Journal of the Korean Data & Information Science Society
2010, **21**(6), 1181–1190

한국데이터정보과학회지

확장된 뷸만-스트라웁 모형에서 신뢰도 추정 연구[†]

이민정[1] · 고한나[2] · 최승경[3] · 이의용[4]

[1234]숙명여자대학교 통계학과

접수 2010년 9월 25일, 수정 2010년 11월 18일, 게재확정 2010년 11월 23일

요 약

보험회사가 보험계약자들의 적절한 보험료를 산출하는 것은 회사나 계약자들 모두에게 중요한 문제이다. 보험료를 결정하는 방법에는 보험회사가 보유하고 있는 매뉴얼에 근거하여 계약자들마다 같은 보험료를 부과하는 방법, 보험계약자 개인이나 집단에 관한 과거 기록에 바탕을 두어 보험료를 차등하여 정하는 방법, 그리고 위 두 가지 방법을 조합하여 보험료를 산정하는 방법 등이 있다. 이중 통계적인 모형에 기반을 둔 보험료 산정 방법에는 최대 정확도 신뢰도 이론이 있고, 이에 바탕을 둔 구체적인 접근법에는 뷸만 모형과 뷸만-스트라웁 모형 등이 있다. 본 논문에서는, 뷸만-스트라웁 모형을 확장하여, 자료 수가 늘어나면서 자료 평균의 변동성이 기대한 것보다 커지는 Hewitt의 아이디어를 반영한 모형에서, 새로운 비모수적 보험료 추정법을 제안한다. 또, 제안된 추정법을 모의자료에 적용해보고, 기존의 뷸만-스트라웁 추정량과 비교해본다.

주요용어: 뷸만 모형, 뷸만-스트라웁 모형, 신뢰도 보험료, 최대 정확도 신뢰도 이론.

1. 서론

보험이라는 제도는 다수의 동질적인 위험 집단을 결합하여 대수의 법칙을 적용시켜 손실의 통계적 예측을 가능하게 하며 다수가 지불한 적은 금액으로 대규모 손실을 부담할 수 있도록 하는 제도이다. 다수의 동질적인 위험집단으로 분류하는 과정에서 대수의 법칙을 적용시킬 수 없을 정도의 자료의 수가 적을 경우, 보험료를 결정하는데 신뢰성을 확보하기 위해 우리는 신뢰도 이론을 적용한다.

보험회사에서 보험료를 산출하는 방법에는 크게 세 가지가 있다. 보험회사가 과거 비슷한 보험계약들에서 산정했던 매뉴얼 (manual)보험료 μ를 이용하는 방법, 보험회사가 보유하고 있는 보험계약자 개인이나 집단에 관한 과거기록 $X_1, X_2, \cdots, X_n$의 평균 $\bar{X}$에 바탕을 두어 보험료를 정하는 방법, 그리고 위 두 가지 방법을 조합하여 보험료를 결정하는 방법이 있다. 이러한 상황에서 매뉴얼 보험료와 보험계약자의 과거 기록 중 어느 것에 더 의존을 해야 하는지를 결정하는데 신뢰도 이론을 사용한다. 신뢰도 이론에는 통계적 기반은 약하지만 아직도 널리 쓰이는 제한된 변동 신뢰도 이론 (limited fluctuation credibility theory)과 통계적 모형에 기초를 둔 최대 정확도 신뢰도 이론 (greatest accuracy credibility theory)이 있다.

[†] 본 연구는 숙명여자대학교 2009학년도 교내연구비 지원에 의해 수행되었음.
[1] (140-742) 서울특별시 용산구 청파동 2가, 숙명여자대학교 통계학과, 석사.
[2] (140-742) 서울특별시 용산구 청파동 2가, 숙명여자대학교 통계학과, 석사과정.
[3] (140-742) 서울특별시 용산구 청파동 2가, 숙명여자대학교 통계학과, 박사.
[4] 교신저자: (140-742) 서울특별시 용산구 청파동 2가, 숙명여자대학교 통계학과, 교수.
E-mail: eylee@sookmyung.ac.kr

최대 정확도 신뢰도 이론은 베이지안 (Bayesian) 추정법에 근거를 둔 신뢰도 이론으로 Bailey (1950), Buhlmann (1967, 1970) 등에 의해 소개되었다. 즉, 보험계약자들은 서로 다른 신뢰도인 위험 모수 (risk parameter) θ를 갖고, θ에 대한 사전 정보는 $\pi(\theta)$라는 확률분포로 주어진다. 그리고 θ가 주어졌을 때, 보험계약자의 보험청구 횟수나 보험청구 금액 등으로 표현되는 보험계약자의 과거기록 X_j는 조건부 분포인 $f_{X|\Theta}(x|\theta)$를 따른다고 가정한다.

최근 국내에서도 보험수리학 분야의 연구가 활발해지고 있다. 이창수 (1997)는 보험료를 결정하는데 신뢰도 기법을 이용하는 선행 연구를 수행하였다. 강중철과 정세창 (2009)은 생명보험 계약자의 보험회사에 대한 만족도, 신뢰도, 충성도에 영향을 주는 요인들을 요인분석과 회귀분석을 사용해서 찾아보았고, 고봉성 등 (2009)은 로지스틱 회귀모형을 이용하여 생명보험사 텔레마케팅을 지원하는 모형을 만들고 효율성을 연구하였다. Shim 등 (2009)은 커널 머신 (kernal machine)을 통해 기존의 신뢰도 추정모형을 변형한 비선형 신뢰도 회귀모형을 제시하였고, 현정민과 차지환 (2010)은 종신보험의 운영에서 예정 수명분포와 예정 이자율의 영향력을 분석하고 비교하였다.

본 논문에서는 최대 정확도 신뢰도 이론 중 뷸만 (Buhlmann) 모형과 뷸만-스트라웁 (Buhlmann-Straub) 모형을 간단히 소개하고, 뷸만-스트라웁 모형을 확장하여, 자료의 수가 늘어나면서 자료 평균의 분산이 독립성 가정하에 기대되는 양보다 커진다 (Hewitt, 1967)는 사실을 반영한 모형에서 새로운 비모수적 보험료 추정법을 제안한다. 또, 제안된 추정법을 모의자료에 적용해보고, 기존의 뷸만-스트라웁 추정량과 비교해본다.

2. 최대 정확도 신뢰도 이론

2.1. 뷸만 모형

뷸만 모형은 통계적인 모형에 기반을 둔 신뢰도 이론 중에서 비교적 단순한 보험료 산정 모형으로 각 보험계약자의 과거 기록 $X_1, X_2, \cdots, X_n$은 $\Theta = \theta$가 주어져 있다는 조건하에 서로 조건부 독립이고, 동일한 분포 $f_{X|\Theta}(x|\theta)$를 따른다고 가정한다. X_j의 조건부 기댓값과 조건부 분산인 가설 평균 (hypothetical mean)과 처리 분산 (process variance)은 각각 $\mu(\theta) = E(X_j|\Theta = \theta)$, $v(\theta) = Var(X_j|\Theta = \theta)$이고, $\mu(\Theta)$의 기댓값과 분산은 $\mu = E[\mu(\Theta)]$와 $a = Var[\mu(\Theta)]$이다. 그리고 $v(\Theta)$의 기댓값은 $v = E[v(\Theta)]$이다. 모든 Θ를 고려한 X_j의 기댓값과 분산은 각각 $E(X_j) = \mu$, $Var(X_j) = v + a$이다.

위의 값들을 이용한 뷸만 신뢰도 보험료 (Buhlmann credibility premium)는

$$P_c = Z\bar{X} + (1 - Z)\mu$$

로 주어진다. 여기서 $Z = \frac{n}{n+k} (k = \frac{v}{a})$는 뷸만 신뢰도 요소, μ는 매뉴얼 보험료 (pure 보험료)에 해당한다.

μ, a, v의 추정을 위해 r명의 보험가입자의 n년 동안의 과거 기록이 있다고 가정하자. 즉, X_{ij}는 i번째 보험계약자가 j년에 청구한 보상액을 나타내는 자료이다. 이를 바탕으로 한 뷸만 신뢰도 모형의 모수인 μ, a, v의 비모수적인 불편추정량은 아래와 같다.

μ의 불편추정량: $\hat{\mu} = \bar{X} = \frac{1}{rn}\sum_{i=1}^{r}\sum_{j=1}^{n} X_{ij}$
v의 불편추정량: $\hat{v} = \frac{1}{r(n-1)}\sum_{i=1}^{r}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2$
a의 불편추정량: $\hat{a} = \frac{1}{r-1}\sum_{i=1}^{r}(\bar{X}_i - \bar{X})^2 - \frac{1}{rn(n-1)}\sum_{i=1}^{r}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2$.

위의 추정량을 사용하여 구하는 보험계약자 i의 뷸만 신뢰도 보험료는 아래와 같다.

$$\widehat{P}_{c(i)} = \hat{Z}\bar{X}_i + (1-\hat{Z})\hat{\mu}, \quad \hat{Z} = \frac{n}{n+\hat{k}}, \quad \hat{k} = \frac{\hat{v}}{\hat{a}}.$$

2.2. 뷸만 - 스트라웁 모형

뷸만-스트라웁 모형은 뷸만 모형을 일반화시킨 모형으로 보험계약자의 과거 보험 가입 기간 동안의 보상청구 횟수 등 보험계약자 개인이 갖고 있는 과거 경험의 수 (크기)를 고려한 모형이다. 뷸만-스트라웁 모형에서는 보험계약자의 과거 기록 X_j가 j번째 년도에서의 평균 보험보상금액이 된다. 각 보상금액은 $\Theta = \theta$가 주어져 있다는 조건하에서 조건부 독립이며 동일한 분포를 따른다고 가정한다. 그리고 이때의 조건부 기댓값과 조건부 분산은 아래와 같다.

$$E(X_j|\Theta=\theta) = \mu(\theta), \quad Var(X_j|\Theta=\theta) = \frac{v(\theta)}{m_j}.$$

여기서 m_j는 j번째 년도에서의 보상건수이며, 보험계약자의 n년 동안의 총 보상건수는 $m = \sum_{j=1}^{n} m_j$이다. 뷸만 모형에서와 마찬가지로, $\mu(\Theta)$의 기댓값과 분산은 각각 $\mu = E[\mu(\Theta)]$, $a = Var[\mu(\Theta)]$이고, $v(\Theta)$의 기댓값은 $v = E[v(\Theta)]$이다. 뷸만 모형은 m_j가 1로 모두 동일한 경우이고, 뷸만- 스트라웁 모형의 특수한 경우가 된다. 모든 Θ를 고려한 X_j의 기댓값과 분산은 $E(X_j) = \mu, Var(X_j) = \frac{v}{m_j} + a$이다.

뷸만-스트라웁 신뢰도 보험료 (Buhlmann-Straub credibility premium)는

$$P_c = Z\bar{X} + (1-Z)\mu$$

이다. 여기서 $Z = \frac{m}{m+k}$ $(k = \frac{v}{a})$는 뷸만-스트라웁 신뢰도 요소로 m의 값에 따라 달라지며, $\bar{X} = \frac{1}{m}\sum_{j=1}^{n} m_j X_j$으로 X_j의 가중평균이다.

μ, a, v의 추정을 위해 r명의 보험가입자의 과거 기록이 있다고 가정하자. 즉, i번째 보험계약자의 n_i년 동안의 과거 기록이 있고, 이 보험계약자 i의 j년에서의 보상건수가 m_{ij}이고, X_{ij}는 i번째 보험계약자가 j년에 m_{ij}번에 걸쳐 청구한 평균보상액이라 하자. i번째 보험가입자의 보상건수의 총합은 $m_i = \sum_{j=1}^{n_i} m_{ij}$이고, 전체 보상건수는 $m = \sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}$이다.

뷸만-스트라웁 신뢰도 보험료를 구하기 위해 필요한 μ, a, v의 비모수적인 불편추정량은 아래와 같다.

μ의 불편추정량: $\hat{\mu} = \bar{X} = \frac{1}{m}\sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}X_{ij}$
v의 불편추정량: $\hat{v} = (\sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2)/(\sum_{i=1}^{r}(n_i-1))$
a의 불편추정량: $\hat{a} = (m - \frac{1}{m}\sum_{i=1}^{r} m_i^2)^{-1}[\sum_{i=1}^{r} m_i(\bar{X}_i-\bar{X})^2 - \hat{v}(r-1)]$.

위의 추정량을 사용하여 구하는 보험계약자의 뷸만-스트라웁 신뢰도 보험료는 아래와 같다.

$$\widehat{P}_{c(i)} = \hat{Z}_i\bar{X}_i + (1-\hat{Z}_i)\hat{\mu}, \quad \hat{Z}_i = \frac{m_i}{m_i+\hat{k}}, \hat{k} = \frac{\hat{v}}{\hat{a}}.$$

뷸만 모형과 뷸만-스트라웁 모형에 대한 자세한 설명은 Klugman 등 (2004)에 잘 나와 있다.

3. 확장된 뷸만-스트라웁 모형

확장된 뷸만-스트라웁 모형은 뷸만-스트라웁 모형에 Hewitt (1967)의 아이디어인 자료의 수가 많아질수록 자료군의 평균의 분산은 독립적인 가정 하에 기대되는 평균의 분산보다 커지는 것을 반영하여 만

든 모형이다. 이 모형 (Klugman 등, 2004)에서 X_{ij}는 i번째 보험계약자의 j번째 년도에서 m_{ij}개의 과거 기록들의 평균이다. X_{ij}에 대해서 $\Theta = \theta_i$가 주어진 상황에서 조건부 기댓값과 조건부 분산은 각각 다음과 같이 정의된다.

$$E(X_{ij}|\Theta = \theta_i) = \mu(\theta_i), Var(X_{ij}|\Theta = \theta_i) = \frac{v(\theta_i)}{m_{ij}} + w(\theta_i).$$

여기서 $w(\theta_i)$는 Hewitt (1967)의 아이디어를 반영하기 위하여 분산에 추가되는 항이다. 즉, 이 모형에서는, 자료 수가 $m_{ij} + m_{ik}$인 자료군의 평균의 분산은 $\frac{v(\theta_i)}{m_{ij}+m_{ik}} + w(\theta_i)$로서, 자료 수가 각각 m_{ij}와 m_{ik}인 두 독립된 자료군을 합친 자료의 평균인 $(m_{ij}X_{ij}+m_{ik}X_{ik})/(m_{ij}+m_{ik})$의 분산 $(v(\theta_i))/(m_{ij}+m_{ik}) + (m_{ij}^2 + m_{ik}^2)/((m_{ij} + m_{ik})^2)w(\theta_i)$보다 커진다.

또한 모든 Θ를 고려한 X_{ij}의 기댓값과 분산, $v(\Theta)$의 기댓값은 각각 $\mu = E[\mu(\Theta)], a = Var[\mu(\Theta)], v = E[v(\Theta)]$이다. 확장된 뷸만-스트라웁 모형에서 분산에 추가한 값인 $w(\Theta)$의 기댓값은 $w = E[w(\Theta)]$이고, 보험계약자 i의 과거 기록에 대한 평균은 $\bar{X}_i = \frac{1}{m_i}\sum_{j=1}^{n_i} m_{ij}X_{ij}$이고, 전체 모든 보험계약자의 과거기록의 평균은 $\bar{X} = \frac{1}{m}\sum_{i=1}^{r} m_i\bar{X}_i$으로 정의한다. 여기서 $m = \sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}$이고, $m_i = \sum_{j=1}^{n_i} m_{ij}$이다.

보험계약자의 확장된 뷸만-스트라웁 신뢰도 보험료는

$$P_{c(i)} = Z_i\bar{X}_i + (1 - Z_i)\mu$$

이다. 여기서 $Z_i = \frac{am_i^*}{1+m_i^*}, m_i^* = \sum_{j=1}^{n_i} \frac{m_{ij}}{v+wm_{ij}},\ i = 1, 2, \cdots, r$이다.

지금부터 확장된 뷸만-스트라웁 신뢰도 보험료를 구하기 위한 모수들의 새로운 비모수적 불편추정량을 아래와 같이 제안한다.

(1) μ의 불편추정량

X_{ij}의 기댓값은 $E(X_{ij}) = E[E(X_{ij}|\Theta)] = E[\mu(\Theta)] = \mu$이고, 보험계약자의 과거 기록의 평균 $\bar{X}_i$의 조건부 기댓값과 기댓값은 각각 $E(\bar{X}_i|\Theta) = \sum_{j=1}^{n_i}\frac{m_{ij}}{m_i}\mu(\Theta) = \mu(\Theta)$, $E(\bar{X}_i) = E[E(\bar{X}_i|\Theta)] = E[\mu(\Theta)] = \mu$이다. 또한 모든 보험계약자에 대한 전체평균의 기댓값은 $E(\bar{X}) = \frac{1}{m}\sum_{i=1}^{r} m_iE(\bar{X}_i) = \frac{1}{m}\sum_{i=1}^{r} m_i\mu = \mu$를 만족한다. 그러므로 μ의 불편추정량은 $\bar{X}$이다. □

w, v, a의 추정량을 구하기 위해 모든 Θ에 대한 X_{ij}의 분산을 정리하면

$$\begin{aligned} Var(X_{ij}) &= E[Var(X_{ij}|\Theta)] + Var[E(X_{ij}|\Theta)] \\ &= E[w(\Theta) + \frac{v(\Theta)}{m_{ij}}] + Var[\mu(\Theta)] = w + \frac{v}{m_{ij}} + a \end{aligned}$$

이고, 이때 보험계약자 i의 과거 기록 평균의 분산은

$$\begin{aligned} Var(\bar{X}_i) &= E[Var(\bar{X}_i|\Theta)] + Var[E(\bar{X}_i|\Theta)] \\ &= Var[\mu(\Theta)] + E[\sum_{j=1}^{n_i}(\frac{m_{ij}}{m_i})^2w(\Theta) + \frac{v(\Theta)}{m_i}] = a + \sum_{j=1}^{n_i}(\frac{m_{ij}}{m_i})^2w + \frac{v}{m_i} \end{aligned}$$

이다. 또한 $\Theta = \theta_i$로 주어졌을 때의 $\bar{X}_i$의 조건부 분산은

$$Var(\bar{X}_i|\Theta = \theta_i) = \sum_{j=1}^{n_i}(\frac{m_{ij}}{m_i})^2Var(X_{ij}|\theta_i) = \sum_{j=1}^{n_i}(\frac{m_{ij}}{m_i})^2[w(\theta_i) + \frac{v(\theta_i)}{m_{ij}}]$$

이다.

(2) w의 불편추정량

w의 추정량을 구하기 위해서 $\hat{v}_i = \frac{1}{n_i-1}\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2$로 놓는다. 여기서 $\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2$의 기댓값은 아래와 같다.

$$E[\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2] = (n_i-1)v + (m_i - \frac{1}{m_i}\sum_{j=1}^{n_i} m_{ij}^2)w.$$

그러므로

$$E[\hat{v}_i] = E[\frac{1}{n_i-1}\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2] = \frac{1}{n_i-1}[(n_i-1)v + (m_i - \frac{1}{m_i}\sum_{j=1}^{n_i} m_{ij}^2)w]$$

이다. 이를 이용하여 $\hat{v}_i$와 $\hat{v}_k$차이의 기댓값을 구해보면

$$\begin{aligned} &E[\frac{1}{n_i-1}\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2 - \frac{1}{n_k-1}\sum_{j=1}^{n_k} m_{kj}(X_{kj}-\bar{X}_k)^2] \\ &= [\frac{1}{n_i-1}(m_i - \frac{1}{m_i}\sum_{j=1}^{n_i} m_{ij}^2) - \frac{1}{n_k-1}(m_k - \frac{1}{m_k}\sum_{j=1}^{n_k} m_{kj}^2)]w \end{aligned}$$

이므로, 이를 w에 대하여 정리하면 아래와 같다.

$$w = \frac{E[\frac{1}{n_i-1}\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2 - \frac{1}{n_k-1}\sum_{j=1}^{n_k} m_{kj}(X_{kj}-\bar{X}_k)^2]}{[\frac{1}{n_i-1}(m_i - \frac{1}{m_i}\sum_{j=1}^{n_i} m_{ij}^2) - \frac{1}{n_k-1}(m_k - \frac{1}{m_k}\sum_{j=1}^{n_k} m_{kj}^2)]}.$$

이때

$$S_{ik} = \frac{(n_k-1)[\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2] - (n_i-1)[\sum_{j=1}^{n_k} m_{kj}(X_{kj}-\bar{X}_k)^2]}{(n_k-1)[m_i - \frac{1}{m_i}\sum_{j=1}^{n_i} m_{ij}^2] - (n_i-1)[m_k - \frac{1}{m_k}\sum_{j=1}^{n_k} m_{kj}^2]}$$

라 하면 $E[S_{ik}] = w(1 \le i \ne k \le r)$이다. w를 추정하는데 ${}_rC_2$개의 S_{ik}의 평균을 이용하는 것이 추정량의 분산이 작아지므로 더 바람직하다. S_{ik}의 평균을 이용한 w의 추정량은

$$\hat{w} = \sum_{1 \le i \ne j \le r} \frac{S_{ik}}{{}_rC_2}$$

이다. □

(3) v의 불편추정량

v의 추정을 위하여 $E[\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2]$의 합을 이용한다. 즉,

$$\sum_{i=1}^{r} E[\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2] = \sum_{i=1}^{r}(n_i-1)v + (m_i - \frac{1}{m_i}\sum_{j=1}^{n_i} m_{ij}^2)w$$
$$= (N-r)v + (m - \sum_{i=1}^{r}\sum_{j=1}^{n_i}\frac{m_{ij}^2}{m_i})w$$

이고, 여기서 $N=\sum_{i=1}^{r} n_i$이다. 이를 v에 대하여 나타내면 v의 추정량은 아래와 같다.

$$\hat{v} = \frac{\sum_{i=1}^{r}[\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2] - \hat{w}(m - \sum_{i=1}^{r}\sum_{j=1}^{n_i}\frac{m_{ij}^2}{m_i})}{N-r}.$$

□

(4) a의 불편추정량

a의 추정을 위하여 $\sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X})^2$의 기댓값을 고려하면

$$E[\sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X})^2] = \sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}Var(X_{ij}) - mVar(\bar{X})$$
$$= (N-1)v + (m - \frac{1}{m}\sum_{i=1}^{r} m_i^2)a + (m - \frac{1}{m}\sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}^2)w$$

이다. 위 식을 a에 대하여 정리하면 a의 추정량은

$$\hat{a} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X})^2] - (N-1)\hat{v} - (m - \frac{1}{m}\sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}^2)\hat{w}}{m - \frac{1}{m}\sum_{i=1}^{r} m_i^2}$$

이다. □

위에서 추정한 추정량을 사용하여 각 보험계약자에 따른 확장된 뷸만-스트라웁 신뢰도 보험료는 아래와 같다.

$$\text{신뢰도 보험료: } \widehat{P_{c(i)}} = \hat{Z}_i\bar{X}_i + (1-\hat{Z}_i)\hat{\mu}$$
$$\hat{Z}_i = \frac{\hat{a}\hat{m_i^*}}{1+\hat{a}\hat{m_i^*}}, \quad \hat{m_i^*} = \sum_{j=1}^{n_i}\frac{m_{ij}}{\hat{v}+\hat{w}m_{ij}}.$$

4. 모의자료에의 적용

확장된 뷸만-스트라웁 모형에서 새로이 제안된 비모수적 불편추정량을 자료에 적용하여, 신뢰도 보험료를 구해보기 위하여 아래와 같이 모의자료를 만들었다. 보험계약자i의 과거 j년도의 기록 X_{ij}는 Hewitt (1967)의 아이디어를 반영하기 위하여 X_{ijk} ($X_{ijk} \sim exp(\theta_i)$)의 연도별 평균에 Y_i ($Y_i \sim exp(2\theta_i)$)를 합하여 만들었다. 여기서 Θ는 $\alpha = 2, \beta = 0.01$인 감마분포를 따른다. 그리고 X_{ijk}는 Θ의 값을 30개 추출한 뒤에 θ_i별로 n_i를 5와 10 사이에서 랜덤 추출하고, 각 n_i별로 m_{ij}를 1과 10사이에서 랜덤 추출하여, $n_i \times m_{ij}$개만큼 추출하였다. 즉, 30명의 보험계약자의 위험 모수 Θ는 감마분포를

따르며, 각 보험계약자 별로 5와 10 사이의 값인 n_i년만큼의 계약기간을 갖고 있으며, j번째 년에서 1과 10사이의 값인 m_{ij}번의 보험청구 횟수를 과거기록으로 갖는다.

모의 자료에 본 논문에서의 제안한 확장된 모형과 기존의 뷸만-스트라웁 모형을 각각 적용하여 구한 μ, a, v, w의 추정량은 표 4.1에 주어져있고, 이때의 신뢰도요소 Z의 값과 신뢰도 보험료는 표 4.2에 주어져있다.

표 4.1 모수의 추정량

	확장된 뷸만-스트라웁 모형의 추정량	뷸만-스트라웁 모형의 추정량
$\hat{\mu}$	108.7045	108.7035
$\hat{a}$	6006.892	6547.317
$\hat{v}$	3826.877	10277.26
$\hat{w}$	3087.113	

표 4.2의 결과를 보면 본 연구에서 제시한 추정량으로 구한 Z의 값이 뷸만-스트라웁의 추정량으로 구한 Z값에 비해서 작은 것을 확인할 수 있다. 이는 확장된 뷸만-스트라웁의 모형에서 보험계약자의 과거 경험인 X_{ij} (m_{ij}개의 평균)의 분산이 자료의 수에 따라 독립이라는 가정 하에 기대되는 양보다 더 커지는 Hewitt의 아이디어를 반영해주었기 때문이다. 즉, 확장된 모형에서는 자료의 수 m_{ij}가 아무리 커져도 자료 평균인 X_{ij}의 분산이 0으로 수렴하지 않는다. 이와 같이, 과거 경험인 X_{ij}의 분산이 기대되는 양보다 커지는 것은 X_{ij}에 대한 불확실성이 커진다는 것이므로 X_{ij}에 대한 가중치인 Z의 값이 확장된 뷸만-스트라웁 모형에서 기존의 뷸만-스트라웁 모형보다 작아지는 것을 확인할 수 있다. 기존의 뷸만-스트라웁의 추정방법은 불확실한 과거 경험에 대한 가중치가 커짐으로 인해 보험회사의 손실로 연결될 수 있다. 반면에 논 본문에서 제안한 추정량을 사용한 확장된 모형에서는 불확실한 것에 대한 가중치를 적게 부여함으로써 신뢰도 보험료를 계산하는데 불확실한 과거 경험의 영향력이 적어진다.

표 4.2 신뢰도 보험료의 추정량

		확장된 뷸만-스트라웁 모형		뷸만-스트라웁 모형	
보험계약자	$\bar{X}_i$	$\hat{Z}_i$	$\hat{P}_c$	$\hat{Z}_i$	$\hat{P}_c$
1	94.643	0.883	96.295	0.945	95.416
2	32.836	0.883	41.726	0.941	37.318
3	170.015	0.934	165.987	0.973	168.372
4	24.148	0.883	34.017	0.956	27.880
5	77.520	0.920	80.007	0.966	78.571
6	51.097	0.908	56.405	0.966	53.081
7	216.187	0.938	209.552	0.973	213.257
8	181.091	0.873	171.878	0.927	175.824
9	6.399	0.923	41.937	0.968	38.736
10	131.361	0.940	129.999	0.974	130.764
11	238.617	0.888	224.115	0.958	233.189
12	367.271	0.906	342.841	0.958	356.468
13	152.127	0.934	149.256	0.972	150.901
14	135.827	0.926	133.815	0.957	134.663
15	94.570	0.881	96.254	0.936	95.473
16	186.963	0.932	181.676	0.970	184.626
17	232.545	0.924	223.166	0.960	227.632
18	111.958	0.882	111.574	0.945	111.779
19	65.216	0.895	69.773	0.933	68.113
20	49.265	0.925	53.741	0.966	51.312
21	21.765	0.892	31.176	0.959	25.303

표4.2 (계속) 신뢰도 보험료의 추정량

		확장된 뷸만-스트라움 모형		뷸만-스트라움 모형	
보험계약자	$\bar{X}_i$	$\hat{Z}_i$	$\hat{P}_c$	$\hat{Z}_i$	$\hat{P}_c$
22	32.624	0.936	37.475	0.971	34.813
23	31.334	0.906	38.645	0.959	34.482
24	94.313	0.929	95.336	0.966	94.798
25	33.609	0.876	42.937	0.924	39.339
26	140.235	0.876	136.334	0.943	138.440
27	22.694	0.926	29.066	0.968	25.474
28	32.140	0.931	37.425	0.966	34.721
29	71.959	0.922	74.824	0.962	73.347
30	76.084	0.936	78.1645	0.970	77.077

표 4.3 m_{ij}에 따른 신뢰도 요소의 값

			확장된 뷸만-스트라움 모형	뷸만-스트라움모형
	m_{ij}	$\bar{X}_i$	$\hat{Z}_i$	$\hat{Z}_i$
보험계약자1	1	264.721	0.579	0.821
보험계약자2	1	22.554	0.508	0.775
보험계약자3	1	17.540	0.463	0.742
보험계약자4	1	106.237	0.584	0.838
보험계약자5	1	85.720	0.546	0.801
보험계약자6	1	117.482	0.546	0.889
보험계약자7	1	41.111	0.579	0.958
보험계약자8	1	128.339	0.608	0.912
보험계약자9	1	135.268	0.632	0.920
보험계약자10	1	42.112	0.463	0.852
$\hat{Z}_i$ 합			5.509	8.509
보험계약자11	2	37.284	0.508	0.870
보험계약자12	2	9.884	0.579	0.932
보험계약자13	2	64.290	0.463	0.896
보험계약자14	2	21.139	0.546	0.823
보험계약자15	2	74.991	0.608	0.939
보험계약자16	2	14.859	0.546	0.941
보험계약자17	2	119.916	0.579	0.948
보험계약자18	2	17.649	0.632	0.908
보험계약자19	2	21.680	0.608	0.954
보험계약자20	2	36.027	0.463	0.820
$\hat{Z}_i$ 합			5.532	9.034

표 4.3은 m_{ij}가 신뢰도 요소 Z_i의 값에 영향을 주는 것을 알아보기 위한 것으로 m_{ij}를 1, 2, 3으로 부여하였다. 표 4.3의 결과를 보면 m_{ij}의 값이 커짐에 따라 Z_i의 값이 대체적으로 더 커지는 경향이 있음을 알 수 있다. 이 특징을 조금 더 쉽게 알아보기 위해 m_{ij}가 같은 값을 갖는 보험계약자들의 Z_i의 값의 합을 구해보았다. 뷸만-스트라움 모형에서는 분산이 m_{ij}만의 영향을 받아서 m_{ij}가 증가함에 따라 빠르게 감소하고 그로 인하여 Z_i가 빠르게 증가한다. 하지만 확장된 뷸만-스트라움 모형에서는 분산에 추가된 $w(\theta_i)$에 의하여 m_{ij}가 증가함에 따라 감소하는 분산의 양이 뷸만-스트라움 모형보다 상대적으로 적고, 따라서 m_{ij}가 증가함에 따라 증가하는 Z_i의 양도 상대적으로 적다는 것을 알 수 있다. 이를 통하여 확장된 뷸만-스트라움 모형이 자료의 수가 많아질수록 자료군의 평균의 분산이 독립적인 가정 하에 기대되는 평균의 분산보다 더 커진다는 Hewitt의 아이디어를 반영하고 있음을 다시 한번 확인할 수 있다.

표 **4.3** (계속) m_{ij}에 따른 신뢰도 요소의 값

			확장된 불만-스트라웁 모형	불만-스트라웁모형
	m_{ij}	$\bar{X}_i$	$\hat{Z}_i$	$\hat{Z}_i$
보험계약자21	3	64.575	0.632	0.966
보험계약자22	3	39.085	0.563	0.990
보험계약자23	3	77.531	0.579	0.958
보험계약자24	3	74.378	0.608	0.963
보험계약자25	3	15.584	0.508	0.945
보험계약자26	3	16.961	0.508	0.954
보험계약자27	3	31.975	0.579	0.965
보험계약자28	3	16.254	0.546	0.960
보험계약자29	3	268.086	0.608	0.969
보험계약자30	3	44.010	0.463	0.945
$\hat{Z}_i$ 합			5.594	9.616

참고문헌

강중철, 정세창 (2009). 생명보험회사에 대한 만족도, 신뢰도, 충성도에 영향을 미치는 요인 분석. <한국데이터정보과학회지>, **20**, 713-717.

고봉성, 이석원, 허정 (2009). 생명보험사 텔리마케팅 효율성 제고에 관한 연구. <한국데이터정보과학회지>, **20**, 673-684.

이창수 (1997). 신뢰도기법을 이용한 자료의 충분성 평가와 보험요율의 조정. <보험개발연구>, **21**, 109-125.

현정민, 차지환 (2010). 종신보험에서의 영향 변수의 영향력 분석에 관한 연구. <한국데이터정보과학회지>, **21**, 71-86.

Bailey, A. (1950). Credibility procedures. *Proceedings of the Casualty Actuarial Society*, **37**, 7-23 and 94-115.

Buhlmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin*, **4**, 199-207.

Buhlmann, H. (1970). *Mathemtical methods in risk theory*, Springer-Verlag, New York.

Hewitt, C. Jr. (1967). Loss ratio distributions-a model. *Proceedings of the Casualty Actuarial Society*, **54**, 70-88.

Klugman, S. A., Panjer, H. and Willmot, G. E. (2004). *Loss models: From data to decision,* 2nd Ed., John Wiley & sons, Hoboken.

Shim, J., Kim, T. Y., Lee, S. and Hwang, C. (2009). Credibility estimation via kernel mixed effects model. *Journal of the Korean Data & Information Science Society*, **20**, 445-452.

Journal of the Korean Data & Information Science Society
2010, **21**(6), 1181–1190

한국데이터정보과학회지

A study on the estimation of the credibility in an extended Buhlmann-Straub model†

Min Jeong Yi[1] · Han Na Go[2] · Seung Kyoung Choi[3] · Eui Yong Lee[4]

[1234]Department of Statistics, Sookmyung Women's University

Received 25 September 2010, revised 18 November 2010, accepted 23 November 2010

Abstract

When an insurer develops an insurance product, it is very critical to determine reasonable premiums, which is directly related to insurer's profits. There are three methods to determine premiums. Frist, the insurer utilizes premiums paid to the similar cases to the current one. Second, the insurer calculates premiums based on policyholder's past records. The last method is to combine the first with the second one. Based on the three methods, there are two major theories determining premiums, Limited Fluctuation Credibility Theory not based on statistical models and Greatest Accuracy Credibility Theory based on statistical models. There are well-known methods derived from Greatest Accuracy Credibility Theory, such as, Buhlmann model and Buhlmann-Straub model. In this paper, we extend the Buhlmann-Straub model to accommodate the fact that variability grows according to the number of data in practice and suggest a new non-parametric method to estimate the premiums. The suggested estimation method is also applied to the data gained from simulation and compared with the existing estimation method.

Keywords: Buhlmann model, Buhlmann-Straub model, credibility premium, greatest accuracy credibility theory.

---

† This research was supported by the Sookmyung Women's University Research Grants 2009.
[1] Master, Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea.
[2] Master's course, Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea.
[3] Doctor, Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea.
[4] Corresponding author: Professor, Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea. E-mail: eylee@sookmyung.ac.kr