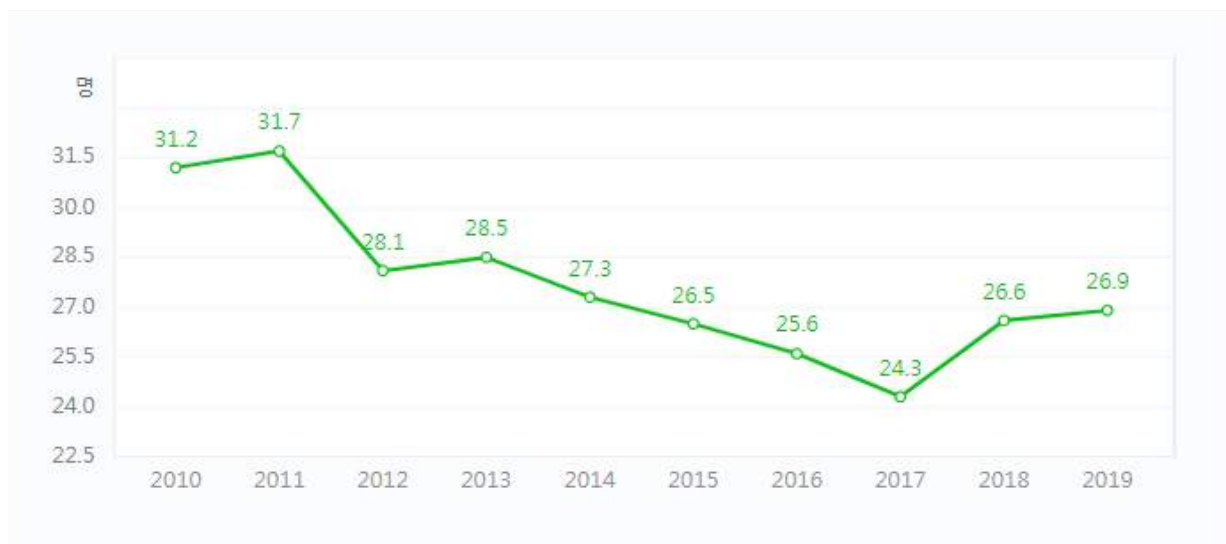


Thompson Sampling을 통한 자살 예방 프로젝트

김태훈

1. 서론

우리 사회가 오래 동안 직면해온, 그러나 그럼에도 해결하지 못한 문제 중 하나가 바로 높은 자살률이다. 아직 2020년의 자살 통계는 발표되지 않았는데, 2003년부터 2019년까지 한국은 2017년 단 한 해를 제외하고는 줄곧 OECD 자살률 1위를 놓치지 않아 ‘자살공화국’의 오명을 쓰고 있다.



10만명당 자살자수 (통계청)

높은 자살률 문제가 해결되지 못하고 있는 것은, 이에 대한 국가 혹은 지방자치단체 차원의 대책이 미비되거나 효과적으로 이루어지지 못하기 때문일 것이다. 소득 수준이 낮은 이들에게 더 많은 복지 혜택이 주어지는 것이 어떤 사회에서든 자연스럽게 받아들여지는 것과 같은 이유로, 자살 충동을 느낄 정도로 한계 상황에 놓인 이들을 더욱 효율적으로 판별하여 이들에게 도움의 손길을 내밀 수 있게 된다면, 한정된 자원 내에서 더 큰 복지를 창출할 수 있는 길이 될 것이다.

높은 자살률 문제가 좀처럼 해결되지 않는 이유는, 이 문제에 있어 ‘선택과 집중’이 제대로 이루어지지 못하고 있기 때문이라고 생각한다. 자살자수가 전체 인구에서 차지하는 비중은 아주 작기 때문에, 자살 위험군을 가려내지 않은 채 전국민에게 동일하고 무차별적인 자살예방대책을 시행하는 것은, 1인당 복지의 밀도를 현격히 낮추는 결과를 가져올 수밖에 없다. 필자는 이러한 문제의식 하에서, 자살예방대책을 시행하기에 앞서 Thompson Sampling 기법을 통해 전국민을 주요 요인에 따라 그룹화 하고 각 그룹의 자살 위험도를 효율적으로 추적하는 방법을 제안하고자 한다.

자살 위험도에 따른 그룹화가 진행되기 위해서는, 자살에 영향을 미치는 주요 feature들에 대한 분석이 선행되어야 한다. 따라서 필자는 통계청의 ‘성별·연령별 자살률’과 ‘자살 충동 이유’ 통계 자료를 바탕으로 어떤 변수들이 자살률의 차이를 가져오는지를 살펴보고, 이를 기준으로 국민들을 그룹화할 것이다. 그 다음에는 Thompson Sampling의 시뮬레이션을 통해 각 그룹의 자살률이 효율적으로 추적될 수 있음을 보임으로써 이 글을 마무리할까 한다.

2. Feature 탐색

2.1. 성별·연령별 자살률

2019 성별 연령별 자살자수(십만명당)					
전체		여성		남성	
0-9세	0	0-9세	0	0-9세	0.1
10-19세	5.9	10-19세	6.4	10-19세	5.5
20-29세	19.2	20-29세	16.6	20-29세	21.6
30-39세	26.9	30-39세	20	30-39세	33.5
40-49세	31	40-49세	17.1	40-49세	44.5
50-59세	33.3	50-59세	15.9	50-59세	50.5
60-69세	33.7	60-69세	14	60-69세	54.2
79-79세	46.2	79-79세	23.5	79-79세	74.6
80세이상	67.4	80세이상	35.5	80세이상	133.4
계	26.9	계	15.8	계	38

성별 연령별 자살률(통계청)

우선 살펴볼 수 있는 것은 남성의 자살률이 여성의 자살률의 2배가 훨씬 넘는다는 점이다. 자살 위험도를 구분함에 있어 성별이 주요한 feature가 될 수 있음을 알 수 있다. 여성의 경우 69세 이전까지는 연령에 따른 자살률 증감의 양상이 뚜렷하지 않다가 이후 점차 증가하는 양상을 보이는 반면, 남성의 경우 연령 증가에 따라 자살률이 꾸준히 증가하는 모습을 보인다. 특히 80세 이상에 접어들어 자살률이 큰 폭으로 증가함을 확인 할 수 있는데, 80세 이상의 경우 배우자를 떠나보내고 홀로 남게 되는 비율이 급증하게 되는 시기라는 점이 자살률의 급증과 무관하지 않아 보인다.

그러나 여성의 경우 노년기에 접어들어도 자살률 증가의 경향이 남성의 경우처럼 뚜렷하지 않은데, 이로써 배우자 사별에 따른 독거 여부가 남녀에게 미치는 영향이 비대칭적이라는 판단을 할 수 있다. 충북보건과학대·단국대 교수팀의 연구에 따르면 독거 남성은 부부가구 남성보다 자살 생각을 2.3배, 부부가구 여성 노인은 남편보다 자살 생각을 2배하는 것으로 밝혀졌는데, 이러한 연구 결과가 이를 뒷받침해 준다.¹⁾ 노년 세대의 강한 가부장적 결혼 문화가, 독거 여부에 대한 남녀의 비대칭적 반응을 가져오는 주요 원인으로 생각된다.

노년 세대의 높은 자살률의 원인이 독거 여부에만 있다고 보기는 힘들다. 노화와 질병, 그리고 노후 대비가 제대로 되어 있지 않았을 경우의 생활수준 급락 등도 주요 원인일 것으로 생각된다. 그러나 노화와 질병의 정도는 계량화하기 힘들고, 더구나 개인의 의료 정보는 국가라 하더라도 마음대로 열람할 수 있는 것이 아니므로 이러한 변수들을 그룹화의 기준으로 삼기는 어렵다.

지금까지 살펴본 결과들을 요약하자면 여성의 경우 노년기에 들어 자살률이 증가하고, 남성의 경우 연령 증가에 따른 자살률 증가가 꾸준히 나타나다가 노년기에 그 양상이 더욱 강해진다고 할 수 있다. 따라서 성별·연령별 자살 위험도를 구분함에 있어 남성의 경우 여성보다 연령에 따른 그룹화가 더 세밀하게 이루어져야 할 것으로 보인다.

1) 김양중, 「'자살 생각' 남성은 아내 없으면 2배, 여성은 남편 있으면 2배」, 한겨레, 2016.10.28., <http://www.hani.co.kr/arti/society/health/767779.html>

2.2. 자살 충동 이유

2020년 자살 충동 이유(%)									
특성별		경제적 어려움	연애 상대와의 관계	신체적·정신적 질환, 장애	직장문제	외로움, 고독	가정불화	학교성적, 진학문제	친구나 동료와의 불화나 따돌림
전체	계	38.2	3.1	19.0	8.7	13.4	11.9	2.3	1.7
성별	남자	43.0	3.4	20.2	11.0	12.8	6.3	1.3	0.8
	여자	34.6	3.0	18.1	7.1	13.9	16.1	3.0	2.3
연령	13~19세	9.3	9.2	12.4	3.0	14.4	11.2	29.7	11.0
	20~29세	23.4	4.2	12.2	23.6	14.8	13.4	3.7	2.4
	30~39세	37.3	4.4	10.8	12.3	15.9	12.9	0.5	2.7
	40~49세	49.0	3.2	9.8	7.5	11.4	16.5	-	1.0
	50~59세	51.9	2.3	19.2	4.3	11.3	10.4	-	0.2
	60세 이상	36.5	1.1	36.8	2.5	13.6	8.5	-	-
	65세 이상	32.5	0.8	41.6	1.0	15.9	6.7	-	-
교육정도	초졸 이하	30.1	1.6	37.0	1.1	14.2	10.5	0.9	3.5
	중졸	34.9	4.9	26.7	3.0	15.6	6.8	7.8	0.1
	고졸	46.7	3.3	16.2	7.0	10.4	11.7	2.5	0.7
	대졸 이상	32.6	2.9	12.1	16.1	15.9	14.7	0.5	2.7
경제활동	취업	46.7	3.7	10.9	10.7	12.2	12.5	0.1	1.8
	실업 및 비경제활동	29.6	2.6	27.2	6.7	14.6	11.4	4.5	1.5

자살 충동 이유(13세 이상 인구)(통계청)

위 자료는 2020년 통계청 조사에서 자살 충동을 느낀 사람들 중 자살 충동의 이유를 조사한 결과이다. 몇 가지 눈에 띄는 점들을 살펴보자.

우선 눈에 띄는 점은 ‘경제적 어려움’이 가장 큰 비중을 차지하는 자살 충동의 이유라는 것이다. 경제적 어려움은 남성, 40~59세, 고졸의 경우 가장 큰 자살 충동의 이유가 되는데, 한창 자식 뒷바라지를 하는 시기의 가장이 가장 큰 경제적 압박을 받는다고 해석할 수 있다.

자살 충동 이유 중 두 번째로 큰 비중을 차지하는 것은, ‘신체적·정신적 질환, 장애’이다. 경제적 어려움을 겪는 사람에 비해 자살 충동을 느낄 정도의 신체적·정신적 질환, 장애를 가진 사람이 훨씬 적을 것임이 분명함에도, 이렇게 큰 비중을 차지한다는 것은, 신체적·정신적 질환, 장애가 자살 충동의 아주 강한 동기가 되고, 자살률의 큰 상승을 가져오는 feature임을 말해준다.

이전의 성별·연령별 데이터를 살펴보았을 때 19세 이하 청소년의 자살률은 성인 자살률의 4분의 1에도 미치지 못했다. 또한, 주요 자살 충동 이유를 살펴보다라도 ‘학교성적, 진학문제’, ‘외로움, 고독’, ‘친구와의 불화나 따돌림’ 등 국가 차원에서 정보를 수집하기 어려운 요인이 대부분이므로 청소년의 경우 분리해서 접근하는 것이 타당해 보인다. 학교성적, 진학문제, 외로움, 고독, 친구와의 불화나 따돌림 등을 상대적으로 잘 파악할 수 있는 공교육 시스템 내에서 자살예방교육이나 상담 등을 통해 문제 해결에 접근하는 것이 바람직할 것으로 판단된다.

국가 차원에서 데이터를 취합할 수 있는 feature와 그렇지 않은 feature들이 구분된다. ‘신체적·정신적 질환, 장애’의 경우 장애인인은 장애 등급을 등록하기 때문에 이미 데이터가 수집되어 있다. 그러나 등록된 장애 여부와 무관한 신체적·정신적 질환은 개인의 의료 정보이므로 국가라 하더라도 임의로 열람하거나 수집할 수 없다. ‘경제적 어려움’의 정도는 주관적일 뿐더러 전국민을 대상으로 한 패널 데이터를 지속적으로 추적해야하는 어려움이 따른다. 따라서 기초수급자, 차상위계층, 신용불량자, 파산 여부 등의 대리변수를 사용하여 그룹을 나누는 것이 적절해 보인다. ‘연애 상대와의 관계’, ‘직장 문제’, ‘친구나 동료와의 불화나 따돌림’ 등은 국가가 데이터를 추적하는 것이 사실상 불가능하므로 우리의 분석에서는 제외하는 것이 적절해 보인다. ‘외로움, 고독’, ‘가정불화’ 역시 파악하기 힘들지만, 자살 충동 이유의 큰 부분을 차지하고 있는 feature일 뿐만 아니라 이혼 여부, 혹은 사별 후 독거 여부가 높은 상관관계를 가질 것으로 생각되어 이를 proxy 변수로 활용할 수 있을 것으로 생각된다.

3. Thompson Sampling Simulation

3.1. Grouping

먼저 성별 연령별 통계에 근거해 그룹을 나눠보자. 19세 이전의 미성년자의 경우 자살률이 성인에 비해 현격히 떨어지고 학교성적, 진학문제 등 주요 자살 동기 또한 정량적으로 파악하기 힘들기 때문에 sampling 대상에서 제외하는 것이 적절하다고 언급한 바 있다. 여성의 경우 20~69세에 이르기까지 자살률에 큰 차이가 없기 때문에 20~69세, 70~79세, 80세 이상의 세 그룹으로 분류하고, 남성의 경우 연령 증가에 따른 자살률 증가의 양상이 뚜렷하기 때문에 20~29세, 30~39세, 40~49세, 50~59세, 60~69세, 70~79세, 80세 이상의 7개 그룹으로 분류하는 것이 타당해 보인다.

자살 충동 이유 통계에서 파악된 주요 feature 가운데 전국민을 대상으로 정량화된 데이터를 얻을 수 있는 것들을 추려내면 ‘경제적 어려움’(기초수급자, 차상위계층 등의 proxy변수 사용), ‘장애 여부’, ‘외로움, 고독’ 혹은 ‘가정불화’(이혼 혹은 사별 등의 proxy변수 사용)이 있다. 이러한 요소들로 위의 성별 연령별 그룹을 다시 세분화하게 된다면, 총 7(성별 연령별 그룹) × 2(경제적 어려움) × 2(장애) × 2(외로움, 고독, 가정불화) = 56개의 세분화된 그룹을 얻게 된다. 이렇게 세분화된 각 그룹에 대해 데이터를 수집하여 Thompson Sampling 기법으로 자살률을 추적한 후, 배정된 예산과 선정된 자살예방대책 하에 자살률이 높은 순으로 집중적인 자살예방방안을 적용하면 될 것이다.

3.2. 데이터 생성

앞서 필자가 제안한 그룹화에 요구되는 데이터는 국가 혹은 지방자치단체 단위에서만 취합 가능한 정보이다. 이 글의 목적이 방법론을 제시하는 것에 있으므로, 여기에서는 현재 주어진 통계 데이터의 제한을 반영하여, 축약된 그룹화 하에서 톰슨 샘플링이 정상적으로 작동함을 보이는데 그 의의를 두겠다. 성별 연령별로 구획된 10개 그룹에 관한 데이터를 생성하여 톰슨 샘플링을 적용해보자.

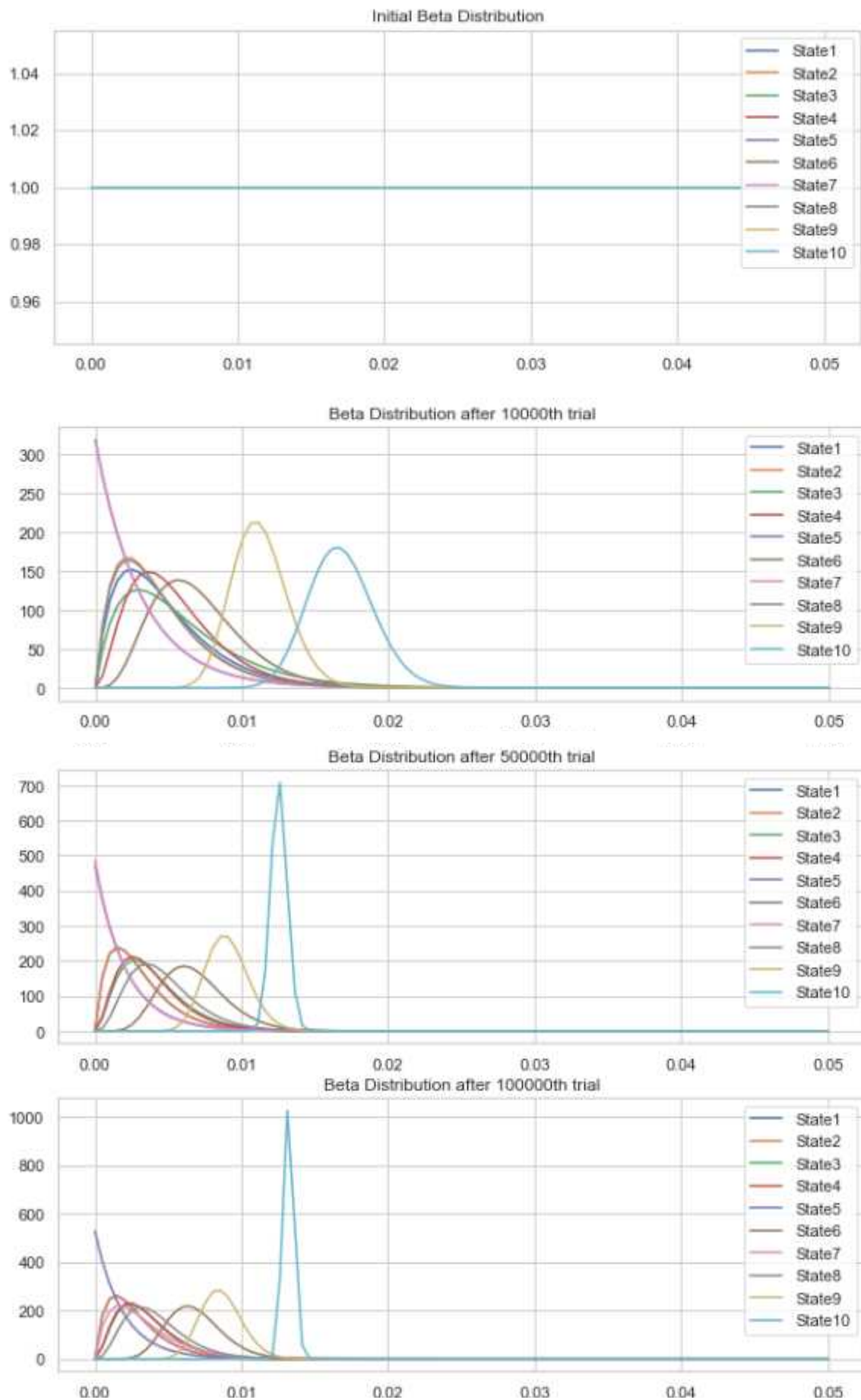
성별 연령별 자살률 데이터는 10만명 당 자살자수를 나타내는데, 비교가 용이하도록 각 그룹 당 샘플의 크기를 10만으로 하여 데이터를 생성하였다. 자살하지 않은 사람은 0, 자살한 사람은 1의 값을 주었는데, 전체 표본에서 1이 차지하는 비중은 아주 작기 때문에, 한 해 동안의 자살자수가 아닌 10년 간 자살자수를 적용하여 1의 비율을 10배로 확대하였다. 여기서는 편의상 각 그룹이 10년 동안 매해 동일한 자살자수를 기록했다고 가정하였는데, 현실에서 이 방법론을 적용함에 있어서는 실제로 수집된 누적 데이터를 활용하면 될 것이다. 여성 20~69세 그룹의 경우 그룹 내의 각 연령대에 동일한 가중치를 적용하여 10만명 당 자살자수를 산정했다.

2019 성별 연령별 자살자수(십만명 당)			
여성		남성	
0-9세	0	0-9세	0.1
10-19세	6.4	10-19세	5.5
20-29세	16.6	20-29세	21.6
30-39세	20	30-39세	33.5
40-49세	17.1	40-49세	44.5
50-59세	15.9	50-59세	50.5
60-69세	14	60-69세	54.2
79-79세	23.5	79-79세	74.6
80세이상	35.5	80세이상	133.4
계	15.8	계	38



10년 간 10만명 당 자살자수	
state1(여성, 20~69)	167
state2(여성, 70~79)	235
state3(여성, 80~89)	355
state4(남성, 20~29)	216
state5(남성, 30~39)	335
state6(남성, 40~49)	445
state7(남성, 50~59)	505
state8(남성, 60~69)	542
state9(남성, 70~79)	746
state10(남성, 80이상)	1334

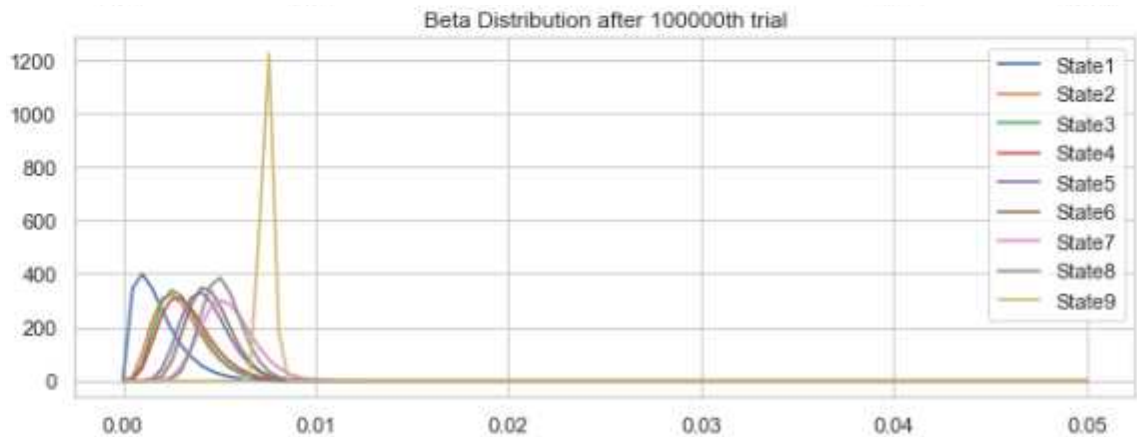
3.3. Thompson Sampling



0, 10000, 50000, 100000회 시행에서의 톰슨 샘플링 결과

10개 그룹에 대한 0, 10000, 50000, 100000회 시행의 톰슨 샘플링 결과를 살펴보자. 가장 자살률이 높은 state10 즉, 80세 이상의 남성 그룹이 가장 높은 자살률을 보임이 뚜렷이 확인되고, 분포 또한 실제 값인 0.013 주변에 밀집되어 있음을 확인할 수 있다. 그러나 나머지 그룹의 경우 분포의 중심을 정확히 알기 어렵고 자살률 고저의 순서 또한 실제와 반드시 일치하지는 않음이 관찰된다. 이러한 현상은 톰슨 샘플링의 메커니즘 상 가장 높은 확률을 가진 그룹이 확인되면, 그 그룹에 높은 가중치를 두고 시행을 반복하기 때문이다.

우리의 목적은 가장 높은 자살률을 보이는 단 하나의 그룹만 뽑는 것은 아니다. 만약 계획된 자살예방정책이 주어진 예산 하에서 세 개 그룹을 대상으로만 시행될 수 있다면, 우리는 가장 높은 자살률을 보이는 세 개의 그룹을 가려내야 한다. 그러나 state10에 해당하는 그룹의 자살률이 압도적으로 높아 계속해서 이 그룹에서만 표본이 추출되면서 다른 그룹의 자살률이 정확히 확인되지 못하고 있다. 이 경우 그 다음으로 높은 자살률을 보이는 그룹과 그 분포를 확인하기 위해서는 state10의 그룹은 제거하고 다시 위의 과정을 반복하면 된다.



state 10 제거 후 100000회 시행에서의 톰슨 샘플링 결과

state10을 제거하고 10만번 시행한 결과 두번째로 높은 자살률을 보이는 state9의 분포가 실제 자살률인 0.00746 주변에 밀집되는 것을 확인할 수 있다. 그러나 나머지 그룹의 자살률은 여전히 정확히 파악되지 않고 있으므로, state9 다음으로 높은 자살률을 보이는 그룹을 가려내기 위해서는 state9 역시 제거한 후 위의 과정을 반복하면 된다.



state 9, 10 제거 후 100000회 시행에서의 톰슨 샘플링 결과

state10과 state9를 모두 제거하자 state8의 자살률이 가장 높고, 그 분포 또한 실제 값인 0.00542 주변에 밀집되어 나타남이 확인된다. state8은 state10과 state9와는 달리 다른 그룹들과 자살률 차이가 크지 않음에도 톰슨 샘플링 기법을 통해 정밀한 구분이 가능함을 알 수 있다. 따라서 우리는 가장 자살률이 높은 세 개의 그룹을 '80세 이상 남성', '70~79세 남성', '60~69세 남성' 즉, 60세 이상의 남성 노인으로 확정하여 이들을 대상으로 하여 자살예방대책을 시행할 수 있게 되었다.

4. 결론

참으로 중요한 철학의 문제는 오직 하나뿐이다. 그것은 자살이다.

인생을 괴로워하며 살 값어치가 있는지에 대해 판단하는 것. 이것이야말로 철학의 근본적인 질문에 대한 응답이다.

노벨 문학상 수상자 알베르 카뮈의 자살에 대한 논평이다. 이처럼 많은 유명한 철학자들과 작가들이 자살에 대한 말들을 쏟아냈지만, 필자가 보기에 실제 이루어지는 자살의 대부분은 그들의 말처럼 고상한 사유(思惟)의 결과물은 아니다. ‘자살’이라는 종착점에 이르는 두 개의 지름길인 ‘우울’과 ‘절망’의 다른 이름은 ‘자기 통제력의 완전한 상실’이므로, 대부분의 경우 자신의 의지로 자살이라는 최선의 옵션을 선택한 것이기 보다는, 주어진 경제적, 신체적, 정신적 여건들에 의해 자살로 내몰린 것이라고 보는 것이 옳다.

필자는 국민의 건강과 안녕을 책임져야 하는 국가가 이러한 자살 문제에 보다 적극적이고 효과적으로 개입하지 않으면 안 된다는 문제의식 하에서 톰슨 샘플링 기법을 통한 대안적 접근 방법을 제안하였다. 이 방법론은 두 가지 점에서 기존의 접근 방법과 차별화된다. 첫째로, 자살 고위험군에 대한 ‘선택’과 ‘집중’이 이루어진다. 기존의 자살 통계는 자살 위험도에 따른 세밀한 그룹화가 이루어지지 않으며, 기껏해야 성별 연령별 자살자수의 정보만을 제공한다. 따라서 이러한 통계에 근거한 자살예방대책 역시 전국민을 대상으로 한 획일적이고 간헐적인 자살예방교육 및 캠페인 수준을 벗어나지 못하고 있다. 필자가 앞서 제안한 56개 그룹으로의 세분화는 성, 연령, 장애 여부, 경제력 수준, 이혼·사별 여부 등 이미 국가 차원에서 수집된 feature data에 의한 것이기 때문에 추가적인 정보 수집을 요구하지도 않는다. 이렇게 세분화된 그룹에 톰슨 샘플링 기법을 적용하여 자살 위험도를 판단한다면, 각 그룹을 타게팅한 효과적인 정책 수립이 가능해진다.

둘째로, 이러한 ‘선택’과 ‘집중’이 이루어진다면, 자살예방대책을 기존의 수동적인 방식에서 보다 적극적인 방식으로 전환해 나갈 수 있다. 자살 위험에 노출된 이들이 전체 인구에서 차지하는 비중은 아주 작기 때문에 기존의 무차별적인 자살예방홍보 방식으로는, 정부의 자살예방 채널들이 자살 고위험군에 효과적으로 홍보되지 못한다. 삶의 의미마저 잃은 사람들이 정부의 자살 예방 채널을 적극적으로 탐색하고 이용하지 않으면 안된다는 것은 아이러니가 아닐 수 없다. 대신 자살 고위험군을 식별하여 이들에게 전화, 문자, 우편물, 배너 광고 등의 홍보 수단을 집중한다면 홍보비용 대비 이용률을 훨씬 높일 수 있을 것이다. 뿐만 아니라 높은 자살률을 보이는 노인, 장애인, 중증 질환자는 정부 채널에 대한 접근성이 떨어질 수밖에 없는데, 이들의 의사를 확인하여 정신과 상담 인력을 정기적으로 파견해주는 식의, 이른 바 ‘찾아가는 서비스’와 같은 적극적인 정책도 시행이 가능할 것이다.

지금까지 톰슨 샘플링 기법에 기반 한 자살예방정책 개선 방향에 대한 필자의 의견을 피력해 보았다. 우리 사회의 구성원 누군가가 자살을 고민할 정도의 큰 고통에 노출되어 있다는 것은 가슴 아픈 일이 아닐 수 없다. 이 글이 ‘자살 공화국’의 오명을 벗어던지는데 조금이라도 보탬이 되었으면 한다.