

이항 반응 자료에 대한 학습곡선의 모형화

이슬지^a, 박만식^{1,a,b}

^a성신여자대학교 통계학과, ^b성신여자대학교 통계연구소

요약

연구자가 같은 작업을 반복적으로 수행할 때, 작업 효율성은 연구에 관련된 지식, 경험, 기술이 축적되면서 향상된다. 결과를 얻기 위해 연구에 투자하는 시간은 같은 작업을 반복함으로써 줄일 수 있다. 이러한 현상을 학습곡선 효과(learning curve effect)라고 일컫는다. 학습곡선(learning curves)은 학습의 변화를 시각적으로 나타낸 것으로 이전의 학습곡선 연구에서는 시간을 일정한 구간으로 나누어 구간별 작업에 대한 숙련도의 평균 차이 여부를 확인하였다. 이러한 방법은 구간을 어떻게 나눌 것인가 하는 기준이 존재하지 않으며, 더욱이 이항 반응 자료로 모형을 적합하기 어려운 문제점을 가지고 있다. 본 연구에서는 이산형 확률변수 중 이항 반응 자료(베르누이자료)에 대한 학습곡선의 통계적 모형에 초점을 맞추고자 한다. 누적확률분포의 특성을 이용하여 모수를 추정하기 위해서 뉴턴-랩슨 방법(Newton-Raphson method)을 사용하였고, 이 연구에서 제안한 모형의 점근적 분포를 구하였다.

주요용어: 학습곡선, 이항반응자료, 누적확률분포, 뉴턴-랩슨 방법, 델타방법.

1. 서론

19세기 독일의 심리학자 헤르만 에빙하우스는 ‘과제를 수행하는 횟수가 증가할수록 같은 일을 하는데 드는 시간이 줄어든다.’라는 학습곡선 효과(learning curve effect)를 주장하였다. 1936년 미국의 라이트 패터슨 공군기지의 연구 결과, 비행기 생산량이 두 배가 될 때마다 노동 시간은 10-15%씩 줄어 든다고 밝혔다. 이 현상을 통해 학습곡선 효과에 대한 주장이 실증적으로 입증되었다. 생산 및 작업관리 분야에서는 노동비용과 생산량의 관계를 설명하기 위해 학습곡선을 사용한다. 학습효과가 보편적인 용어로 대중화되고, 과거의 경험을 통해 같은 실수를 반복하지 않는다는 의미가 강해지면서 의학 분야에서는 수술의 경험횟수에 따른 수술의 성공여부를 설명하기 위해 개념화되었다. 학습곡선은 전형적으로 초기의 시도 후에 반복 학습으로 인하여 학습 정보의 축적으로 학습 효율이 증가하는 형태를 보 이다가 어느 시점 이후로는 완만해지는 형태를 보인다.

연구자가 시행횟수에 따른 학습시간의 자료에서 학습시간이 급격하게 감소하는 시점을 경험을 토 대로 학습극복시점으로 정하여 학습극복 전,후의 차이를 비교하는 것이 일반적인 학습곡선에 대한 분석방법이다. 최유신 등 (2006)은 조기위암에 시행된 복강경 보조 원위부 위절제술의 학습곡선 극복 시 점을 연구자가 경험적으로 정하고, 카이제곱 검정과 t -검정을 이용하여 학습곡선 극복 시점 전후의 임 상적인 차이를 연구하였고, 송수진 등 (2003)은 수술시간을 학습곡선 전·후로 비교하여 내시경하 자 궁근중절제술이 개복하 자궁근중절제술과 유의한 차이가 없다는 것을 카이제곱 검정과 독립표본 t -검 정을 통하여 보였다. 이와 동일한 분석방법을 이용한 논문은 Kim 등 (2005), Schauer 등 (2002), Fer guson 등 (2005), Sim 등 (2006)이 있다. 김학선 등 (2003)은 흉곽 내시경을 이용한 최소 침습 방법의

이 논문은 2011년도 성신여자대학교 학술연구조성비 지원에 의하여 연구되었음.

¹ 교신저자: (136-742) 서울특별시 성북구 보문로 34다길 2, 성신여자대학교 자연과학대학 통계학과, 조교수.

E-mail: mansikpark@sungshin.ac.kr

적추 기형 교정에 대하여 수술의 난이도와 증례수에 따른 학습곡선을 통해 수술이 능숙해지는 시점을 제시하였고, 오윤경 등 (2010)은 동일 기관에서의 세 명의 산부인과 전문의의 복강경하 전자궁절제술의 학습곡선을 비교하여 평균 변환점 연구하였다. 이와 같은 분석방법을 이용한 논문은 Ballantyne 등 (2005)이 있다. Cook 등 (2004)은 무작위 통제 시행(randomized controlled trial; RCT)에서 새로운 수술 기법에 대한 학습곡선을 평가하기 위해 계층적 베이즈 모형(hierarchical Bayes model)을 이용하였다. 이 때 학습곡선의 효과가 존재하도록 수술 결과를 조절하기 위하여 계층적 베이즈 모형을 적용하였다. Lim 등 (2011)은 로봇을 이용한 자궁절제술과 복강경 자궁절제술의 효율성을 비교하기 위하여 각 수술 기법에 대한 수술시간의 학습곡선을 회귀모형을 이용하여 적합시켰다. Smith 등 (2004)은 상태공간 모형(state-space model)을 이용하여 학습곡선에 적합시켰다. 이 때, EM-알고리즘을 이용하여 모수를 추정하였다. Adler와 Clark (1991)은 두 곳의 전자 장비 회사의 제조 부서의 데이터를 활용하여 생산성을 향상시키는 모형을 구축하였다. 이 모형은 공정변화와 직원교육을 통하여 누적된 경험을 변수로 이용하였다. 동일한 작업의 반복에서 학습을 발생시키는 전통적인 경험 변수들의 역할과 공정변화, 직원교육과 같이 경영에 의한 변수들의 역할이 학습곡선이 발생하는 과정에 각각 어떠한 차이를 보이는지 기술하였다. Williams와 Vivarelli (2000)는 정규분포를 따르는 학습곡선의 신뢰구간을 제시하였다. 분석 데이터의 표본수에 따른 기대오차가 학습곡선의 형태를 보이며, 이를 이용하여 정규분포를 따르는 학습곡선의 초기 양상을 설명하였다. Smunt (1999)는 연속적인 학습은 로그-선형(log-linear) 학습곡선 모형에 근사하다는 것을 이용하여 제조업 데이터에서 생산성 경향을 분석하였다. 학습곡선은 생산 계획과 가격을 예측하는 도구로 산업분야에서 활용되기도 한다. Lieberman (1984)는 37개의 화학 제품으로 학습곡선과 가격 양상에 대한 가설을 검증하고 시장집중이 가격의 유연성과 가격 변화에 대한 학습곡선에 강한 영향을 미친다는 것을 연구하였다. Mazzola와 McCardle (1996)은 비용함수가 학습곡선의 경향을 따를 때 독점적인 생산 계획의 최적화를 모형화하였다. 이 때, 베이지안(Bayesian) 방법을 이용하여 학습곡선의 최적값을 결정하였다.

반응변수가 이산형인 경우 학습곡선을 적용할 수 있는 방법은 주로 누적합을 이용하고 있다. 순차적으로 일어나는 공정이 잘 관리되고 있다는 가정이 기각되는지를 판단하는 누적합(cumulative sum; CUSUM) 분석은 산업분야에서 품질관리를 위해 널리 이용되는 통계적 기법이다. 최근에는 의학분야에서도 시간에 따른 새로운 수술 기법의 습득 정도를 누적합 분석에 이용하여 수술의 품질평가에 이용하고 있다. 이는, 어떤 수술 기법이 정해진 수준까지 도달하는지의 평가를 시각적으로 쉽게 이해할 수 있다는 장점이 있다. 학습곡선의 누적합 분석은 목표값으로부터의 차이를 순차적으로 누적시키면서 나타냄으로써 미리 정해둔 기준값으로부터 벗어나는 경우를 감지하는 방법이다. 이러한 관리도를 만들기 위해서는 허용되지 않는 실패율과 허용되는 실패율, 제1종 오류 및 제2종 오류의 확률들을 미리 정해두어야 한다. 이들 변수로부터 한계값과 변수값을 계산할 수 있다. 누적합을 계산하기 위한 기본적인 통계량, C_t 은 다음과 같다.

$$C_t = C_{t-1} + (Y_t - Y_0), \quad t = 1, 2, \dots, T.$$

여기서, $C_0 = 0$ 이고, Y_t 는 시점 t 에서 관측한 실패여부(실패이면 1, 성공이면 0)이며 Y_0 는 연구자에 의해 고려되는 잠정적인 실패율을 의미한다. 이를 통해 알 수 있듯이 연속되는 시술이 성공한 경우 t 시점에서의 누적합값은 직전의 누적합값 C_{t-1} 으로부터 Y_0 만큼 줄어들게 되고, 실패하게 되는 경우 $1 - Y_0$ 만큼 직전 시점의 누적합값에 더해지게 된다. 결국, 성공한 시술에는 그래프의 방향이 음(-)의 방향으로 향하게 되고, 실패한 시술에는 양(+)의 방향으로 움직이게 된다. 이러한 누적합값의 형태와 더불어 연구자에 의해 추가적으로 주어지게 되는 조건들(허용되지 않는 실패율, 허용할 수 있는 실패율, 제1종 오류, 그리고 제2종 오류)에 의해 일종의 가설검정을 수행할 수 있게 된다. 이때 귀무가설은 ‘아직 능숙하지 않은 상태’이며, 대립가설은 ‘이미 능숙한 상태’이다. 따라서 누적합이 아래쪽 한계값에 도달

한다면, 이것은 귀무가설을 기각하고 대립가설을 선택하는 것이 통계학적으로 의의가 있다는 결론을 내릴 수 있게 된다.

학습곡선의 누적합 분석을 정리하면 시행횟수에 따른 학습의 성공 여부에 대한 자료에서 위의 분석방법을 이용하면 학습이 능숙해지는 지점을 정할 수 있다. 서강열 등 (2010)은 새로운 수술기법을 배우는 교육생이 얼마나 빨리 적절한 수준의 심폐소생술을 시행하게 되는지에 대하여 심폐소생술 교육을 받은 의과대학생들의 흉부압박의 능숙도(proficiency)를 누적합 분석을 이용해서 개인별로 능숙한 시술에 이르는 정도의 차이를 비교 분석하였다. Tekkis 등 (2005)는 복강경 결장 수술과 직장 수술에서 왼쪽 절개와 오른쪽 절개의 수술 결과에 대한 차이를 학습곡선을 사용하여 비교하였다. 이 때, 학습곡선을 평가하기 위한 방법으로 위험요인이 조정된 누적합(Risk-adjusted CUSUM) 분석을 이용하였다. 이와 동일한 분석 방법을 이용한 논문은 Biau 등 (2008), Lee 등 (2006), Forbes 등 (2004), 그리고 Mazzola와 McCardle (1996)이 있다.

이상의 언급된 방법론들은 현실적으로 쉽게 적용할 수 있다는 장점이 있지만 연구자의 주관적인 견해를 바탕으로 학습이 능숙해지는 지점을 정하기 때문에 연구자의 경험과 연구 과정에 의존하게 된다는 무시할 수 없는 단점을 가지고 있다. 이는 연구자에 따라 그 결과와 그에 따른 해석이 상이할 수 있고, 결론적으로 일반화가 용이하지 않다는 한계를 가지고 있다. 이에 본 연구에서는 좀더 객관적으로 과학적인 방법을 동원하여 이를 극복해 보고자 한다. 즉, 이항 반응 자료의 형태를 취하는 자료에 대해 특정한 누적확률분포함수(cumulative probability distribution function)의 특성을 이용하여 통계적 모형화를 통해 기존 연구에서의 학습곡선 적합의 한계를 보완하고자 한다.

이 논문의 순서는 다음과 같이 전개하고자 한다. 2절에서는 누적확률분포의 특성을 이용한 학습곡선 모형을 제안하고, 3절에서는 제안한 모형에 대한 여러가지 시나리오(scenario)에 맞추어 모의실험을 수행하고 그 결과를 자세히 설명하고자 한다. 4절에서는 Lim 등 (2002)이 누적합 학습곡선 방법을 통해 분석한 방사선 전문의의 유방생체 검사 수행자료를 본 연구에서 제안한 방법으로 분석한다. 마지막으로 5절에서는 결론 및 앞으로의 연구방향에 관하여 논의한다.

2. 누적확률분포를 이용한 학습곡선의 모형화

지금까지의 학습곡선은 작업을 능숙하게 성취할 수 있는 시점을 경험을 바탕으로 연구되어 왔다. 의학분야에서는, 임상학적으로 몇 번의 수술을 하였을 때 수술에 능숙도를 보이는지가 주요 관심대상이 된다. 따라서 본 연구에서는 누적확률분포의 특성을 이용하여 학습곡선의 통계적 모형화를 하고자 한다.

2.1. 모형 구축

본 연구에서 관심 있는 사건은 특정 수술의 실패 여부(실패는 1이고, 성공은 0)이고 이에 대한 학습곡선의 모형 구축은 그림 1를 바탕으로 진행하고자 한다. 이 때, 작업 수행 시 실험 초기의 실패 확률을 θ_2 , 더 이상 줄일 수 없는, 능숙한 상태(안정화 단계)에 도달했을 때의 실패 확률을 θ_1 이라 정의한다. 시행횟수 $t = 1, \dots, T$ 에 대하여 시점 t 에서의 작업의 실패 여부, Y_t 는 실패 확률이 $0 < p_t(\theta) < 1$ 인 베르누이 분포(Bernoulli distribution)를 이룬다. 즉,

$$Y_t \sim \text{Bernoulli}(p_t(\theta))$$

이고, 모형의 단순화를 위해 Y_t 는 독립이라고 가정한다. 여기서, $\theta = (\theta_1, \theta_2, \dots, \theta_p)^T$ 이고,

$$p_t(\theta) = \theta_1 + (\theta_2 - \theta_1) \cdot \left[1 - F(t | \theta_3, \dots, \theta_p) \right] \quad (2.1)$$

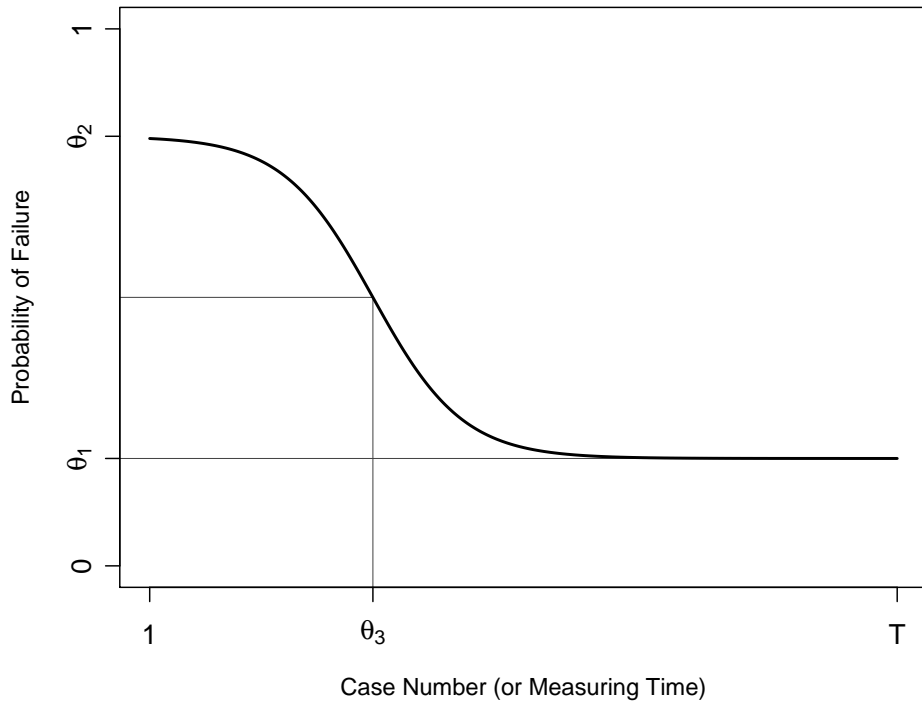


그림 1: 누적확률분포를 이용한 학습곡선의 모형화

이라 정의한다. 여기서, 실험 초기의 실패확률과 능숙한 상태의 실패확률에 대하여 다음과 같은 제약 조건을 고려할 수 있다.

$$0 < \theta_1 \leq \theta_2 < 1.$$

이 때, 식 (2.1)의 $F(t|\cdot)$ 는 다음과 같이 정의하고자 한다. $t = 1, 2, \dots, T$ 에 대하여

$$F(t|\theta_3, \dots, \theta_p) = \int_{-\infty}^t f(x|\theta_3, \dots, \theta_p) dx \quad (2.2)$$

이다. 여기서, $f(x|\cdot)$ 는 어떤 확률밀도함수(probability density function)이고 $\theta_3, \dots, \theta_p$ 는 $f(\cdot)$ 에 관여하는 모수들이다. 즉, 식 (2.2)의 $F(\cdot)$ 는 주어진 확률밀도함수, $f(\cdot)$ 의 누적인 형태로 나타내어지므로 누적확률분포함수가 되고 어떠한 확률밀도함수를 사용하느냐에 따라 이에 대응하는 누적분포함수의 형태가 달라지게 된다. 일반적으로 확률변수의 범위(range)가 실수전체($-\infty, \infty$)에 해당되는 경우, 식 (2.2)의 t 가 증가한다 하더라도 누적분포함수의 특성으로 인해 식 (2.2)의 $F(t|\cdot)$ 이 1을 가질 수 없다. 따라서 특정 수술의 시행횟수를 아무리 늘려도 식 (2.1)의 $p_t(\theta)$ 가 능숙한 상태의 실패 확률인 θ_1 에 이를 수 없게 된다. 즉,

$$\lim_{t \rightarrow \infty} (1 - F(t|\theta_3, \dots, \theta_p)) = 0$$

이다. 현실적으로 시행횟수를 무한히 많이 늘릴 수 없고 $p_t(\theta)$ 가 능숙한 상태의 실패 확률에 도달할 수 없기 때문에 본 연구에서는 능숙 정도가 $100 \times k\%$ 인 시행횟수(능숙 시점)를 λ_k 로 정의한다. 즉,

$$k = F(\lambda_k|\theta_3, \dots, \theta_p) = \int_{-\infty}^{\lambda_k} f(x|\theta_3, \dots, \theta_p) dx$$

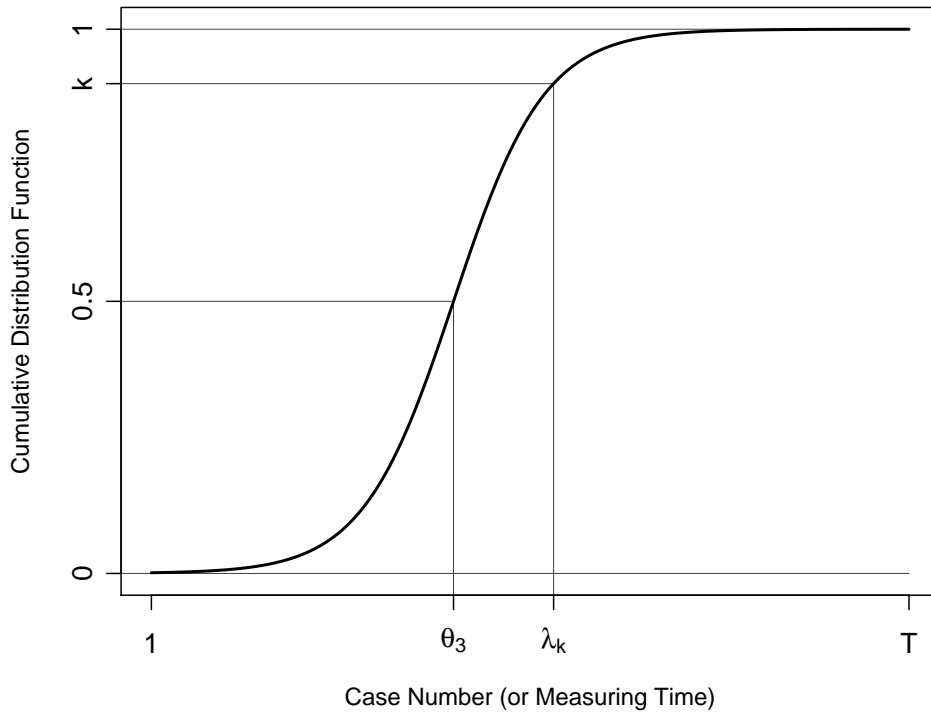


그림 2: 누적확률분포함수(F)와 100 × k% 능숙 시점(λ_k)

이 된다. 앞서 언급한 바와 같이 식 (2.1)과 식 (2.2)의 누적확률분포함수는 일반적인 확률밀도함수에 의해 결정될 수 있다.

본 연구에서는 학습곡선의 모형화를 위해 로지스틱분포(logistic distribution)를 이용하고자 한다. 로지스틱 분포의 가장 큰 장점은 위치모수를 기준으로 대칭을 이룬다. 따라서 본 연구에서 정의한 더 이상 줄일 수 없는 실패 확률(θ_1), 실험 초기의 실패확률(θ_2)를 가늠할 수 있는 지표로 실험 초기의 실패 확률인 θ_2 와의 차이를 계산하면 실험 초기와 실험 종료 사이에 얼마나 향상 되었는지 파악이 가능하다. 이 차이의 절반, 즉, 전체의 학습 향상도의 절반이 일어나는 시점이 위치모수로 정의한 θ_3 값과 일치하게 된다. 또한, 로지스틱 분포의 경우 누적확률분포함수를 프로그램으로 구현하는 것이 비교적 용이하다는 장점을 가지고 있다.

다음은 로지스틱분포의 확률밀도함수이다. $-\infty < t, \theta_3 < \infty$ 와 $0 < \theta_4 < \infty$ 에 대하여

$$f(t|\theta_3, \theta_4) = \frac{1}{\theta_4} \frac{e^{-\frac{(t-\theta_3)}{\theta_4}}}{\left[1 + e^{-\frac{(t-\theta_3)}{\theta_4}}\right]^2}. \tag{2.3}$$

여기서, θ_3 는 위치모수(location parameter)이고 θ_4 는 규모모수(scale parameter)이다. 식 (2.3)의 누적확률분포함수는 그림 2와 같은 형태를 띄게 되고 다음과 같이 표현된다.

$$F(t|\theta_3, \theta_4) = \int_{-\infty}^t f(x|\theta_3, \theta_4) dx = \left[1 + e^{-\frac{(t-\theta_3)}{\theta_4}}\right]^{-1}. \tag{2.4}$$

따라서 로지스틱 확률밀도함수를 고려하는 경우, 능숙도가 $100 \times k\%$ 인 시행횟수(능숙 시점), λ_k 를 다음과 같이 구할 수 있다.

$$\lambda_k = \theta_3 + \theta_4 \times \ln\left(\frac{k}{1-k}\right). \quad (2.5)$$

이러한 능숙 시점은 오직 분포함수에 관계된 모수(θ_3, θ_4)에 의해서만 영향을 받는 것을 알 수 있다. 본 연구에서는 여러 확률분포들 중에서 로지스틱확률분포에 근거하여 진행하고자 한다.

2.2. Newton-Raphson 방법을 이용한 모수추정

로지스틱분포를 이용한 학습곡선 모형의 모수 θ 를 참값에 가까운 값으로 추정하기 위하여 본 논문에서는 뉴턴-랩슨 방법(Newton-Raphson method)을 이용하였다. 뉴턴-랩슨 방법은 주어진 현재의 추정값으로부터 다음 단계의 추정값의 차이가 반복 과정을 거치면서 수렴할 때까지 진행된다. 여기서 행렬의 과중한 계산 부담을 줄이기 위해 스코어(scoring) 방법을 사용한 후 뉴턴-랩슨 방법으로 바꾸는 알고리즘을 사용하였다. 이 때, 스코어(scoring) 방법은 다음과 같이 정의되는 반복 추정모형을 사용한다.

$$\hat{\theta}^{(m)} = \hat{\theta}^{(m-1)} + [\mathcal{J}^{(m-1)}]^{-1} \cdot U^{(m-1)}$$

여기서, $\mathcal{J}^{(m-1)}$ 은 $\theta = \theta^{(m)}$ 을 사용하여 계산된 정보행렬(Information matrix)이다.

작업의 실패 확률을 다음과 같이 정의할 때

$$p_t(\theta) = \theta_1 + (\theta_2 - \theta_1) \cdot [1 - F(t|\theta_3, \theta_4)] \equiv p_t.$$

우리가 추정해야 할 모수는 $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)^T$ 이며 로지스틱분포의 누적확률분포함수는 식 (2.4)와 같다. 본 연구에서는 Y_t 가 베르누이 분포를 따른다는 가정에 근간하여 모형화를 진행하나, 베르누이 분포를 따르는 Y_t 를 확장시켜 Y_t 가 이항분포를 따르는 경우의 모형화로 일반화가 가능하기 때문에 이 절에서는 확률변수 Y_t 가 이항분포(binomial distribution)인 $B(n_t, p_t)$ 를 따른다고 가정하여 전개하도록 한다. 확률변수 Y_t 가 이항분포(binomial distribution)인 $B(n_t, p_t)$ 를 따른다고 할 때, 이항분포의 확률질량 함수(probability mass function)는

$$f(y_t|p_t) = \binom{n_t}{y_t} p_t^{y_t} (1-p_t)^{n_t-y_t}, \quad y_t = 0, 1, 2, \dots, n_t, \quad t = 1, \dots, T$$

이고, 우도함수(likelihood function)의 로그(log)는

$$l(p_t) = \sum_{t=1}^T \log f(y_t|p_t) = \sum_{t=1}^T \left\{ y_t \log p_t + (n_t - y_t) \log(1 - p_t) + \log \binom{n_t}{y_t} \right\}$$

로 단순화된다. 이것을 이용하여 $p_t(\theta)$ 에 대한 1차 편미분(partial differentiation)은

$$l(p_t) = \sum_{t=1}^T \frac{\partial \log f(y_t|p_t)}{\partial p_t} = \sum_{t=1}^T \left(\frac{y_t}{p_t} - \frac{n_t - y_t}{1 - p_t} \right) = \sum_{t=1}^T \left(\frac{y_t - n_t p_t}{p_t(1 - p_t)} \right)$$

이다. 다음으로 $p_t(\theta)$ 의 스코어 통계량(score statistic)은 각각의 모수에 대한 편미분을 하여 구한다.

$$U(\theta|p_t) = \sum_{t=1}^T \frac{\partial l(\theta|p_t)}{\partial \theta} = \{U_i(\theta|p_t)\}_{i=1, \dots, 4}$$

여기서, U_i 의 원소는 다음과 같다.

$$U_i = \sum_{t=1}^T \frac{\partial l(p_t)}{\partial \theta_i} = \sum_{t=1}^T \frac{\partial l(p_t)}{\partial p_t} \cdot \frac{\partial p_t}{\partial \theta_i}$$

이 때, p_t 에 대한 로그 우도함수의 1차 편미분은

$$\frac{\partial l(p_t)}{\partial p_t} = \frac{y_t - n_t p_t}{p_t(1 - p_t)}$$

이고, θ_i 각각에 대한 p_t 의 편미분은 다음과 같이 표현될 수 있다.

$$\frac{\partial p_t}{\partial \theta_i} = \begin{cases} F(t|\theta_3, \theta_4), & i = 1, \\ [1 - F(t|\theta_3, \theta_4)], & i = 2, \\ \left(\frac{\theta_2 - \theta_1}{\theta_4}\right) \exp\left(-\frac{t - \theta_3}{\theta_4}\right) \{F(t|\theta_3, \theta_4)\}^{-2}, & i = 3, \\ \frac{(\theta_2 - \theta_1)(t - \theta_3)}{\theta_4^2} \exp\left(-\frac{t - \theta_3}{\theta_4}\right) \{F(t|\theta_3, \theta_4)\}^{-2}, & i = 4. \end{cases}$$

이를 바탕으로 θ 를 추정하기 위하여 정보행렬(\mathcal{J})을 구한다.

$$\mathcal{J} = \{I_{jk}\}_{j,k=1,\dots,4} = \left\{ \frac{\partial U_j}{\partial \theta_k} \right\}_{j,k=1,\dots,4}$$

여기서, $i, j = 1, \dots, 4$ 에 대하여 I_{jk} 는

$$I_{jk} = \frac{\partial U_j}{\partial \theta_k} = \frac{\partial U_j}{\partial p_t} \cdot \frac{\partial p_t}{\partial \theta_k} = \frac{\partial^2 l(p_t)}{\partial p_t^2} \cdot \frac{\partial p_t}{\partial \theta_j} \cdot \frac{\partial p_t}{\partial \theta_k}$$

로 계산될 수 있으며, 이 때, p_t 에 대한 로그 우도 함수의 2차 편미분은 다음과 같다.

$$\frac{\partial^2 l(p_t)}{\partial p_t^2} = \frac{-n_t p_t^2 + (2p_t - 1)y_t}{\{p_t(1 - p_t)\}^2}.$$

2.3. 델타 방법(Delta method)에 의한 실패 확률의 근사 신뢰구간 추정

델타 방법은 근사 신뢰구간(asymptotic confidence interval)을 계산하기 위한 일반적인 접근법으로 최대우도함수를 이용한다. 이 방법은 분석적으로 분산을 계산하기 복잡한 경우 테일러 전개에 기반을 두어 함수의 평균과 분산으로 간단하게 계산할 수 있다.

$p_t(\hat{\theta})$ 에 대한 $100 \times (1 - \alpha)\%$ 근사 신뢰구간은 다음과 같다.

$$p_t(\hat{\theta}) \pm Z_{\frac{\alpha}{2}} \cdot \sqrt{\text{Var}\{p_t(\hat{\theta})\}}$$

여기서, $p_t(\hat{\theta})$ 의 기대값은 델타 방법에 의하여

$$E\{p_t(\hat{\theta})\} = p_t(\theta) + \sum_{i=1}^4 \frac{\partial p_t}{\partial \theta_i} \cdot E(\theta_i - \hat{\theta}_i)$$

이고, $p_t(\hat{\theta})$ 의 분산은

$$\widehat{\text{Var}}\{p_t(\hat{\theta})\} = \sum_{i=1}^4 \left(\frac{\partial p_t}{\partial \theta_i}\right)^2 \widehat{\text{Var}}(\hat{\theta}_i) + 2 \sum_{i>j}^4 \frac{\partial p_t}{\partial \theta_i} \frac{\partial p_t}{\partial \theta_j} \widehat{\text{Cov}}(\hat{\theta}_i, \hat{\theta}_j)$$

가 된다.

이상과 같이 제 2절에서는 본 연구에서 제안한 학습곡선의 효과에 대한 통계적 모형화에 대해 자세히 설명하였다. 주어진 모형의 적합과 모수에 대한 추정을 통해 어느 정도의 반복적 경험과 훈련을 거치게 되면 해당 작업에 대해 능숙해지는지를 명확히 설명할 수 있다.

3. 모의실험

누적확률분포의 특성을 이용한 학습곡선의 통계적 모형이 수술의 성공 여부에 대해 얼마나 잘 적합하는지를 살펴보기 위해 3절에서 다른 추정방법을 근간으로 모의실험을 수행하였다(표 1 참조). 여기서, 더 이상 줄일 수 없는 실패 확률(θ_1)은 0.2, 실험 초기의 실패 확률(θ_2)은 0.8로 고정시키고 학습곡선의 변곡점 시점을 결정하는 θ_3 은 100, 150, 250으로 정하고 학습곡선의 변동폭을 결정하는 θ_4 는 10, 30, 50으로 정하였다. 이와 같은 다양한 시나리오(scenario) 하에서 다음과 같은 절차에 따라 각 모수들을 추정하였다.

3.1. 모의실험 방법

각 시나리오마다 모의실험의 절차는 동일하며 다음과 같다.

- (1) 로지스틱 분포를 이용하여 시행횟수 $t = 1, \dots, T (= 500)$ 에 대한 실패 확률 $p_t(\theta)$ 를 계산한다.
- (2) 단계 (1)에서 구한 실패확률이 $p_t(\theta)$ 인 베르누이 분포를 따르는 Y_t 를 생성한다.
- (3) 단계 (2)에서 생성한 자료를 이용하여 3절의 추정방법으로 모수를 추정한다.
- (4) 단계 (1)부터 단계 (3)까지의 과정을 200개의 자료에 각각 적용한다. 이 때, 실패 확률은 다음과 같이 표현할 수 있다.

$$\mathbf{p}^{(i)} = \{p_t^{(i)}\}_{t=1}^T, \quad i = 1, \dots, N (= 200).$$

모의실험에서 사용되는 각 자료로부터 모수의 추정값 뿐 만 아니라 각 시점별 추정확률을 구하게 된다. 결국 전체 200개의 자료들로부터 각 시점별 200개씩의 실패확률의 추정값들을 얻게 된다. 이 추정값들을 이용해서 p_t 에 대한 $100 \times (1 - \alpha)\%$ 경험적 신뢰구간(empirical confidence interval)을 다음과 같이 구할 수 있다.

$$\left(\{\mathbf{p}_t\}_{(\frac{N\alpha}{2} + 1)}, \{\mathbf{p}_t\}_{(N - \frac{N\alpha}{2})} \right),$$

여기서 $\{\mathbf{p}_t\}_{(i)}$ 는 시점 t 에서 얻은 200개의 실패확률들을 크기 순으로 나열하였을 때 i 번째로 작은 값을 의미한다. 모수의 추정과 결과 제시를 위해 SAS 통계프로그램(Version 9.2)과 R 프로그램(Version 2.14.1)을 사용하였다.

표 1: 로지스틱 확률분포를 이용한 학습곡선 모형의 추정 결과 ($T = 500$)

Scenario	Parameter	Truth	Mean	SE	MAD
1.	θ_1	0.2	0.195	0.002	0.013
	θ_2	0.8	0.811	0.004	0.036
	θ_3	100	99.323	0.550	4.899
	θ_4	10	11.302	0.436	3.626
2.	θ_1	0.2	0.197	0.002	0.016
	θ_2	0.8	0.793	0.006	0.065
	θ_3	100	102.240	1.149	11.170
	θ_4	30	26.346	0.907	8.775
3.	θ_1	0.2	0.200	0.002	0.020
	θ_2	0.8	0.751	0.008	0.090
	θ_3	100	114.005	1.834	18.921
	θ_4	50	40.678	1.466	16.095
4.	θ_1	0.2	0.199	0.002	0.017
	θ_2	0.8	0.801	0.003	0.024
	θ_3	150	150.113	0.503	4.716
	θ_4	10	10.744	0.474	4.155
5.	θ_1	0.2	0.195	0.002	0.018
	θ_2	0.8	0.810	0.004	0.040
	θ_3	150	149.137	1.113	11.830
	θ_4	30	29.898	0.867	8.273
6.	θ_1	0.2	0.198	0.002	0.025
	θ_2	0.8	0.786	0.006	0.067
	θ_3	150	154.391	1.651	18.077
	θ_4	50	44.847	1.358	13.223
7.	θ_1	0.2	0.196	0.002	0.018
	θ_2	0.8	0.803	0.002	0.019
	θ_3	250	249.651	0.480	4.567
	θ_4	10	11.503	0.443	3.587
8.	θ_1	0.2	0.193	0.003	0.025
	θ_2	0.8	0.807	0.003	0.023
	θ_3	250	250.410	0.943	9.664
	θ_4	30	30.474	0.967	7.869
9.	θ_1	0.2	0.196	0.004	0.032
	θ_2	0.8	0.803	0.004	0.031
	θ_3	250	252.874	1.578	13.344
	θ_4	50	48.615	1.374	10.653

3.2. 모의실험 결과

표 1은 시나리오에 따른 모수 $\theta_1, \theta_2, \theta_3, \theta_4$ 의 참값과 추정값들의 평균(Mean), 표준오차(SE), MAD의 결과를 정리한 것이다. 여기서 MAD는 각 시점마다의 중위절대편차(Median Absolute Deviation; MAD)로 중위절대편차는 다음과 같이 정의한다.

$$MAD = \text{median} |\theta_i - \hat{\theta}_i|.$$

MAD 값을 중심으로 살펴보면 모수 θ_1, θ_2 은 대체로 낮은 값으로 나타났으며 이는 추정값과 참값의 차이가 거의 나지 않음을 나타내며 학습곡선의 모양에 크게 영향을 받지 않는 값을 확인할 수 있다. 그에 비해 θ_3, θ_4 는 학습곡선의 모양에 영향을 받으며, 특히 $\theta_4 = 50$ 인 경우에 추정값과 참값의 차이가 크게 나타났다. 이는, 학습의 일어나기까지 걸리는 시간을 나타내는 지표인 θ_4 의 값이 클수록 참값과

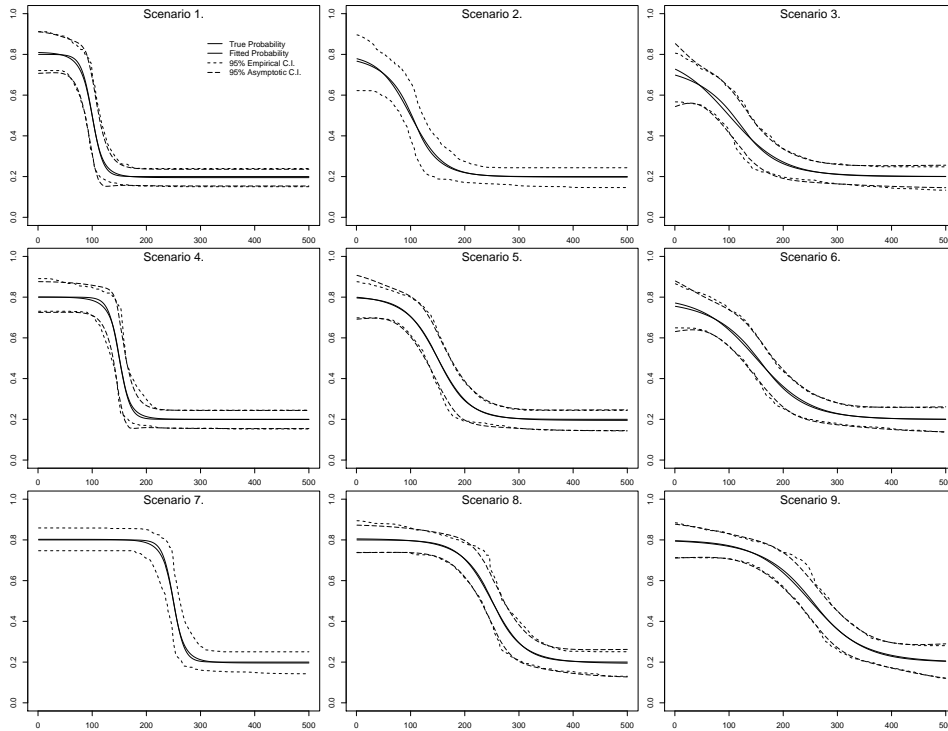


그림 3: 시나리오에 따른 관측시점별 참확률과 추정확률, 그리고 95% 신뢰구간들

추정치 사이의 차이가 크게 나타남을 알 수 있다. 시나리오 1번을 기준으로 살펴보면 시나리오 2번과 3번을 비교하였을 때 $\theta_4 = 10$ 인 경우가 $\theta_4 = 30, 50$ 인 경우보다 SE와 MAD가 각각 0.436과 3.626으로 작은 값을 보였다. 또한 $\theta_3 = 100$ 으로 고정되었을 때 $\theta_4 = 50$ 일 때보다 $\theta_4 = 10$ 인 경우가 θ_3 의 SE와 MAD의 값이 각각 0.550, 4.899로 작은 값을 나타냈다. 다시 말해서 θ_3 가 고정되었을 때 θ_4 가 작을수록 θ_4 의 값이 안정적이며, θ_3 또한 안정적인 값을 나타낸다는 것을 알 수 있다. 시나리오 1번을 시나리오 4번, 7번과 비교하였을 때 $\theta_3 = 100$ 인 경우가 $\theta_3 = 150, 250$ 인 경우보다 SE와 MAD가 각각 0.550과 4.899로 약간 큰 값을 나타내고 있으나 θ_4 의 SE와 MAD 값의 차이가 크지 않았다. 즉, θ_4 가 고정되었을 때 θ_3 의 값의 변화가 모형에 미치는 영향력이 미미하다는 것을 알 수 있다.

그림 3은 작업의 성공 여부가 베르누이 분포인 데이터를 200번 반복하여 얻은 학습곡선의 적합 모형과 95% 신뢰수준 하에서의 각 시점별 경험적 신뢰구간, 95% 신뢰수준 하에서의 각 시점별 근사 신뢰구간의 평균, 참값을 이용한 실패확률과 각 추정치들을 이용한 평균 적합선을 나타내었다. 95% 경험적인 신뢰구간은 200번을 반복하여 구한 t 시점의 실패확률을 오름차순으로 정렬하여 상위 여섯번째 값을 t 시점의 95% 경험적인 신뢰구간의 하한으로 정의하고, 하위 여섯번째 값을 t 시점의 95% 경험적인 신뢰구간의 상한으로 정의한다. 95% 근사 신뢰구간의 평균은 3.3절의 델타 방법을 이용하여 t 시점에서의 근사 신뢰구간을 200번 반복하여 각각 구한 후, 그 값들의 평균으로 나타내었다. 참값을 이용한 실패확률과 각 추정값들을 이용한 평균 적합선의 차이가 거의 나타나지 않았고, 95% 근사 신뢰구간의 평균과 95% 경험적인 신뢰구간과의 차이가 거의 보이지 않을 정도로 일치하는 형태를 보였다. θ_3 의 값은 학습곡선의 변곡점의 위치를 나타내며 θ_4 는 학습곡선의 휘어짐 정도를 나타낸다. θ_4 가 커질수록 학습곡선의 완만하게 휘어지며 이는 동일한 θ_3 값을 가질 때, 시작 시점에 비해 학습이 늦게 형성되는

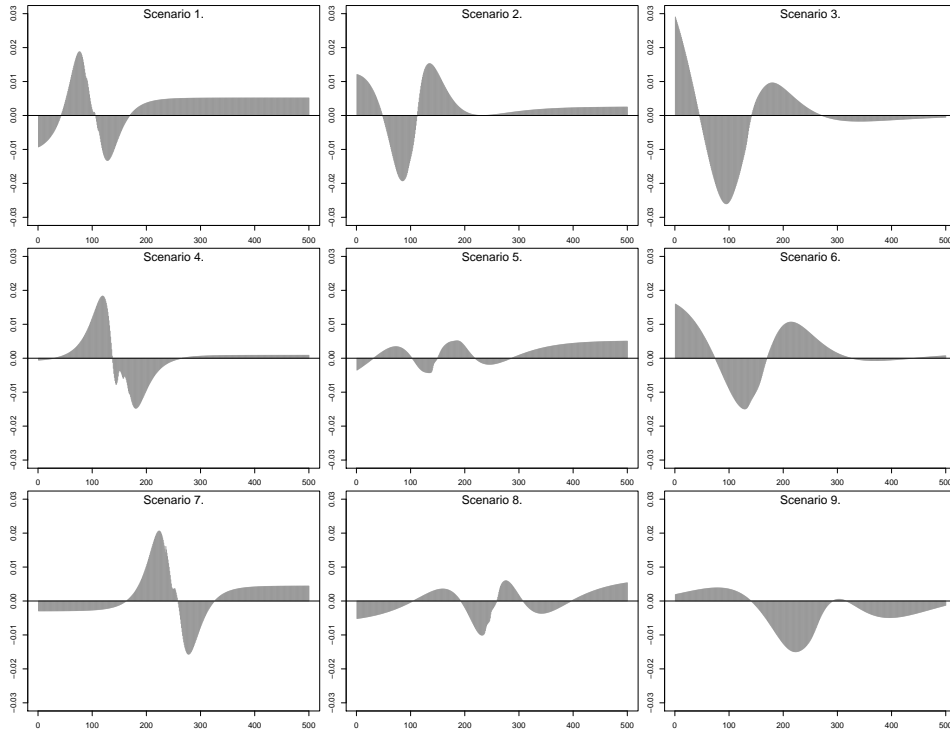


그림 4: 시나리오에 따른 관측시점별 참확률과 추정확률의 차이

것을 확인할 수 있다.

그림 3은 작업의 성공여부에 대한 실패확률의 전체적인 윤곽을 나타내고 있다. 이를 세부적으로 들여다 보기 위해 각 시나리오마다 관측시점별 참(실패)확률과 추정확률의 차이를 알아 보았다. 그림 4는 시점별 참확률과 추정확률 간의 간격을 나타낸 도표로서 학습곡선이 휘어지는 변곡점(학습이 형성되는 중간 지점)이 나타나는 시점(θ_3) 이전과 이후에서 변동이 크게 나타나고 있음을 알 수 있다. 즉, 추정값과 참값의 차이는 학습이 시작되는 시점과 변곡점에 도달하는 구간과 변곡점으로부터 안정화 단계에 이르는 구간에서 크게 나타났다. 시나리오 5번과 8번의 경우에는 참확률과 추정확률의 차이가 다른 모의실험의 결과에 비해 비교적 작다는 것을 확인할 수 있으며 이런 경우는 안정적인 학습곡선의 형태를 보이고 있다. 그리고 다른 조건에서의 모의실험 결과에서는 참확률과 추정확률의 차이가 비교적 큰 파동을 그리고 있음을 알 수 있다. 모의실험마다 정해진 모수들의 값에 따라 차이를 나타내는 파동의 형태가 달라진다. 예를 들면, 시나리오 1번과, 4번, 그리고 7번의 경우와 같이 학습효과가 단기간에 일어나는, 즉 변동성의 정도를 의미하는 θ_4 가 작은 경우($\theta_4 = 10$)에는 학습효과가 나타나지 않는 초기 단계에서는 참확률보다 높게 추정되다가 학습효과가 발생되기 시작하는 단계에서는 참확률보다 낮게 추정되며 변곡점을 지나서 안정화 단계에 접어들기 전까지 다시 과대추정되는 경향을 보여주고 있다. 이후 안정화 단계에 접어든 이후에는 다소 과소추정되는 형태를 보이고 있다. 그러나 학습효과가 비교적 긴 시간동안 발생하는 경우(예를 들면 시나리오 3번, 6번, 그리고 9번)에는 앞의 경우와 반대되는 경향을 보이고 있다.

표 2는 각 시나리오에 따른 추정결과를 바탕으로, 식 (2.5)를 이용하여 구한 $100 \times k\%$ 능숙시점(λ_k)의 결과이다. 시나리오 1번에 의한 결과를 살펴보면, 능숙도(k)의 크기가 클수록 λ_k 의 SE의 값

표 2: 추정 결과를 이용한 $100 \times k\%$ 능숙시점(λ_k)

Scenario	k	λ_k	Mean	SE	MAD
1.	0.90	121.972	124.155	1.024	10.917
	0.95	129.444	132.600	1.312	14.811
	0.99	145.951	151.256	1.990	23.431
2.	0.90	165.916	160.128	1.974	23.725
	0.95	188.333	179.814	2.565	30.701
	0.99	237.853	223.302	3.970	47.844
3.	0.90	209.861	203.383	2.987	38.800
	0.95	247.221	233.779	3.942	49.315
	0.99	329.756	300.925	6.218	74.389
4.	0.90	171.972	173.719	1.077	12.903
	0.95	179.444	181.747	1.401	16.694
	0.99	195.951	199.482	2.150	26.867
5.	0.90	215.916	214.829	1.959	26.885
	0.95	238.333	237.170	2.525	36.043
	0.99	287.853	286.522	3.866	56.007
6.	0.90	259.861	252.929	3.014	36.850
	0.95	297.221	286.439	3.912	47.778
	0.99	379.756	360.467	6.030	72.366
7.	0.90	271.972	274.925	1.080	11.984
	0.95	279.444	283.520	1.384	16.295
	0.99	295.951	302.508	2.086	24.563
8.	0.90	315.916	317.368	2.327	26.115
	0.95	338.333	340.139	3.002	34.375
	0.99	387.853	390.443	4.545	51.407
9.	0.90	359.861	359.692	3.312	43.801
	0.95	397.221	396.018	4.244	56.807
	0.99	479.756	476.266	6.406	83.835

과 MAD의 값이 증가함을 알 수 있다. 능숙도가 90%인 시점은 약 122번째에서 나타날 것이고, 능숙도가 95%인 시점은 약 129번째에서 나타날 것이며, 약 146번을 시도하였을 때 능숙도가 99%에 도달할 것임을 알 수 있다. 모의실험 결과 능숙도가 90%인 시점은 평균적으로 약 124번째에서 나타났고, 능숙도가 95%인 시점은 약 133번째에서 나타났으며, 능숙도가 99%인 시점은 약 151번째에서 나타났음을 확인할 수 있다. 모수의 참값이 변화함에 따라 능숙시점의 변동성 정도를 평가하기 위해 SE와 MAD를 측정한 결과, 학습곡선의 변곡점(θ_3)보다 변곡점 자체의 변동폭(θ_4)이 능숙시점의 변동성에 많은 영향을 미치고 있음을 알 수 있다. 그 예로 시나리오 1번, 4번, 그리고 7번을 비교해 보면, 동일한 능숙도 하에서 계산된 능숙시점의 변동성은 변곡점이 증가하더라도 거의 유사하다는 것을 알 수 있다. 하지만 시나리오 1번, 2번, 그리고 3번을 비교하면, 즉 변곡점은 동일하나 변동폭을 달리한 경우, 고정된 능숙도 하에서 변동폭이 증가할수록 SE와 MAD가 커진다는 것을 알 수 있다. 따라서, λ_k 의 변동성은 결국 θ_4 에 의해 크게 좌우된다.

지금까지의 결과는 각 자료의 최종 관측시점(시행횟수)을 500으로 가정한 것이다. 표 3은 표 1과 마찬가지로 각 시나리오에 따른 모수의 참값과 추정값들의 평균(Mean), 표준오차(SE), MAD의 결과를 정리하였으나 최종 관측시점을 50으로 고려한 결과이다. 시나리오 1번을 기준으로 살펴보면 시나리오 2번, 그리고 3번을 비교하였을 때 $\theta_4 = 1$ 인 경우가 $\theta_4 = 3, 5$ 인 경우보다 SE와 MAD가 각각 0.159와 0.748으로 작은 값을 보였다. 또한 $\theta_3 = 10$ 으로 고정되었을 때 $\theta_4 = 5$ 인 경우보다 $\theta_4 = 1$ 인 경우가 θ_3 의 SE와 MAD의 값이 각각 0.358, 2.071으로 상대적으로 작은 값을 나타냈다. 이는 θ_3 가 고정되었을

표 3: 로지스틱확률분포를 이용한 학습곡선모형의 추정 결과 ($T = 50$)

Scenario	Parameter	Truth	Mean	SE	MAD
1.	θ_1	0.2	0.163	0.005	0.056
	θ_2	0.8	0.911	0.010	0.200
	θ_3	10	9.831	0.358	2.071
	θ_4	1	2.421	0.159	0.748
2.	θ_1	0.2	0.153	0.006	0.066
	θ_2	0.8	0.898	0.011	0.200
	θ_3	10	10.334	0.633	2.531
	θ_4	3	4.408	0.313	1.619
3.	θ_1	0.2	0.148	0.007	0.073
	θ_2	0.8	0.858	0.013	0.200
	θ_3	10	12.743	0.685	3.771
	θ_4	5	6.653	0.559	3.167
4.	θ_1	0.2	0.157	0.005	0.058
	θ_2	0.8	0.876	0.008	0.111
	θ_3	15	15.308	0.225	1.980
	θ_4	1	2.538	0.193	0.763
5.	θ_1	0.2	0.158	0.006	0.063
	θ_2	0.8	0.877	0.011	0.200
	θ_3	15	14.864	0.534	3.060
	θ_4	3	3.953	0.223	1.512
6.	θ_1	0.2	0.137	0.007	0.079
	θ_2	0.8	0.872	0.011	0.200
	θ_3	15	16.153	0.776	4.059
	θ_4	5	5.491	0.303	2.697
7.	θ_1	0.2	0.160	0.006	0.057
	θ_2	0.8	0.836	0.007	0.070
	θ_3	25	25.037	0.241	1.482
	θ_4	1	1.697	0.151	0.957
8.	θ_1	0.2	0.141	0.007	0.075
	θ_2	0.8	0.846	0.008	0.085
	θ_3	25	25.589	0.363	3.036
	θ_4	3	3.796	0.247	1.854
9.	θ_1	0.2	0.139	0.008	0.100
	θ_2	0.8	0.841	0.009	0.096
	θ_3	25	25.809	0.488	4.302
	θ_4	5	4.893	0.269	2.705

때 θ_4 가 작을수록 θ_4 의 변동성이 작으며, θ_3 또한 안정적인 값을 나타내고 있다. 이는 장기적인 시행횟수($T = 500$)에 대한 모의실험의 결과와 비슷한 양상을 보이고 있다. 시나리오 1번을 4번, 그리고 7번과 비교하였을때 $\theta_3 = 10$ 인 경우가 $\theta_3 = 15, 25$ 인 경우보다 SE와 MAD가 각각 0.358과 2.071으로 약간 큰 값을 나타냈으며, θ_4 가 고정되었을때 $\theta_3 = 15$ 인 경우에 대체로 안정적인 값을 나타냈다. 즉, 장기적인 시행횟수($T = 500$)에 대해서는 θ_4 가 고정되었을때 θ_3 의 값의 변화가 모형에 미치는 영향력이 미미하였으나, 단기적인 시행횟수($T = 50$)인 경우에는 θ_4 가 고정되었을때 학습곡선의 변곡점의 시점이 실제 확률이 감소하기 시작하는 시점과 학습이 최종적으로 형성되는 시점의 평균 시점에서 안정적인 형태를 보인다는 것을 확인할 수 있다.

이상의 모의실험을 통해 학습곡선의 모형화를 위해 사용된 통계적 모형의 성능을 평가하였다. 본 연구에서는 로지스틱분포를 이용한 누적확률함수의 형태를 학습곡선의 모형을 구축하는 데에 사용하

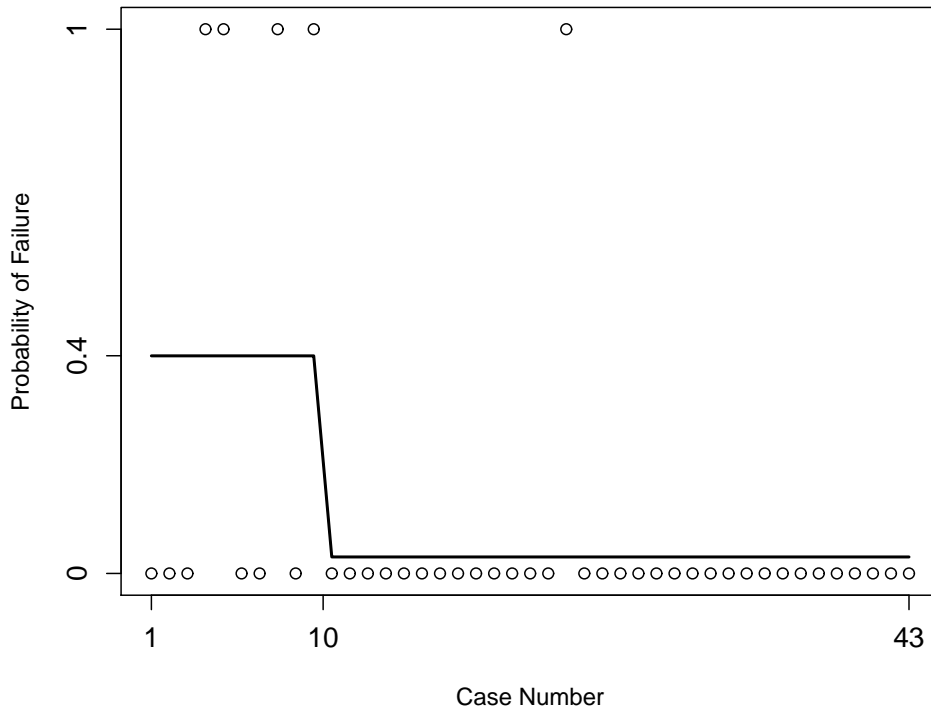


그림 5: 유방 생체검사 자료와 모형적합 결과

였으며 각 모수의 추정값과 추정된 (공)분산으로부터 관측시점별 실패확률을 계산하였다. 이 논문에서는 포함되어 있지 않으나 초기 실패확률(θ_2)과 안정화단계에서의 실패확률(θ_1)에 대한 모의실험도 진행하였으며 그 결과 추정에 대해서는 큰 영향을 주지 않았음을 확인하였다.

4. 실증연구

본 연구에서 제안한 학습곡선의 통계적 모형을 실제자료에 적용하였다. 분석결과에 대한 객관성을 부여하기 위해 실증연구에서 사용된 자료는 Lim 등 (2002)에서 언급된 방사선 전문의의 유방생체검사(breast biopsy)자료이다 (그림 3 참조). 이 자료는 한 명의 방사선 전문의가 43번의 반복된 검사시행을 통해 관측하게 된 유방촬영술에 의한 유방생체검사의 성공 여부에 대한 것이다. 그림 5를 통해 각 검사 시행 후의 성공 여부를 확인해 보면 10번째 검사시행 이전에 4번의 실패를 이 방사선 전문의가 경험하였고 그 이후에는 오직 한 번의 실패를 경험하였음을 알 수 있다. Lim 등 (2002)은 이 자료에 누적합분석을 적용하여 학습효과가 완성된 시점을 10회라고 판단하였다. 이 때, 용인되지 않을 실패율을 5%로, 용인되는 실패율을 2%로 설정하였다.

본 연구에서 제안한 학습곡선에 대한 통계적 모형을 주어진 자료에 적용한 추정 결과는 표 4에 제시되어 있다. 이 표를 통해 알 수 있듯이, 초기의 실패 확률(θ_2)은 약 40%, 안정화단계에서의 실패 확률(θ_1)은 약 3%, 학습이 일어나는 지점(변곡점)은 약 10.5회로 추정되었다. 또한 학습곡선의 변동폭(θ_4)은 0.053으로 나타났으며, 이는 그림 5에서 확인할 수 있듯이 초기의 다소 빈번했던 검사 실패가 10여 회의 경험을 통해 급속하게 개선되었음에 기인한다. 이로 인해 학습효과의 완성에 이르는 시간이 매우 짧은 것으로 나타났다.

표 4: 실제 자료 (Lim 등, 2002)를 이용한 적합

Parameter	Estimate	Standard error	p-value
θ_1	0.030	0.021	0.159
θ_2	0.400	0.110	0.001
θ_3	10.522	3.758	0.008
θ_4	0.053	0.451	0.906

Lim 등 (2002)의 누적합을 이용한 결과와 본 논문에서 제안한 모형의 학습이 일어나는 시점의 결과가 각각 10회와 10.5회로 큰 차이가 나진 않았으나 Lim 등 (2002)의 누적합분석은 학습이 일어나는(혹은 완성되는) 시점을 구하기 위해서는 경험적이고 주관적인 값들이 필요하다는 단점이 존재한다.

5. 결론

기존의 학습곡선에 대한 연구는 학습이 일정 수준 이상으로 능숙해 지는 시점이 언제인지를 찾기 위한 목적으로 이루어졌으며 이를 위해 대부분의 연구에서 통계적 모형에 의한 적합보다는 연구자의 경험이나 사전 사례분석의 결과를 토대로 진행되어 왔다. 따라서 동일한 자료가 주어졌다 하더라도 연구자마다 학습곡선이 완성되어지는 (또는 극복되어지는) 시점을 각기 설정할 가능성이 높다. 결과적으로 연구의 결론이 달라지게 되므로 일반화가 어렵다는 단점을 가지고 있다. 이에 본 연구에서는 누적확률분포의 특성을 이용하여 이항반응자료(성공/실패)에 대한 학습곡선의 효과를 통계적 모형으로 모형화하고자 하였다. 이를 통해 특정 작업의 능숙한 정도(이하 능숙도)가 일정 수준 이상이 되는 시점이 언제인지를 모형에 근거하여 찾고자 하였다. 또한 특정 작업을 최초 시도하였을 때 소요되는 시간과 학습곡선의 효과가 완성이 되었을 때 소요시간 등을 예측하였고 소요시간에 대한 근사적인 신뢰구간을 구성하였다. 이를 통해 특정 시점의 소요시간이 어느 정도인지, 그리고 특정한 작업을 얼마만큼 반복적으로 수행해야 원하는 소요시간이 달성되는지 등의 정보를 제안된 통계적 모형으로부터 얻고자 하였다.

본 연구에서 고려하는 모수는 네 가지로, 더 이상 줄일 수 없는 실패 확률(θ_1), 실험 초기의 실패 확률(θ_2), 위치모수(θ_3), 규모모수(θ_4)이다. 여기서 θ_1 은 실험을 진행하는 사람의 한계를 가늠할 수 있는 지표로 실험 초기의 실패확률인 θ_2 와의 차이를 계산하면 실험 초기와 실험 종료 사이에 얼마나 향상되었는지 파악이 가능하다. 다음으로 θ_3 는 학습 곡선의 변곡점을 나타내는 값으로 는 시행 초기에 빠르게 형성되는 학습 능숙 정도가 완만해지기 시작하는 시점을 의미하고 이를 통하여 시작부터 추정된 값까지 걸린 시간을 통해 학습효율이 좋은 시간이 얼마인지 유추할 수 있다. 마지막으로 학습의 변동폭을 결정하는 θ_4 는 값이 클수록 능숙해지기 까지 많은 시간이 소요됨을 의미하므로 θ_4 의 값이 절대적으로 쓰이기보다는 나머지 모수들이 고정되었을 때(혹은 비슷할 때) θ_4 의 값을 비교하여 학습이 능숙해지기 까지 걸리는 효율을 비교할 수 있다.

모수를 참값에 가까운 값으로 추정하기 위하여 뉴턴-랩슨 방법을 사용하였고, 델타 방법으로 최대 우도함수를 이용하여 신뢰구간을 추정하였다. 9가지 시나리오에 따라 모의 실험을 한 결과에 의하면 학습이 일어나는 시점부터 학습이 종료되는 시점까지의 변동폭이 큰 경우는 적합이 잘 이루어지지 않으며, 특히 학습곡선이 휘어지는 변곡점에서의 시점(θ_3)과 학습이 이루어지는 기간(θ_4)이 참값과 추정값 사이의 차이에 영향을 미쳤다. 즉, 학습곡선의 변곡점 시점을 결정하는 θ_3 를 고정하였을 때, 학습곡선의 변동폭을 결정하는 θ_4 의 값이 작을수록 θ_3 와 θ_4 의 값이 안정적이고, θ_4 를 고정하였을 때는 θ_3 의 값의 변화가 모형에 미치는 영향력이 미미하다는 사실을 알 수 있었다. 작업에 완벽하게 능숙해지는 상태를 100%로 볼 때, 작업에 $100 \times k\%$ 정도 능숙해지는 시점을 구할 수 있고, 실패 확률에 대한 95% 신뢰구간을 통하여 특정 시점에서 가능한 실패 확률의 범위와, 특정 실패 확률이 나타날 수 있는 시행횟

수의 범위를 구할 수 있었다. Lim 등 (2002)이 분석한 유방생체검사에 대해 본 논문에서 제안한 모형을 적용한 결과 학습이 일어나는 시점은 Lim 등 (2002)의 결과와 본 논문에서 제안한 모형의 결과와 큰 차이를 나타내지 않았으나, 본 논문은 기존의 학습곡선을 이용한 분석에 비하여 연구자의 주관적인 견해가 배제된 객관적인 학습곡선의 모형화가 가능함을 보였다.

궁극적으로 주어진 모형의 적합과 모수에 대한 추정을 통해 어느 정도의 반복적 경험과 훈련을 거치게 되면 해당 작업에 대해 능숙해지는지를 명확히 설명할 수 있다. 또한 이 정보를 통해 앞으로 어떤 작업을 배우려는 수련생들의 교육과 훈련에 도움이 되는 지침이 될 수 있을 것이다. 향후에는 이항분포 뿐 만 아니라, 보다 다양한 확률분포를 고려한 학습곡선의 구축을 연구하고자 하며, 모수의 추정에 있어서 베이지안(Bayesian)기법을 적용하고자 한다.

참고 문헌

- 김학선, 최윤락, 이환모, 문성환, 김경희, 박진오, 하중원, 신동은 (2003). 흉곽 내시경을 사용한 척추 기형 수술시 학습곡선, <대한척추외과학회지>, **10**, 261-268.
- 서강열, 손유동, 안지운, 안희철, 조준휘 (2010). 학습곡선-누적합 분석을 이용한 흉부압박의 능숙도 평가, <대한응급의학>, **21**, 293-298.
- 송수진, 김철호, 김성희, 박은영, 박근식, 문화숙, 김경서, 주보선, 김상갑 (2003). 내시경하 자궁근종절제술의 learning curve, <대한산부회지>, **46**, 2345-2351.
- 오윤경, 황효순, 이경옥, 송승훈, 이재관, 허준용, 신정호 (2010). 세 명의 산부인과 전문의의 별 복강경하 전자궁절제술의 학습곡선, <Korean Journal of Obstetrics and Gynecology>, **53**, 927-933.
- 최유신, 박도중, 이혁준, 김민찬, 김형호, 양한광, 한호성, 이건욱 (2006). 조기위암에 시행된 복강경 보조 원위부 위절제술의 학습곡선 극복 시점 및 극복 전후의 비교, <대한외과학회지>, **70**, 370-374.
- Adler, P. S. and Clark, K. B. (1991). Behind The learning curve: A sketch of The learning process, *Management Science*, **37**, 267-281.
- Ballantyne, G. H., Ewing, D., Capella, R. F., Capella, J. F., Davis, D., Schmidt, H. J., Wasielewski, A. and Davies, R. J. (2005). The learning curve Measured by Operating Times for Laparoscopic and Open Gastric Bypass: Roles of Surgeon's Experience, Institutional Experience, Body Mass Index and Fellowship Training, *Obesity Surgery*, **15**, 172-182.
- Biau, D. J., Williams, S. M., Schlup, M. M., Nizard, R. S. and Porcher, R. (2008). Quantitative and individualized assessment of the learning curve using LC-CUSUM, *British Journal of Surgery*, **95**, 925-929.
- Cook, J. A., Ramsay, C. R. and Fayers, P. (2004). Statistical evaluation of learning curve effects in surgical trials, *Clinical Trials*, **1**, 421-427.
- Ferguson, G. G., Ames, C. D., Weld, K. J., Yan, Y., Venkatesh, R. and Landman, J. (2005). Prospective evaluation of learning curve for laparoscopic radical prostatectomy: Identification of factors improving operative times, *Adult Urology*, **66**, 840-844.
- Forbes, T. L., DeRose, G., Kribs, S. W. and Harris, K. A. (2004). Cumulative sum failure analysis of the learning curve with endovascular abdominal aortic aneurysm repair, *Journal of Vascular Surgery*, **39**, 102-108.
- Kim, M. C., Jung, G-J. and Kim, H-H. (2005). learning curve of laparoscopy-assisted distal gastrectomy with systemic lymphadenectomy for early gastric cancer, *World Journal of Gastroenterology*, **47**, 7508-7511.
- Lee, J. H., Ryu, K. W., Lee, J-H., Park, S. R., Kim, C. G., Kook, M. C., Nam, B-H., Kim, Y. W. and Bae, J-M. (2006). learning curve for Total Gastrectomy with D2 Lymph Node Dissection: Cumulative sum analysis for qualified surgery, *Annals of Surgical Oncology*, **13**, 1175-1181.

- Lieberman, M. B. (1984). The learning curve and pricing in the chemical processing industries, *Rand Journal of Economics*, **15**, 213–228.
- Lim, P. C., Kang, E. and Park, D. H. (2011). A comparative detail analysis of the learning curve and surgical outcome for robotic hysterectomy with lymphadenectomy versus laparoscopic hysterectomy with lymphadenectomy in treatment of endometrial cancer: A case-matched controlled study of the first one hundred twenty two patients, *Gynecologic Oncology*, **120**, 413–418.
- Lim, T. O., Soraya, A., Ding, L. M. and Morad, Z. (2002). Assessing doctors' competence: Application of CUSUM technique in monitoring doctor' performance, *International Journal for Quality in Health Care*, **14**, 251–258.
- Mazzola, J. B. and McCardle, K. F. (1996). A Bayesian approach to managing learning curve uncertainty, *Management Science*, **42**, 680–692.
- Schauer, P., Ikramuddin, S., Hamad, G. and Gourash, W. (2002). The learning curve for laparoscopic Roux-en-Y gastric bypass is 100 cases, *Surgical Endoscopy*, **17**, 212–215.
- Sim, H. G., Yip, S. K., Lau, W. K., Tan, Y. H., Wong, M. Y. and Cheng, C. W. (2006). Team-based approach reduces learning curve in robot-assisted laparoscopic radical prostatectomy, *International Journal of Urology*, **13**, 560–564.
- Smith, A. C., Frank, L. M., Wirth, S., Yanike, M., Hu, D., Kubota, Y., Graybiel, A. M., Suzuki, W. A. and Brown, E. N. (2004). Dynamic analysis of learning in behavioral experiments, *The Journal of Neuroscience*, **24**, 447–461.
- Smunt, T. L. (1999). Log-linear and non-log-linear learning curve models for production research and cost estimation, *International Journal of Production Research*, **37**, 3901–3911.
- Tekkis, P. P., Senagore, A. J., Delaney, C. P. and Fazio, V. W. (2005). Evaluation of the learning curve in Laparoscopic Colorectal Surgery, *Annals of Surgery*, **242**, 83–91.
- Williams, C. K. I. and Vivarelli, F. (2000). Upper and Lower Bounds on the learning curve for Gaussian Processes, *Machine Learning*, **40**, 77–102.

Statistical Modeling of Learning Curves with Binary Response Data

Seul Ji Lee^a, Man Sik Park^{1,a,b}

^aDepartment of Statistics, Sungshin Women's University

^bInstitute of Statistics, Sungshin Women's University

Abstract

As a worker performs a certain operation repeatedly, he tends to become familiar with the job and complete it in a very short time. That means that the efficiency is improved due to his accumulated knowledge, experience and skill in regards to the operation. Investing time in an output is reduced by repeating any operation. This phenomenon is referred to as the learning curve effect. A learning curve is a graphical representation of the changing rate of learning. According to previous literature, learning curve effects are determined by subjective pre-assigned factors. In this study, we propose a new statistical model to clarify the learning curve effect by means of a basic cumulative distribution function. This work mainly focuses on the statistical modeling of binary data. We employ the Newton-Raphson method for the estimation and Delta method for the construction of confidence intervals. We also perform a real data analysis.

Keywords: Learning curve, binary response, cumulative distribution function, Newton-Raphson method, Delta method.

This work was supported by the Sungshin Women's University Research Grant of 2011.

¹ Corresponding author: Assistant Professor, Department of Statistics, Sungshin Women's University, 249-1 Dongseondong 3-ga, Seongbuk-Gu, Seoul 136-742, Korea. E-mail: mansikpark@sungshin.ac.kr