

因果律 및 變數間 關係 分析에 관한 研究

方 演 根*

〈目 次〉

- | | |
|----------------|-----------------------|
| I. 序 論 | III. 變數의 類型 및 因果關係 分析 |
| II. 因果律에 대한 立場 | IV. 結 語 |

I. 序 論

많은 경우 우리는 연구과제를 수행하면서 연구주제와 관련된 이론과 관심에 입각해 변수들을 설정하고 이들간의 관계를 분석하는 작업을 수행하게 된다. 또한 분석결과로부터 변수간에 원인과 결과를 확인하고 이를 통해 향후의 시사점을 제시하게 된다. 이러한 과정에서 대부분 명확히 밝히지는 않지만 우리는 어떤 입장의 인과율을 채택하게 되고 이에 근거하여 변수간의 관계를 설명하게 된다. 본고에서는 그동안 명확히 취급하지 않았던 인과율과 변수간의 관계를 분석하는 기본적인 방법을 살펴보고자 한다.

II. 因果律에 대한 立場

1. 인과율을 보는 수준

인과율(causality)을 논하기에 앞서 우리가 인지(cognition)할 수 있는 수준을 존재적 수준(ontic level), 인식적 수준(epistemic level), 개념적 수준(conceptual level)의 셋으로 나눈 코크 모형(Koch, 1994)을 설명하는 것이 향후의 논의 전개에 도움이 된다고 생각한다.

* 한국철도기술연구원 책임연구원

코크 모형은 우리의 지각(perception)과는 독립적으로 실재(reality) 세계가 존재한다고 가정하고 이를 인지의 존재적 수준(ontic level) 또는 존재하는 세계(ontic world)라 한다. 존재적 실재(ontic reality)는 우리의 감각이 접근할 수 있는 경험을 생산하는 것으로 본다. 이러한 경험들은 인지의 인식적 수준(epistemic level)을 형성한다. 개념적 수준(conceptual level)은 인식과 존재적 실재 사이를 중재할 때 사용된다. 개념적 수준은 인식에 영향을 미치기도 하지만, 역으로 개념은 경험된 인식에 기초하여 형성된다.

존재적 수준에서의 인과율은 무엇이 실재, 객관적인 세계에서 원인이 되는가, 원인이 되는 것으로 실재 세계에서 무엇을 취하는가 하는 것이 관심사가 된다. 개념적 수준에서의 인과율은 원인형성, 무엇이 원인이고 결과인가, 원인과 결과간의 관계 특성에 관해 현재 확립되어 있는 개념은 무엇인가가 관심사가 된다. 인식적 수준에서는 우리가 지금 까지 아는 한 실재 원인이 되는 것은 무엇인가, 원인과 결과라고 할 때 무엇을 관찰할 수 있거나, 발견할 수 있거나, 확립할 수 있거나, 합리적으로 믿을 수 있는가가 관심사가 된다(Mackie, 1985).

2. 인과율의 폐기

근대적인 의미에서의 인과율은 험(Hume, 1711-1776)에 의해 제시되었는데, 두 事象(events)이 시간적으로 계속하여 일어나고(succession in time) 공간적으로 근접(contiguity)하여 있으면 인과적으로 연계되어 있다고 본다. 또한 사람들은 원인과 결과간에 필연적인 연계(necessary connection)가 있다고 생각한다고 주장한다. 이러한 필연적인 연계를 암으로써 선형적(a priori)으로, 원인으로부터 결과를 추론하는 것이다. 그러나 이러한 필연적 연계가 존재하는지는 증명될 수 없어 가상적인 것(fictitious)으로 남아있는 실정이다. 따라서 인과율은 증명할 수 있는 사실의 문제가 아니라 인간의 지각(perception)의 문제라고 할 수 있다. 과학은 가상적 존재가 아니라 인지적 현상(epistemic phenomena)을 다루어야 한다고 주장하는 논리실증주의자(logical empiricist)들은 (Russell, 1913) 인과율이 더 이상 사용되어서는 안 된다고 주장하게 된다. 이들은 인과율대신에 상관관계(correlation)와 분할표(contingency tables)를 사용하고자 하였다.

3. 인과율의 부활

논리실증주의자들이 폐기 처분하였던 인과율은 물리학자 프랭크(Frank, 1884-1966)에 의해 그 유용성이 부활되는데, 그도 인과율을 실재 세계를 지배하는 근본적인 법칙으로 받아들이지는 않았지만, 우리 주변의 일상세계를 구조화하는 개념으로는 받아들이고 있다(Frank, 1988). 인과율에 대한 인간의 이해를 가상적인 필연적 연계로 설명하는 것이 아니라, 우리가 우리 자신을 변화의 원천(originator)으로 경험할 때 일어나는 행동의 성찰(self-reflection)에 기반하여 설명한다. “p가 q의 원인이다”라고 주장할 때의 의미는 p를 행함으로써 q를 초래할 수 있다는 것이다. 그러나 이러한 원인에 대한 정의는 일상 생활에서 우리가 사용하는 ‘원인’ 및 우리가 조작할 수 없는 事象을 포함하지 못한다는 비판을 받고 있다.

이러한 입장에서 일보 전진한 것이 인간 동기부여의 관점에서 원인을 개념화한 것이다 (Ströker, 1992). 이 관점에서 원인은 현상의 특성이 아니라 현상을 관찰하는 인간의 특성이 된다. 이 인과율 입장은 두 가지 동기요인을 갖고 있는데, 하나는 어떤 목적을 달성하기 위한 전략을 추구함으로써 경험적으로 인식적인 지식을 확장하는 것이고, 다른 하나는 현존하는 개념이 설명하지 못하는 현상에 대한 설명을 추구함으로써 개념적 이해를 확장하고자 하는 것이다.

이들은 필연적 연계(necessary connection)를 특정 조건법적 문장(counter-factual statement)으로 대체한다. 조건법적 문장의 眞偽를 전통적인 논리로 결정할 수 없는 문제를 극복하기 위해, 루이스(Lewis, 1973)는 가능한 세계(possible worlds)라는 아이디어를 고안하였다. 실재에 가장 근접한 세계를 구성하는데 근접성(contiguity), 시간 연속성(time succession), 조작의 가능성(possibility of manipulation) 등이 사용될 수 있고 이 중 몇 개는 반드시 사용되어야 한다(Mackie, 1985).

이들이 설명하는 인과율은 다음과 같다. 만약 not-c 사상 및 not-e 사상이 眞인 가능한 세계가 존재하고, 이 가능한 세계가 not-c 사상이 眞인 모든 가능한 세계 중에서 가장 실재에 가까운 세계라고 하면, 관찰된 두 개의 사상 c와 e는 인과적으로 의존한다고 하고, 이에 기초하여 만약 c가 없었다면, e도 존재하지 않았을 것이라고 결정한다. 만약 c(또는 비슷한 c들의 집합인 C)에 이어 e(또는 비슷한 e들의 집합인 E)가 발생하고, not-c(또는 Not-C)에 이어 not-e(또는 Not-E)가 발생한다면, c(또는 C)는 e(또는 E)의 원인

이고, 이는 c (또는 C)가 e (또는 E)에 대해 필요 충분조건이라는 것을 의미한다.

그러나 C가 E의 원인이라고 했음에도 불구하고 c가 e를 결과하지 않을 수도 있는 점(원인이 비슷하면 결과도 비슷할 것이나, 결과가 비슷하다고 하여 원인이 반드시 비슷하다고는 할 수 없다)을 반영하여, 인과율에 대한 확률적 정의가 사용된다(Suppes, 1970). 원인이 반드시 결과를 초래할 필요는 없으며, 다만 어떤 상황하에서도 결과가 나타날 확률만을 높이면 되는 것이다. 확률적 인과율 정의는 시간적 연속을 필요로 한다.

III. 變數의 類型 및 因果關係 分析

이러한 인과율에 대한 논의를 염두에 두고 먼저 연구대상이 되는 변수들간에 어떤 유형이 있을 수 있는지 살펴보기로 한다.

1. 변수의 유형

변수는 기본적으로 독립변수와 종속변수, 그리고 독립변수와 종속변수간의 관계를 변화시키는 조절변수(moderator)로 나뉘어진다고 할 수 있다.

조절변수의 존재여부는 체계적인 변동(systematic variation)에 대한 Chi-Square test를 통해 판단하기도 하는데(Hunter et al. 1982), Chi-square가 유의하지 않으면 독립변수와 종속변수간에는 조절변수가 없다고 본다. 즉, 독립변수와 종속변수간의 관계가 많은 표본에서 변화가 적어 일정하다고 하면 조절변수가 없다고 판단한다. 헌터가 제시하는 판별식은 다음과 같다.

$$x_{k-1}^2 = \frac{N}{(1-r^2)^2} S_r^2 \quad \dots \quad (1)$$

N : 총조사개체의 수

r^* : 표본크기를 반영한 평균 상관관계

S²_r : 상관관계의 분산

또한 조절변수가 미치는 영향력을 ANOVA(analyses of variance) 분석에서 독립변수와 독립변수가 처한 조건을 규정하는 요인간의 상호작용으로 표시된다(Baron & Kenny, 1986, p. 1174). 조절변수는 연구자에 의해 측정되고, 조작되고, 선택되는 변수로 언제 독립변수와 종속변수간에 상관관계가 존재하는지를 밝힌다.

독립변수와 종속변수간의 관계를 설명하는데 도움이 되나 초기 연구모형에 설정되지 않고, 관찰되지 않은 외생변수를 매개변수(mediator 또는 intervening variable)라고 하는데, 이는 종속변수 변화에 대한 기준의 독립변수의 설명력이 약할 경우, 또는 강하다 하더라도 다른 인과관계를 모색할 때 발견되는 종속변수와 강한 관계를 갖고 있는 변수이다. 매개변수를 다시 독립변수로 설정하여 연구를 진행한다. 매개변수를 통제할 때 독립변수와 종속변수의 관계가 사라지면, 매개변수는 독립변수와 종속변수를 연계하고 있는 변수라고 할 수 있다.

통제변수는 진정한 독립변수와 종속변수간의 관계를 밝히기 위해 통제되는 변수이다. 통제되는 변수에는 독립변수와 종속변수보다 시간적으로 앞서서 이들 두 변수에 영향을 미친다고 가정된 선행변수(antecedent variable)가 있다. 이 변수를 통제한 상황 하에서 독립변수와 종속변수간의 관계가 유지되지 않으면 독립변수와 종속변수로 설정된 두 변수간에는 인과관계가 없었고 괴상적 관계(spurious relationship)만 있었다고 판단한다. 통제변수의 일부 영역에서만 두 변수간에 독립변수와 종속변수의 관계가 성립되는 경우도 있다. 연구자가 종속변수에 대한 특정변수의 영향력을 제거하고자 할 때 이 변수를 통제변수로 설정한다.

독립변수, 조절변수, 통제변수가 종속변수에 영향을 미친다고 볼 수 있으며 종속변수와 이들 사이에 매개변수가 있을 수 있다.

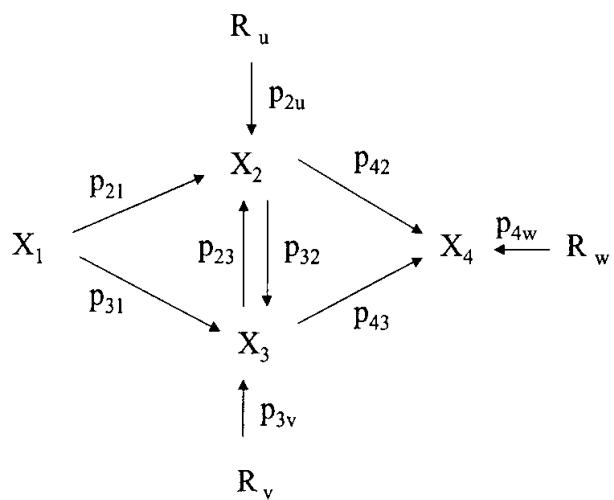
2. 인과관계 분석

인과관계모형은 여러 가지 형태로 사회과학에서 많이 사용되고 있으나, 본고에서 다루는 인과관계분석은 부문간 비교(cross-sectional) 및 등간격 척도를 사용하는 자료(interval-level data)를 바탕으로 하는 연구에 한정하고자 한다.

1) 인과관계 모형 설정

어떤 과제 또는 문제에 관해 인과적 접근방법을 사용하여 변수간에 인과관계를 설정하고 이를 화살표를 사용하여 그림으로 표시하는 것은 가설을 좀더 명확히 드러내고 연구 과제에 대해 추가적인 통찰을 얻기 쉽게 한다는 점에서 유용하다.

〈그림 III-1〉 인과관계 모형



위 그림에서 변수를 연결하는 화살표는 가설을 나타내는 것이고, 화살표의 방향은 인과관계를 나타내는데 이는 이론과 실재에 바탕을 두고 결정하게 된다. 가설은 변수간의 관계의 크기를 측정하여 검증하게 된다. 관계의 크기가 작을 경우에는 인과관계가 약하다고 추론하게 된다. 일반적으로 뒤에서 언급하게 되는 경로계수(path coefficient)의 값이 0.05 또는 0.1 이하인 경우 인과관계가 없다고 결정한다(Asher, 1983, p.25). 실제 관계의 크기를 측정하기 위해선 만족할만한 지표가 개발되고, 적절한 자료의 수집 등이 가능하여야 한다.

위 모형을 좀 더 설명하면, X_1 은 모형 내에서 측정되는 다른 변수들의 영향을 받지 않기 때문에 외생변수(exogenous variable)라고 규정한다. X_2, X_3, X_4 는 모형 내 다른 변수들의 영향을 받기 때문에 내생변수(endogenous variables)라고 규정한다. R_i 는

잔여변수(residual variables) 또는 오류항(error terms)이라고 하는데 실제 측정되는 않으나 내생변수에 영향을 미치는 요인들이다. p_{ij} 는 한 변수(X_i)의 다른 변수(X_j)에 대한 영향력을 나타내는 가중치가 된다. 이를 경로계수(path coefficients)라고 부른다.

각 내생변수에 대해서는 구조 방정식(structural equation)이 가능한데, 다음과 같이 표현한다.

$$X_2 = p_{21}X_1 + p_{23}X_3 + p_{2u}R_u \quad \dots \dots \dots \quad (2)$$

$$X_3 = p_{31}X_1 + p_{32}X_2 + p_{3v}R_v \quad \dots \dots \dots \quad (3)$$

구조방정식은 선형으로 설정하고 상수항을 포함하지 않는데, 이는 모든 변수 및 오류항을 표준화하기 때문이다. 경로계수 p_{ij} 는 모형에서 변수를 연계하는 인과구조를 나타낸다. 변수간에 상호 영향을 주고받는 것이 없는 순환(recursive model)의 경우에는 일반 최소자승 회귀(ordinary least square regression)가 경로계수에 대한 적절한 추정치를 제공하게 된다. 그러나 변수간에 상호작용이 있는 경우(reciprocal model) 그리고 잔여변수와 설명변수간에 상관관계가 있는 경우는 일반 최소자승 회귀법이 경로계수를 추정하는데 적합하지 않다(Asher, 1983, p.15).

2) 인과관계의 조건

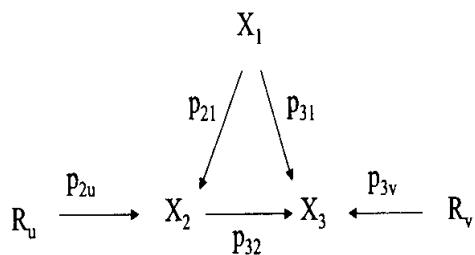
우리는 앞서 두 변수간에 인과관계가 존재한다고 추론할 수 있으려면, 두 변수가 공변동(covariation)을 하여야 하고, 두 변수간에 시간적인 순서가 있어야 한다는 것을 언급하였다. 그러나 여기에 추가하여야 할 조건 하나는 두 변수보다 앞선 선행변수의 영향력을 제거하더라도 두 변수간에 공변동이 없어지지 않아야 한다는 것이다.

그러나 현실적으로 두 변수의 모든 선행변수의 영향력을 제거한다는 것은 불가능하다. 따라서 우리가 이론적으로 경험적으로 중요하다고 생각할 수 있는 범위 내에서 이러한 작업이 이루어질 수밖에 없다. 단지 분석 결과 오류항(error terms)이 변동의 많은 부분을 설명하면, 중요한 변수가 누락되었다고 추론할 수 있을 뿐이다. 이런 점에서 인과 관계 추론은 한계를 지니고 있다고 할 수 있다.

3) 인과관계 제3조건의 분석

위에서 언급한 인과관계의 제3조건, 즉 선행변수의 영향력을 통제해도 두 변수간의 상관관계가 존재하여야 인과관계를 추론할 수 있다는 조건을 분석하기 위한 작업을 살펴보기로 한다. 만약 선행변수를 통제하였을 때 두 변수간의 상관관계가 사라지면 두 변수간의 관계는 피상적(spurious) 관계였다고 한다. 세 변수간에 다음과 같은 관계를 설정한다.

〈그림 III-2〉 3변수간 인과관계도



위 모형에서 내생변수인 X_2 , X_3 에 대한 구조 방정식은 다음과 같다.

$$X_2 = p_{21}X_1 + p_{2u}R_u \quad \dots \dots \dots \quad (5)$$

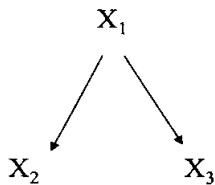
$$X_3 = p_{31}X_1 + P_{32}X_2 + P_{3v}R_v \quad \dots \dots \dots \quad (6)$$

위 모형도에서 잔여변수들은 서로 상관관계가 없고, 또한 각 구조방정식에서 설명변수와 잔여변수간에는 상관관계가 없다고 가정한다. 이러한 가정은 검증될 수 있는 것은 아니며, 이론이나 경험에 기초하여 설정할 뿐이다.

위에서 제시한 두 구조방정식에서 식(2)에 X_1 을 곱하여 기대값을 구하고, 식을 단순화시킨 다음, 다시 식(2)에 X_2 를 곱하여 기대값을 구하고 식을 단순화시키어 얻은 두 개의 식으로부터 p_{32} 을 계산하면 다음과 같다.

이 결과는 p_{32} 가 표준화된 회귀계수 $B_{32 \cdot 1}$ 과 같고, 표준화된 회귀계수의 분자는 부분상관계수(partial correlation coefficient) $r_{32 \cdot 1}$ 의 분자와 같다는 것을 보여준다. 따라서 부분상관계수의 값이 영이 될 때 인과관계가 없어진다고 할 수 있다는 것을 알 수 있다. 위 식으로부터 p_{32} 가 영이 되는 조건은, 즉 인과관계가 없어지는 조건은 분자가 영이 되는 $r_{23} = r_{12} \cdot r_{13}$ 인 경우이다. 이를 그림으로 표시하면 다음과 같이 두 가지 경우가 발생한다.

〈그림 III-3〉 피상적 관계



〈그림 III-4〉 발전적 관계



피상적 관계(spurious relation)와 발전적 관계(developmental relation) 중 어느 경우가 선택되는가 하는 것은 즉, X_2 가 X_1 에 영향을 미치는가 아니면 X_1 이 X_2 에 영향을 미치는가는 통계적으로 해답이 나올 수 있는 사항이 아니고, 이론적 또는 경험적 통찰에 의해 결정될 사항이다.

IV. 結 語

연구에 필요한 조사분석을 하면서 우리가 선정하여야 할 변수들의 범주, 즉 유형들은 무엇이 있으며, 이들간의 관계를 인과적으로 설명할 때, 어떤 인지 수준에 입각하고 있는지, 그리고 그 관계가 피상적인 관계인지 여부를 파악하는 것은 가장 기본적인 연구자의 일이라고 할 수 있을 것이다. 연구를 둘러싼 논의를 진행함에 있어 그 토론의 질을

높이고 어느 정도 정형화하는데 기여했으면 하는 것이 필자의 바램이다. 더욱 자세한 논의는 연구방법론을 본격적인 연구대상으로 할 때 적합하다고 판단하여 본고에서는 이에 관한 부분을 제외하였다.

참 고 문 헌

- Asher H. B.(1983). *Causal Modeling*, 2nd. ed, Sage Publications, p. 15.
- Baron & Kenny(1986). "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology*, Vol. 51, pp. 1173- 1182.
- Hume, D.(1995). "Enquiry Concerning Human Understanding." In James Fieser(ed.): *The Writings of David Hume*, Internet Release.
- Frank, P.(1988). *Das Kausalgesetz un seine Grenzen*. Suhrkamp, Frankfurt.
- Koch, G.(1994), "Kausalitat, Determinismus und Zufall in der wissenschaftlichen Naturbeschreibung." In the series, *Erfahrung und Denken*, Band 75, Berlin: Duncker & Humblot.
- Mackie, J. L.(1985). "Logic and Knowledge, Selected Papers," *Causation in Concept, Knowledge and Reality*, Vol. 1, pp. 178-191.
- Russell, B.(1913), "On the Notion of Cause," *Proceedings of the Aristotelian Society, New Series*, Vol. 13, pp. 1-26.
- Sondhauss U. M.(1998), "Influence of Philosophical Concepts of Causality on Causal Modelling in Statistical Research," presented at the Workshop, *Causal Networks: from Inference to data mining* CaNew '98.
- Ströker, E.(1992). "Warum-Fragen. Schwierigkeiten mit einem Modell für kausale Erklärungen," In R. Bubner (ed.), *Kausalität*, Neue Hefte für Philosophie, 32/33, Vandenhoeck & Ruprecht, Gottingen, pp. 105-129.

Suppes, P.(1970), *A Probabilistic Theory of Causality*. Amsterdam:
North-Holland.