

Suggestions of Partial Credibilities for Proper Non-Life Insurance Premium

Myung Joon Kim^a · Jung-Ah Choi^b · Yeong-Hwa Kim^{c,1}

^aDepartment of Business Statistics, Hannam University

^bDepartment of Statistics, Graduate School of Chung-Ang University

^cDepartment of Applied Statistics, Chung-Ang University

(Received January 7, 2013; Revised March 6, 2013; Accepted March 29, 2013)

Abstract

Credibility theory is one of important theories in actuarial science to produce proper insurance premium. In this paper, new partial credibilities are proposed and introduced with widely accepted credibility theories such as rule of relative exposure volume, square root rule, Bühlmann credibility and Bühlmann-Straub credibility. Also, with credibilities estimated by current and newly suggested, the performance of the accuracy for estimating the risk is compared through real data analysis and we show that the newly suggested methods are improving the performance by reducing the error.

Keywords: Bühlmann, Bühlmann-Straub, credibility, insurance premium, square root rule.

1. 서론

소비자보호원이 2011년 각 회사별 자동차 보험료를 비교한 결과, 최저 보험료와 최대 보험료가 약 77% 이상 차이 나는 것으로 조사된 바 있으며, 가격 자유화 이후 회사별 보험료의 격차는 점차 확대되어 가는 추세이다. 따라서 위험을 평가하는 단위를 적정하게 구분하고, 해당 그룹에 적정한 보험료를 합리적으로 추정하는 것은 보험가입자와 보험회사 모두에게 중요한 문제라 할 수 있다. 보험가입자에게는 합리적인 가격으로 자동차보험을 선택해야 하는 문제가 있으며, 보험회사에게는 적정한 보험료를 책정하여 많은 소비자들에게 선택을 받아야 하는 문제가 있다. 보험료는 동일위험 그룹으로 분류된 집단별로 대수의 법칙에 의해 경제적 손실의 대가로서 예측, 추정된다. 이렇게 추정된 보험료는 보험업법에서 규정하고 있는 ‘보험소비자로부터의 총 보험료는 보험사고 발생으로 보험회사가 지급해야 하는 보험금 및 보험회사의 영업비 총액이 같아야 한다’는 수지 균등의 원칙에 따라 전체 총 평균을 동일하게 하는 과정을 거쳐 최종 보험료가 확정되게 된다. 따라서, 합리적인 보험료 산출을 위해서는 적절한 고객분류(customer segmentation)와 분류된 고객 그룹에 대한 합리적인 가격책정(pricing)이 중요한 요소라 할 수 있다. 고객분류는 위험도에 따라 고객들을 여러 그룹으로 분류하는 것을 의미하며, 그룹내 편차는 작게, 그룹간 편차는 일정 수준 이상으로 분류되어야 적절한 고객분류라 할 수 있다. 가격책정은 과거의

This paper has been supported by the 2012 Hannam University Research Fund.

¹Corresponding author: Professor, Department of Applied Statistics, Chung-Ang University, 221 Heuksuk-Dong, Dongjak-Gu, Seoul 156-756, Korea. E-mail: gogators@cau.ac.kr

데이터를 바탕으로 각 그룹의 위험도를 정확하게 예측하여 수치화하는 것을 의미한다. 자동차보험은 보험기간이 1년이라는 특징으로 인하여 매년 새로운 고객을 유치하고자 하는 가격 경쟁이 치열하게 전개되고 있으며, 이는 위험그룹을 세분화하여 우량집단에 대한 매출을 극대화하고자 하는 노력으로 이어지고 있다. 위험 그룹의 세분화는 그룹내 위험의 동질성을 극대화한다는 장점과 더불어 해당 그룹의 표본수가 점차 감소한다는 단점을 동시에 가지고 있다. 따라서 부분신뢰도를 고려하여야 하는 단위 수가 점차 증가되는 추세이며, 이에 대한 적절한 적용 방식이 보험회사에게는 중요한 요소로 재 부각되고 있는 상황이다. 또한 보험사의 특정 그룹의 매출이 급증하여 해당 계층이 급속도로 증가하는 경우를 흔히 볼 수 있으며, 회사의 신설 또는 급성장 등의 변화를 겪는 경우에도 위험도의 적정성에 대한 판단은 어려운 문제이다. 이에 본 논문에서는 고객분류 결과 완전신뢰도를 부여할 수 없는, 즉 표본의 수가 충분하지 않거나, 표본의 수가 급속도로 증가하는 특정 그룹에 대하여 어떻게 적정 위험도를 산출할 것인지에 대하여 논의하려고 한다. 기존에 활용하고 있는 신뢰도 산출방식의 소개와 더불어 본 연구에서는 새로운 신뢰도 산출 방식을 제안하고, 현실적인 위험도 추정에 더욱 적절하게 판단할 수 있는 신뢰도 적용 방식을 실증자료 분석을 통해 비교해 보고자 한다.

2. 신뢰도(Credibility)

2.1. 신뢰도의 기본 개념

전 세계적으로 손해보험료 산출에 널리 사용되고 있는 통계적 방법의 한가지인 신뢰도 이론(credibility theory)의 개념은 보험료 산출에 있어 현재까지 수집된 데이터의 활용 방식에 대한 해결책으로 CAS(미국 손해보험협회)에 의하여 제시되어 보험회사의 가격정책에 적용되기 시작하였다. 과거의 보험청구 경험에 의한 보험요율 산출시 경험 통계에 대한 신뢰의 정도에 따라 일정한 수정을 가하는 방법은 통계자료가 절대적으로 부족하여 보험요율의 산출이 어려운 경우에 있어 필요한 보험요율 산출방법이다. 신뢰도(credibility)는 현재까지 수집된 데이터를 얼마나 믿을 수 있는가에 대한 개념으로 신뢰도가 높을수록 현재의 데이터를 믿는 정도가 높아지는 것을 의미한다. 신뢰도의 범위는 '0'과 '1'사이이며, 데이터가 적어서 보험요율을 정하는데 사용하기 어려울 경우 신뢰도는 '0'에 가까워지며, 데이터가 충분할 경우 신뢰도는 '1'이 되며, 이를 완전신뢰도(full credibility)라고 한다.

Mowbray (1914)는 얼마나 많은 시행의 결과가 있어야 현재의 데이터를 믿을 수 있는지를 연구하였고, Whitney (1918)는 새로운 위험도 산출을 위하여 현재의 신뢰도 추정치와 관측된 새로운 데이터를 어떻게 조합하여 사용할 것인가에 대한 연구에 대하여 다음과 같은 신뢰도의 기본 개념을 정립하였다.

$$\text{New Estimate} = \text{Credibility} \times \text{Observed Data} + (1 - \text{Credibility}) \times \text{Prior Mean},$$

여기서 Credibility항이 당해년도 데이터에 의해 결정되는 신뢰도($0 \leq \text{Credibility} \leq 1$)가 되는 것이며, 당해년도에 관측된 데이터에 어떤 수준의 가중치를 부여하는가에 대한 문제인 것이다. 일반적으로 Credibility는 관측된 데이터에 할당된 신뢰도를, $(1 - \text{Credibility})$ 는 신뢰도의 보완적 의미를 가지며, 현재 관측치의 평균(observed data)과 과거 실적에 근거하여 적용 중인 사전평균(prior mean)의 가중평균 형태로 표현되며, 이러한 관점은 베이저안 패러다임과 동일하다고 볼 수 있다. 또한 신뢰도에 관하여 최근 주로 연구되는 방법은 크게 두 가지로 구분할 수 있다.

먼저 변동(fluctuation)을 제한하는 방법은 변동을 제한하는 방법은 데이터에 포함되어 있는 랜덤 변동(random fluctuation)의 영향을 제한하는 고전적인 방법이다.

다음으로 정확도(accuracy)를 최대한 높이는 방법은 추정오차를 최소화하는 방법이다. 이러한 방법에 근거하여 고객의 경력에 따라 신뢰도를 부여하는 여러 가지 절차가 Bühlmann (1967, 1969),

Bühlmann-Straub (1970), Hachemeister (1975), Frees (2003) 등에 의하여 개발되었다.

본 논문에서는 위험도가 증가하는 추세가 있는 반면에 자료의 관측치가 부족하여 현재의 위험도를 제한적으로 적용할 수 밖에 없는 단점을 기존의 신뢰도 적용 방식을 보완하고자 새로운 유효대수법칙, 새로운 제곱근 법칙을 제안하고자 한다. 또한 Kim 등 (2011)과 Kim과 Lee (2010)에서 사용되었던 기존의 유효대수 법칙(the rule of relative exposure volume)과 제곱근 법칙(the square root rule), Bühlmann 신뢰도와 Bühlmann-Straub 신뢰도를 기준으로 실제 자료를 통해 위험도 추정에 있어 그 정확성을 비교하고자 한다.

2.2. 기존의 부분신뢰도 적용방식

위험도를 판단하는 데 있어서 계약 건수 또는 사고 발생이 극히 일부분인 계층에 대하여서는 완전신뢰도(full credibility)를 부여하는 것이 합리적이지 않다. 따라서 집단 내에 관측치가 적게 존재하는 경우에는 위험도에 대한 정확한 판단이 어렵고, 산출된 위험도에 대한 신뢰성 문제로 인하여 부분신뢰도(partial credibility) 방식을 사용해왔다.

N_{full} 을 완전신뢰도인 경우의 필요한 보험료 유효대수라하고, 현재 데이터에서 관측된 보험료 유효대수를 N 이라고 하자. 먼저 $N \geq N_{full}$ 인 경우에는 완전신뢰도 $Z = 1$ 을 부여한다. 그러나 $N < N_{full}$ 인 경우에는 $0 \leq Z < 1$ 인 부분신뢰도를 결정하여 부여하게 된다. 부분신뢰도 가운데 가장 널리 알려진 이론이 유효대수 법칙과 제곱근 법칙이며, 이는 현재 국내 보험사들이 주로 적용하고 있는 부분신뢰도이고, 미국 손해보험협회에서 보험요율 산정에 사용하고 있는 방법이기도 하다. 또한 오차를 최소화하여 최소제곱 신뢰도라고도 불리는 Bühlmann 신뢰도와 사전 확률에 대한 정보를 활용하는 Bühlmann-Straub 신뢰도는 최대 정확도를 가지는 신뢰도로 널리 알려져 있다.

2.2.1. 유효대수 법칙 유효대수 법칙(the rule of relative exposure volume)은 Whitney (1918)가 처음으로 제안하였고, Perryman (1932), Longley-Cook (1960) 등의 연구에서 주로 사용되기 시작한 방식으로 전년도의 유효대수와 당해년도의 유효대수의 함수로 정의되며, 전년도와 당해년도의 표본 중 당해년도가 가지는 가중치를 부분신뢰도로 부여하는 방식으로 부분신뢰도 Z 는 다음과 같이 정의된다.

$$Z = \frac{N_1}{N_0 + N_1},$$

여기서 N_0 은 전년도 해당계층의 유효대수를, N_1 당해년도 해당계층의 유효대수를 의미한다. 이 방식은 계산과 적용이 매우 간단하다는 장점이 있는 반면 해당 계층(cell)의 분포만을 고려하는 제한적인 성질을 가지고 있다.

2.2.2. 제곱근 법칙 현재 보험 현업에서 가장 많이 사용되는 부분신뢰도로 제곱근 법칙(the square root rule)을 들 수 있다. 제곱근 법칙은 해당 계층의 위험도를 완전신뢰도를 만족하는 유효대수와 비례되도록 적용하는 방식으로 고안되었으며 부분신뢰도 적용에 있어 가장 보수적인 적용 방식이라 할 수 있다. 제곱근 법칙에서 신뢰도 Z 는 다음과 같이 결정된다.

$$Z = \sqrt{\frac{N}{N_{full}}},$$

여기서 N 은 당해년도의 보상건수, N_{full} 은 완전신뢰도를 만족하는 충분한 표본의 수를 의미한다. 완전신뢰도를 부여하지 못하는 집단에서 부분신뢰도 적용을 통한 위험도의 추정의 수정은 극히 제한적이며 적용범위가 제한적인 만큼 과거에 적용되었던 위험도를 그대로 적용하게 되는 단점을 가지고 있다.

2.2.3. Bühlmann 신뢰도 신뢰도 Z 를 계산하기 위한 사전정보가 필요하지 않거나 많은 요인들이 필요하지 않은 접근법을 Bühlmann (1967, 1969)이 제안하였다. 최대정확도를 갖는 신뢰도로 자주 사용되는 Bühlmann 신뢰도는 최소제곱 신뢰도라고도 불리며, 서로 독립인 두 추정량의 제곱오차 (squared error)를 사용하여 부분신뢰도를 구하는 방법으로서 부분신뢰도 Z 는 다음과 같이 정의된다.

$$Z = \frac{N}{N + K}, \quad K = \frac{EPV}{VHM}.$$

위 식에서 N 은 해당 계층의 보상건수, K 는 Bühlmann 신뢰도 모수, EPV(the expected value of the process variance)는 과정분산의 기대치이고 VHM(variance of the hypothetical means)은 가설평균의 분산을 의미하게 되며, 여기서 Bühlmann 신뢰도 모수인 K 를 추정해야 한다. EPV는 전체 위험그룹에 걸쳐 있는 과정분산의 기대값으로서 평균 급간 변동(average class variance)이라고도 하며, VHM은 그룹내에 존재하는 동질성과 이질성의 통계적 척도로서 급간 평균사이의 변동(variance between class means)이라고도 한다. 총분산은 EPV, VHM의 두 부분으로 분해되는데, 개념의 이해를 돕기 위해 임의의 확률변수 X 에 대하여 EPV와 VHM을 정의하면 다음과 같다.

확률변수 X 의 분산 $\text{Var}(X)$ 는 임의의 확률변수 θ 를 사용하여 다음과 같이 표현할 수 있다.

$$\text{Var}(X) = E[\text{Var}(X|\theta)] + \text{Var}[E(X|\theta)],$$

여기서 $E[\text{Var}(X|\theta)] = EPV$, $\text{Var}[E(X|\theta)] = VHM$ 이며, 따라서 다음과 같이 표현할 수 있다.

$$\begin{aligned} EPV &= E[\text{Var}(X|\theta)] \\ &= E[E(X^2|\theta)] - E\{[E(X|\theta)]^2\} \\ &= E(X^2) - E\{[E(X|\theta)]^2\} \\ VHM &= \text{Var}[E(X|\theta)] \\ &= E\{[E(X|\theta)]^2\} - \{E[E(X|\theta)]\}^2 \\ &= E\{[E(X|\theta)]^2\} - [E(X)]^2 \end{aligned}$$

따라서 $\text{Var}(X) = EPV + VHM$ 이며, 실제적으로 VHM은 초기에 알려지지 않는 것이며 이러한 경우에 앞에서 정의한 총분산공식(total variance formula)을 사용하여 추정하게 된다.

2.2.4. Bühlmann-Straub 신뢰도 Bühlmann-Straub (1970)이 제안한 Bühlmann-Straub 신뢰도는 Bühlmann 신뢰도의 일반화된 형태로 Bühlmann 신뢰도가 과거 보험기간 동안의 여러가지 보상건수 또는 보상금액의 분포를 고려하지 못하는 단점을 보완하고자 제시한 방법이다. Bühlmann 신뢰도에서는 사전확률을 임의로 지정하여 사용(보통 동일한 사전확률 부여)하였지만 Bühlmann-Straub 신뢰도에서는 보상 건수에 대해 가중치를 두어 사전확률을 사용한다. Bühlmann-Straub 신뢰도의 기본 원리를 요약하면 다음과 같다 (Hong과 Chun, 2006).

$j = 1, 2, \dots, n+1$ 에 대하여 j 번째 보험기간 동안 단일 보험가입자의 보험금을 X_j 라 하면, n 년의 사전기간 동안 $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ 이 주어졌을 때, $n+1$ 번째 확률변수 X_{n+1} 의 기대값을 추정하기 위하여 다음과 같은 가정을 설정한다.

- (1) $j = 1, 2, \dots, n+1$ 에 대하여 각 X_j 의 분포는 모수 θ 에 의존한다.
- (2) θ 가 주어졌을 때, X_1, X_2, \dots, X_{n+1} 은 조건부 독립이고 동일한 분포를 따른다.

$j = 1, 2, \dots, n + 1$ 에 대하여 $\mu(\theta) = E(X_j|\theta)$ 라 정의한다. 여기서 m_j 는 j 번째 보험기간동안의 보상 건수를 나타내며 기지의 상수이다. Bühlmann-Straub 모형하에서의 목적은 $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ 이 주어진 $\mu(\theta)$ 의 조건부 기대값 $E(X_{n+1}|X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ 의 점추정 값을 구하는 것이며 Mayerson (1964)은 이러한 경우에 $E(X_j|\theta) = \mu(\theta)$ 임을 증명하였다. 결과적으로 $\mu(\theta)$ 의 점추정량은 다음과 같이 나타낼 수 있다.

$$Z\bar{X} + (1 - Z)\mu, \quad \bar{X} = \frac{\sum_j m_j X_j}{\sum_j m_j},$$

여기서 μ 는 사전평균(prior mean)이다. Bühlmann-Straub 신뢰도도 Bühlmann 신뢰도와 마찬가지로 신뢰도 모수는 K 를 추정하여 사용해야 하며, 그 형태는 Bühlmann 신뢰도와 동일하게 정의된다.

본 논문의 실증자료 분석에서는 Bühlmann 신뢰도에서의 EPV와 VHM을 EPV₁와 VHM₁, Bühlmann-Straub 신뢰도에서의 EPV와 VHM을 EPV₂와 VHM₂라 표기하고, 이에 근거하여 계산되는 얻어지는 Bühlmann 신뢰도의 모수를 K_1 , Bühlmann-Straub 신뢰도의 모수를 K_2 라 표기하기로 한다.

2.3. 새로운 부분신뢰도 제안

현재 가장 널리 활용되고 있는 유효대수 법칙과 제공근 법칙은 위험도 추정에 있어 제한적인 특징을 내포하고 있기 때문에 해당 집단의 위험도를 산출, 적용하는 상황에 따라 왜곡된 결과를 초래할 수 있다. 예를 들어, 현재 국내 신설 직판사의 경우 매출이 급증하여 해당 계층이 급속도로 증가하는 경우를 흔히 볼 수 있다. 이때, 한 계층이 급증하게 되면 전년도와 당해년도의 위험도의 적정성에 대한 판단보다는 당해년도에 가중치를 더 부여하게 된다. 해당계층의 관측치가 충분히 많지 않아 적은 계약건수의 변화에도 민감하게 변동되는 계층은 당해년도의 우연적인 몇 건의 대형 사고들로 인하여 실제 위험도가 아닌 왜곡된 위험도로 수렴되어 적용되는 문제점을 가지고 있다. 회사의 신설 또는 급성장 등의 변화를 겪는 경우에 고전적인 유효대수 법칙은 안정적인 위험도의 판단에 취약한 부분이 드러나게 되는 것이다.

해당 계층의 유효대수가 충분하지 못한 경우 그 계층의 위험도의 판단에 확신을 가지지 못할 때에는 과거의 위험도를 그대로 적용하는 것이 바람직하다고 할 수 있다. 그러나 일정 시간이 지나 해당 계층의 위험도가 어느 정도 판단이 가능한 경우에도 적정 위험도가 반영이 제한적일 수 밖에 없다는 문제점이 있다. 예를 들어, 과거 3년전 어떤 계층의 위험도가 1.2였다고 가정하자. 이후 3년여 동안 2.0이라는 위험도가 계속되어 왔으나, 해당 계층의 당해년도 부분신뢰도가 10% 수준일 경우 과거 수준의 위험도 1.2에 90%의 비중이 지속적으로 적용된다. 이렇게 되면 당해년도 및 익년도의 적정 위험도 2.0으로의 수렴이 계속 지연될 수 밖에 없고, 따라서 해당 계층에 대한 위험도가 지속적으로 실제로는 2.0으로 실현되고 있으나 이보다 과소하게 적용될 가능성이 존재하며, 반대의 경우에는 위험도가 떨어지지 않고 지속적으로 과대하게 적용하게 되는 가능성이 존재하게 되며, 이는 보험료 부과와 형평성 문제를 야기시켜 해당 계층의 위험도가 비해당계층으로 전가되는 현상을 초래하게 되는 것이다. 따라서 본 논문에서는 현재 자동차 보험업계 현업에서 가장 널리 활용되고 있는 유효대수 법칙 방식과 제공근 법칙 방식이 가지고 있는 제한적인 성질을 보완하여 아래와 같은 적용 방식을 제안하고자 한다. 이는 다양한 사업환경에서 보험회사가 각자의 상황에 맞는 방식을 취사선택할 수 있는 여러 가지 대안을 제시함으로써, 가장 적절한 위험도를 추정, 산출하여 고객에게 최적의 보험료를 부과할 수 있다는 데에 그 의미를 둘 수 있다.

2.3.1. 새로운 유효대수 법칙

회사의 신설 또는 급성장 등의 변화를 겪는 경우에 새롭게 제시되는 이

방식은 안정적인 위험도의 판단에 취약한 부분을 보완할 수 있는 방식으로 다음과 같이 정의한다.

$$Z = \frac{\frac{N_1}{T_1}}{\frac{N_0}{T_0} + \frac{N_1}{T_1}},$$

여기서 T_0 는 전년도 전체 유효대수, T_1 은 당해년도 전체 유효대수를 의미하며, N_0 는 전년도 해당계층의 유효대수, N_1 은 당해년도에서 해당계층의 유효대수를 의미한다. 즉, 해당 계층이 당해년도에 차지하고 있는 비율에 대한 가중치를 고려하여 계산되는 방식이다. 새로운 유효대수 법칙은 해당계층의 유효대수 뿐만 아니라 전년도와 당해년도의 전체 유효대수를 포함하여 기존의 유효대수 법칙보다 상대적으로 당해년도에 치우치지 않는 가중치를 적용하게 된다. 즉, 기존의 유효대수 법칙에 비해 당해년도의 위험도에 과도하게 치우치는 문제점을 보완할 수 있으며, 해당 계층이 전체적인 계약 포트폴리오에서 차지하는 비중을 고려할 수 있다는 장점을 가지게 된다.

2.3.2. 새로운 제공근 법칙 앞서 언급하였듯이 기존의 제공근 법칙은 완전신뢰도를 부여하지 못하는 집단에서 부분신뢰도가 지속적으로 낮은 비중을 차지하는 경우, 위험도 판단에 있어 과거의 위험도에 수렴하는 현상이 나타나게 되며, 위험도의 추세로 판단되는 적정 위험도가 반영이 지연되는 현상을 가지고 있다. 따라서 이와 같은 위험도 수렴의 지연 현상을 보완하고자 다음과 같은 새로운 제공근 법칙을 제안한다.

$$Z = \sqrt{\frac{N + N'}{N_{full}}},$$

여기서 N 은 전년도 보상건수, N' 은 당해년도 보상건수를 의미하며, N_{full} 은 완전신뢰도를 부여하기 위해 충분히 필요한 표본의 수를 의미한다. 새롭게 제안하는 제공근 법칙은 당해년도의 보상건수 뿐만 아니라 과거 년도의 보상건수를 추가로 고려해 주는 방식이다. 새로운 제공근 법칙은 해당 계층에 대한 위험도가 지난 몇 년간 해당계층의 위험도가 증가했을 때 그 위험도의 적용이 적절하다고 판단되는 경우 그 위험도를 수렴 속도를 좀 더 빠르게 진행 시킬 수 있다는 장점을 가진다고 할 수 있다.

3. 실증 자료 분석

현재 국내 자동차보험사에서는 다양한 고객의 정보를 이용하여 사고 발생시 고객에게 지급되는 보험금을 기준으로 고객이 납부하는 보험료를 책정한다. 따라서 고객 정보의 정확한 분석을 통한 고객 집단별 보험료의 정확한 추정은 보험사에 있어 가장 중요한 일이라 할 수 있다. 본 장에서는 전년도, 당해년도 데이터를 통해 부분신뢰도를 해당 방식별로 산출하고, 위험도 추정에 있어 그 적합성을 익년도 데이터를 사용하여 서로 비교해보고자 한다.

3.1. 데이터의 구성

분석에 사용된 자료는 국내 손해보험사 S사의 실제 자료이다. 설명 변수로서 고객의 사고 유무에 따른 할인할증, 운전경력, 운전자 한정 특약, 차종, 연령대, 성별을 사용하였다. 변수의 수준에 따라 고객 집단을 나누어 사고 1건당 발생하는 손해액, 즉 보험사에서 지급하는 보험금을 집단별로 계산하고 신뢰도를 적용하여 새로운 보험료를 책정한다. 설명변수와 그 수준은 Table 3.1과 같으며, 종속 변수는 보험료이다.

Table 3.1. Explanatory variables

설명변수명	범주수	범주 내용
할인할증	3	사고(할증), 신규, 무사고(할인)
운전경력	4	1~2년, 3~4년, 5~6년, 7년이상
운전자한정특약	4	부부, 1인, 가족 형제, 기타
차종	3	소형, 중대형, 대형(SUV, 다인승)
연령대	3	20~39세, 40~59세, 60세이상
성별	2	남, 여

Table 3.2. Data arrangement

연도	총관측수	평균청구액(원)
year1	310,454	44,030
year2	310,280	52,650
year3	310,717	54,080

실제 데이터는 2009년 유효한 자동차보험 계약의 대물담보 실적에서 추출된 932,880건이며, 이를 Uniform(0, 1) 난수생성을 통해 세 집단으로 분류하였다. 첫 번째 집단 year1은 전년도 데이터로 가정하였고, 두 번째 집단 year2는 당해년도 데이터로 가정하였으며, 세 번째 집단 year3는 익년도 데이터로 간주하였다. 즉, 전년도 데이터와 당해년도의 데이터는 익년도 보험료를 산출하는데 필요한 데이터로 사용하였고, 익년도의 데이터는 여러 가지 신뢰도를 사용하여 산출한 보험료가 적절한 지를 판단하는 가상의 데이터로 사용하였다. 전체 고객은 사고유무에 따른 할인할증, 운전경력, 운전자한정, 차종, 연령대, 성별의 수준에 따라 864개의 집단으로 분류하였다.

Table 3.2를 보면 약 93만개의 데이터가 랜덤하게 year1, year2, year3의 세 집단으로 나누어 진 것을 확인할 수 있다. 본 연구에서 데이터를 랜덤하게 나누어 본 결과 평균청구액을 살펴보면, 전년도에 비해 당해년도가 약 1.19배, 전년도에 비해 익년도가 약 1.23배 늘어난 것을 확인할 수 있다. 평균청구액이 늘어나는 것으로 나타난 것은 데이터의 분류방법이나 난수생성방법 등에 따라 달라질 수 있다. 본 연구에서 데이터의 분류에서 이러한 경향이 나타난 것으로 볼 때 기존의 부분신뢰도 방식보다 새롭게 제안하는 부분신뢰도 방식이 익년도 보험료 추정에 정확성을 보일 것으로 예측되며, 특히 유효대수를 만족하지 못하는 집단에 대한 보험료 추정에서 새로운 부분신뢰도의 정확성이 두드러질 것이라 생각된다. 본 연구에 사용된 실증자료는 자동차보험의 대물담보로써 매년 인건비와 부품대의 상승으로 인하여 지속 위험도가 증가하는 특징을 가지고 있으며, 이는 표본의 수가 적은 그룹에 영향을 크게 미치게 된다. 따라서 본 연구는 특정 그룹의 표본의 수가 급증되거나, 위험도의 변경이 일어나는 경우 등에 있어서 보다 합리적인 대안을 제시하고자 한다. 신뢰도의 적용방식은 보험업 감독규정에서 보험사가 합리적인 방식을 선택할 수 있도록 규정하고 있는 바, 다양한 방식의 제안을 통하여 보험사가 각자의 상황에 맞는 합리적인 적용 방안을 취사선택할 수 있도록 하는데 그 의의를 두고자 한다. 본 장에서는 전년도 year1과 당해년도 year2를 이용하여 완전신뢰도에 필요한 유효대수, 새로운 부분신뢰도 방법, 기존의 부분신뢰도 방법을 이용해 신뢰도를 구한 후 새로운 보험료를 추정해서 실제 year3의 평균 청구액과 비교해 새로운 부분신뢰도의 제안이 의미 있는 것인지 논의하고자한다.

3.2. 위험도와 완전신뢰도의 경우에 필요한 유효대수 추정

먼저 6개의 설명변수의 각 수준에 따라 현행상대도와 위험도를 구하는 과정이 필요하다. 여기서 현행상대도는 과거 실적에 의거하여 산출되어 적용 중인 위험도를 의미하며, 위험도는 현재 데이터 실적을 바

Table 3.3. Risks & Sample sizes (when $Z = 1$)

변수	수준	현행상대도	위험도	표본수
할인할증	사고(할증)	1.4859445275	1.377710251	3169.90
	신규	1.4422449836	1.445574028	3549.64
	무사고(할인)	0.9043783608	0.910257052	4399.58
운전경력	1~2년	1.4479521207	1.527592968	4070.88
	3~4년	1.1551397565	1.004949939	3515.99
	5~6년	1.0161815613	0.949053480	3474.26
	7년이상	0.9245784209	0.936434732	4301.62
⋮	⋮	⋮	⋮	⋮
성별	남	1.0529299769	1.074476093	4405.14
	여	0.9174872080	0.885062633	3662.69

탕으로 하여 추정하고자 하는 위험도를 의미한다. 따라서 전년도 데이터를 사용하여 현행상대도를, 당해년도 데이터를 사용하여 위험도를 구할 수 있으며, 아래와 같은 과정을 거쳐 해당 계층별로 위험도가 계산된다. 현행상대도 C_{ij} 가 i 번째 변수의 j 번째 수준에서의 위험도라고 정의된 경우, 할인할증의 첫 번째 수준(사고(할증))에 해당하는 현행상대도 C_{11} 은 ‘year1의 고객 중 사고(할증)를 낸 고객의 평균손해액’을 ‘year1의 평균손해액’으로 나눈 값이므로 $C_{11} = 65422.686303/44027.677408 = 1.4859445$ 와 같이 계산된다. 또한 당해년도 위험도 D_{ij} 를 i 번째 변수의 j 번째 수준에서의 값이라 하면, 할인할증의 첫 번째 수준(사고(할증))에 해당하는 위험도 D_{11} 은 ‘year2의 고객 중 사고(할증)를 낸 고객의 평균손해액’을 ‘year2의 평균손해액’으로 나눈 값이므로 $D_{11} = 72531.03585/52646.074018 = 1.37771$ 과 같이 구해진다. 실제 데이터에 대하여 계산된 현행상대도와 위험도는 Table 3.3에 제시되어 있다.

완전신뢰도를 부여하기 위하여 필요한 표본의 수는 빈도에 대한 포아송 분포 가정을 통하여 다음과 같이 정의되며,

$$N_{full} = \left(\frac{z_p}{k} \right)^2 \left\{ \frac{\text{Var}(F)}{E(F)} + \frac{\text{Var}(S)}{[E(S)]^2} \right\}.$$

실제 데이터분석에서는 현재 국내 자동차 보험회사에서는 오차허용한계 k 는 0.05, 신뢰수준 p 는 0.9를 주로 사용하므로 여기서도 그 값들을 사용하였으며, 그 결과 완전신뢰도를 부여하는데 필요한 표본의 수는 Table 3.3에 제시되어 있다.

3.3. 부분신뢰도 추정

완전신뢰도를 부여하기 위하여 필요한 최소한의 유효대수가 확보되어 있는 경우는 100% 신뢰도($Z = 1$)를 부여하지만 그렇지 않은 경우에는 100%보다 작은 부분신뢰도를 부여하게 된다.

유효대수 법칙에 따른 부분신뢰도를 Z_1 , 제공근 법칙에 따른 부분신뢰도를 Z_2 , Bühlmann 신뢰도를 Z_3 , Bühlmann-Straub 신뢰도를 Z_4 , 새롭게 제안하는 유효대수 법칙에 따른 부분신뢰도를 Z_5 , 새롭게 제안하는 제공근 법칙에 따른 부분신뢰도를 Z_6 라 하면, 각각의 부분신뢰도는 아래와 같이 정의된다.

$$Z_1 = \frac{N_1}{N_0 + N_1}, \quad Z_2 = \sqrt{\frac{N}{N_{full}}}, \quad Z_3 = \frac{N_1}{N_1 + K_1}, \quad K_1 = \frac{EPV_1}{VHM_1},$$

$$Z_4 = \frac{N_1}{N_1 + K_2}, \quad K_2 = \frac{EPV_2}{VHM_2}, \quad Z_5 = \frac{N_1/T_1}{N_0/T_0 + N_1/T_1}, \quad Z_6 = \sqrt{\frac{N + N'}{N_{full}}}$$

Table 3.4. Estimates of 6 partial credibilities

변수	수준	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
할인할증	사고(할증)	0.502562	0.835070	0.843992	0.84399	0.502702	0.874107
	신규	0.503781	0.619361	0.611584	0.611582	0.503921	1
	무사고(할인)	0.499122	0.812792	0.748393	0.748391	0.499262	1
운전경력	1-2년	0.505305	1	0.955347	0.955347	0.505445	0.989768
	3-4년	0.501724	0.616654	0.684033	0.68403	0.501865	0.87371
	5-6년	0.496125	0.559350	0.662718	0.662716	0.496266	0.809207
	7년이상	0.499338	1	0.950781	0.950781	0.499479	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
성별	남	0.498872	1	0.940274	0.940273	0.499013	1
	여	0.501391	1	0.910723	0.910722	0.501531	1

Table 3.4는 위와 같이 정의된 부분신뢰도를 실제 데이터에 적용하여 얻은 결과이다.

3.4. 부과보험료 추정 및 부분신뢰도 비교

부과보험료를 추정하기 위하여 반영위험도를 먼저 계산하여야 한다. 반영위험도를 계산하기 위하여 2.1절에서 소개하였던 Whitney (1918)의 제안에 따라 전년도 자료와 당해년도 자료를 다음과 같이 조합하여 사용한다.

$$\text{New Estimate} = \text{Credibility} \times \text{Obs.} + (1 - \text{Credibility}) \times \text{Prior Mean.}$$

i 번째 변수의 j 번째 수준에 대하여 반영위험도를 R_{ij} , 현행상대도를 C_{ij} , 위험도를 D_{ij} 라 하면 위 식은 다음과 같이 표현된다.

$$R_{ij} = Z \times D_{ij} + (1 - Z) \times C_{ij}.$$

예를 들어, 할인할증의 첫 번째 수준(사고(할증))의 반영위험도 R_{11} 은 다음과 같이 계산된다.

$$\begin{aligned} R_{11} &= Z \times D_{11} + (1 - Z) \times C_{11} \\ &= Z \times 1.37711025 + (1 - Z) \times 1.48594452. \end{aligned}$$

위 식과 같이 각 변수의 수준마다 반영위험도를 구한 뒤, 6개 변수의 수준 조합인 864개 집단에 대한 반영위험도의 곱을 계산한다. 864개 집단에 대하여 얻어진 반영위험도의 값과 당해년도의 총평균 손해액을 곱해 부과보험료를 산출한다. 예를 들어, ‘할인할증 = 1(사고(할증)), 운전경력 = 2(3~4년), 운전자한정특약 = 2(1인), 차종 = 1(소형), 연령대 = 3(60세 이상), 성별 = 1(남)’인 집단의 부과보험료(New Rate)는 다음과 같이 계산한다.

$$\text{New Rate} = \text{Mean.Loss} \times R_{11} \times R_{22} \times R_{32} \times R_{41} \times R_{53} \times R_{61},$$

여기서 Mean.Loss는 총평균손해액으로서 당해년도의 전체손해액을 전체가입자수로 나눈 값이며, 864개 집단에 대한 부과보험료는 Table 3.5에 제시하였다. 각 집단에 대한 보험료는 6가지로 나타나는데 이것은 신뢰도를 6가지 방법에 의해서 구했기 때문이다. 또한 Table 3.5의 부과보험료 추정치는 추정된 부과보험료의 총합이 당해년도의 전체손해액과 동일하게 적용되도록, 즉 보험업계에서 적용하는 수지균등의 원칙(Off-Balance) 과정을 거친 결과이다.

Table 3.5. Premiums for each group by credibilities

집단	표본수	신뢰도					
		Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
1	30	96902.5958	94049.0716	94819.1847	94819.1951	96901.5392	93766.3890
2	383	97745.4929	95792.5571	96503.6527	96503.6460	97745.5012	97450.1007
3	231	61186.6152	59995.9372	60453.5909	60453.5860	61186.6649	61092.5673
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
864	321	53883.2066	53456.8528	53678.2545	53678.2528	53883.2391	53788.4403

Table 3.6. MSEs for estimated premiums

신뢰도	평균 오차 제곱합
Z_1	4.695453E11
Z_2	4.695049E11
Z_3	4.694911E11
Z_4	4.694911E11
Z_5	4.695452E11
Z_6	4.694192E11

3.5. 위험도의 추정 정합성 비교

앞에서 산출된 결과를 바탕으로 익년도(year3) 실제 데이터와의 격차를 확인함으로써 위험도 추정에 대한 정교성을 검증할 수 있다. 즉, year1 데이터와 year2 데이터를 통하여 year3의 위험도를 추정하고, 추정된 결과를 year3의 실제 데이터와의 오차를 확인하는 것이다. 다음 Table 3.6은 익년도 데이터 864개 집단의 전체 데이터를 사용하여 각 부분신뢰도에 따라 추정된 보험료와 익년도(year3)에 발생한 실제 손해액 사이의 오차 제곱합의 평균값을 계산한 결과이다. 즉, 익년도 데이터 864개 집단 중 완전신뢰도를 부여하지 못하는 변수의 수준들을 포함하는 집단에 대해 각 부분신뢰도에 따라 추정된 보험료와 익년도(year3)에 발생한 실제 손해액 사이의 오차 제곱합의 평균값을 계산한 결과이다.

계산 결과를 작은 순서대로 정리해 보면 $Z_6 < Z_3 \leq Z_4 < Z_2 < Z_5 < Z_1$ 이다. 즉, 본 논문에서 제안하려고 하는 새로운 유효대수 법칙에 따른 부분신뢰도 Z_5 가 기존의 유효대수 법칙에 따른 부분신뢰도 Z_1 보다 평균 오차 제곱합의 값이 더 작게 나타났음을 알 수 있다. 따라서 위험도가 증가하는 추세의 데이터에서 뿐만 아니라 일부 완전신뢰도를 부여하지 못하는 집단에 대해서도 기존의 유효대수 법칙에 따른 부분신뢰도보다 본 논문에서 새롭게 제안하는 유효대수 법칙에 따른 부분신뢰도가 부과보험료 추정에 있어 정확도가 높다고 할 수 있다.

또한 새롭게 제안하는 제곱근 법칙에 따른 부분신뢰도 Z_6 가 기존의 제곱근 법칙에 따른 부분신뢰도 Z_2 보다 평균 오차 제곱합이 작게 나타난 것과 함께, 모든 부분신뢰도 가운데 가장 작은 평균 오차 제곱합을 보이는 것을 확인할 수 있다. 이것은 완전신뢰도를 부여하지 못하는 집단에 대해 기존의 제곱근 법칙에 따른 부분신뢰도보다 새로운 제곱근 법칙에 따른 부분신뢰도가 부과보험료 추정에 있어 정확도가 월등히 높아진 것으로 해석할 수 있다.

평균 오차 제곱합이 작다는 것은 부분신뢰도 적용방법을 통해 해당 계층에 대한 위험도를 추정하여 부과보험료를 부과하는데 있어 그 정확도 및 형평성을 개선시킨다는 것을 의미하게 된다. 완전신뢰도를 부여하지 못하는 집단들을 선택하여 평균 오차 제곱합을 계산해 본 결과 Table 3.6에서 유효대수 법칙과 제곱근 법칙에 있어 새롭게 제안된 법칙에 따른 부분신뢰도가 기존의 부분신뢰도보다 평균 오차 제곱합이 작게 나타나는 결과를 보였다. 이것은 앞서 언급한 대로 일부 완전신뢰도를 부여할 수 없는 집단에

대한 위험도 추정에 있어 기존의 방법들 보다 정확도를 개선시켜 적정한 보험료 산출에 기여한다는 것을 의미한다고 할 수 있다.

4. 결론

본 논문에서는 새로운 유효대수 법칙에 따른 부분신뢰도와 새로운 제공근 법칙에 따른 부분신뢰도를 제안하고 Kim 등 (2011)에서 사용한 기존의 4가지의 부분신뢰도를 이용해 부과보험료를 추정하였다. 여러 가지 부분신뢰도에 의해 추정한 부과보험료와 손해액과의 평균 오차 제곱합을 계산하였다. 그 결과, 새로운 제공근 법칙에 따른 부분신뢰도를 바탕으로 완전신뢰도를 부여할 수 없는 집단들을 대상으로 $Z_6 < Z_3 \leq Z_4 < Z_2 < Z_5 < Z_1$ 의 순서대로 부분신뢰도의 정확성을 보였다. 즉, 새로운 제공근 법칙에 따른 부분신뢰도가 기존의 제공근 법칙에 따른 부분신뢰도보다 정확도가 월등히 높은 것을 확인할 수 있었으며, 또한 새로운 유효대수 법칙에 따른 부분신뢰도 역시 기존의 유효대수 법칙에 따른 부분신뢰도보다 정확도가 높게 나왔다. 이것은 앞서 논의한 특정 계층의 관측치가 증가하여 위험도 판단에 적정성을 따져보아야 할 때 본 논문에서 제안하는 새로운 유효대수 법칙에 따른 부분신뢰도가 기존의 방식보다 위험도 추정의 정확도면에서 보다 좋은 접근 방식임을 보여주는 것이다. 또한 매년 위험도가 지속적으로 증가하는 상황에서 완전신뢰도를 부여하기 어려운 일부 계층에 대해 적절한 신뢰도를 부여하는 방법에 대해 본 논문이 제안하는 새로운 제공근 법칙에 따른 부분신뢰도가 기존의 방식보다 정확도가 높은 결과를 보였다. 그러므로 위험도 산출에 있어 본 논문에서 제안하는 새로운 유효대수 법칙에 따른 부분신뢰도와 새로운 제공근 법칙에 따른 부분신뢰도가 적절한 것으로 판단된다.

본 논문에서 제안하는 새로운 두 가지의 신뢰도는 기존의 신뢰도보다 보험청구 사건, 사고의 다양화와 특정인을 대상으로 하는 보험상품에 적용이 용이할 것으로 보인다. 새로운 두 가지의 부분신뢰도의 적용에 대하여 통계적, 수학적 당위성을 이론적으로 증명할 수는 없으나, 그 상황에 맞도록 적용될 수 있는 다양한 방식이 보완, 제시되는 것이 적정한 위험도를 판단하고 적정 보험료를 부과하는데 도움이 될 수 있다는 것에 본 논문의 제안에 의의를 두고자 한다.

기존의 완전신뢰도를 충족하는 표본의 수를 구하는 과정에 가정하고 있는 분포에 대한 연구는 무정보적 사전분포 및 공액 사전분포 등을 통한 베이지안 추론의 연구 분야가 될 수 있을 것으로 판단되며 향후 연구과제로 제시하는 바이다.

References

- Bühlmann, H. (1967). Experience rating and credibility I, *ASTIN Bulletin*, **4**, 199–207.
- Bühlmann, H. (1969). Experience rating and credibility II, *ASTIN Bulletin*, **5**, 157–165.
- Bühlmann, H. and Straub, E. (1970). Glaubwürdigkeit für schandensätze, *Bulletin of the Swiss Association of Actuaries Communications in Statistics, Theory and Methods*, **70**, 111–133.
- Frees, E. W. (2003). Multivariate credibility for aggregate loss models, *North American Actuarial Journal*, **7**, 13–37.
- Hachemeister, C. A. (1975). Credibility for regression models with application to trend, *Credibility, Theory and Application*, Academic Press, New York, 129–163.
- Hong, J. and Chun, H. (2006). *Credibility Theory*, Free Academy, Seoul.
- Kim, Y.-H., Kim, M. J. and Kim, M. J. (2011). Estimating the automobile insurance premium based on credibilities, *The Korean Journal of Applied Statistics*, **24**, 279–292.
- Kim, Y.-H. and Lee, H. (2010). A comparison study for the pricing of automobile insurance premium based on credibility, *Communications of the Korean Statistical Society*, **17**, 713–724.
- Longley-Cook, H. (1960). *An Introduction to Credibility Theory*, XLIX PCAS, 194.

- Mayerson, Allen L. (1964). A Bayesian View of Credibility, *Proceedings of the Casualty Actuarial Society*, **1**, 85–104.
- Mowbray, A. H. (1914). How extensive a payroll exposure is necessary to give a dependable pure premium, *Proceedings of the Casualty Actuarial Society*, **1**, 24–30.
- Perryman, F. S. (1932). *Some Notes on Credibility*, XIX PCAS, **65**.
- Whitney, A. W. (1918). The theory of experience rating, *Proceedings of the Casualty Actuarial Society*, **4**, 274–292.

적정 손해보험료 산정을 위한 부분신뢰도 제안

김명준^a · 최정아^b · 김영화^{c,1}

^a한남대학교 비즈니스 통계학과, ^b중앙대학교 대학원 통계학과, ^c중앙대학교 응용통계학과

(2013년 1월 7일 접수, 2013년 3월 6일 수정, 2013년 3월 29일 채택)

요약

신뢰도 이론은 적정하고 합리적인 보험료를 산정하기 위한 보험통계학의 이론 중 하나이다. 본 논문에서는 신뢰도의 개념과 함께 기존에 널리 사용되고 있는 여러 가지 부분신뢰도(유효대수 법칙, 제공근 법칙, 불만, 불만-스트라움)를 소개하고, 새로운 부분신뢰도(새로운 유효대수 법칙, 새로운 제공근 법칙)를 제안하였다. 또한 실제 자료를 이용하여 기존의 방법과 새로운 방법으로 부분신뢰도와 부과보험료를 추정하였으며, 추정된 부과보험료와 전체 손해액의 차이를 비교하여 정확성을 알아보았다. 실제 자료를 바탕으로 분석한 결과, 본 논문에서 제안하는 새로운 부분신뢰도를 적용하여 부과보험료를 산출한 것이 기존의 방법보다 오차가 개선되는 것을 확인할 수 있었다.

주요용어: 보험료, 불만 신뢰도, 불만-스트라움 신뢰도, 신뢰도.

이 논문은 2012년도 한남대학교 학술연구조성비 지원에 의하여 연구되었음.

¹교신저자: (156-756) 서울특별시 동작구 흑석동 221, 중앙대학교 경영경제대학 응용통계학과, 교수.

E-mail: gogators@cau.ac.kr