

**자동차사고 재발생의
확률밀도함수분석과 활용방안**
**A Study on Probability Density Function Analysis
and Application of Car Reaccident**

이 공 섭*
Lee, Gong-seob
김 영 민**
Kim, young-min

Abstract

Due to the increasing of the number of cars and bad road conditions, car accidents are increasing every year in Korea. When a person meets a car accident, it is necessary for him to analyze and determine whether applying insurance or not, because standard discount rate and special increasing rate change with accident types and the amount of accident expenditure. When we consider insurance rate that includes more than ten elements, we need a decision making. In this paper, S insurance company investigated previous car causers in 1988, 1989, 1990 to 1996 with 600,000 real data. We investigate probability density functions and cumulative distribution functions for each year using ARENA software. We can apply the results of this study to various accidents that occur under uncertainty in our life. I hope that this paper contribute to strengthening competitive power of companies and developing new insurance rate systems in future.

1. 서 론

복잡하고 다기능화된 현대사회에 적응하기 위해서 이제 자동차가 필수품으로 자리를 잡아가고 있다. 또한 급속한 경제성장과 생활수준향상으로 자동차사용이 급증하여 1997년 7월 현재 1006만대(표1)[3]에 이르고 있지만, 교통안전의식 결여와 교통환경의 개선과 제도개선 등에 대한 투자가 이에 따르지 못하여 교통사고가 빈발하고있다. 1996년 한해동안 발생한 교통사는 약 265,052건[1]이고 사망자수는 12,653명으로 1일평균 726건이 발생하고 사망자도 35명에 이르고 있다. 이는 차량1만대당 12.2명으로 노르웨이의 1.2명의 10.2배(표2)[2]에 이르는 실정이다. 이로 인한 보험회사의 보험금 지급규모가 전체 보험금 5조6,480억원의 58.8%인 3조3,233억원[4]에 이르고 있다.

표1 국내 자동차 보유현황 (1997년 7월말 현재, 단위:천대)

승용차	승합차	화물차	특수차	합계
7,300	696	2,033	35	10,064

* 유한전문대학 공업경영과 교수

** 인하대학교 산업공학과 교수

자동차사고를 보험처리하면 우량할인, 불량할증요율, 특별할증요율이 보상금액에 따라 상이하게 보험료가 적용[7,8]된다. 현행의 개인용 자동차보험료 산출요소가 10여가지에 이르고 개인별 담보종목이 다르기 때문에 일률적으로 일정금액이면 현금처리가 유리[13]하다는 것은 잘못된 판단이다. 일반 보험가입자가 복잡한 보험료를 정확히 계산하기가 매우 어렵고 더군다나 향후사고를 고려해서 보험처리여부를 결정하기란 거의 불가능하다.

본 연구에서는 보험처리여부의 의사결정을 하기 위해서 근거가 되는 1차사고가 난후 특정 기간내에 재차 사고가 날 확률밀도함수를 구하고[14,15,16,17] 그 함수의 적합도검정[9]을 하였다. 1차사고를 낸 사람이 재차사고가 날 조건부확률을 구하면 보다 정확한 보험료를 산출할 수 있어 보험처리여부의 판단자료로 활용할 수 있도록 하였다.

표2 세계 각국의 교통사고 발생상황

국 가	인구(천명)	자동차(천대)	발생건수	사 망 자			부상자(명)	치사율	부상률
				(명)	10만명당	1만대명당			
한 국	44,850	8,469	248,865	10,323	23.0	12.2	331,747	4.1	133.3
독 일	81,000	51,000	2,270,818	9,814	12.1	1.9	526,229	0.4	58.6
이탈리아	57,140	57,140	170,000	6,600	11.6	1.9	240,000	3.9	141.2
프 랑 스	58,027	24,210	132,949	8,412	14.5	3.5	181,403	6.3	136.4
노르웨이	4,369	2,441	8,321	304	7.0	1.2	11,328	3.7	136.1
미 국	260,341	192,337	2,128,223	40,676	15.6	2.1	3,215,000	2.2	151.6
일 본	125,341	69,291	761,789	10,679	8.5	1.5	922,677	1.4	121.1
캐 나 다	7,126	4,228	35,502	824	11.6	1.9	48,685	2.3	137.1
영 국	56,753	24,601	234,101	3,650	6.4	1.5	315,189	1.6	134.6

본 연구에서는 자동차의 소유와 형태에 따라 개인용 자동차, 업무용 자동차, 영업용 자동차로 구분하는데 전체 자동차의 72.6%[6]를 차지하는 개인용 자동차를 연구대상으로 하였다. 개인용 자동차보험료의 사고경력은 자동차의 대체와 보험가입회사의 변경에는 관계없이 주민등록번호로 추적할 수 있기때문에 S보험회사에 가입한 60여만명을 대상으로 88년, 89년, 90년에 1차사고를 낸 사람을 선택하여 각각 재차 사고유무를 96년도까지 추적조사 하였다. 이를 Interaccident Time으로 환산하여 Arena를 이용하여 확률밀도함수를 구하고 그에 해당하는 누적분포함수를 Minitab[10]을 이용하여 확률값을 구하였다. 통계적 의사결정법과 경제성공학 기법[5]을 활용하여 보다 합리적이고 정확한 의사결정을 도모할 수 있는 자료를 제시하였다. 사례로 자동차 사고에 대한 연구를 하였지만 각종 상해보험과 불확실성하에서 일어나는 모든 사고의 분석에도 확대적용 할 수 있다.

2. 이론적 고찰

2.1 통계적 의사결정론

모집단의 특징을 나타내는 미지의 모수를 추정하거나 검정하려고 할 때 우리는 제한된 표본정보만을 가지고 추론할 수밖에 없으며, 따라서 이러한 추론의 결과는 항상 불확실성을 내포하고 있다. 동시에 추론과정에 내제된 불확실성은 장기상대도수(long-run relative frequency)개념에 입각한 객관적 확률로 환산하여 표현할 수 있게 된다. 추측통계학의 분석과정은 모수에 대한

추정치를 결정하거나 모수에 대한 특정가설의 채택여부를 결정하는 일련의 통계적 의사결정과정이다.

본절에서는 전통적인 통계적 추론과정에서 한걸음 더 나아가 별도의 기준이나 정보를 활용하여 의사를 결정하는 여러 가지 방법과 이들의 활용방안을 다루고자하며, 이러한 과정에서의 분석내용을 통계적 의사결정론(statistical decision theory)[12]이라고 한다. 의사결정은 그 과정이 불확실성하에서 이루어지는지의 여부에 따라 아래 그림1과 같이 구분될 수 있으며, 본절에서는 불확실성하에서의 의사결정과정중 EMV와 의사결정분기도표 활용 및 베이지정리 활용[11]을 주분석대상으로 삼았다.

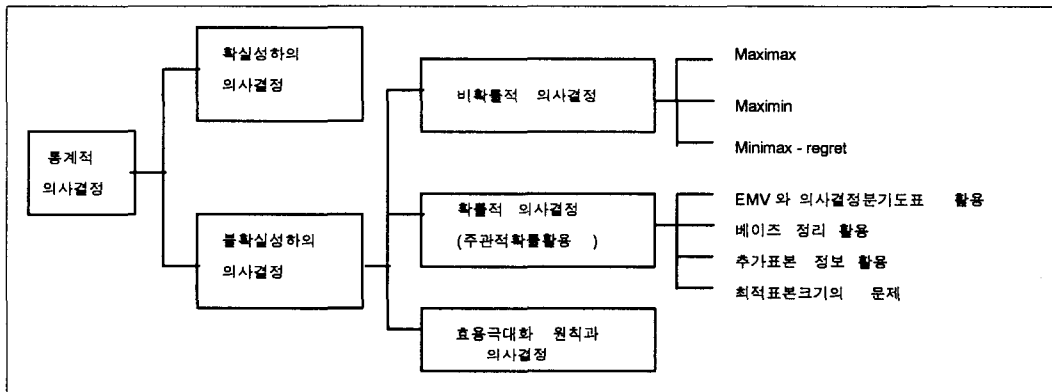


그림1 통계적 의사결정의 분류

일반적으로 문제의 해결을 위해, 첫째로 주어진 문제를 주의깊게 인식하여 취할 수 있는 모든 행위(action)와 발생가능한 현상을 확인하고, 둘째로 이러한 현상들이 발생가능한 확률을 산정한 다음, 마지막으로 고려된 각 행위의 결과로 발생할 잠재적인 경제적귀결(potential economic consequences)을 파악하게 된다. 그리고 이렇게 취득한 정보를 종합적으로 활용함으로써 최선의 의사결정을 수행할 수 있는 것이다.

일반적으로 의사결정 방법으로 많이 사용되는 의사결정 분기도표를 보충처리여부에 적용한 예를 제시하였다.(그림4 참조)

2.2 적합도 검정

모집단의 확률분포의 모양을 가설로 설정한 경우, 표본자료의 분포를 분석하여 모집단의 확률분포에 대한 가정의 타당성을 검정하는 것을 적합도검정(goodness of fit)이라 한다. 검정의 대상이 되는 분포는 정규분포나 포아송분포 등 어떤 분포도 가능하다. 적합도 검정에 많이 사용되는 기법이 카이제곱 검정과 Kolmogorov-Smirnov 검정이다.

2.2.1 카이제곱 적합도검정

이산확률분포에 대한 적합도검정은 근본적으로 연속확률분포에 대한 적합도검정과 동일하다. 카이제곱검정에 의한 적합도 검정절차[9]를 요약하면 다음과 같다.

Ho : 모집단은 특정의 확률분포를 따른다.

H1 : 모집단은 특정의 확률분포를 따르지 않는다.

<검정통계량>
$$\chi^2 = \sum \frac{[n_i - E(n_i)]^2}{E(n_i)}$$

여기서 $E(n_j) = n * p_j$
 n_j 는 각 계급의 도수
 <기각영역> $X^2 > \chi^2 \alpha$
 ($\chi^2 \alpha$ 는 자유도가 $df = k-m-1$ 인 χ^2 분포에서 구해진다.
 k 는 계급의 수, m 은 표본에서 추정하는 모수의 수)
 <가정> 각 계급의 기대치 $E(n_j)$ 는 5와 같거나 크다.

2.2.2 콜모고로브 스미르노브 적합도검정

적합도 검정방법 중 다른 하나는 Kolmogorov-smirnov검정[9]이다. 이 두 검정은 자료의 형태에 따라 보통 4가지로 구분한다.

자료의 형태

- 1) 명목척도(nominal scale) : 자료가 분류만의 의미를 가지고 있는 경우
 예) 남자, 여자, 대도시, 중소도시, 읍면
- 2) 순위척도(ordinal scale) : 자료끼리 비교해서 순위를 부여할 수 있는 경우
 예) 프로야구 팀의 순위, 선호하는 상표의 순위
- 3) 등간척도(interval scale) : 자료끼리 일정한 간격으로 배열할 수 있는 경우
 예) 온도, 기간별 사고건수
- 4) 비율척도(rating scale) : 자료끼리의 비율을 계산할 수 있는 경우
 예) 체중, 키

위의 자료들은 명목척도에서 순위척도, 등간척도, 비율척도로 갈수록 측정의 정확성이 증가한다고 할 수 있다. 그리고 비율척도를 등간척도로, 순위척도, 명목척도 등의 하향변환은 가능하지만 그 반대는 불가능하다.

이러한 자료 중에서 Kolmogorov-Smirnov검정을 할 수 있는 경우는 순위척도 이상인 경우이며, 명목척도인 경우에는 카이제곱검정을 하여야 한다. 그리고 카이제곱검정은 각 칸의 기대도수가 5 이상이어야 한다는 제약이 있지만 Kolmogorov-Smirnov검정에는 이러한 제약이 없다.

3. 각 사고유형별 확률밀도함수 및 누적분포함수 분석

1차대물사고가 발생(51,382명)한후 2차대물사고와 대물·대인·자차사고가 난 경우(12,796명)를 S보험회사에 가입한 60만명의 실제 데이터를 사용하여 분석하였다. 사고유형의 확률밀도함수를 각각 구하여 Histogram으로 그림2와 그림3에 나타내었고 해당되는 함수의 적합도를 검정하였다. 또한 1년, 2년, 3년이내에 재차사고가 날 확률을 기간별 누적분포함수를 이용하여 각각 구했고 이들을 표3에 정리하여 제시하였다.

3.1 대물/대물 사고의 확률밀도함수 및 적합도검정

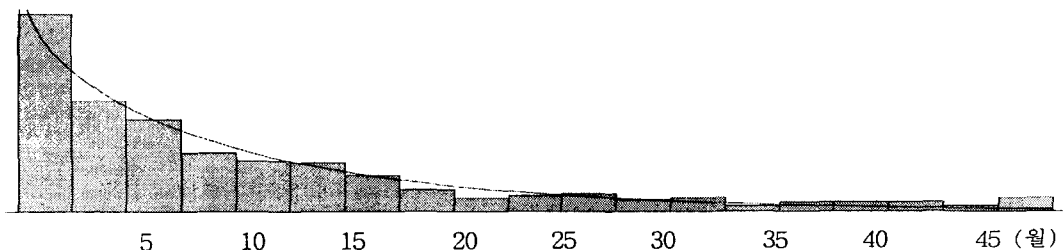


그림2 대물/대물사고의 Histogram

첫 번째는 1차대물사고를 낸 사람이 2차사고로 대물사고가 발생한 예를 분석하였다. 분석결과 이유형의 사고는 Weibull분포가 적합하다고 나타났고, 1차사고후 1년 이내에 재차사고가 난 확률은 66%로 나타났다. 분석결과를 다음 표3에 정리하였다.

3.2 대물/대물 · 대인 · 자차 사고의 확률밀도함수 및 적합도검정

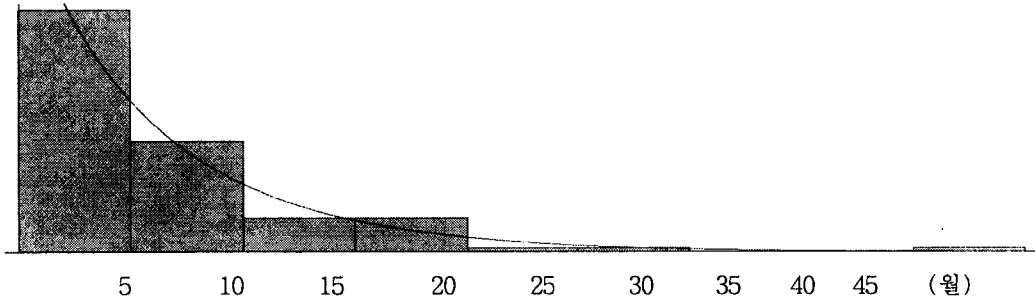


그림3 대물/대물 · 대인 · 자차사고 유형의 Histogram

두 번째는 1차대물사고후 2차대물 · 자차 · 대인사고가 발생한 예를 분석하였다. 분석결과 이유형의 사고도 Weibull분포가 적합하다고 나타났고, 1,2차 사고를 낸 사람중에서 1차사고후 다시 1년 이내에 이 유형의 재차사고가 난 확률은 82%로 나타났다. 각 사고유형별 분석결과를 표3에 정리하였다.

표3 사고유형별 확률밀도함수, 적합도검정, 누적분포함수 결과

항 목	대물/대물	대물/대물 · 대인 · 자차
Distribution	Weibull	Weibull
Expression	Weibull(11, 0.92)	Weibull(6.76, 0.96)
Chi Square Test		
Test statistic	19.1	1.02
p - value	0.0245	< 0.005
Kolmogorov-Smirnov Test		
Test statistic	0.0474	0.0879
p - value	> 0.15	> 0.15
기간별 누적분포함수		
F(12) : 1년 이내	0.661537	0.823576
F(24) : 2년 이내	0.871240	0.965777
F(36) : 3년 이내	0.949033	0.993132

4. 확률밀도함수의 활용방안

실제로 우리주변에서 일어나는 각종사고 발생을 미리 알아서 이에 대처할 필요가 있지만 이러한 사고들을 사전에 예측하여 대처하기란 현실적으로 쉽지않다. 그러나 각종사고가 발생하는 확률밀도함수를 알수 있다면 가능할 것이다. 본연구에서는 이를 해결하기 위하여 불확실한 상황에서 일어나는 각종사고들이 어떠한 분포를 하는지를 Arena라는 software를 이용하여 필요한 데이터를 입력하여 확률밀도함수를 구했고 구한 확률밀도함수의 적합도를 검정하였다. 또

한 확률밀도함수의 기간별 누적분포함수를 구함으로써 해당 기간별 확률값을 산출하였다.

이러한 확률값을 여러 가지에 이용할 수 있겠지만 본 연구에서는 사례연구로 자동차보험에 가입한 사람이 사고가 발생했을 때 현재 발생한 사고처리를 보험으로 처리할 것인지 현금처리할 것인지를 의사결정 하는데 적용할 수 있는 방법을 의사결정도표(그림2)로 제시하였다. 1차 사고 발생시 현재의 사고만 가지고 판단을 할 것이 아니라 본 연구에서 제시한 조건부확률을 이용하여 재차사고가 날 확률을 구하고 각 사고유형에 따라 보험요율이 달라지는 것을 감안하여 보다 올바른 의사결정을 할 수 있을 것이다. 이예를 의사결정 자료표(표4)로 제시하였다.

본 연구에서 사례연구로 제시한 교통사고 보험처리 문제 뿐만 아니라 불확실성하에서 발생하는 각종 사고 등에도 적용할 수 있을 것이다.

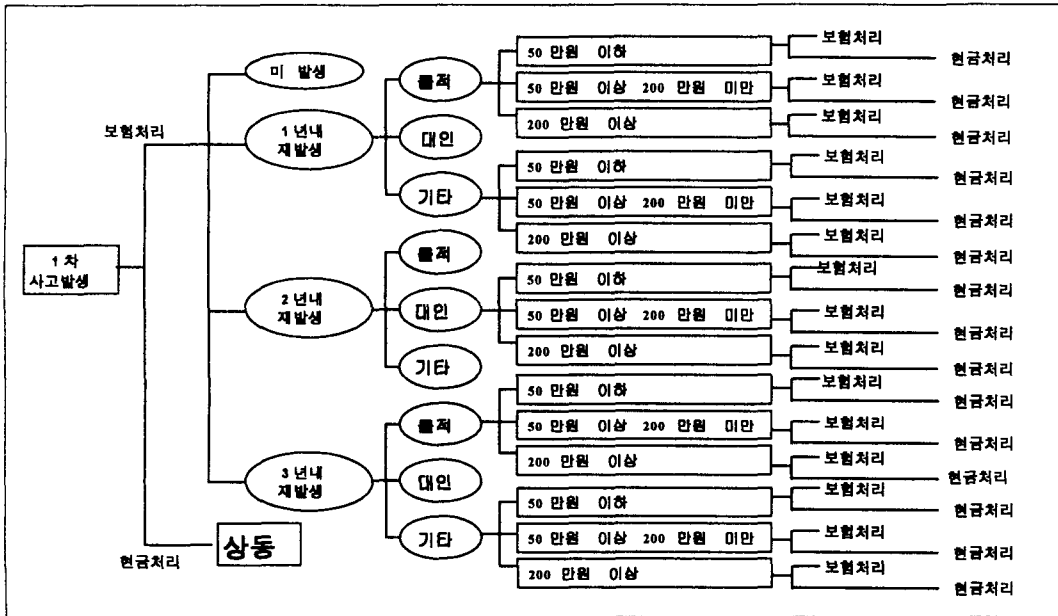


그림4 의사결정 분기도표

표 4 종합 의사결정 자료표

1차	2차		사고발생시					
			1년내		2년내		3년내	
	현금	보험	현금	보험	현금	보험		
현금	455,960	737,832	261,738	399,156	160,867	233,072		
보험	993,052	479,603	479,603	651,325	286,016	366,866		

1차	2차		미 발생시	합계
	현금소계			
	현금처리	보험처리		
현금처리	878,565	1,370,060	1,285,623	3,534,248
보험처리	1,758,671	2,081,285	2,768,386	6,608,342
보험처리와 현금처리의 차액:3,074,000(원)				

1차 사고만 고려 : 2,606,000(원) > 1,557,000(원):보험처리유리(8년)
 2차 사고만 고려 : 2,606,000(원) < 3,074,000(원):현금처리 유리(8년)

5. 결 론

자동차 사고의 보험처리시 1차사고후 재차사고가 3년 이내에 발생할 때 사건회수, 사고내용, 사고원인에 따라 할인.할증에 차이가 있어 각 사고유형을 고려해야만 한다. 1차사고후 1년, 2년, 3년 이내에 재차사고가 날 확률을 각 사고유형별 확률밀도함수를 각각 구하고 그 적합도 검정을 행하여 가장 적합한 확률밀도함수를 찾았다. 또한 일정 기간별 누적분포함수를 이용하여 구한 확률 값으로 보험수가를 추정하고 보험처리를 할 것인가를 결정할 수 있는 의사결정의 기초자료를 제시하였다. 본 연구에서 조사한 결과 1차사고후 2차 사고가 발생하는 확률밀도함수는 대부분 Weibull분포를 하고 있었고, 1차사고를 낸 사람이 2차사고로 1년 이내에 사고가 발생할 확률은 대물인 경우에는 16.40%($0.66 * 0.249$), 대물·대인·자차사고인 경우에는 20.4%($0.82 * 0.2439$)로 나타났다.

또한 1년 이내, 2년 이내, 3년 이내에 재차사고가 날 확률값을 이용하여 보험료를 재계산하여 이를 현재가치로 환산하고 현재발생금액과 비교하면 보험처리가 유리한지 현금처리가 유리한지를 결정할 수 있다.

이러한 방법을 활용한다면 일반적으로 예측할 수 없는 불확실한 상황하에서도 여러가지 자연현상의 사건들을 그에 해당하는 확률밀도함수를 찾아 적합도를 검정하고 검정된 확률밀도함수를 이용하여 향후 1차사건이 발생했다는 조건하에 2차사건이 발생할 확률을 예측한다면 보다 올바른 의사결정을 할 수 있을 것이다.

본 연구에서는 자동차사고중 1차사고가 난후 재차사고가 날 확률에 적용하여 보험처리를 할 것인가의 여부에 대한 의사결정에 관한 방안을 제시하였지만 이러한 자동차사고 뿐만 아니라 각종 사고나 불확실성하에서 일어날 수 있는 사건들에도 본 연구와 같이 적합한 확률밀도함수를 찾아 조건부 확률을 이용한다면 여러분야에 확대적용 할 수 있을 것이다.

향후 연구과제로 1차 사고유형별과 2차 사고유형과의 관계를 확률적으로 규명하여 보험수가의 합리적인 산정을 위한 연구가 이루어졌으면 한다.

참 고 문 헌

1. 경찰청, '95 교통사고 발생현황', 1996
2. 교통안전관리공단, '교통안전(월간)', 1997. 7
3. 교통안전관리공단, '교통안전(월간)', 1997. 9
4. 교통안전관리공단, '교통안전(월간)', 1996. 11
5. 김영휘외3인, '경제성공학', 청문각, 1997
6. 대한손해보험협회, '손해보험(월간)', 1996. 9
7. 대한손해보험협회, '자동차보험약관집', 1997.
8. 보험개발원, '자동차보험요율서', 1996.
9. 안상형외1인, '현대통계학', 학현사, 1995
10. 연세대학교 응용통계학과, '미니탭', 자유아카데미, 1996
11. 이공섭, '응용통계학', 아카데미아, 1996
12. 이종원, '경제경영통계학', 박영사, 1997
13. 황충하, 자동차사고에서 보험처리여부에 따른 비용분석, 인하대학교 산업대학원 석사논문, 1994.
14. Systems Modeling Corporation, 'Arena Template Reference Guide', 1994
15. Systems Modeling Corporation, 'Arena User's Guide', 1994
16. Systems Modeling Corporation, 'SIMAN V Reference Guide', 1994
17. Systems Modeling Corporation, 'Variable Guide', 1994