \frac{f(x,y)}{f_1(x) f_2(y)} \quad (1)$$

여기서, $f(x,y)$는 확률변수(또는 확률벡터) X와 Y의 결합확률(밀도)함수이고, $f_1(x)$는 확률변수(또는 확률벡터) X의 주변확률(밀도)함수이고, $f_2(y)$는 확률변수(또는 확률벡터) Y의 주변확률(밀도)함수이다.

이 상호정보는 결합확률(밀도)함수 $f(x,y)$와 주변확률(밀도)함수의 곱 $f_1(x)f_2(y)$ 사이의 Kullback-Leibler divergence이다. 상호정보는 다음과 같은 성질을 갖고 있다.

1. $I_{XY} \geq 0$
2. $I_{XY} = 0$이면 두 확률변수(확률벡터)는 서로 독립이 되고, $I_{XY} > 0$ 이면 두 확률변수(확률벡터)는 서로 종속이 된다.

두 변수가 범주형 변수일 때는 다음과 같이 상호정보의 추정량이 주어진다.

$$\hat{I}_{XY} = \sum_i \sum_j \frac{n_{ij}}{n} \ln \frac{n_{ij} n}{n_{i+} n_{+j}} \quad (2)$$

여기서, n은 분할표 상의 총도수, n_{ij}는 범주형 변수 X에서 i번째 범주와 범주형 변수 Y에서

1) (608-737) 부산광역시 남구 대연3동 599-1 부경대학교 수리과학부 통계학전공, 교수
E-mail: dhjang@pknu.ac.kr

j번째 범주에 해당하는 칸(cell)의 도수, n_{i+}는 변수 X에서 i번째 범주의 도수, n_{+j}는 변수 Y에서 j번째 범주의 도수이다.

(1)식을 이용하여 상호정보 $I(x,y)$를 구한 후 상호정보의 값을 지수화하기 위하여 $I(x,y)$를 $\lambda = \sqrt{1-e^{-2I(x,y)}}$로 변환시키면 $0 \le \lambda \le 1$이 된다. 상호정보를 계산하기 위해서는 결합확률(밀도)함수 $f(x,y)$, 주변확률(밀도)함수 $f_1(x)$와 $f_2(y)$의 형태를 알아야 한다. 이 함수들의 형태를 알 수 없는 경우 주어진 자료를 이용하여 이 함수들의 형태를 추정하게 되는 데 크게 두 가지 방법이 주로 쓰인다.

1. 이변량 히스토그램을 이용하는 방법　　2. 이변량 커널추정량을 이용하는 방법

이변량 히스토그램을 이용하는 방법은 다시 두 가지 방법(equidistant cells과 equiprobable cells)으로 나뉜다. 상호정보의 추정 방법들에 대한 연구들이 최근까지 활발히 진행되고 있다(예로, Darbellay(1999), Darbellay와 Vajda(1999), Harrold외 2인(2001), Chelikani외 2인(2003), Kraskov외 2인(2004), Zhou외 3인(2005) 등이 있다.). 상호정보는 통계학 분야뿐만이 아니라 학제간 연구가 최근까지 활발히 진행되고 있다(예로, Kraskov외 3인(2003), Sugimachi외 3인(2003), Shiraishi외 3인(2003), Dhillon외 2인(2003), Stogbauer외 3인(2004), Kojadinovic(2004), Dionisio외 2인(2004), Yin(2004), Wang외 2인(2005), Huang와 Chow(2005), Kojadinovic(2005) 등이 있다.).

2. 수리통계학 교육에서 상호정보의 활용

2.1 확률변수의 독립성

두 개의 확률변수 X와 Y의 독립성은 다음과 같이 정의한다.

$$f(x,y) = f_1(x)f_2(y) \quad \forall\, x \in X, y \in Y \;\leftrightarrow\; X\text{와 } Y\text{는 독립}$$

여기서, $f(x,y)$는 확률변수 X와 Y의 결합확률(밀도)함수이고, $f_1(x)$는 확률변수 X의 주변확률(밀도)함수이고, $f_2(y)$는 확률변수 Y의 주변확률(밀도)함수이다.

위의 등식이 성립하지 않을 때 우리는 두 개의 확률변수 X와 Y는 종속이라고 한다. 두 개의 확률변수 X와 Y가 종속일 때 상호정보를 이용하면 두 개의 확률변수 X와 Y 사이의 종속의 정도를 측정할 수 있게 된다. $\lambda = 0$이면 두 개의 확률변수 X와 Y는 독립이 되고, $\lambda > 0$면 두 개의 확률변수 X와 Y는 종속이 되며 λ값이 1에 가까울수록 두 개의 확률변수 X와 Y의 종속의 정도가 점점 강해지며 $\lambda = 1$이면 두 개의 확률변수 X와 Y는 완전 종속이 된다.

예 2.1(Rohatgi & Saleh(2001), 121p)) 두 개의 확률변수 X와 Y의 결합확률밀도함수가 다음과 같으면

$$f(x,y)=\begin{cases}\frac{1+xy}{4} & ,\ |x|<1,\ |y|<1\\ 0 & ,otherwise\end{cases}$$

주변확률(밀도)함수는 각각 $f_1(x)=\frac{1}{2}, |x|<1$과 $f_1(y)=\frac{1}{2}, |y|<1$이 된다. 그러므로 두 개의 확률변수 X와 Y는 종속이 된다. $I(x,y)$를 구하면 $\frac{\pi^2}{16}-\frac{5}{4}+ln(2)\approx 0.060$이 되고 $\lambda=0.336$이 된다.

예 2.2 두 개의 확률변수 X와 Y의 결합확률밀도함수가 다음과 같으면

$$f(x,y)=\begin{cases}x+y & ,0<x<1,0<y<1\\ 0 & ,otherwise\end{cases}$$

주변확률(밀도)함수는 각각 $f_1(x)=x+\frac{1}{2}, 0<x<1$과 $f_1(y)=y+\frac{1}{2}, 0<y<1$이 된다. 그러므로 두 개의 확률변수 X와 Y는 종속이 된다. $I(x,y)$를 구하면 $\frac{10}{3}ln(2)-\frac{9}{4}ln(3)+\frac{1}{6}\approx 0.005$가 되고 $\lambda=0.102$가 된다.

예 2.3(Hogg와 Tanis(2001, 235p)) 두 개의 확률변수 X와 Y의 결합확률밀도함수가 다음과 같으면

$$f(x,y)=\begin{cases}\frac{3}{16}xy^2 & ,0\le x\le 2,0\le y\le 2\\ 0 & ,otherwise\end{cases}$$

주변확률(밀도)함수는 각각 $f_1(x)=\frac{x}{2}, 0\le x\le 2$와 $f_2(y)=\frac{3y^2}{8}, 0\le y\le 2$가 된다. 그러므로 두 개의 확률변수 X와 Y는 독립이 된다. $I(x,y)$를 구하면 0이 되고 $\lambda=0$이 된다.

예 2.4(Hogg와 Tanis(2001, 235p)) 두 개의 확률변수 X와 Y의 결합확률밀도함수가 다음과 같으면

$$f(x,y)=\begin{cases}\frac{3}{2} & ,x^2\le y\le 1,0\le x\le 1\\ 0 & ,otherwise\end{cases}$$

주변확률(밀도)함수는 각각 $f_1(x)=\frac{3}{2}(1-x^2), 0\le x\le 1$과 $f_2(y)=\frac{3\sqrt{y}}{2}, 0\le y\le 1$이 된다. 그러므로 두 개의 확률변수 X와 Y는 종속이 된다. $I(x,y)$를 구하면 $-ln(2)-\frac{1}{2}ln(3)+\frac{7}{3}\approx 1.091$가 되고 $\lambda=0.942$가 된다.

예 2.5(Hogg와 Tanis(2001, 265p)) 두 개의 확률변수 X와 Y의 결합확률밀도함수가 다음

과 같으면

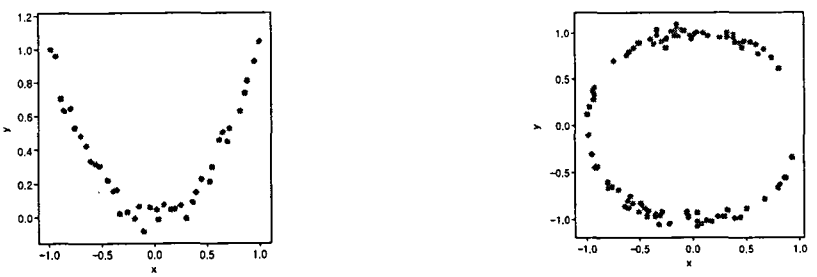
$$f(x,y)=\begin{cases}\frac{1}{2}e^{-y} & ,-y<x<y, 0<y<\infty \\ 0 & ,otherwise\end{cases}$$

주변확률(밀도)함수는 각각 $f_1(x)=\frac{1}{2}e^{-|x|}, -\infty<x<\infty$와 $f_2(y)=ye^{-y}, 0<y<\infty$가 된다. 그러므로 두 개의 확률변수 X와 Y는 종속이 된다. $I(x,y)$를 구하면 $\gamma\approx 0.577$이 되고 $\lambda=0.826$이 된다. 그런데 두 확률변수 X와 Y의 상관계수를 구하면 0이 된다.

2.2 표본상관계수

모상관계수에 대한 추정량으로서 n개의 이변량자료 $(x_i,y_i), i=1,2,...,n$을 이용하여 우리는 표본상관계수를 구할 수 있다. 이 표본상관계수를 통하여 두 변수 X와 Y 사이의 선형관계의 정도를 측정할 수 있다. 그러나 이 표본상관계수는 두 변수 X와 Y 사이의 비선형관계는 알 수가 없다. 이 때 상호정보를 이용하면 두 변수 X와 Y 사이의 선형관계 뿐만이 아니라 두 변수 X와 Y 사이의 비선형관계도 알 수 있게 된다. 즉, 상호정보를 일반화상관계수(global correlation coefficient)로 사용할 수 있다.

예 2.6 다음 2개의 산점도에서 표본상관계수를 구하면 (a)는 0.029, (b)는 0.077로 거의 0이라 할 수 있다. 즉, 두 변수 X와 Y 사이에 선형관계가 거의 없다 할 수 있다. 그러나 (a)는 2차 곡선 $y=x^2$, (b)는 원 $x^2+y^2=1$의 비선형관계를 나타내고 있다. 표본상관계수는 이러한 비선형관계를 나타내지 못 한다. 그러나 이변량 히스토그램을 이용하는 방법으로 상호정보를 구하여 보면 (a)는 1.279, (b)는 0.631로 계산되고 이를 이용하여 λ를 계산하여 보면 (a)는 0.961, (b)는 0.847이 되어 비선형관계가 강하게 있음을 알 수 있다.



(a) (b)

그림 2.1 비선형관계를 나타내는 산점도

3. 결론

본 논문을 통하여 수리통계학 교육 시 상호정보를 이용하는 방법을 두 가지 경우로 나누어 살펴보았다. 상호정보를 이용하는 학제 간 사례가 많아지므로 상호정보의 개념과 활용 예들을 개발하여 수리통계학 교육에 활용할 필요가 있다.

참고문헌

Chelikani, S., Purushothaman, K. and Duncan, J. S. (2003), Support Vector Machine Density Estimator as a Generalized Parzen Windows Estimator for Mutual Information Based Image Registration, *Leture Notes in Computer Science*, Vol. 2879, 854-861.

Darbellay, G. A. (1999), An Estimator of the Mutual Information Based on a Criterion for Independence, *Computational Statistics and Data Analysis*, Vol. 32, 1-17.

Darbellay, G. A. and Vajda, I. (1999), Estimation of the Information by an Adaptive Partition of the Observation Space, *IEEE Transaction on Information Theory*, Vol. 45, 1315-1321.

Dhillon, I. S., Mallela, S. and Kumar, R. (2003), A Divisive Information-Theoretic Feature Clustering Algorithm for Text Classification, *Journal of Machine Learning Research*, Vol. 3, 1265-1287.

Dionisio, A., Menezes, R. and Mendes, D. A. (2004), Mutual Information: A Measure of Dependency for Nonlinear Time Series, *Physica A*, Vol. 344, 326-329.

Harrold, T. I. Sharma, A. and Sheather, S. (2001), Selection of a Kernel Bandwidth for Measuring Dependence in Hydrologic Time Series Using the Mutual Information Criterion, *Stochastic Environmental Research and Risk Assessment*, Vol. 15, 310-324.

Hogg, R. V. and Tanis, E. A. (2001), *Probability and Statistical Inference*, Prentice Hall, Upper Saddle River.

Huang, D. and Chow, T. W. S. (2005), Effective Feature Selection Scheme Using Mutual Information, *Neurocomputing*, Vol. 63, 325-343.

Hutter, M. and Zaffalon, M. (2005), Distribution of Mutual Information from Complete and Incomplete Data, *Computational Statistics and Data Analysis*, Vol. 48, 633-657.

Kojadinovic, I. (2004), Agglomerative Hierarchical Clustering of Continuous Variables Based on Mutual Information, *Computational Statistics and Data Analysis*, Vol. 46, 269-294.

Kraskov, A., Stogbauer, H., Andrzejak, R. G. and Grassberger, P. (2003), http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=2003q.bio....11037K&db_key=PRE

Kraskov, A., Stogbauer, H. and Grassberger, P.(2004), Estimating Mutual Information, *Physical Review E*, Vol. 69, 066138/1-066138/16.

Rohatgi, V. K. and Saleh, A. K. (2001), *An Introduction to Probability and Statistics*, John-Wiley, New York.

Shiraishi, Y., Morii, M., Uyematsu, T. and Sakaniwa, K. (2003), Some Properties of a Nonlinear Function: Linear Complexity, Mutual Information, and Correlation

Immunity, *Electronics and Communications in Japan*, Vol. 86, 44-56.

Stogbauer, H., Kraskov, A., Astakhov, S. and Grassberger, P. (2004), Least Dependent Component Analysis Based on Mutual Information, *Physical Review E*, Vol. 70, 1-18.

Sugimachi, T., Ishino, A., Takeda, M. and Matsuo, F. (2003), A Method of Extracting Related Words Using Standardized Mutual Information, *Leture Notes in Computer Science*, Vol. 2843, 478-485.

Wang, Q., Shen, Y. and Zhang, J. Q. (2005), A Nonlinear Correlation Measure for Multivariable Data Set, *Physica D*, Vol. 200, 287-295.

Yin, X. (2004), Canonical Correlation Analysis Based on Information Theory, *Multivariate Analysis*, Vol. 91, 161-176.

Zhou, G., Yang, L., Su, J. and Ji, D. (2005), Mutual Information Independence Model Using Kernel Density Estimation for Segmenting and Labeling Sequential Data, *Leture Notes in Computer Science*, Vol. 3406, 155-166.