**Mathematical modeling of AIDS proliferation**

1. **개요**

질병관리본부의 ‘2016 HIV/AIDS 신고 현황’에 따르면 2016년까지 국내 누적 감염자가 11,439명, 약 1천명의 새 감염자가 신고 접수 되고 있다. HIV/AIDS의 위험성이 더욱 위협적으로 증가하고 있는 가운데, 우리는 HIV/AIDS 감염 확산모델에서 고려되는 질병의 원인 요소들의 영향력을 추정하여, HIV/AIDS 효과적인 억제 방안을 모색하고자 한다. 질병 확산 모델에 영향을 주는 시스템 계수들을 추정하기 위해 본 연구에서는 변수 추정 알고리듬을 통해 HIV/AIDS 확산 모델을 추정한다.

1. **수학적 모델링**

질병 확산의 수학적 모델에서 사용되는 시스템 변수와 매개변수는 figure 1, 2 와 같으며 감염 확산 모델의 미분 방정식은 figure 3 과 같이 표현되며 블록 선도는 figure 4와 같다.

|  |  |
| --- | --- |
|  | **연간 누적된 남성 이성애자** |
|  | **연간 누적된 여성 이성애자** |
|  | **연간 누적된 남성 동성애자** |
|  | **연간 누적된 감염된 남성 동성애자** |
|  | **연간 누적된 감염된 남성 이성애자** |
|  | **연간 누적된 감염된 여성 이성애자** |
|  | **연간 누적 사망자** |
| **Figure 1. System Variables** | |

|  |  |  |
| --- | --- | --- |
|  | **안정성, 위생교육(예방교육)** | |
|  | **성 정체성 교육 ( 에서 로 가는 것을 방지)** | |
|  | **탈동성애 교육** | |
|  | **PreP 효과** | |
|  | **매년 에서 로 전환되는 비율** | |
|  | **AIDS 감염자와 이성간 성관계시 감염률** | |
|  | **AIDS 감염자와 남성간 성관계시 감염률** | |
|  | **이성간 성관계 횟수 평균** | |
|  | **남성간 성관계 횟수 평균** | |
|  | **감염률과 성관계 횟수에 따른 연간 감염률** | |
| **Figure 2. System Parameters** | | |
|  | | **(1)** |
|  | | **(2)** |
|  | | **(3)** |
|  | | **(4)** |
|  | | **(5)** |
|  | | **(6)** |
|  | | **(7)** |
| **Figure 3. System Equations** | | |

|  |
| --- |
| ***KakaoTalk_Photo_2017-10-19-02-44-39.png*** |
| **Figure 4. System Block Diagram** |

1. **매개변수 추정**
   1. **가정**

추정하려는 질병확산모델은 다음과 같은 가정을 가진다.

|  |
| --- |
| **1. 2016년 이전 의 영향력은 무시** |
| **2. 출생률과 사망률은 알려져있다.** |
| **3. 매년, 동성애자의 수는, 11세 이상 남성 수의 0.05% 둔다.** |
| **4. 매년 에서 0.02% 로 둔다.** |
| **5. 처음11세가 되는 모든 사람은 감염되지 않은 이성애자 이다.** |
| **6. 국내HIV/AIDS 실제 감염자 수는 신고자 수 \* 10 으로 계산한다.** |

* 1. **상태 공간 방정식**

Figure 3 에 표현 되어 있는 수학적 모델에서 시스템 파라미터를 바로 추정하기보다는, 모델링의 용이성을 위해 figure 4 와 같이 각 시스템 변수들 앞의 수식을 시스템 계수 로 표현하였다.  **가 추정된 이후에 질병 확산 모델의 실제 파라미터를 추정하고자 한다.**

|  |  |
| --- | --- |
|  | **(8)** |
|  | **(9)** |
|  | **(10)** |
|  | **(11)** |
|  | **(12)** |
|  | **(13)** |
|  | **(14)** |
| **Figure 4. Reduced System Equations** | |

시스템의 상태 변수를 식(15) 같이 표현 하면

|  |  |
| --- | --- |
|  | **(15)** |

시스템의 상태 공간 방정식은 다음과 같이 표현된다.

|  |  |
| --- | --- |
|  | **(15)** |
| **,** | **(16)** |
| **Figure 5. State Space Equation** | |

1. **확장 칼만 필터를 이용한 파라미터 상태 추정 기법 (SLAC)**

확장칼만필터는 시스템의 입력에 대한 출력과 이미 알려진 참조 데이터에서 발생하는 오차와 변수의 불확정성에 대한 지표인 분산(Variance) 를 통해 최적의 시스템 상태 변수를 추정하는 알고리듬이다. 이 때, 를 시스템 상태 변수에 추가함으로서 값을 추정하고자 하는데 이것이 확장 칼만 필터를 이용한 SLAC(Simultaneous Localization and Auto Calibration) 기법 이다.

* 1. **확장칼만필터 알고리듬**확장칼만필터의 기본적인 알고리듬과 표기에 대한 설명은 다음과 같으며, 자세한 수식과 설명은 다음 장에서 설명하겠다.

|  |  |  |  |
| --- | --- | --- | --- |
| **Step** | **Calculation** | **Meaning** | 추정한 상태 변수 예측된 상태 변수  참조 값(측정값)  시스템 입력칼만 이득추정된 상태 공분산 행렬 예측된 상태 공분산 행렬 시스템 행렬 공분산 행렬 |
| 1 |  | 추정된 공분산 |
| 2 |  | 칼만 이득 계산 |
| 3 |  | 상태 변수 추정 값 |
| 4 |  | 예측된 공분산 |
| 5 |  | 상태 변수 예측 값 |

* 1. **SLAC 알고리듬 및 방정식**

**파라미터를 추가한 상태 변수( 는 다음과 같으며**

|  |  |
| --- | --- |
|  | **(17)** |

**시스템 모델은 다음과 같이 표현된다.**

|  |  |
| --- | --- |
|  | **(18)** |

**이때, 시스템 함수 와 시스템 입력 는 다음과 같다.**

|  |  |
| --- | --- |
|  | **(19)** |
|  | **(20)** |

**확장 칼만 필터 알고리듬의 행렬은 측정 변수()로부터 시스템의 상태 변수()의 차원으로 사상해주는 역할을 한다.**

**또한, , 행렬은 순서대로 측정값과 시스템 입력에 대한 오차공분산을 나타내는 행렬이다.**

**는 다음과 같이 표현된다.**

|  |  |
| --- | --- |
|  | **(21)** |
|  | **(22)** |
|  | **(23)** |
|  | **(24)** |

**다음으로, 알고리듬4번째 단계 ”** **에 등장하는 행렬은 시스템 함수 을 각각 시스템의 상태변수( 와 시스템의 입력( 의Jacobian 행렬 (다변수 벡터 함수의 도함수 행렬) 이다.**

**의미적으로는, 시점에서 시스템 함수 의 순간 변화량을 도출함으로서 상태 변수의 공분산 행렬을 업데이트 하는데 이용된다.**

|  |  |
| --- | --- |
|  | **(25)** |
|  | **(26)** |

1. **모의실험 결과**
   1. **Scenario 1**

|  |
| --- |
| 1. **\*** 2. **\*** 3. **연간 출산율/사망률은 통계청 자료를 참고** 4. **92년 이후 에이즈 신고자 수는 질병관리본부 통계 참고** 5. **92년 이전 에이즈 누적 신고자 수 = 200명 가정(성비 1:1) \***   **(\* 표시는 가정값)** |
| **Figure 6. Simulation preset (Scenario 01)** |

**Figure 6와 같은 설정을 한 후, SLAC 알고리듬을 통해**  를 추정하였고 결과는 다음과 같다.

|  |
| --- |
| KALMAN/groups_except_infected_groups_scence3.png |
| **Figure 7. Groups [Infected groups excluded] (Scenario 01)** |

|  |
| --- |
| KALMAN/infected_groups_scence3.png |
| **Figure 8. Infected Groups (Scenario 01)** |

의 경우에는 대체로 일치하는 경향을 보이지만, 남성 동성애자 중 에이즈에 감염된 수 ( 에서 2000 년 이후부터 오차가 크게 발생한다. 남성동성애자 그룹은 인원이 지속적으로 증가하고 있지만, 2016 에 가까워 지면서 변화율이 낮아지다가 음수로 바뀌게 되는데, 이것은 시간이 지날수록 남성 동성애자 에이즈 감염된 수 가 많아지면서 더욱더 많은 동성애자 감염자가 발생하는 것으로 추측할 수 있다.

|  |
| --- |
| **KALMAN/fit_corr_scence3.png** |
| **Figure 9. Fitted Data Correlation (Scenario 01)** |

HIV/AIDS 감염자수, 남성 동성애자 그리고 감염된 남성 동성애자 수를 정확하게 파악할 수 있는 통계치가 없기 때문에 어느정도 오차가 발생할 수 밖에 없다. 하지만, 실제로 가정한 값과 모델을 통해 모의실험을 한 값 사이의 상관계수를 확인해보면 Figure 8 과 같으며 질병 확산 경향성이 매우 높게 나타남을 확인 할 수 있다.

|  |
| --- |
| **KALMAN/parameter_est_scence3.png** |
| **Figure 10. Parameter Estimation (Scenario 01)** |

Figure 8 은SLAC 알고리듬의 파라미터 추정 추이를 나타낸 그래프이다. 질병 확산 모델은 시간에 따라 시스템의 특성이 바뀌는 시변 시스템 (Time-Variant System) 이기 때문에, 추정 값이 지속적으로 변화 할 수 밖에 없다. 하지만, 알고리듬 내부 파라미터(변수 표준편차, 초기값) 등을 적절하게 설정하지 못하거나, 참조 데이터의 신뢰성이 떨어질 경우 추정 파라미터값이 발산하거나, 수렴하더라도 최종 파라미터 추정 값의 최적화 성능에 큰 영향을 미치게 된다. 따라서, 본 시나리오에서는 개략적으로 파라미터의 존재 범위를 추정한 후, 직접 보정하는 작업을 거쳤다. 값은 전체 년도 추정치의 평균값으로 취하고 상수로 유지하였다.

SLAC알고리듬 자체적으로는 실제 값을 잘 추종하지만 이것은 칼만이득을 통한 추가적인 보정이 들어간 결과물이기 때문에, SLAC 을 통해 추정한 파라미터를 질병 모델에 대입하더라도 같은 성능을 낼 수 없다.

1. **결론 & 향후 연구 계획**

칼만필터를 활용한 파라미터 추정은 시불변 선형 시스템에서 신뢰도가 결정된 데이터들이 주어질 수록 좋은 성능을 발휘하게 되어있다. 하지만, 질병 확산모델에서는 시스템의 특성이 시간에 따라 변화하며, 신뢰도가 확정적이지 않은 자료를 기반으로 하기 때문에 좋은 추정 성능을 내지 못하였다.

결론적으로, 파라미터의 분포 구간 정도를 추정 할 수 있었고, 파라미터 변화 추이를 통해 감염률의 변화 양상 정도를 추정해 볼 수 있었다.

전 장의 모의실험에서 소개한 시나리오 외 더 많은 시나리오에 대해 모의실험을 해보았지만, AIDS 감염자 수, 질병에 의한 사망률, 국내 동성애자 수를 다르게 가정함에 따라 파라미터 값이 상이하였다. 따라서, 시나리오 상황을 확정하고 난 후, 그리고 파라미터 추정 기법의 정확성을 좀더 확보한 후에 추가적으로 다른 시나리오를 진행할 예정이다.

향후 계획은 칼만 필터 알고리듬외 동적 시스템 파라미터 추정 기법에 대해 조사를 해볼 예정이다. 또한, 질병 확산 모델 시스템 자체를 시불변 선형 시스템으로 가정을 하고, 파라미터 입력에 대한 피팅 오차를 출력으로 가지는 함수를 통해 파라미터 변화에대한 최소값을 찾는 수치 해석 알고리듬또한 구현 해볼 예정이다.