

A Statistical Model for Severe Acute Respiratory Syndrome

Yeon Woong Hong¹⁾

Abstract

The severe acute respiratory syndrome(SARS) is a novel infectious disease with global impact. The rapid worldwide spread of SARS has led to 30 countries reporting cases of July 13, 2003. In this paper, we develop a statistical model for SARS-caused-death data under some assumptions. The model developed is a continuous time Markov process with a constant intensity for each stage.

Keywords : SARS, 통계모형, 수명분포

1. 서론

2002년 11월 1일부터 2003년 7월 13일까지 30개국에서 8,439명의 추정환자가 발생하여 813명의 사망자를 발생시켜 9.63%의 치사율²⁾을 나타내고 있는 중증급성호흡기증후군(Severe Acute Respiratory Syndrome : SARS)은 세계보건 당국의 비상경계령하에도 불구하고 진정되는 기미가 보이지 않고 있는데 이는 SARS를 확실하게 배제할 수 있는 진단방법과 치료제가 없게 때문이다. 세계보건기구(WHO)가 잠정적으로 정의한 SARS는 다음과 같이 증상에 따라 의심환자(Suspect case), 추정환자(Probable Case) 및 배제기준, 재분류기준으로 구분하고 있다.

1) 경북 영주시 풍기읍 교촌리 동양대학교 인터넷산업공학부 교수
E-mail : ywhong@phenix.dyu.ac.kr

2) WHO, 발표기관에 따라 차이가 있으며 본 연구에서는 발생환자수 대비 사망자수의 비율로 산출하였음

<표 1> SARS 환자 사례의 정의

환자구분	증 상
의심환자 (Suspect case)	1) 38도 이상의 고열이 있고, 기침 또는 호흡곤란이 있으면서 발병하기 10일전에 다음 중 한가지 혹은 그 이상에 폭로력이 있는 경우 ① 사스 의심 또는 추정환자와 밀접한 접촉 ³⁾ 을 한 경우 ② 사스감염위험지역 ⁴⁾ 의 여행력(공항 환승 포함)이 있는 경우 ③ 사스감염위험지역에 거주한 적이 있는 경우 2) 설명되지 않은 호흡기 질환으로 사망하였으나 부검이 시행되지 않았으며 발병하기 10일전에 다음 중 한가지 이상의 폭로력이 있는 경우 ① 사스 의심 또는 추정환자와 밀접한 접촉을 한 경우 ② 사스감염위험지역의 여행력(공항 환승 포함)이 있는 경우 ③ 사스감염위험지역에 거주한 적이 있는 경우

<표 1> SARS 환자 사례의 정의(계속)

환자구분	증 상
추정환자 (Probable Case)	1) 의심환자이면서 흉부방사선 소견상 ① 폐렴에 합당한 침윤 소견이 있거나 ② 호흡곤란증후군(Respiratory Distress Syndrome) 소견을 보이는 경우 2) 사스 코로나바이러스 양성인 의심환자 3) 부검소견상 원인불명의 호흡곤란증후군 병리소견을 보인 의심환자
SARS 환자	1) PCR(polymerase chain reaction) 양성 확진(다음 중 1개) ① 적어도 두종류의 다른 검체에서 양성인 경우 (예를들어 비인두검체와 대변검체) ② 동일한 검체라도 임상 경과동안 2일 이상 간격을 두고 채취한 검사에서 양성인 경우(예를들어 비인두 검체를 2-3일간격으로 검사) ③ 원 검체에 대해 2회 검사를 반복하여 양성인 경우 2) 항체검사 양성 : ELISA, IFA (다음 중 1개) ① 항체가 양전된 경우(급성기에 음성이었는데 회복기에서 양성으로 전환된 경우) ② 급성기와 회복기 항체가 4배이상 증가한 경우 3) 바이러스 분리 ① PCR에 의해 확진된 검체에 대해 세포배양검사에서 바이러스가 분리

본 연구에서는 SARS 사망 데이터에 대한 통계모형을 개발하여 SARS 데이터를 분석할 수 있게 하고, 아울러 SARS에 대한 의학통계적인 접근을 시도하고자 한다. 이를 위하여 본 연구에서는 데이터 구조 및 사망이 발생하기까지 단계별로 필요한 통

3) 밀접한 접촉력 : 사스 의사 또는 추정 환자를 치료, 간호했거나, 같이 살거나, 호흡기분비물 혹은 체액을 직접적으로 접촉한 경우

4) 사스 감염위험지역 : 최근 사스의 지역내 전파가 있는 지역으로 중국 본토, 홍콩, 대만, 싱가포르, 캐나다 토론토, 몽골 울란바토르(5월 2일 현재)

계적인 가정을 전제로 통계모형을 제안하였다.

2. 통계모형

2.1 단계별 증상

중국에서 발생해 홍콩, 싱가포르 등 전세계로 확산되고 있는 전염병인 SARS는 홍콩의 Amoy Gardens에서 집단적으로 발병하였는데, 이들을 관찰한 결과 SARS 환자들은 3단계 임상경과를 보였다고 한다. 제1단계에서는 발열, 근육통 등 증상을 보여 몇 일 안에 회복이 되었으며, 감염력이 커지는 두 번째 주는 발열, 설사, 산소부족 증상들이 재발되며, 이들 중 약 20%가 3단계로 진행되어 호흡곤란증후군이 발생하여 기계호흡이 필요하게 되었다고 한다(국립보건원 홈페이지[1]). 결국 임상적인 관찰에 따르면 SARS로 인한 사망자는 제1단계부터 제2, 3단계를 지나서 사망에 이르기까지 다단계의 진행과정을 거친다는 것을 알 수 있으며, 이를 정리하면 <표 2>와 같다.

<표 2> SARS로 인한 사망까지의 단계 및 증상

단계	증상
제1단계	발열, 근육통 등 증상을 보임
제2단계	발열, 설사, 산소부족 증상이 재발됨
제3단계	호흡곤란증후군이 발생하여 기계호흡
제4단계	사망

2.2 가정 및 관찰 데이터 구조

통계모형을 개발하기 위하여 다음과 같은 가정을 한다.

가정 1: SARS로 인한 사망자의 경우 관찰 기간중 증상이 호전되어 관찰대상에서 제외되는 환자는 없다. 이러한 가정은 본 연구가 모형의 단순화를 위하여 사망통계만 다루기 때문이다.

가정 2: 어떤 단계에서 다음 단계로의 진행률(force of progression)은 일정하다.

가정 3: 단계 i 에 있는 시간은 단계 j 에 있는 시간에 영향을 미치지 않는다. 즉 서로 독립이다.

가정 5: SARS에 감염되어 사망에 이르기까지의 몇 주 또는 몇 개월의 관찰기간동안 환자들의 증상은 <표 3> 및 <표 4>와 같이 정리될 수 있다.

<표 3> 관측구간별 증상단계별 환자수 분포

	증상단계1	증상단계2	증상단계3
관측구간 1	n_{11}	n_{12}	n_{13}
관측구간 2	n_{21}	n_{22}	n_{23}
...
관측구간 k	n_{k1}	n_{k2}	n_{k3}

<표 4> 관측구간별 증상단계별 SARS사망자 및 증상악화자의 분포

	증상단계1	증상단계2	증상단계3
관측구간 1	x_{11}	x_{12}	x_{13}
관측구간 2	x_{21}	x_{22}	x_{23}
...
관측구간 k	x_{k1}	x_{k2}	x_{k3}

그러면 n_{ij} 는 관측구간 i 에서 증상단계 j 에 속한 환자의 수이고, x_{ij} 는 n_{ij} 명의 환자 가운데 SARS로 사망했거나 증상이 악화된 환자의 수를 의미한다. 단계 4는 확정적인 상태이므로 모형의 설정과정에서는 포함되지 않는다. k 는 관측구간의 수를 나타내며 통계적으로 x_{ij} 와 x_{ij}' 의 차이가 유의할수록 결정하는 것이 바람직하며 SARS의 경우 발병 시점부터 사망 시점까지의 기간이 단기인 측면에 비추어 3 또는 4가 적절하다고 사료된다. 병원이나 보건당국에서 <표 3> 및 <표 4>의 양식으로 SARS 데이터를 관리하고 있는지 확인 할 수 없었지만 사망자의 진료기록을 참고하면 이러한 형태로 정리할 수 있을 것으로 기대된다.

2.3 통계모형

가정 2에 따라 λ_j , $j=1,2,3$ 을 단계 j 에서 그 다음 단계로의 진행률이라하고, T_j 를 환자가 단계 j 에 머무는 시간을 나타내는 확률변수라고 하면, i 번째 관측구간에서 단계 j 상태의 환자가 t 시간 이상 단계 j 에 있을 확률은 $P(T_j > t) = \exp(-\lambda_j t)$ 로 지수분포를 따른다. 또한 가정 3으로부터 T_1, T_2, T_3 는 서로 독립이며 지수분포의 무기억성(memoryless property)에 따라 현 단계에 머문 시간은 다음 단계에 머물 시간에 영향을 미치지 않는다.

그러면 SARS로 사망할 때까지 소요되는 총시간은 $T_0 = T_1 + T_2 + T_3$ 이므로 T_0 의 분포는 Everitt과 Hand(1981, p. 79)로부터 다음과 같이 구할 수 있다.

$$f_{T_0}(t) = \frac{\lambda_2\lambda_3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} f_{T_1}(t) + \frac{\lambda_1\lambda_3}{(\lambda_1-\lambda_2)(\lambda_3-\lambda_2)} f_{T_2}(t) + \frac{\lambda_1\lambda_2}{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)} f_{T_3}(t) \quad (1)$$

여기서 $f_{T_j}(t)$ 는 T_j 의 확률밀도함수이다. 식(1)로부터 제1단계의 환자가 t 시간내에 사망할 확률, 평균 여명 등 관심있는 모수 등을 추정할 수 있을 것이다.

2.4 모수의 추정

이제 단계 j 에서 다음 단계로의 SARS진행률 λ_j 를 추정하자. 증상의 각 단계에 머무는 시간이 서로 독립이므로 우선 증상단계 j 에 대해서만 알아본다. 이를 위하여 p_{ij} 를 관측구간 i 에서 증상단계 j 에 잔류할 확률이라하면 $q_{ij}=1-p_{ij}$ 는 관측구간 i 에서 증상단계 j 에서 적어도 한 단계 진행될 확률이다. 우도함수는

$$L(\lambda_j) = \prod_{i=1}^k \binom{n_{ij}}{x_{ij}} (1-p_{ij})^{x_{ij}} p_{ij}^{n_{ij}-x_{ij}} \quad (2)$$

$$= \prod_{i=1}^k \binom{n_{ij}}{x_{ij}} (1 - \exp(-\lambda_j t))^{x_{ij}} \exp(-\lambda_j t (n_{ij} - x_{ij}))$$

이므로 대수우도함수를 λ_j 에 대하여 편미분하면

$$\frac{\partial \ln L(\lambda_j)}{\partial \lambda_j} = \sum_{i=1}^k \left(\frac{x_{ij} t \exp(-\lambda_j t)}{1 - \exp(-\lambda_j t)} - (n_{ij} - x_{ij}) t \right) \quad (3)$$

이므로 Newton-Raphson방법을 사용하여 $\partial \ln L(\lambda_j) / \partial \lambda_j = 0$ 을 만족하는 λ_j 의 최우 추정치 $\hat{\lambda}_j$ 을 구할 수 있을 것이며, 이는 n_{ij} , x_{ij} 및 t 에 영향받음을 알 수 있다. 여기서 n_{ij} 와 x_{ij} 는 <표 3> 및 <표 4>에서 구할 수 있으나, t 는 $p_{ij} = P(T_j > t) = \exp(-\lambda_j t)$ 의 관계로부터 사전에 정해진 관측시간(prespecified observation time)으로 본 연구에서는 증상단계가 j 이고 관측구간이 i 일 때의 t 를 각 관측구간의 중간시점(midpoint)이라 가정하고 t_{ij}^0 라고 나타낸다. 그러면 식(3)에서 t 가 t_{ij}^0 로 대체되고 다음의 우도방정식을 다시 얻을 수 있게된다.

$$\sum_{i=1}^k \left(\frac{x_{ij} t_{ij}^0 \exp(-\lambda_j t_{ij}^0)}{1 - \exp(-\lambda_j t_{ij}^0)} - (n_{ij} - x_{ij}) t_{ij}^0 \right) = 0 \quad (4)$$

식(4)로부터 $\lambda_j(j=1,2,3)$ 의 최우추정치를 얻을 수 있다. 또한 최우추정량 $\hat{\lambda}_j$ 의 점근분산

$$AsVar(\hat{\lambda}_j) = -1/E \left[\frac{\partial^2 \ln L}{\partial \lambda_j^2} \right]$$

는 $\lambda_j = \hat{\lambda}_j$ 을 대입하여 추정할 수 있다.

2.5 수치 예제

관측구간의 수를 3, 즉 $k=3$ 이라하고, 가상의 데이터가 <표 5> 및 <표 6>과 같이 얻어졌다고 하자. 6일부터 36일까지를 3 구간으로 나누어 관측되었으며 각 관측구간별로 3가지의 증상단계로 분류할 수 있다고 한다. <표 5>는 관측구간별 증상단계별 환자수를 나타내고 있으며 <표 6>은 <표 5>를 구성하던 환자 가운데 증세가 악화되었거나 사망한 사람을 나타내고 있다. 앞 절에서 언급한 방법으로 모수를 추정하여보자.

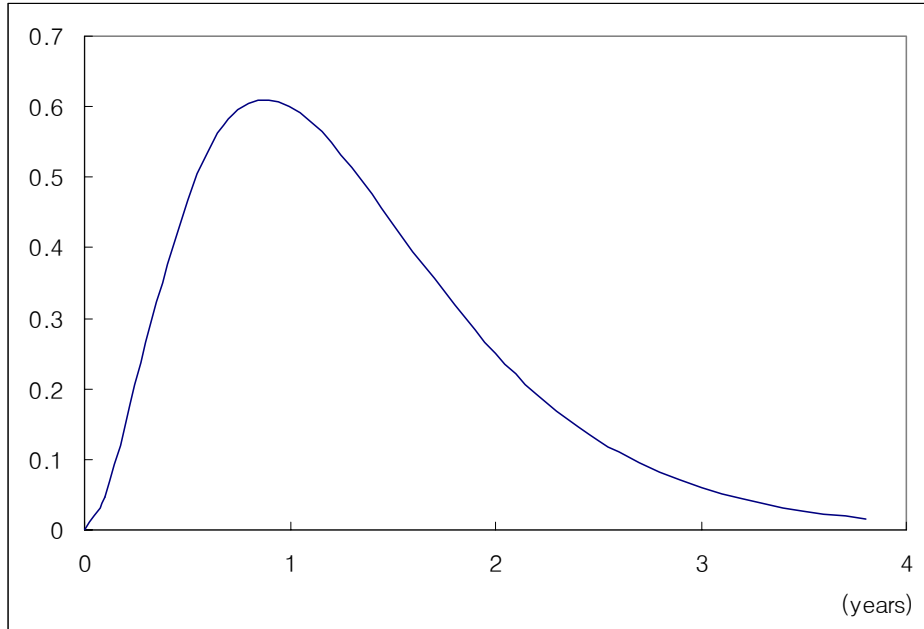
<표 5> 관측구간 및 증상단계별 환자수

증상 관측구간	단계1	단계2	단계3	합계
6일-12일	14	18	51	83
12일-24일	21	20	29	70
24일-36일	3	5	19	27
합 계	38	43	99	180

<표 6> 관측구간별 증상단계별 SARS사망자 또는 증상악화자 분포

증상 관측구간	단계1	단계2	단계3	합계
6일-12일	1	2	4	7
12일-24일	2	2	2	6
24일-36일	1	0	2	4
합 계	4	4	8	17

관측기간을 각 관측구간의 중간지점까지라고 가정하고 시간을 연(year)단위로 하면 $t_{1j}^0 = (6일 + 12일) / (2 \times 365일/년) = 0.025년$ 이고 $t_{2j}^0 = 0.049년$, $t_{3j}^0 = 0.082년$ 이다. 이들 값과 <표 5> 및 <표 6>의 x_{ij} 와 n_{ij} 를 식(4)에 대입하여 Newton-Raphson방법으로 풀면 $\hat{\lambda}_1 = 2.632$, $\hat{\lambda}_2 = 2.266$ 이고 $\hat{\lambda}_3 = 1.966$ 이다. 이를 식(1)에 대입하면 SARS 사망자의 수명밀도함수는 <그림 1>과 같이 추정됨을 알 수 있다.



<그림 1> 추정된 밀도함수

3. 결론

본 연구에서는 중증급성호흡기증후군으로 인한 사망 데이터에 대하여 적용가능한 통계모형을 제안하였다. SARS에 감염되어 사망에 이르기까지 환자의 증상에 따라 3 단계로 나누었다. 각 단계에서 잔류하는 시간의 분포는 지수분포를 따르며, 환자 증상의 변화를 여러 관측구간으로 나누어 기간별로 증상별로 집계될 수 있다는 가정하에 모형화하였다. 제안된 모형을 활용하면 SARS 환자의 수명분포 등 다양한 기초적인 연구와 SARS보험개발 등에 활용될 수 있을 것이다. 이를 계기로 향후에는 치료기간 중의 회복환자 통계까지 포함하는 보다 일반화된 모형의 개발이 기대되며, 실제 데이터를 이용하여 제안된 모형의 효용성을 확인할 필요가 있다고 사료된다. 끝으로 본 연구는 SARS에 대한 의학적인 정의조차 명확하지 않을 뿐만 아니라 현재 의학적인 연구가 활발히 진행중이라는 측면에서 통계적인 접근이 제한적일 수 있지만 SARS와 같이 전염성이 강한 유사한 질병에 대한 통계적인 접근은 가능하다고 판단된다.

참 고 문 헌

1. 국립보건원, http://www.nih.go.kr/html/kor_html/index.html, (accessed July 13, 2003).
2. Everitt, B. and Hand, D.(1981), *Finite Mixture Distributions*, Chapman and Hall.

3. WHO. Cumulative number of reported probable cases of severe acute respiratory syndrome(SARS). <http://www.who.int/csr/sars/en/> (accessed July 13, 2003).

[2003년 4월 접수, 2003년 7월 채택]