

# Causal inference from nonrandomized data: key concepts and recent trends

Young-Geun Choi<sup>a,1</sup> · Donghyeon Yu<sup>b</sup>

<sup>a</sup>Data R&D Center, SK Telecom; <sup>b</sup>Department of Statistics, Inha University

(Received December 03, 2018; Revised January 21, 2019; Accepted January 22, 2019)

---

## Abstract

Causal questions are prevalent in scientific research, for example, how effective a treatment was for preventing an infectious disease, how much a policy increased utility, or which advertisement would give the highest click rate for a given customer. Causal inference theory in statistics interprets those questions as inferring the effect of a given intervention (treatment or policy) in the data generating process. Causal inference has been used in medicine, public health, and economics; in addition, it has received recent attention as a tool for data-driven decision making processes. Many recent datasets are observational, rather than experimental, which makes the causal inference theory more complex. This review introduces key concepts and recent trends of statistical causal inference in observational studies. We first introduce the Neyman-Rubin's potential outcome framework to formalize from causal questions to average treatment effects as well as discuss popular methods to estimate treatment effects such as propensity score approaches and regression approaches. For recent trends, we briefly discuss (1) conditional (heterogeneous) treatment effects and machine learning-based approaches, (2) curse of dimensionality on the estimation of treatment effect and its remedies, and (3) Pearl's structural causal model to deal with more complex causal relationships and its connection to the Neyman-Rubin's potential outcome model.

Keywords: causal inference, average treatment effect, conditional treatment effect, propensity score, structural causal model

---

## 1. 서론

과학적 연구의 연구 주제 및 가설은 많은 경우 인과적 질문(causal question)을 포함한다. 예를 들어 “전염병의 예방을 위한 특정 치료법은 얼마나 효과적인가?”, “특정 건설정책으로 부동산 가격의 상승을 막을 수 있었는가?” 혹은 미시적으로 “지금 웹사이트를 방문한 고객에게 어떤 광고를 보여주어야 가장 구매 확률이 높을 것인가?” 등의 질문을 생각할 수 있다.

Fisher RA가 실험계획법을 체계화한 이래로, 인과관계를 규명하는 가장 확실한 방법은 임의실험이었다. 하지만 현실에서 다루게 되는 대부분의 자료는 임의실험보다는 비실험 관찰로 수집된 자료이고, 이에 대한 인과성 추론을 위한 이론 개발이 요구되었다. 비실험 관찰자료로부터의 인과성 추론은 흡연과

---

This work was supported by Inha University Research Grant (INHA-55456).

<sup>1</sup>Corresponding author: Data R&D Center, SK Telecom, Euljiro 65, Jung-gu, Seoul 04539, Korea.

E-mail: [yg.choi@sk.com](mailto:yg.choi@sk.com)

폐암의 인과관계 논쟁에서 급격하게 발전하였다 (Cornfield 등, 2009). 특히 Rubin D와 그의 연구 그룹은 1970년대 이래로 관찰자료로부터의 인과적 효과의 측정을 가능케 하는 개념들을 확립하여 왔다 (자세한 종설은 Rubin (2005)에서 확인할 수 있다). 하지만 관찰자료로부터 인과적 관계가 추론이 가능한지에 대한 논쟁은 여전히 계속되고 있다 (Dawid, 2000).

인과 추론은 임상실험과 정책결정에서 주로 이용되었으나, 이른바 빅데이터 시대의 도래로 가용한 관측 자료가 폭발적으로 증가함에 따라 인과 추론에 대한 잠재적 응용가치와 수요 또한 지속적으로 증가하고 있다. 연간 인과 추론에 관련된 논문도 1983년 38편에서 2013년에는 3844편으로 크게 증가하였고, 보건학과 경제학에서는 이미 인과 추론의 이론과 응용 논문이 통계학 학술지에서보다 더 많은 논문이 출판되고 있다 (An과 Ding, 2018). 최근 인과 추론의 개념들은 마케팅과 개별 맞춤형료에서의 개인별 효과 측정 및 실험대상 선정의 최적화에 활용되고 있다 (Chan 등, 2010; Chakraborty와 Murphy, 2014; Varian, 2016).

본 종설논문의 목적은 관찰자료로부터의 인과적 추론을 위한 핵심적 개념들과 최근 연구 동향을 소개하는 데 있다. 본 논문의 구성은 다음과 같다. 2절에서는 인과적 질문을 평균처치효과로 수식화하기 위하여 Neyman-Rubin 잠재 결과(potential outcome) 모형을 소개한다. 3절에서는 평균처치효과를 추정하는 방법으로 성향 점수 기반 추정법과 회귀분석 기반 추정법을 소개한다. 4절에서는 최근 연구 동향 중 주목할 만한 흐름 중 조건부 처치효과의 머신러닝 기반 추정방법들, 고차원 자료에서의 ‘차원의 저주’가 처치효과 추정에 미치는 영향 및 극복 방안들, 그리고 Pearl의 구조적 인과 모형을 논한다. 5절에서는 간단한 설정 하에서 (조건부) 평균처치효과의 계산을 직접 R로 구현하여 본다.

## 2. 기호와 가정: Neyman-Rubin 잠재 결과 모형

인과 추론의 핵심적인 가정을 논하기 위하여, 가장 단순한 설정에서 시작한다. 처치가 적용되기 전의 공변량(pre-treatment covariate)은  $X \in \mathbb{R}^p$ 로, 공변량과는 별도로 의사결정권자가 개입(intervention)하여 바꿀 수 있는 확률변수는 처치(treatment) 혹은 행위(action)로 정의하여  $A$ 로 나타낸다. 통제된 임의실험의 경우  $A = 1$ 을 실험군에,  $A = 0$ 을 대조군에 각각 대응시킬 수 있다. 통제된 임의실험의 자료가 아니더라도, 추후에 자료를 바탕으로 최적의 개입행위를 결정하여야 하는 변수에  $A$ 가 대응될 수 있다. 예를 들어 정책(policy), 광고 및 처방의 선택이  $A$ 로 표현될 수 있다. 한편 개입행위 후에 관찰된 관심대상 결과값(outcome)은  $Y$ 로 나타낸다.  $(X, A, Y)$ 는 알려지지 않은 분포  $\mathbb{P}$ 를 따를 때, 우리는 확률표본  $\{(X_i, A_i, Y_i)\}_{i=1}^n$ 를 관찰하였다고 가정한다. 예를 들어 진료정보  $X_i$ 를 가진 환자에 대하여 의사가 임의로 특정 약을 복용시키지 않거나( $A_i = 0$ ) 복용시킨( $A_i = 1$ ) 후의 예후( $Y_i$ )를 관찰하였다고 상정할 수 있다.

인과적 질문을 해석하는 방식 중 하나는 반사실적 결과(counterfactual outcome)들을 비교하는 것인데, Neyman-Rubin의 잠재 결과 모형으로 구체화될 수 있다. 예를 들어 “복지정책은 거주민들의 소득을 향상시키는가?”가 관심 질문이라고 하자. 이에 대답하는 이상적인 방법은, 동일한 개체 둘을 상정하여 각각을 실험군(복지정책을 적용함)과 대조군(복지정책을 적용하지 않음)에 배정하고, 두 개체의 추후 소득을 비교하는 것이다. 잠재 결과 모형에서는, 단일 개체  $i$ 에 대하여 만약 실험군과 대조군에 배정하였을 경우의 결과값을 각각  $Y_i(0)$ ,  $Y_i(1)$ 로 나타낸다. 주어진 정책은 아래와 같이 정의되는 평균처치효과(average treatment effect; ATE)에 의하여 평가될 수 있다:

$$\tau := \mathbb{E}[Y(1) - Y(0)]. \quad (2.1)$$

처치효과  $Y(1) - Y(0)$ 는 한 개체가 실험군으로 배정된 상황의 결과값과 대조군으로 배정된 상황의 결과값의 차이를 나타내며, 그 차이의 모집단에 대한 기댓값이 평균처치효과이다.

하지만 현실에서 단일 개체의 복제는 불가능하다. 따라서 각  $i$ 에 대하여  $Y_i(0)$  또는  $Y_i(1)$ 만 관찰된다. 나머지 하나는 현실에서는 결측이나 “만약 처치를 바꾸었다면 관찰하였을 수 있는” 반사실적 결과값이 된다. 이러한 구조적 결측 상황에서도  $\tau$ 의 추정이 가능하려면 아래의 가정들이 필요하다.

- 무시가능성(ignorability) 혹은 교란변수 비측정 불가능성(no-unmeasured-confounders) 가정:  $\{Y(0), Y(1)\} \perp A|X$ .
- 양수성(positivity) 혹은 중첩성(overlapping): 어떤 양수  $c$ 가 존재하여, 모든  $x$ 에 대하여  $\mathbb{P}(A = 1|X = x) \geq c$ .
- 일치성(consistency):  $Y = Y(A)$ .

상기 가정들의 직관적 의미는 “관찰된 데이터가 국소적으로는 임의실험의 특성을 가짐”이다. 무시가능성은 공변량이 동일한 개체들에 대하여는 이미 잠재 결과(잠재 결과의 분포)가 실험 분배 기전에 상관 없이 정해져 있음을 뜻한다. 만일 임의실험에서 데이터를 얻었다면 무시가능성이 만족된다고 주장할 수 있으나, 관찰자료로부터의 무시가능성을 증명하기는 불가능하다. 양수성은 임의의  $X = x$ 를 공변량으로 취하는 개체들에 대하여 처치군과 대조군이 고르게 분배되어 있음을 뜻한다. 역시 임의실험에서는 만족된다고 주장할 수 있으나, 관찰자료에 대하여는 가정이 만족되는지 검토가 필요하다. 일치성은 처치분배 기전이 오류없이 기록되었음을 보장한다. 한편, 무시가능성과 양수성에 의하여 평균처치효과를 아래와 같이 조건부 기댓값으로 나타낼 수 있다:

$$\begin{aligned}\tau &= \mathbb{E}[Y(1)|A = 1] - \mathbb{E}[Y(1)|A = 0] \quad (\text{무시가능성}) \\ &= \mathbb{E}[Y|A = 1] - \mathbb{E}[Y|A = 0] \quad (\text{일치성}).\end{aligned}\tag{2.2}$$

### 3. 평균처치효과의 추정법

평균처치효과를 추정하는 방법에는 크게 (1) 성향점수(propensity score) 기반 추정법과 (2) 회귀분석 기반 추정법이 있으며, 두 방법을 혼합한 이중 로버스트 방법도 존재한다.

#### 3.1. 성향점수 기반 추정법

먼저 성향 점수에 의한 보정의 필요성을 논의한다. 직관적으로 식 (2.1)의  $\tau$ 는 실험군과 대조군의 결과값들의 차이이므로, 관찰된 값들의 평균을 이용하여  $\tau$ 를 추정할 수 있다고 생각할 수 있다. 예를 들어, 다음의 추정량을 고려할 수 있다:

$$\hat{\tau}^{\text{Naive}} = \frac{1}{n_0} \sum_{i=1}^n A_i Y_i - \frac{1}{n_1} \sum_{i=1}^n (1 - A_i) Y_i,$$

여기서  $n_k$ 는 ( $k = 0, 1$ ) 대조군/실험군 배정된 개체의 수 ( $n_k = \sum_i I(A_i = k)$ )이다.  $\hat{\tau}^{\text{Naive}}$ 에 의한 추정은 임의실험에서 처치효과를 추정하는 가장 전형적인 방법이다. 그러나 관찰연구 자료에서  $\hat{\tau}^{\text{Naive}}$ 로부터  $\tau$ 를 추정한다면, 특정 공변량을 가지는 개체가 표본에서 다수를 차지할 경우 해당 집단의 결과값이 과대표될 수 있는 점에서 편향된 추정을 야기할 수 있다. 따라서 표본론의 사후층화 절차와 같이 적절한 가중치 보정이 필수적이다. 인과 추론에서 가중치 보정에 대응되는 개념은 “성향 점수 보정”인데, 구체적인 절차는 아래와 같이 시행한다.

성향점수는 주어진 공변량에 대하여 실험군에 대응될 확률로 정의되고  $\pi(x) := \mathbb{P}(A = 1|X = x)$ 으로 나타낼 수 있다. 만일 참 성향점수를 알고 있다면, 평균 처치효과의 불편추정량(unbiased estimator)은

$(1/n) \sum_{i=1}^n \{(A_i/\pi_i)Y_i - (1-A_i)/(1-\pi_i)Y_i\}$ 에 대응한다. 여기서  $\pi_i = \mathbb{P}(A_i = 1|X_i = x_i)$ 이다. 물론 참 성향점수는 대개의 경우 알려져 있지 않으므로, 참 성향점수에 대한 적당한 모형가정과 추정 과정을 거치게 된다. 이를테면 다음의 로지스틱 모형을 가정하고,

$$\pi_i = \frac{\exp(x_i^T \alpha)}{1 + \exp(x_i^T \alpha)}, \quad (3.1)$$

최대 가능도 방법으로 추정량  $\hat{\pi}_i$ 을 고려할 수 있다. 성향점수를 추정하였다는 가정 하에, 평균처치효과에 대하여 다음과 같은 역확률 추정량(inverse probability estimator; IPWE)을 고려할 수 있다.

$$\hat{\tau}^{\text{IPWE}} = \frac{1}{n} \sum_{i=1}^n \left( \frac{A_i}{\hat{\pi}_i} Y_i - \frac{1-A_i}{1-\hat{\pi}_i} Y_i \right). \quad (3.2)$$

관찰연구 데이터로부터 안정된 성향점수를 추정하는 노력은 지속되었다. 앞서 서술된 가능도 기반 성향점수 추정 외에 가장 대중적으로 사용되던 방법으로는 짝짓기(matching) (Rosenbaum과 Rubin, 1985; Abadie와 Imbens, 2006)와 층화(stratification) (Rosenbaum과 Rubin, 1984)이다. 짝짓기의 기본적인 아이디어는 실험군과 대조군에서 공변량이 비슷한 개체들만 짝짓고, 짝지어진 쌍들의 결과 값(outcome)을 비교하는 것이다. 짝짓기 방법의 장점으로는 처치효과 측정에서 교란변수의 영향에 덜 민감하다는 점이 있고, 단점으로는 동일구간의 분산을 얻기 위하여 더 많은 표본을 필요로 할 수도 있다. 층화 방법은 공변량의 공간을 몇 수준의 영역으로 나누어 해당 영역에 속한 실험군과 대조군의 효과 크기를 비교한다. 층화 방법의 장점으로는 작은 크기의 표본에 대하여도 처치효과의 계산이 가능하다는 점이 있으나, 단점으로는 공변량으로부터의 교란이 완벽히 통제되지 않고 편향된 추정량을 내놓을 가능성이 크다. 두 방법에 관한 자세한 증설은 Lunceford와 Davidian (2004), Imbens (2004), Caliendo와 Kopeinig (2008) 및 Stuart (2010)에 소개되어 있으며, 국내 문헌 중에는 Huh (2014)의 3장에도 간단히 소개되어 있다. 관련된 R 패키지로는 MatchIt를 참조할 수 있다 (Ho 등, 2011). 한편, 최근의 관찰 자료에서는 짝짓기와 층화 방법론에서의 한계점들을 극복할 수도 있을 정도로 표본크기가 충분히 큰 경우가 많으나, 공변량의 개수도 많아 차원의 저주 문제가 새롭게 발생하게 된다. 고차원 공변량 관찰자료에서의 인과 추론에 관련된 최근의 연구 동향은 4.2절에서 자세히 논의하도록 한다.

### 3.2. 성향점수의 추정

식 (3.2)에서처럼 역확률가중치 방법으로 평균처치효과를 구할 경우에는 추정량의 성능이 성향점수 모형의 선택과 추정 방법에 따라 크게 좌우된다. 본 절에서는 성향점수의 추정에 대한 최근의 논의들을 요약한다.

Imai와 Ratkovic (2014)는 최대가능도 기반의 성향점수 대신, 공변량의 분포를 균형되게 만드는 성향점수(covariate balancing propensity score; CBPS)의 사용을 제안하였으며 R 패키지 CBPS에 해당 절차를 구현하였다 (Fong 등, 2018). CBPS는 로지스틱 모형 (3.1) 하에서, 아래 방정식의 해  $\hat{\alpha}^{\text{CB}}$ 를 이용하여  $\pi_{\hat{\alpha}^{\text{CB}}}(x)$ 를 성향점수의 추정량으로 사용한다:

$$\frac{1}{n} \sum_{i=1}^n \left( \frac{A_i}{\pi_{\alpha}(X_i)} X_i - \frac{1-A_i}{1-\pi_{\alpha}(X_i)} X_i \right) = 0.$$

위 식은 식 (3.2)에서  $Y$ 를  $X$ 로 대체한 것처럼 해석할 수도 있는데, 보다 근본적인 원리는 로지스틱 모형의 최대가능도 추정량  $\hat{\alpha}^{\text{MLE}}$ 가 다음의 스코어 방정식을 만족시키는 사실에서 온 것이다:

$$\frac{1}{n} \sum_{i=1}^n \left( \frac{A_i}{\pi_{\alpha}(X_i)} \frac{\partial}{\partial \alpha} \pi_{\alpha}(X_i) - \frac{1-A_i}{1-\pi_{\alpha}(X_i)} \frac{\partial}{\partial \alpha} \pi_{\alpha}(X_i) \right) \Bigg|_{\alpha=\hat{\alpha}^{\text{MLE}}} = 0.$$

Imai와 Ratkovic (2014)는 위 식을 참 성향 점수의 “균형화 성질(balancing property)”라 명명하였는데, 이 관점에서  $\hat{\alpha}^{CB}$ 는 위 식의  $(\partial/\partial\alpha)\pi_\alpha(X_i)$ 을  $X_i$ 로 대체한 버전에서 균형화 성질을 만족시키는  $\alpha$ 라고 말할 수 있다. CBPS 방법은 성향점수가 바르게 적시되지 않은 상황에서 최대가능도 기반 방법보다 로버스트한 효과크기 추정량을 제공함이 경험적으로 알려진 바 있다. 이후 Fan 등 (2016)은  $\hat{\alpha}^{CB}$ 의 일치성을 논의하였으며, Ning 등 (2017)은 균형 방정식을 고차원 자료로 확장하였다. 한편, Zhao (2019)과 Zhao와 Percival (2017)은  $\pi(x) = \mathbb{P}(A = 1|X = x)$ 에 로지스틱 모형이 아닌 다른 모형을 고려한 경우에 균형 방정식의 확장 방법을 논의하였다.

성향점수가 낮거나 너무 높은 개체의 존재가 존재할 경우 양수성 가정에 위배될 가능성이 커지게 되며, 아무리 성향점수 모형을 맞게 적시하여도 추정량의 분산이 큰 값을 취하게 되고 유한표본 성능이 저하된다 (Kang과 Schafer, 2007; Ponomareva와 Powell, 2010). 따라서 효과 크기 추정 절차를 수행하기 이전에 먼저 가지치기(trimming)를 고려할 수 있는데, 추정된 성향 점수가 특정 구간, 이를 테면  $[a_1, a_2]$ 에 속하는 표본들만 남겨 이후 분석을 시행하는 것이다 (Crump 등, 2009). Yang과 Ding (2018)은 가지치기 절차에서 추가되는 불확실성을 최종 추정량의 분산에 반영하였다.

### 3.3. 회귀분석 기반 추정법

결과값에 대한 회귀분석 기반 방법은 직관적인 이해가 용이하다. 논의의 편의성을 위하여 다음과 같은 선형모형을 고려한다:

$$\mathbb{E}(Y|A, X) = A\gamma + \beta_0 + X^T\beta_1 + AX^T\beta_2.$$

식 (2.2)에 의하여  $\tau = \mathbb{E}[Y|A = 1] - \mathbb{E}[Y|A = 0]$ 이므로, 위 선형 모형에 맞추어 정리하면 아래와 같은 식을 얻을 수 있다.

$$\tau = \gamma + \mathbb{E}(X)^T\beta_2.$$

따라서, 적절한 표준화 절차를 이용하여  $\mathbb{E}(X) = 0$  가정을 만족시키면, 선형 모형이 참이라는 전제 하에  $\gamma$ 는 평균처치효과  $\tau$ 와 동일하며, 따라서 기준에 정립된 회귀계수 추정법들을 사용하여  $\hat{\gamma}$ 를 평균처치효과와의 추정량으로써 사용할 수 있다. 더 일반적으로는,  $\mathbb{E}(Y|A = a, X = x) = Q(a, x)$ 으로 서술하였을 때,  $Q$ 의 모형족을 올바르게 선택다면 (모형족이 참 생성모형을 포함하고 있다면)  $Q$ 의 추정량을 통하여 평균처치효과  $\tau$ 를 추정할 수 있다. 회귀분석 기반 방법은 성향점수의 모형화와 추정을 필요로 하지 않는 장점이 있으나, 반대급부로  $Q$ 의 모형족을 적시하여야  $\tau$ 를 일치되게 추정할 수 있다.

성향점수 기반 추정법은 성향점수 모형의 적시를 요구하며, 회귀분석 기반 추정법은 결과값에 대한 모형의 적시를 요구한다. Robins 등 (1994)와 Bang과 Robins (2005)는 두 모형을 혼합하는 이중 로버스트 추정량(doubly robust estimator)을 제안한 바 있다. 이중 로버스트 추정량은 성향점수 모형족 혹은 결과값 회귀모형족 둘 중 하나라도 참 모형을 포함하고 있을 경우에는 일치성을 가진 추정량을 얻을 수 있는 장점이 있다. 또한 두 모형족 모두 참 모형을 포함하고 있는 경우에는 제안된 추정량이 효율성(efficiency)을 달성할 수 있음이 증명되었다. 그러나, 성향점수 및 결과값 회귀모형 둘 다 올바르게 적시된 경우에는, 이중 로버스트 추정량이 성향점수 기반 추정량이나 회귀분석 기반 추정량보다 더 편향된 추정치를 제공할 수 있음이 예증된 바 있다 (Kang과 Schafer, 2007).

정리하자면, 성향점수/회귀분석 기반 방법론 및 이중 로버스트 방법론이 안정적인 추정량을 제공하기 위해서는 결과값 함수  $Q$  및 성향점수  $\pi$ 의 모형족을 올바르게 적시하여야 한다. 최근에는 유연한 모형족을 가정하고 임의의 머신러닝 기법으로  $Q$  또는  $\pi$ 를 추정하여 방법론이 고려되어 왔는데, 관련 연구 동향은 4.1절에서 다시 논의하도록 한다.

## 4. 최근 연구 동향

### 4.1. 조건부 평균 처치효과 및 머신러닝의 활용

전통적으로 평균 처치효과는 신약 혹은 공공정책이 모집단에 평균적으로 제공하는 효용을 추정하는 지표였다. 개별 맞춤처방(personalized treatment regime)이나 개별 광고 타겟팅을 위하여는, 개인별로 기대되는 처치효과를 다르게 추정하고 다른 결정을 내리는 것이 더 자연스럽다. 소개된 기호들을 조합하면, 다음과 같이 정의되는 조건부 평균 처치효과(conditional average treatment effect; CATE) 및 개별 맞춤 법칙(individualized treatment rule; ITR)을 고려할 수 있다:

- 조건부 평균 처치효과(CATE):

$$\begin{aligned}\tau(x) &:= \mathbb{E}[Y(1) - Y(0)|X = x] \\ &= \mathbb{E}[Y|A = 1, X = x] - \mathbb{E}[Y|A = 0, X = x] \quad (\because \text{무시가능성/일치성})\end{aligned}\tag{4.1}$$

- 개별 맞춤 법칙(ITR):

$$\delta(x) = \begin{cases} 1, & \text{if } \tau(x) \geq 0, \\ 0, & \text{if } \tau(x) < 0. \end{cases}$$

식 (4.1)의 가장 오른쪽 수식을 참고하면,  $\tau(x)$ 를 추정하는 가장 직관적인 방법은  $\mathbb{E}[Y|A = a, X = x] = Q(a, x)$ 의 모형화에 기반한 회귀분석 기반 추정임을 알 수 있다. Qian과 Murphy (2011)은  $Q$ 에 대한 별점화된 선형 모형 기반에 기반하여 추정량  $Q$ 를 얻은 뒤  $\hat{\tau}(x) = \hat{Q}(1, x) - \hat{Q}(0, x)$ 로 추정할 것을 제안하였다. Zhao 등 (2017)는 Robinson (1988)의 공식에 기반하여  $\hat{\tau}(x)$ 을 별점화 선형모형으로 추정하고 신뢰구간을 선택적 추론(selective inference) 관점에서 계산하였다. Robinson의 공식은  $Q$ 를 추정하지 않고도  $\tau(x)$ 를 직접 추정할 수 있는 근거를 제공하나 대신 성향점수  $\pi(x)$ 의 모형화를 필요로 하는 단점이 있다. 한편,  $\delta(x)$ 를 추정하는 방법으로는  $Q(a, x)$ 나  $\tau(x)$ 의 추정 후 간접적으로 계산하는 방법도 있으나,  $\delta(x)$ 가 취하는 값은 이진값이므로 이진분류 문제의 관점에서 추정하였을 때 경험적으로 성능 향상의 여지가 있다 (Zhao 등, 2012; Zhang 등, 2012). 그러나 이진분류 문제로 재서술하여  $\delta(x)$ 를 추정하는 경우에도  $\pi(x)$ 의 모형화를 필요로 한다. 정리하면, 어떠한 방법을 적용하더라도  $\pi(x)$ 와  $Q(a, x)$  둘 중 하나에 대한 모형화는 반드시 필요하다.

최근에는 비모수적 모형과 심층 신경망 등으로  $\pi(x)$ 와  $Q(a, x)$ 에 최대한 유연한 함수공간을 상정하는 시도들이 제안되어 왔다. Künzel 등 (2017)은 랜덤포레스트나 Bayesian additive regression tree (BART) 등의 임의의 머신러닝 기법들을 적용하여  $\hat{\tau}(x)$ 를 추정할 때의 최소최대 수렴속도(minimax convergence rate)를 계산하였다. Wager와 Athey (2018)은  $Q(1, x)$ 과  $Q(0, x)$ 를 각각 랜덤포레스트로 따로 추정하는 방법을 제안하였고, 부가적으로 랜덤포레스트 기반 분류기에 대한 일반적인 신뢰구간을 유도한 뒤, 이를 바탕으로  $\hat{\tau}(x)$  및  $\hat{\tau}$ 의 신뢰구간을 제공하였다. Belloni 등 (2017), Chernozhukov 등 (2017) 및 Nie와 Wager (2017)은 임의의 머신러닝 기법을 기반으로 (조건부) 평균 처치효과를 계산할 때, 후술할 교차 추정 방법을 이용하여  $\hat{\tau}(x)$  및  $\hat{\tau}$ 에 대한  $\sqrt{n}$ -효율적인 신뢰구간을 계산하였다. 한편, M. van de Laan의 연구그룹은 임의의 머신러닝 기법으로 성향점수와 회귀분석 모형 모두를 추정한 뒤 이른바 “targeted learning”의 절차를 거쳐 (조건부) 처치효과를 추정하며, 역시  $\sqrt{n}$ -효율적인 신뢰구간을 제공한다 (van der Laan과 Rose, 2011, 2018). R 패키지 `drtmle`에는 targeted learning을 이용한 (조건부) 평균처치효과의 이중 로버스트 추론 방법론이 구현되어 있다 (Benkeser, 2018).

#### 4.2. 차원의 저주

인과 추론에서도 차원의 저주는 여전히 발생한다. 직관적으로는, 가능한 모든 종류의 교란변수를 공변량  $X$ 에 포함하면 무시가능성 가정을 만족시킬 가능성이 더 높아질 수도 있다. 문제는 양수성 가정의 위반인데, 공변량의 차원이 증가할수록 단위 개체의 근접 이웃 개체들을 찾기가 힘들게 되고, 따라서 어떤 공변량  $x_0$ 에 대하여  $\mathbb{P}(A = 1|X_0 = x_0) = 1$  또는 0일 가능성이 더 커지게 된다. D'Amour 등 (2017)는 강한 양수성이 성립하고 공변량이 정규분포를 따르는 경우의 모집단의 성질을 연구하였고, 성향점수  $\mathbb{P}(A = 1|X)$ 에 대한 강한 가정 (이러면면 회소성 가정)이 없는 한,  $p \rightarrow \infty$ 일수록 비현실적인 성질들이 성립함을 증명하였다.

공변량이 고차원일 경우의 관찰연구는 주로 참 성향점수가 최소 로지스틱 선형 모형을 가정하는 경우 이루어졌다. Farrell (2015)은 성향점수를 벌점화된 가능도 방법으로 추정한 뒤 평균처치효과에 대한 이중 로버스트 추정량 및 신뢰구간을 계산하였고, Belloni 등 (2017), Chernozhukov 등 (2017a) 및 Nie와 Wager (2017)는 조건부 평균처치효과 및 더 다양한 추정대상에 일반적인 추론 절차를 제안하였다.

고차원 공변량이 존재할 때의 처치효과 추론은 기본적인 고차원 추론 이론의 확장을 요구한다. 먼저 일반적인 고차원 회소선형모형에서 관심대상 회귀계수에 대한 점근적 신뢰구간을 구축하는 문제를 고려하여 보면, 관심대상 회귀계수에 대한 스코어 방정식에서 비관심대상( nuisance) 모수 추정으로부터의 변동이 제거되어야 함이 지적된 바 있다 (Van De Geer 등, 2014; Ning과 Liu, 2017). 이와 같은 사고의 흐름을 인과 추론에 적용한다면, 관심 대상의 모수는 (조건부) 평균처치효과가 되고 성향 점수 모형의 모수는 비관심대상( nuisance)의 모수가 되므로, 평균처치효과와 점근적 신뢰구간의 구축 과정에서 성향 점수 추정으로부터의 변동을 제거하는 추가적인 보정 절차가 필요하게 된다. 그러나 성향점수와 평균처치효과는 서로 다른 손실함수로부터 얻어지는 통계량이므로, 고차원 선형모형에서의 회귀계수들처럼 관심대상/비관심대상 모수들이 동일한 손실함수로부터 추정되는 상황과는 다르고, 더 복잡한 계산을 필요로 한다.

이와는 다르게 Belloni 등 (2017), Chernozhukov 등 (2017) 및 Nie와 Wager (2017)은 교차 추정을 고려하였다. 교차 추정은 표본을 두 집합으로 나누어, 한 집합은 성향점수의 추정에만 사용하고, 다른 집합으로 처치효과 계산 절차를 수행한다. 그리고 두 집합의 역할을 바꾸어 동일 절차를 수행한 후, 추정된 두 개의 처치효과 통계량의 평균을 최종 추정량으로 사용한다. 교차 추정 절차를 사용하는 성향점수 추정으로부터의 변동이 처치효과 추정으로부터의 변동이 독립이 되어, 상기 언급된 보정절차가 간단해지는 장점이 있다.

#### 4.3. 구조적 인과 모형

사회과학에서 인과관계의 효과를 연구하는 전통적인 주요 연구방법은 구조방정식(structural equation model)이다. 예를 들어 Figure 4.1을 대표하여  $Z = U_Z$ ,  $X = Z\alpha + U_X$ ,  $Y = \beta/X + U_Y$  식의 모수적 모형을 고려할 수 있다. 구조방정식은 기본적으로 모수적 가정과 분포 가정이 강하다는 지적을 받아왔으나, 보다 근본적으로는 ‘개입’의 개념을 명확하게 설명하지 못한다는 한계가 있었고, 이에 Neyman-Rubin의 잠재 결과 모형이 ‘개입’의 개념을 효과적으로 도입한 이후 빠르게 구조방정식을 대체하여 왔다. Pearl J.은 구조방정식을 방향성 비순환 그래프(directed acyclic graph; DAG) 기반의 비모수적 구조적 인과 모형으로 확장하였고, 구조방정식 모형에 특정 개입의 효과를 수식화하는 이른바 ‘do-계산법(do-calculus)’ 기법을 제안한 공로로 2011년 Turing Award를 수상하였다. 전통적 구조방정식과 비교하면, 비모수적 구조적 인과 모형은 DAG으로 표현된 인과관계 하에서 데이터셋이 생성됨을 가정한다는 점에서는 동일하나, 각 화살표의 인과관계를  $Y = \beta/X + U_Y$  처럼 닫힌 형태의 방정식으로 나타낼

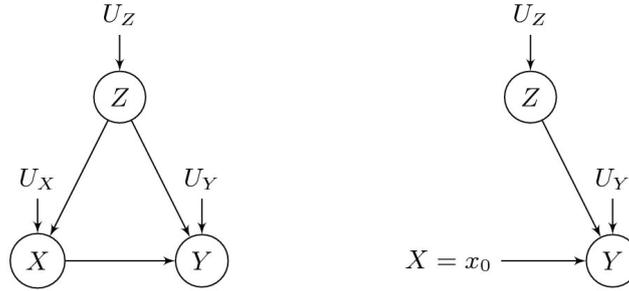


Figure 4.1. 구조적 인과모형식에 대응되는 방향성 비순환 그래프들. 좌측: (4.2), 우측: (4.3).

필요가 없이  $Y = f(X, U_Y)$ 처럼 입력/출력만 적시하여도 충분하다는 점에서 가장 큰 차이가 있다. 그럼에도 불구하고 특정 변수에의 개입이 다른 변수에 영향을 미치는지 그래프이론을 이용하여 판단할 수 있고, 만약 영향을 미친다면 그 크기를 do-계산법을 이용하여 유도해 낼 수 있다. 구조적 인과모형에 대한 자세한 설명은 Pearl (2009a) 및 Pearl (2009b)를 참조할 수 있고, 직관적이고 쉬운 설명은 Pearl 등 (2016) (한국어 번역판은 Kim (2018))를 참조할 수 있다.

쉬운 예제를 통하여 인과모형을 방향성 비순환 그래프로 나타내고 개입의 효과를 do-계산법으로 나타내는 방법을 알아보자. 먼저 다음과 같은 함수 관계는 Figure 4.1의 좌측의 방향성 비순환 그래프에 대응된다.

$$Z = f_Z(U_Z), \quad X = f_X(Z, U_X), \quad Y = f_Y(X, Z, U_Y), \quad (4.2)$$

여기서  $X, Y, Z$ 는 관심 대상의 확률 변수이며,  $U_X, U_Y, U_Z$ 는 독립적인 오차항들을 나타낸다.  $\text{do}(\ )$  연산자는 실험자가 인위적으로 특정 변수를 고정시키는 개입(처치)행위를 나타내는 데 사용된다. 이를테면,  $\text{do}(X = x_0)$  연산은 확률변수  $X$ 를  $X = x_0$ 처럼 외생적으로 고정시키는 상황을 가리킨다.  $\text{do}(X = x_0)$  연산을 적용한 후 식 (4.2)는 다음과 같이 변형된다:

$$Z = f_Z(U_Z), \quad X = x_0, \quad Y = f_Y(x_0, Z, U_Y). \quad (4.3)$$

대응되는 방향성 비순환 그래프는 Figure 4.1의 우측처럼 나타나게 된다. 동일 그림의 좌측과 비교하면,  $X = x_0$ 으로 고정될 뿐만 아니라  $Z$ 에서  $X$ 로의 인과 경로(causal path)도 끊기게 된다.

위의 두 모형을 이용하여, 방향성 비순환 그래프를 이용하여 관찰자료로부터 인과 효과를 계산하는 방법을 알아보자. 예를 들어, 관찰된 자료  $(X, Y, Z)$ 는 식 (4.2)를 따르는 확률분포  $\mathbb{P}$ 의 표본인 상황에서, 가상의 개입 식 (4.3)에서의  $Y$ 의 분포를 계산하여 보자. 편의상 식 (4.3) 하에서의  $(Y, Z)$ 의 결합분포를  $\mathbb{P}(Y, Z | \text{do}(X = x_0))$ 이라고 정의하고, 간단히  $\mathbb{P}_m(Y, Z)$ 이라 나타내자. 우리의 관심 문제는  $\mathbb{P}$ 로부터  $\mathbb{P}_m(Y) = \mathbb{P}(Y | \text{do}(X = x_0))$ 을 계산하는 것이다. 식 (4.2)과 식 (4.3)를 비교하면  $\mathbb{P}$ 과  $\mathbb{P}_m$ 에 대한 다음 세 가지 사실을 이끌어낼 수 있다.

- 개입모형 (4.3) 하에서  $X = x_0$ 는 고정되어 있으므로  $\mathbb{P}_m(Z = z | X = x_0) = \mathbb{P}_m(Z = z)$ 이다.
- 개입 전후에  $Z$ 의 분포에 영향이 없으므로  $\mathbb{P}_m(Z) = \mathbb{P}(Z)$ 이다.
- $Y$ 가  $(X, Z, U_Y)$ 에 반응하여 계산되는 원리는 개입에 영향을 받지 않으므로  $\mathbb{P}_m(Y | Z, X) = \mathbb{P}(Y | Z, X)$ 이다.

위 셋을 Bayes 정리와 조합하면, 아래와 같이 계산이 가능하다(편의상 이산확률변수를 가정하였다).

$$\begin{aligned}\mathbb{P}(Y = y|\text{do}(X = x_0)) &= \mathbb{P}_m(Y = y|X = x_0) \\ &= \sum_z \mathbb{P}_m(Y = y|X = x_0, Z = z) \cdot \mathbb{P}_m(Z = z|X = x_0) \\ &= \sum_z \mathbb{P}_m(Y = y|X = x_0, Z = z) \cdot \mathbb{P}_m(Z = z) \\ &= \sum_z \mathbb{P}(Y = y|X = x_0, Z = z) \cdot \mathbb{P}(Z = z)\end{aligned}$$

수식의 가장 마지막 행은  $\mathbb{P}$ 에 의해 서술되므로 자료에 의한 경험적 추정이 가능하고, 따라서 관찰된 자료로부터  $\mathbb{P}(Y = y|\text{do}(X = x_0))$ 를 계산할 수 있게 된다.

한편, 구조적 인과모형은 잠재 결과 모형을 포괄하는 개념인데, 간단히 설명하면 다음과 같다. 예를 들어 식 (4.2) 모형을 잠재 결과 모형에 대응시켜  $Z$ 를 공변량,  $X$ 를 처치,  $Y$ 를 결과값이라 생각하여 보자. 그러면 무시가능성과 일치성 가정 하에서 평균처치효과  $\tau = \mathbb{E}(Y|X = x_1) - \mathbb{E}(Y|X = x_0)$ 는  $\mathbb{E}(Y|\text{do}(X = x_1)) - \mathbb{E}(Y|\text{do}(X = x_0))$ 과 동일함을 증명할 수 있다. 자세한 설명은 Pearl 등 (2016)의 4장에 소개되어 있다. 잠재 결과 모형과 비교하면, 구조적 인과모형 하에서는 변수간의 인과관계가 더 복잡한 경우에도 특정 개입의 효과를 추적할 수 있는 장점이 있다.

## 5. R에서의 구현

본 절에서는 간단한 가상실험 데이터를 생성하여 (조건부) 평균처치효과 추정량들을 직접 계산하고 비교하여 본다. 먼저 아래와 같이  $X$ ,  $A$ ,  $Y$ 를 생성하자.  $n = 500$ ,  $p = 10$ 으로 설정하였으며, 각  $X_{ij}$  ( $i = 1, \dots, n$ ,  $j = 1, \dots, p$ )는  $[-1, 1]$ 에서의 균등분포에서 추출하였다. 참 성향점수는  $\pi(x) = \mathbb{P}(A = 1|X = x) = \exp\{0.5(x_1 + x_2)\} / [1 + \exp\{0.5(x_1 + x_2)\}]$ 을 고려하였으며, 결과값은  $Y = X_1 + X_3 + 0.3A(1 + X_1 + X_4)$ 로 계산하였다. 이 때 참 평균처치효과는  $\tau = 0.3$ 임을 쉽게 유도할 수 있다.

```
n = 500 ; p = 10 ;
set.seed(10) ; # fix randomness
inv_logit = function(t) exp(t) / (1 + exp(t)) ;
x = matrix(runif(n*p, min=-1, max=1), nrow=n) ;
propensity = inv_logit( 0.5*(x[,1] + x[,2]) ) ;
a = rbinom(n, size=1, prob=propensity) ;
y = x[,1] + x[,3] + a*0.3*(1 + x[,1] + x[,4]) ;
```

성향점수/회귀분석 기반의 추정량들은 아래와 같이 계산할 수 있다. 논의의 간단함을 위해 추론 절차는 생략하였으나,  $\hat{\tau}^{\text{Naive}}$ 와 비교하여 다른 추정량이 참값에 더 가까이 근사됨을 쉽게 확인할 수 있다.

```
# naive estimator
tau_naive = mean(y[a == 1]) - mean(y[a == 0])
# propensity-score based estimators
### IPW by a logistic regression
pi_logistic = glm(a ~ x, family=binomial(link = "logit"))$fitted.values
tau_IPW_logistic = mean(y*a/pi_logistic - y*(1-a)/(1-pi_logistic))
```

```

### IPW by the covariate-balancing propensity score (CBPS) by Imai and Ratkovic (2014)
require(CBPS)
pi_CBPS = CBPS(a ~ x)$fitted.values
tau_IPW_CBPS = mean(y*a/pi_CBPS - y*(1-a)/(1-pi_CBPS))
# regression-based estimators
mylm = lm(y ~ x + a*x, data=list(a,x,y))
tau_regression = coef(mylm) ["a"]
print(c(tau_naive, tau_IPW_logistic, tau_IPW_CBPS, tau_regression))
# should be [1] 0.4884083 0.3018749 0.3051994 0.3000000

```

동일 설정 하에서, 조건부 평균처치효과를  $\hat{\tau}(x) = \hat{Q}(1, x) - \hat{Q}(0, x)$ 로부터 계산하여 보자. 자료 생성 모형으로부터  $\tau(x) = 0.3(1 + x_1 + x_4)$ 이므로, 추정량의 성능은  $\tau(x_i)$ 와  $\hat{\tau}(x_i)$ 의  $xy$ -그래프를 그려봄으로써 직관적으로 파악할 수 있다.

```

tauX_true = 0.3*(1 + x[,1] + x[,4]) ; # true CATE
## estimating Q by a linear model
tauX_lm = cbind(1,x,1,x) %*% coef(mylm) - cbind(1,x,0*x) %*% coef(mylm)
plot(tauX_true, tauX_lm) ; curve(eval(x), add=T, col="red")
## estimating Q by a random forest
library(randomForest)
myrf = randomForest(y=y, x=cbind(x, a, a*x)) ;
tauX_rf = predict(myrf, cbind(x, 1, 1*x)) - predict(myrf, cbind(x, 0, 0*x))
plot(tauX_true, tauX_rf) ; curve(eval(x), add=T, col="red")

```

## 6. 요약 및 제언

통계적 인과 추론의 목표는 인간의 인위적인 개입에 대한 효과를 정확하게 추정하는 것이다. 임의실험이 아닌 관찰자료로부터 효과를 측정하기 위하여는 관찰자료 자체에서는 검증할 수 없는 가정을 하는 점에서 근본적인 한계가 있으나, 임의실험이 윤리적으로 불가능하거나 고비용을 소모하는 상황에서도 개입의 효과를 측정하기 위한 개념들을 제공한다는 점에서 유용하다.

통계적 인과 추론에서 중요하게 다루어진 주제이지만 지면상의 이유로 생략된 부분도 존재한다. 예를 들자면, 성향점수가 잘못 적시된 시나리오에서 처치효과 추정량의 변화를 검토함으로써 추정량의 안정성을 측정하는 기법들(민감도 분석, sensitivity analysis), 인과 경로에 중개변수가 존재하는 경우의 처치효과 추정 기법(중개 분석, mediation analysis), 두 번 이상의 처치에 대한 의사결정이 필요한 경우의 처치효과 추정 기법(교차 분석, interaction analysis), 표본 내 개체 간에 처치효과가 간접적 영향을 미치는 시나리오에서 간접 효과(spillover effect)를 보정하는 기법들이 연구되어 왔다. 이들에 대한 자세한 소개는 Rosenbaum (2017) 및 Imbens와 Rubin (2015)에서 찾아볼 수 있다.

## References

- Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects, *Econometrica*, **74**, 235–267.
- An, W. and Ding, Y. (2018). The landscape of causal inference: perspective from citation network analysis, *The American Statistician*, **72**, 265–277.

- Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models, *Biometrics*, **61**, 962–973.
- Belloni, A., Chernozhukov, V., Fernandez-Val, I., and Hansen, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica*, **85**, 233–298.
- Benkeser, D. (2018). rtmle: Doubly-Robust Nonparametric Estimation and Inference. *R package* version 1.0.4. <https://CRAN.R-project.org/package=drtmle>
- Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching, *Journal of Economic Surveys*, **22**, 31–72.
- Chakraborty, B. and Murphy, S. A. (2014). Dynamic treatment regimes, *Annual Review of Statistics and Its Application*, 1(1):447–464.
- Chan, D., Ge, R., Gershony, O., Hesterberg, T., and Lambert, D. (2010). Evaluating online ad campaigns in a pipeline. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '10)*, 7–15.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., and Newey, W. (2017). Double/debiased/Neyman machine learning of treatment effects, *American Economic Review*, **107**, 261–265.
- Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, E. L. (2009). Smoking and lung cancer: Recent evidence and a discussion of some questions. *International Journal of Epidemiology*, **38**, 1175–1191.
- Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects, *Biometrika*, **96**, 187–199.
- D’Amour, A., Ding, P., Feller, A., Lei, L., and Sekhon, J. (2017). Overlap in observational studies with high-dimensional covariates, *arXiv preprint arXiv:1711.02582*, 1–31.
- Dawid, A. P. (2000). Causal inference without counterfactuals, *Journal of the American Statistical Association*, **95**, 407–424.
- Fan, J., Imai, K., Liu, H., Ning, Y., and Yang, X. (2016). *Improving covariate balancing propensity score: a doubly robust and efficient approach* (Technical Report), Princeton University, Princeton.
- Farrell, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *Journal of Econometrics*, **189**, 1–23.
- Fong, C., Ratkovic, M., and Imai, K. (2018). CBPS: Covariate Balancing Propensity Score. *R package* version 0.19. <https://CRAN.R-project.org/package=CBPS>
- Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, **42**, 1–28.
- Huh, M. (2014). *Applied Data Analytics*, Freedom Academy, Seoul.
- Imai, K. and Ratkovic, M. (2014). Covariate balancing propensity score, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76**, 243–263.
- Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: a review, *Review of Economics and Statistics*, **86**, 4–29.
- Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*, Cambridge University Press, Cambridge.
- Kang, J. D. Y. and Schafer, J. L. (2007). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, **22**, 523–539.
- Kim, M. (2018). *Causal Inference in Statistics: A Primer*, Kyowoo Book, Seoul.
- Künzel, S. R., Sekhon, J. S., Bickel, P. J., and Yu, B. (2017). Meta-learners for estimating heterogeneous treatment effects using machine learning, *arXiv preprint arXiv: 1706.03461*, 1–39.
- Lunceford, J. K. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study, *Statistics in Medicine*, **23**, 2937–2960.
- Nie, X. and Wager, S. (2017). Quasi-oracle estimation of heterogeneous treatment effects. *arXiv preprint arXiv: 1712.04912*, 1–43.
- Ning, Y. and Liu, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models, *The Annals of Statistics*, **45**, 158–195.
- Ning, Y., Peng, S., and Imai, K. (2017). *High dimensional propensity score estimation via covariate balancing* (Technical Report), Cornell University, 1–47.

- Pearl, J. (2009a). Causal inference in statistics: an overview, *Statistics Surveys*, **3**, 96–146.
- Pearl, J. (2009b). *Causality*, Cambridge University Press, Cambridge.
- Pearl, J., Glymour, M., and Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*, Wiley.
- Ponomareva, M. and Powell, J. (2010). Irregular identification, support conditions, and inverse weight estimation, *Econometrica*, **78**, 2021–2042.
- Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules, *The Annals of Statistics*, **39**, 1180–1210.
- Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed, *Journal of the American Statistical Association*, **89**, 846–866.
- Robinson, P. M. (1988). Root-n-consistent semiparametric regression, *Econometrica*, **56**, 931.
- Rosenbaum, P. (2017). *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press.
- Rosenbaum, P. R. and Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score, *Journal of the American Statistical Association*, **79**, 516–524.
- Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, **39**, 33–38.
- Rubin, D. B. (2005). Causal inference using potential outcomes, *Journal of the American Statistical Association*, **100**, 322–331.
- Stuart, E. A. (2010). Matching methods for causal inference: a review and a look forward, *Statistical Science*, **25**, 1–21.
- Van De Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, **42**, 1166–1202.
- van der Laan, M. J. and Rose, S. (2011). *Targeted Learning*, Springer Series in Statistics, Springer New York.
- van der Laan, M. J. and Rose, S. (2018). *Targeted Learning in Data Science*, Springer Series in Statistics, Springer International Publishing, Cham.
- Varian, H. R. (2016). Causal inference in economics and marketing. In *Proceedings of the National Academy of Sciences*, **113**, 7310–7315.
- Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests, *Journal of the American Statistical Association*, **113**, 1228–1242.
- Yang, S. and Ding, P. (2018). Asymptotic inference of causal effects with observational studies trimmed by the estimated propensity scores. *Biometrika*, (March), 1–7.
- Zhang, B., Tsiatis, A. A., Davidian, M., Zhang, M., and Laber, E. (2012). Estimating optimal treatment regimes from a classification perspective, *Stat*, **1**, 103–114.
- Zhao, Q. (2019). Covariate balancing propensity score by tailored loss functions. *The Annals of Statistics*, **47**(2):965-993.
- Zhao, Q. and Percival, D. (2017). Entropy balancing is doubly robust. *Journal of Causal Inference*, **5**(1):1–21.
- Zhao, Q., Small, D. S., and Ertefaie, A. (2017). Selective inference for effect modification via the lasso. *arXiv preprint arXiv: 1705.08020*, 1–21.
- Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning, *Journal of the American Statistical Association*, **107**, 1106–1118.

# 비실험 자료로부터의 인과 추론: 핵심 개념과 최근 동향

최영근<sup>a,1</sup> · 유동현<sup>b</sup>

<sup>a</sup>SK텔레콤 Data 기술원, <sup>b</sup>인하대학교 통계학과

(2018년 12월 03일 접수, 2019년 1월 21일 수정, 2019년 1월 22일 채택)

## 요약

과학적 연구에서 핵심적인 연구 주제 또는 가설은 대부분 인과적 질문(causal question)을 포함한다. 예를 들어, 전염병 예방을 위한 치료법의 효과 연구, 특정 정책의 시행으로 인한 효용(utility)의 평가에 대한 연구, 특정 사용자 대상으로 노출된 광고의 종류에 따른 광고의 효과성에 대한 연구는 모두 인과 관계(causal relationship)의 추론이 요구된다. 이러한 인과 관계를 다루는 통계적 인과 추론(statistical causal inference)의 주요 관심사 중 하나는 모집단에 일종의 개입(정책 혹은 처치)을 적용한 후 개입의 효과를 정확하게 추정하는 것이다. 인과 추론은 임상실험과 정책결정에서 주로 이용되었으나, 이른바 빅데이터 시대의 도래로 가용한 관측자료가 폭발적으로 증가하였고 이로 인하여 인과 추론에 대한 잠재적 응용가치와 수요가 지속적으로 증가하고 있다. 하지만 가용한 대부분의 자료는 임의실험 기반의 자료와 달리 개입이 임의로 분배되지 않은 비실험 관측자료이다. 따라서, 본 논문은 비실험 관측자료로부터 개입의 효과를 추정하기 위한 인과 추론의 핵심 개념과 최근의 연구동향을 소개하고자 한다. 이를 위하여 본문에서는 먼저 개입의 효과를 Neyman-Rubin의 잠재 결과(potential outcome) 모형으로 나타내고, 개입의 효과를 추정하는 여러 접근법 중 특히 성향점수(propensity score) 기반 추정법과 회귀모형 기반 추정법을 중점적으로 소개한다. 최근 연구동향으로는 (1) 평균 효과 크기 추정을 넘어선 개인별 효과 크기의 추정, (2) 효과크기 추정에 있어서 자료 규모의 증대로 인한 차원의 저주가 야기하는 난제들과 이에 대한 해결방안들, (3) 복합적 인과관계를 반영하기 위한 Pearl의 구조적 인과 모형(structural causal model) 및 잠재 결과 모형과의 비교의 3가지 주제로 구분하여 소개한다.

주요용어: 인과 추론, 평균 처치 효과, 개별 처치 효과, 성향 점수, 구조적 인과 모형

이 논문은 2017년도 인하대학교의 지원에 의하여 연구되었음 (INHA-55456).

<sup>1</sup>교신저자: (04539) 서울특별시 중구 을지로 65, SK텔레콤 Data 기술원. E-mail: yg.choi@sk.com