The Korean Journal of Applied Statistics (2014) DOI: http://dx.doi.org/10.5351/KJAS.2014.27.6.1017
27(6), 1017–1028

Bayes Risk Comparison for Non-Life Insurance Risk Estimation

Myung Joon Kim[a] · Ho Young Woo[b] · Yeong-Hwa Kim[c,1]

[a]Department of Business Statistics, Hannam University
[b]Department of Statistics, Graduate School of Chung-Ang University
[c]Department of Applied Statistics, Chung-Ang University

(Received August 12, 2014; Revised October 14, 2014; Accepted October 15, 2014)

Abstract

Well-known Bayes and empirical Bayes estimators have a disadvantage in respecting to overshink the parameter estimator error; therefore, a constrained Bayes estimator is suggested by matching the first two moments. Also traditional loss function such as mean square error loss function only considers the precision of estimation and to consider both precision and goodness of fit, balanced loss function is suggested. With these reasons, constrained Bayes estimators under balanced loss function is recommended for non-life insurance pricing.; however, most studies focus on the performance of estimation since Bayes risk of newly suggested estimators such as constrained Bayes and constrained empirical Bayes estimators under specific loss function is difficult to derive. This study compares the Bayes risk of several Bayes estimators under two different loss functions for estimating the risk in the auto insurance business and indicates the effectiveness of the newly suggested Bayes estimators with regards to Bayes risk perspective through auto insurance real data analysis.

Keywords: Bayes risk, insurance risk, balanced loss function, constrained Bayes estimator.

1. 서론

손해보험의 여러 영역 가운데 자동차 보험은 의무보험, 즉 자동차를 소유한 사람은 법적으로 가입을 반드시 해야한다는 특징과 단기보험, 즉 보험기간이 6개월 또는 1년 등 단기간으로 시작과 종료를 반복한다는 특징을 가지고 있다. 이러한 특징이 자동차 보험산업에서 시사하는 바는, 첫째 분석할 자료, 데이터의 양이 풍부하다는 것이다. 국내 자동차 등록대수가 1천 8백만대를 넘어서고 있다는 통계는 결국 의무보험이라는 특징으로 인하여 분석할 데이터의 양이 자동차의 수 만큼 확보된다는 점이며, 둘째로는 보험기간이 단기간에 종료된다는 것은 과거 데이터 분석을 통한 통계적 분석 결과를 적용하고, 이에 대한 적정성 검증을 보험기간 이후, 즉 단기간 내에 바로 검증할 수 있기 때문에 이를 바탕으로 분석 방식 및 적용 방식 등을 지속적으로 수정 보완할 수 있다는 점이다.

이러한 특성들로 인하여 통계적 이론의 접목 및 활용 방식에 대한 연구가 가장 활발히 진행되는 분야가 바로 자동차 보험이며, 그 중에서 보험업의 본질 분야인 위험도의 추정과 관련된 사항들이 통계적 분석

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

의 핵심이라고 할 수 있다. 이는 선진국에서 정기적으로 개최되는 손해보험 세미나에서 가격 산출, 즉 위험도 추정과 관련된 주제의 세미나(predictive modeling seminar)가 별도로 분리되어 운영된다는 것에서 그 중요성을 확인할 수 있으며, 국내에서도 위험도 추정의 한 주제인 신뢰도 적용방식과 관련한 연구가 Kim 등 (2011)에 의하여 제시되었다.

위험도의 추정은 보험상품에 대한 가격을 결정하게 되며, 이는 보험회사의 경쟁력 문제로 직결된다. 따라서 각 보험사는 이 분야에 있어 선진적인 통계기법의 도입과 적용에 매우 적극적이며 다양한 다변량 분석방식과 모델링 방식이 현재도 이미 활용되고 있는 실정이다. 현재 가장 널리 사용되는 가격모델 기법은 GLM(generalized linear model) 방식이며, 업계에서는 향후 베이지안 추론 기법의 적용으로 발전될 것으로 예상하고 있다. 이에 Kim과 Kim (2013, 2014)은 일반적인 베이즈 추정량(Bayes estimator; 이하 BE로 표기) 뿐만 아니라 베이즈 추정량의 단점을 극복하고자 Ghosh (1992)에 의해 제안된 constrained 베이즈 추정량(이하 CB로 표기)의 위험도 추정 및 활용방식에 대한 연구결과를 제시하였다. 또한 정도(precision)만을 고려하는 전통적인 제곱평균 손실함수의 문제점을 보완하고자 Zellner (1988, 1992)가 제안한 균형 손실함수 하에서의 활용 방안에 대한 연구 결과도 제안하였다.

보험상품의 원가를 결정하는 위험도라는 모수가 갖는 중요성으로 인하여 대부분의 연구는 모수 추정에 집중해 있는 이유로 인하여, 활용하고자 하는 베이즈 추정량에 대한 오차, 즉 베이즈 위험(Bayes risk)에 대한 추정 및 검증에 대한 연구는 미미한 수준이다. 그러나 고려 대상이 되는 여러 가지 손실함수 하에서 추정, 활용되는 다양한 베이즈 추정량들의 베이즈 위험이 정확한 수식으로 도출, 증명되지 않고 근사적(asymptotic) 형태를 나타내는 경우가 있기 때문에 새롭게 제안되는 적정 추정량들에 대한 오차 검증 작업은 필수 불가결한 요소라고도 할 수 있다. 근사적으로 증명된 오차로 인하여 모수 추정량의 불확실성의 증가가 가능하기 때문이다.

따라서 본 연구는 두가지 손실함수, 즉 제곱평균 손실함수와 균형 손실함수 하에서 여러 가지 베이즈 추정량들을 고려하고 이에 대한 베이즈 위험을 각각 추정 비교함으로써 모수의 추정뿐만 아니라 이에 대한 오차의 적정성을 검증하여 제시하고자 한다. 비교 대상이 되는 추정량은 널리 알려진 일반적인 베이즈 추정량(BE), 경험적 베이즈 추정량(empirical Bayes estimator; 이하 EB로 표기)과 CB(constrained Bayes estimator) 추정량, constrained empirical 베이즈 추정량(이하 CEB로 표기) 등이며, 이론적으로 도출 증명된 결과를 토대로 실제 보험사의 손실 금액 자료인 실증 자료를 분석하여 그 효용성을 증명해 보고자 한다.

2. 다양한 베이즈 추정량의 베이즈 위험

2.1. CB의 기본 개념

베이지안 기법은 모수의 동시 추정, 소지역 추정 등의 분야에서 이미 널리 알려진 이론이기는 하나 보험업계에서는 아직 활발한 적용과 활용이 되고 있지는 않다. 그러나 위험도 모수 추정에 대한 추정에 있어 베이지안 기법의 필요성과 향후 미래 적용 기법에 대한 당위성은 이미 잘 알려져 있다.

일정 보험기간을 주기로 반복적인 모수의 추정이 필요한 손해보험 분야에서는 위험도라는 위치 모수(location parameter)의 추정과 더불어 특정 사고로 인한 왜곡 현상 방지와 안정적인 위험도 모수 추정을 위하여 산포 모수(dispersion parameter)에 대한 고려가 필요하다. 이러한 요구 조건을 만족하는 베이즈 추정량이 바로 Ghosh (1992)에 의해 제안된 CB이며, 이후 다변량 공간에서의 해당 추정량에 대한 연구 결과도 Ghosh와 Kim (2002)에 의하여 증명되고 제안되었다.

기존에 알려져 있는 일반적인 베이즈 추정량(BE)은 베이즈 위험을 과다축소(overshrink)하는 문제점이

있음이 지적되어 왔으며, 이는 추정하고자 하는 모수를 θ_i, 베이즈 추정량을 θ_i^B 라고 할 경우 다음과 같이 증명된다.

$$\sum_{i=1}^{m}\left\{\theta_i^B(\boldsymbol{X})-\bar{\theta}^B(\boldsymbol{X})\right\}^2 < E\left\{\sum_{i=1}^{m}(\theta_i-\bar{\theta})^2\middle|\boldsymbol{X}\right\}.$$

이를 보완하고자 제안된 것이 CB이며, 기본 개념은 일차 적률(first moment)과 이차 적률(second moment)을 동시에 일치시키는 추정량을 찾고자 하는 것이며, 이를 수식으로 표현하면 다음과 같다.

$$(1)\ E\left(\bar{\theta}\middle|\boldsymbol{x}\right) = m^{-1}\sum_{i=1}^{m} t_i(\boldsymbol{x}) = \bar{t}(\boldsymbol{x}),$$

$$(2)\ E\left\{\sum_{i=1}^{m}(\theta_i-\bar{\theta})^2\middle|\boldsymbol{x}\right\} = \sum_{i=1}^{m}\left\{t_i(\boldsymbol{x})-\bar{t}(\boldsymbol{x})\right\}^2.$$

여기서 $\boldsymbol{t}(\boldsymbol{x}) = (t_1,\ldots,t_m)^T$는 위의 두 조건을 만족시키는 추정량을 의미하며 $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_m)^T$는 추정하고자 하는 모수 벡터(parameter vector)를 나타낸다. 따라서 두 가지 조건을 동시에 만족하면서 산출된 함수 $\boldsymbol{t}(\boldsymbol{x})$가 CB가 되며, 산출된 결과를 이용하여 베이즈 위험을 계산할 수 있다. 본 연구에서는 일반적인 베이즈 추정량 BE와 경험적 베이즈 추정량 EB뿐만 아니라 보험상품 가격산출에 적정한 것으로 제안된 추정량 CB와 이것의 베이즈 위험이 연구 대상이 된다.

2.2. 균형 손실함수(Balanced loss function)의 기본 개념

전통적으로 알려져 있는 손실함수인 제곱평균 손실함수는 추정한 모수의 정확도(precision)만을 고려하는 함수이다. 분석 상황에 따라서는 모수와 모수 추정량과의 거리 뿐만 아니라, 실제 데이터와 모수와 거리를 동시에 고려할 필요가 있다. 이러한 제곱평균 손실함수의 한계를 극복하고자 Zellner (1988, 1992)에 의해 제안된 손실함수가 균형 손실함수(balanced loss function)이다. 균형 손실함수는 제곱평균 손실함수가 다루고 있는 정도(precision) 외에 추가적으로 적합도(goodness of fit)를 동시에 고려하는 손실함수로서 다음과 같이 정의된다.

$$L(\boldsymbol{\theta},\boldsymbol{t}) = w\|\boldsymbol{X}-\boldsymbol{t}(\boldsymbol{X})\|^2 + (1-w)\|\boldsymbol{\theta}-\boldsymbol{t}(\boldsymbol{X})\|^2.$$

여기서 $\|\cdot\|$는 유클리디안 거리(norm)를 의미하며, w $(0 \le w \le 1)$는 알려져 있는 가중치(weight)로 가정한다. 주어진 손실함수에서 첫번째 항은 적합도(goodness of fit)를 고려하는 항이고, 두번째 항은 정도(precision)를 고려하는 항임을 알 수 있으며, 가중치의 정도에 따라 두가지 고려 사항 중 어떠한 것에 가중치를 더 부여할 것인지를 결정하게 되는 것이다. 참고로 가중치가 0이 되는 경우 첫번째 항이 고려되지 않는 제곱평균 손실함수가 된다.

본 연구에서는 전통적인 제곱평균 손실함수 하에서 뿐만 아니라, 앞서 살펴 본 균형 손실함수를 추가로 고려하여 추정된 추정량의 베이즈 위험을 도출하게 될 것이다. 또한 균형 손실함수에서는 가중치 w의 값에 의하여 오차의 정도가 변화될 수 있는 개연성이 있으므로 여러 가중치 값의 조건에 따른 베이즈 위험을 계산하여 비교함으로써, 실제 보험산업 현장에서 다양한 활용과 적용을 하는데 참고할 수 있도록 하고자 한다.

2.3. 제곱평균 손실함수 하에서의 베이즈 위험

먼저 전통적인 손실함수인 제곱오차 손실함수 하에서의 결과들을 살펴보기로 하며, 두개의 섹션으로 구

분하여 첫번째는 일반적인 베이즈 추정량 BE와 경험적 베이즈 추정량 EB에 대한 베이즈 위험을 확인하고, 두번째는 Ghosh (1992)에 의해 제안된 CB와 CEB에 대한 베이즈 위험을 확인한다.

2.3.1. BE와 EB의 베이즈 위험 베이즈 추정량 BE는 사전적 정보에 대한 활용을 가정하고 있다. 따라서 최대 우도 추정량(maximum likelihood estimator)과 사전 평균(prior mean)의 조합으로 이루어진다. 일반적으로 잘 알려져 있는 정규 분포를 가정하는 경우, 즉, $X_i|\theta_i$는 독립적인 $N(\theta_i, A), i = 1, \ldots, m$ 분포이고, 모수 θ_i는 $i.i.d.N(\mu, A)$ 분포가 되며, $X_i \sim i.i.d.N(\mu, 1+A)$이고, 사후 분포(posterior distribution)는 $\theta_i|X_i$는 $i.i.d.N((1-B)X_i + B\mu, 1-B)$로 도출된다. 여기서 B는 $(1+A)^{-1}$을 나타낸다. 따라서, 제곱평균 손실함수하에서의 BE를 $\hat{\boldsymbol{\theta}}^B$라 하면 $\hat{\boldsymbol{\theta}}^B$는 다음의 사후 평균(posterior mean; PM)과 같게 된다.

$$\hat{\boldsymbol{\theta}}^B = (1-B)\boldsymbol{X} + B\mu\mathbf{1}_m,$$

여기서 $\boldsymbol{X}$는 $(X_1, X_2, \ldots, X_m)^T$인 벡터를 의미하며 $\mathbf{1}_m$은 구성요소가 모두 1인 벡터를 의미한다. 이렇게 도출된 BE에 대한 베이즈 위험 BR$(\hat{\boldsymbol{\theta}}^B)$는 다음과 같이 도출된다.

$$\mathrm{BR}\left(\hat{\boldsymbol{\theta}}^B\right) = m^{-1}E\left\|\hat{\boldsymbol{\theta}}^B - \boldsymbol{\theta}\right\|^2 = \mathrm{Var}\left(\hat{\boldsymbol{\theta}}^{\mathrm{PM}}\right) = 1 - B.$$

EB를 구하기 위해서는 모수 중 알려져 있지 않은 μ와 A를 주변 분포(marginal distribution)로부터 추정해야 한다. $\boldsymbol{X} \sim N(\mu\mathbf{1}_m, B^{-1}\boldsymbol{I}_m)$이므로, $S = \sum_{i=1}^m (X_i - \overline{X})^2/(m-1)$로 가정하는 경우, $S \sim B^{-1}\chi^2_{m-1}/(m-1)$이 된다. 따라서 μ를 $\overline{X}$로 B를 $\widehat{B} = S^{-1}$로 추정하는 경우 EB를 $\hat{\boldsymbol{\theta}}^{\mathrm{EB}}$라 하면 $\hat{\boldsymbol{\theta}}^{\mathrm{EB}}$는 다음과 같으며

$$\hat{\boldsymbol{\theta}}^{\mathrm{EB}} = \left(1 - \widehat{B}\right)\boldsymbol{X} + \widehat{B}\overline{X}\mathbf{1}_m,$$

이에 대한 베이즈 위험 $BR(\hat{\boldsymbol{\theta}}^{EB})$는 다음과 같다.

$$\begin{aligned}
\mathrm{BR}\left(\hat{\boldsymbol{\theta}}^{\mathrm{EB}}\right) &= m^{-1}E\left\|\hat{\boldsymbol{\theta}}^{\mathrm{EB}} - \boldsymbol{\theta}\right\|^2 \\
&= m^{-1}E\left\|\hat{\boldsymbol{\theta}}^{\mathrm{EB}} - \hat{\boldsymbol{\theta}}^B\right\|^2 + m^{-1}E\left\|\hat{\boldsymbol{\theta}}^B - \boldsymbol{\theta}\right\|^2 \\
&= m^{-1}E\left\|\hat{\boldsymbol{\theta}}^{\mathrm{EB}} - \hat{\boldsymbol{\theta}}^B\right\|^2 + 1 - B = 1 - B + \frac{3B}{m} + o\left(m^{-1}\right).
\end{aligned}$$

위 수식의 첫째 항인 $m^{-1}E\|\hat{\boldsymbol{\theta}}^{\mathrm{EB}} - \hat{\boldsymbol{\theta}}^B\|^2$은 정확한 계산이 불가하여 Kim과 Kim (2005)에 의하여 근사적인 결과가 증명되어 제시되었으며, 이는 $1/m$보다 차수(order)가 높은 항들은 0으로 수렴됨을 가정한 결과이다.

2.3.2. CB와 CEB의 베이즈 위험 BE의 사후분포의 오차에 대한 과도한 축소(over shrinkness)로 인하여 제안된 CB는 Louis (1984)와 Ghosh (1992)에 의해 제안된 이후, 진화되어 왔으며 이를 $\hat{\boldsymbol{\theta}}^{\mathrm{CB}}$라 하면 제곱평균 손실함수하에서 다음과 같이 정의된다.

$$\begin{aligned}
\hat{\boldsymbol{\theta}}^{\mathrm{CB}} &= a_B\left[(1-B)\boldsymbol{X} + B\mu\mathbf{1}_m\right] + (1-a_B)\left[(1-B)\overline{X} + B\mu\right]\mathbf{1}_m \\
&= a_B(1-B)\left(\boldsymbol{X} - \overline{X}\mathbf{1}_m\right) + \left[(1-B)\overline{X} + B\mu\right]\mathbf{1}_m.
\end{aligned}$$

앞서 가정한 동일한 분포함수하에서 모수 μ와 B의 추정량을 이용하여 구해지는 CEB를 $\hat{\boldsymbol{\theta}}^{\mathrm{CEB}}$라 하면 $\hat{\boldsymbol{\theta}}^{\mathrm{CEB}}$는 다음과 같은 수식으로 정의할 수 있다. 다음 수식에서 a_B는 $a_B^2 = 1 + 1/(1-B)S$로 정의되는 함수이며, a_{EB}는 μ와 B의 추정량이 적용된 함수로 이해할 수 있다.

$$\hat{\boldsymbol{\theta}}^{\mathrm{CEB}} = a_{\mathrm{EB}}\left[\left(1-\widehat{B}\right)\boldsymbol{X} + \widehat{B}\overline{X}\mathbf{1}_m\right] + (1-a_{\mathrm{EB}})\overline{X}\mathbf{1}_m.$$

이와 같이 추정된 통계량 CB와 CEB에 대한 각각의 베이즈 위험 BR($\hat{\boldsymbol{\theta}}^{\mathrm{CB}}$)와 BR($\hat{\boldsymbol{\theta}}^{\mathrm{CEB}}$)는 Ghosh 등 (2004)의 연구 결과에 따라 각각 다음과 같이 표현할 수 있다.

$$\begin{aligned}
\mathrm{BR}\left(\hat{\boldsymbol{\theta}}^{\mathrm{CB}}\right) &= m^{-1}E\left\|\hat{\boldsymbol{\theta}}^{\mathrm{CB}} - \boldsymbol{\theta}\right\|^2 \\
&= m^{-1}E\left\|\hat{\boldsymbol{\theta}}^{\mathrm{CB}} - \hat{\boldsymbol{\theta}}^{B}\right\|^2 + m^{-1}E\left\|\hat{\boldsymbol{\theta}}^{B} - \boldsymbol{\theta}\right\|^2 \\
&= m^{-1}E\left\|\hat{\boldsymbol{\theta}}^{\mathrm{CB}} - \hat{\boldsymbol{\theta}}^{B}\right\|^2 + 1 - B = 1 - B + A_1(B) + m^{-1}[A_2(B) - A_1(B)] + o\left(m^{-1}\right) \\
\mathrm{BR}\left(\hat{\boldsymbol{\theta}}^{\mathrm{CEB}}\right) &= m^{-1}E\left\|\hat{\boldsymbol{\theta}}^{\mathrm{CEB}} - \boldsymbol{\theta}\right\|^2 \\
&= m^{-1}E\left\|\hat{\boldsymbol{\theta}}^{\mathrm{CEB}} - \hat{\boldsymbol{\theta}}^{B}\right\|^2 + m^{-1}E\left\|\hat{\boldsymbol{\theta}}^{B} - \boldsymbol{\theta}\right\|^2 \\
&= m^{-1}E\left\|\hat{\boldsymbol{\theta}}^{\mathrm{CEB}} - \hat{\boldsymbol{\theta}}^{B}\right\|^2 + 1 - B \\
&= 1 - B + A_1(B) + m^{-1}[B + A_3(B) - A_1(B)] + o\left(m^{-1}\right)
\end{aligned}$$

위 식에서 $A_1(B)$, $A_2(B)$, $A_3(B)$는 각각 $B^{-1}(1-B)[1-(1-B)^{1/2}]^2$, $B(1-B)^{3/2}/2$, $B(1-B)^{-1/2}/2$로 정의되며, 해당 추정량의 베이즈 위험은 근사적으로만 계산이 가능하다.

2.4. 균형 손실함수하에서의 베이즈 위험

정도(precision)와 적합도(goodness of fit)를 동시에 고려하는 균형 손실함수에서는 앞서 도출된 여러 베이즈 추정량이 다른 형태를 띠게 되며, 베이즈 위험도 균형 손실함수하에서 다르게 도출된다. 보험 상품의 위험도 측정에 있어 매년 검증, 확인되는 모수로의 수렴도 중요하지만 정책, 제도 등의 변경으로 인한 변화가 있을 경우 데이터 자체에 대한 오차도 고려할 필요가 있으므로 균형 손실함수를 고려하는 것은 보험업계에 의미있는 시도라고 할 수 있다.

2.4.1. BE와 EB의 베이즈 위험 균형 손실함수에서는 제곱평균 손실함수에서와는 달리 베이즈 추정량은 사후 평균으로 도출되지 않는다. 앞서 가정한 동일한 정규 분포하에서 균형 손실함수를 최소화하는 BE를 $\hat{\boldsymbol{\theta}}^{\mathrm{NB}}$라 하면 $\hat{\boldsymbol{\theta}}^{\mathrm{NB}}$는 다음과 같이 정의된다.

$$\begin{aligned}
\hat{\boldsymbol{\theta}}^{\mathrm{NB}} &= w\boldsymbol{X} + (1-w)\hat{\boldsymbol{\theta}}^{B} \\
&= w\boldsymbol{X} + (1-w)\left[(1-B)\boldsymbol{X} + B\mu\mathbf{1}_m\right] \\
&= [1-(1-w)B)]\boldsymbol{X} + (1-w)B\mu\mathbf{1}_m
\end{aligned}$$

위의 베이즈 추정량의 형태는 데이터 자체와 기존의 제곱평균 손실함수하에서 도출된 베이즈 추정량의 가중평균 형태의 모습을 나타내며, $w = 0$인 경우 균형 손실함수는 제곱평균 손실함수와 동일하게 되므로 베이즈 추정량도 제곱평균 손실함수하에서 도출된 통계량과 일치하게 된다. 또한 알려져 있지 않은

모수 μ와 B를 주변 확률분포로부터 추정하면 다음과 같이 균형 손실함수하에서의 EB인 $\hat{\boldsymbol{\theta}}^{\mathrm{NEB}}$를 구할 수 있으며

$$\hat{\boldsymbol{\theta}}^{\mathrm{NEB}} = \left[1-(1-w)\widehat{B}\right]\boldsymbol{X} + (1-w)\widehat{B}\overline{X}\mathbf{1}_m,$$

균형 손실함수하에서 BE인 $\hat{\boldsymbol{\theta}}^{\mathrm{NB}}$에 대한 베이즈 위험 BR($\hat{\boldsymbol{\theta}}^{\mathrm{NB}}$)는 다음과 같다.

$$\begin{aligned}\mathrm{BR}\left(\hat{\boldsymbol{\theta}}^{\mathrm{NB}}\right) &= wm^{-1}E\left\|\boldsymbol{X}-\hat{\boldsymbol{\theta}}^{\mathrm{NB}}\right\|^2 + (1-w)m^{-1}E\left\|\hat{\boldsymbol{\theta}}^{\mathrm{NB}}-\boldsymbol{\theta}\right\|^2 \\ &= w(1-w)^2B + (1-w)m^{-1}\left[E\left\|\hat{\boldsymbol{\theta}}^{\mathrm{NB}}-\hat{\boldsymbol{\theta}}^{B}\right\|^2 + E\left\|\hat{\boldsymbol{\theta}}^{B}-\boldsymbol{\theta}\right\|^2\right] \\ &= w(1-w)^2B + (1-w)\left[w^B+1-B\right] = (1-w)[1-(1-w)B],\end{aligned}$$

여기서 $w=0$ 인 경우 제곱평균 손실함수하에서의 베이즈 위험 $1-B$와 일치함을 확인할 수 있다. 같은 방법으로 균형 손실함수 하에서 EB인 $\hat{\boldsymbol{\theta}}^{\mathrm{NEB}}$에 대한 베이즈 위험 BR($\hat{\boldsymbol{\theta}}^{\mathrm{NEB}}$)는 쉬바르츠 부등식 등의 성질을 이용하여 근사적으로 다음과 같이 계산할 수 있다.

$$\begin{aligned}\mathrm{BR}\left(\hat{\boldsymbol{\theta}}^{\mathrm{NEB}}\right) &= wm^{-1}E\left\|\boldsymbol{X}-\hat{\boldsymbol{\theta}}^{\mathrm{NEB}}\right\|^2 + (1-w)m^{-1}E\left\|\hat{\boldsymbol{\theta}}^{\mathrm{NB}}-\boldsymbol{\theta}\right\|^2 \\ &= (1-w)[1-(1-w)B] + \frac{3(1-w)B}{m} + o\left(m^{-1}\right),\end{aligned}$$

여기서도 $w=0$ 이면 제곱평균 손실함수하에서의 베이즈 위험과 일치함을 확인할 수 있다.

2.4.2. CB와 CEB의 베이즈 위험 연구 과정 중에 증명된 CB의 특징은 제곱평균 손실함수와 균형 손실함수하에서 동일한 추정량이 도출된다는 것이다. 즉, 균형 손실함수의 가중치에 영향을 받지 않는다는 점인데, 이는 해당 추정량의 로버스트(robust)한 특성을 나타내는 것이다. 따라서 균형 손실함수 하에서의 CB인 $\hat{\boldsymbol{\theta}}^{\mathrm{NCB}}$와 CEB인 $\hat{\boldsymbol{\theta}}^{\mathrm{NCEB}}$는 다음과 같이 제곱평균 손실함수하에서 도출한 결과와 동일한 형태로 나타난다.

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{\mathrm{NCB}} &= a_B(1-B)\left(\boldsymbol{X}-\overline{X}\mathbf{1}_m\right) + \left[(1-B)\overline{X}+B\mu\right]\mathbf{1}_m \\ \hat{\boldsymbol{\theta}}^{\mathrm{NCEB}} &= a_{\mathrm{EB}}\left[\left(1-\widehat{B}\right)\boldsymbol{X} + \widehat{B}\overline{X}\mathbf{1}_m\right] + (1-a_{\mathrm{EB}})\,\overline{X}\mathbf{1}_m.\end{aligned}$$

그러나 이에 대한 베이즈 위험은 적용된 손실함수하에서 다른 형태로 계산된다. 먼저 CB에 대한 베이즈 위험 BR($\hat{\boldsymbol{\theta}}^{\mathrm{NCB}}$)는 다음과 같이 근사적인 값으로 계산된다.

$$\begin{aligned}\mathrm{BR}\left(\hat{\boldsymbol{\theta}}^{\mathrm{NCB}}\right) &= wm^{-1}E\left\|\boldsymbol{X}-\hat{\boldsymbol{\theta}}^{\mathrm{NCB}}\right\|^2 + (1-w)m^{-1}E\left\|\hat{\boldsymbol{\theta}}^{\mathrm{NCB}}-\boldsymbol{\theta}\right\|^2 \\ &= w + A_1(B) + m^{-1}\left[(1-2w)(1-B) - A_1(B) + A_2(B)\right] + o\left(m^{-1}\right),\end{aligned}$$

여기서 $A_1(B) = 2B^{-1}(1-B) - 2B^{-1}(1-B)^{1/2}][1-(1-w)B]$을 $A_2(B) = \frac{B}{2}(1-B)^{1/2}[1-(1-w)B]$을 나타내며, 베이즈 위험이 균형 손실함수에서 정의한 가중치에 영향을 받는 것을 확인할 수 있다. 또한 CEB에 대한 베이즈 위험 BR($\hat{\boldsymbol{\theta}}^{\mathrm{NCEB}}$)는 다음과 같다.

$$\begin{aligned}\mathrm{BR}\left(\hat{\boldsymbol{\theta}}^{\mathrm{NCEB}}\right) &= wm^{-1}E\left\|\boldsymbol{X}-\hat{\boldsymbol{\theta}}^{\mathrm{NCEB}}\right\|^2 + (1-w)m^{-1}E\left\|\hat{\boldsymbol{\theta}}^{\mathrm{NCEB}}-\boldsymbol{\theta}\right\|^2 \\ &= w + C_1(B) + m^{-1}\left[3 - 2wB - C_1(B) + 2(1-(1-w)BC_2(B)\right] + o\left(m^{-1}\right),\end{aligned}$$

여기서 $C_1(B), C_2(B)$는 $2B^{-1}[1-B(1-B)^{1/2} + (1-w)(1-B)^{1/2}$, $1-(2-B)(1-B)^{-1/2} + \frac{1}{4}B(1-B)^{-3/2}$로 각각 정의된다. 두 결과 모두 $w=0$인 경우 제곱평균 손실함수의 결과와 동일함을 확인할 수 있으며, 이는 Ghosh 등 (2008)에 의하여 제안된 증명을 기반으로 하고 있다.

2.5. 손해보험 산업에서의 위험도 추정 모형과 베이즈 위험

손해보험 산업에서 가격을 결정하는 위험도의 추정은 회사 경쟁력의 핵심 요소이므로 많은 이론적 발전을 통해 폭넓게 활용되고 있다. 위험도의 추정 요인 중 대표적인 것은 위험도를 결정짓는 변수(rating factor)를 발굴하고 해당 변수의 수준(level)을 적정하게 구분(segmentation)하는 것이다. 즉, 위험의 수준이 유사한 대상을 동일한 그룹으로 분류하면서, 위험의 수준이 다르다고 판단되는 몇 개의 그룹을 만드는 것이다. 이러한 개념은 통계학의 ANOVA 모형과 매우 유사한 개념이 된다. 즉, 동일한 그룹의 변동인 급내 변동(within variance)은 최소화하면서 급간 변동(between variance)를 최대화하는 시도와 일맥 상통한다고 할 수 있다.

따라서 앞서 살펴 본 베이즈 추정량들에 대하여 ANOVA 모형에 적용시킨 결과의 도출이 가능하고, 이를 보험상품의 각 그룹별 위험도 추정에 적용하여 활용할 수 있다. 뿐만 아니라 적용된 추정량의 베이즈 위험을 계산함으로써 추정하여 적용된 통계량의 적합성에 대한 평가가 가능해지는 것이다. 이를 위하여 다음과 같은 모형을 고려해 볼 수 있다.

위험도의 결과를 나타내는 손실금액을 Y_{ij}라고 하는 경우를 ANOVA 모형에 적용하게 되면 $Y_{ij} = \theta_i + e_{ij}$ (i=1,...,m; j=1,...,k)가 되고, $\theta_i = \mu + \alpha_i$로 표시할 수 있다. 여기서 e_{ij}와 α_i는 상호 독립임을 가정하며 각각 $e_{ij} \sim iid\ N(0, \sigma^2)$, $\alpha_i \sim iid\ N(0, \tau^2)$ 분포임을 가정한다. 따라서 베이지안 추론 프레임에서 $Y_{ij}|\theta_i \sim iid\ N(\theta_i, \sigma^2)$, $\theta_i \sim iid\ N(\mu, \tau^2)$, $i = 1, \ldots, m$인 것을 쉽게 확인할 수 있다.

또한 최소 충분성(minimal sufficiency) 조건을 통하여 그룹 평균인 $X_i = \frac{1}{k}\sum_{j=1}^{k} Y_{ij} = \overline{Y_i}$와 급내 변동인 $\mathrm{SSW} = \sum_{i=1}^{m}\sum_{j=1}^{k}(Y_{ij} - \overline{Y_i})^2$로 제한하여 고려할 수 있으며, $X_1, \ldots, X_m$와 SSW는 상호 독립적으로 각각 $X_i \sim iid\ N(\mu, \tau^2 + \sigma^2/k)$, $\mathrm{SSW} \sim \sigma^2\chi^2_{m(k-1)}$ 분포를 따른다는 것을 알 수 있다. 여기서 앞서 이론적 증명에서 활용한 표기를 따라 정의하면, $X_i \overset{iid}{\sim} N(\mu, \sigma^2/(kB))$가 되며, $B = (\sigma^2/k)/(\sigma^2/k + \tau^2) = \sigma^2/(\sigma^2 + k\tau^2)$와 같이 표기할 수 있다. 그리고 급간 변동 $\mathrm{SSB} = k\sum_{i=1}^{m}(X_i - \bar{X})^2$로 정의하는 경우 CB에서 활용되는 함수 a_B는 다음과 같이 정의할 수 있다.

$$a_B^2 = 1 + \frac{\sigma^2}{(1-B)\mathrm{MSB}},$$

여기서 MSB는 급간 변동의 평균인 $\mathrm{SSB}/(m-1)$을 의미한다. 또한 ANOVA 모형에서 EB의 모수를 추정하는데 사용되는 통계량으로 σ^2는 MSW로, $\widehat{B}$는 $\min[(m-3)/(m-1), (m-3)\mathrm{MSW}/(m-1)\mathrm{MSB}]$로 대체하여 정의할 수 있으며, 이에 따라 $a_{\mathrm{EB}}^2 = 1 + \mathrm{MSW}/(1-\widehat{B})\mathrm{MSB}$로 정의할 수 있다. 또한 주어진 ANOVA 모형에서 연구의 주요 관심 대상 중 하나인 균형 손실함수하에서의 CB와 CEB에 대한 베이즈 위험은 각각 다음과 같이 계산된다.

$$\mathrm{BR}\left(\hat{\boldsymbol{\theta}}^{\mathrm{NCB}}\right) = (1-w)(1-B) + \frac{\sigma^2}{k}A_1(B) + \frac{\sigma^2}{km}[A_2(B) - A_1(B)] + o\left(m^{-1}\right),$$
$$\mathrm{BR}\left(\hat{\boldsymbol{\theta}}^{\mathrm{NCEB}}\right) = (1-w)(1-B) + \frac{\sigma^2}{k}A_1(B) + \frac{\sigma^2}{km}[(1-w)B - A_1(B) + A_3(B) + A_4(B)] + o\left(m^{-1}\right),$$

여기서 정의된 $A_1(B)$, $A_2(B)$, $A_3(B)$, $A_4(B)$는 각각 $B^{-1}[1-(1-w)B][2-B-2(1-B)^{1/2}]$, $\frac{B}{2}(1-B)^{1/2}[1-(1-w)B]$, $2(2-B)[1-2(1-B)^{1/2}]$, $\frac{B}{k-1}[4/k - k(1-B)^{-3/2}/2]$으로 정의된 함수이다.

실증 분석에서는 보험상품의 위험도 추정에 적합한 ANOVA 모형에 대하여 다양한 베이즈 추정량들의 베이즈 위험을 실제로 계산하여 비교한 것을 바탕으로, 기존의 단점을 극복하고자 새롭게 제안된 베이즈 추정량들을 활용하고 적용하는데 제한적이지 않음을 증명함으로써 보험업계에서 보다 다양한 선택이 가능하도록 하고자 한다.

Table 3.1. Data structure (n denotes the sample size of each group and here $n = 1,000$)

N	Group			
	Small size car	Medium size car	Large size car	SUV & Van
1	y_{11}	y_{21}	y_{31}	y_{41}
2	y_{12}	y_{22}	y_{32}	y_{42}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_{1n}	y_{2n}	y_{3n}	y_{4n}

3. 실증 자료 분석

자동차 보험의 위험도 평가는 사고의 확률을 나타내는 빈도(frequency)와 사고의 정도를 나타내는 심도(severity) 등 크게 두 가지로 구분된다. 사고의 확률은 사고 여부, 사고 발생 건수와 같이 이산형 자료에 대한 추정이므로 본 연구에서는 ANOVA 모형에 적합한 연속형 자료로 구성되어 있는 심도에 대한 추정을 주요 내용으로 한다. 즉, 사고 발생 시 위험도를 평가하고자 하는 요율 요소별로 손해액의 분포가 어떻게 이루어지는지에 대한 적정 규모를 추정하는 과정으로 이해할 수 있다.

보험사 입장에서는 사고에 따른 손해 정도가 큰 그룹을 위험집단으로, 적은 그룹을 우량집단으로 분류하고 이에 대한 적정 위험도를 추정하여 보험료를 결정하게 된다. 따라서 이러한 위험도의 추정은 보험사의 보험료 결정에 직결되는 매우 중요한 과정이라고 할 수 있으며, 보험 기간별로 적용된 위험도의 적정성을 평가하고 수정, 보완하여 새로운 위험도를 추정하고 이를 근거로 새로운 보험료를 결정하게 된다.

본 장에서는 국내 보험사의 실제 데이터를 기반으로 하여, 자동차 보험의 보험 대상, 즉 보험 목적물인 차량의 종류별로 해당 위험도에 대한 추정과 베이즈 위험을 비교, 분석함으로써 예측, 적용하고자 하는 모수의 추정량으로서의 적정성 여부를 확인해 보고자 한다.

3.1. 데이터의 구성

실증 분석에 사용된 자료는 국내 손해보험사 S사의 차량 사고와 관련된 실제 손해 결과 자료이며, 자동차 보험에서 사고의 종류에 따라 정의하는 담보하는 기본 5개 담보-대인,대물,자기신체사고,무보험,자기차량- 중 대물 담보의 결과를 활용하는 것으로 한다. 이는 사고 발생 건수의 크기와 사고의 조작여부의 배제가 가능한 담보라는 이유이다.

위험도를 평가하는 요율 요소는 차량의 종류로 소형(Small size car), 중형(Medium size car), 대형(Large size car), SUV 및 Van(SUV & Van) 등 보험업계에서 일반적으로 통용되는 4가지 기준을 고려하며, 각 그룹별로 1,000개의 표본을 무작위 추출하여 총 4,000개의 표본을 사용하게 된다. 추가적으로 모형의 정규성 가정을 위하여 로그 변환을 실시하여 가정의 적정성을 만족하도록 한다. 따라서 2장에서 논의된 이론적 근거의 표기법에서 벡터 요소의 수를 나타내는 i는 1,2,3,4가 되는 것이고, 각 그룹별 데이터의 개수를 나타내는 j는 $1, 2, \ldots, 1000$이 되는 것이며, 이를 적용하고자 하는 ANOVA 모형으로 요약하면 Table 3.1과 같다.

3.2. 조건별 베이즈 위험과 베이즈 위험의 추정

연구 모형에서 추정하고자 모수 $\boldsymbol{\theta}$는 각 차량 종류별 평균 손실금액으로 이루어진 벡터로서, 본 연구에서는 무작위 추출 대상 집단인 1년간의 손해 자료 전체에서 계산된 평균 손실금액으로 가정을 한다. 보험회사의 입장에서 다년간 경험을 통하여 축적된 자료를 바탕으로 내부적으로 알려져 있는 모수의 정보,

Table 3.2. Estimates when $w = 0.0$ and Index

Estimates	Group				Index
	Small size car	Medium size car	Large size car	SUV & Van	
$\boldsymbol{\theta}$	476227.6	545406.7	360198.6	579786.3	
$\hat{\boldsymbol{\theta}}^{B}$	479158.6	566669.1	364075.6	578674.1	1.512762
$\hat{\boldsymbol{\theta}}^{\mathrm{EB}}$	479301.4	565542.0	365580.3	577322.6	1.515160
$\hat{\boldsymbol{\theta}}^{\mathrm{CB}}$	479052.3	567575.5	362911.9	579731.3	1.510755
$\hat{\boldsymbol{\theta}}^{\mathrm{CEB}}$	479233.6	566118.6	364834.8	577994.9	1.514588
$\hat{\boldsymbol{\theta}}^{\mathrm{NB}}$	479158.6	566669.1	364075.6	578674.1	1.512762
$\hat{\boldsymbol{\theta}}^{\mathrm{NEB}}$	479301.4	565542.0	365580.3	577322.6	1.515160
$\hat{\boldsymbol{\theta}}^{\mathrm{NCB}}$	479052.3	567575.5	362911.9	579731.3	1.510755
$\hat{\boldsymbol{\theta}}^{\mathrm{NCEB}}$	479233.6	566118.6	364834.8	577994.9	1.514588

즉 회사별로 추정하고자 하는 위험도 원가의 참값(true cost)일 수도 있으나, 본 연구에서는 1년간의 손해 자료가 모집단임을 가정하여 표본으로 추출된 정보를 통하여 모집단을 추정하는 방식으로 이해할 수 있다.

따라서 ANOVA 모형에서 추정하고자 하는 대상 모수인 차량 종류별 평균 손실금액 벡터에 대하여 앞서 논의한 4가지 추정량 BE, EB, CB, CEB와 해당 추정량의 베이즈 위험을 계산하게 된다. 이 추정량들과 해당 베이즈 위험은 제곱평균 손실함수와 균형 손실함수하에서 각각 도출되게 된다. 즉, 4가지의 대표적 베이즈 추정량들과 해당 베이즈 위험이 두 가지 손실함수 하에서 계산되어 총 8가지의 조합별 결과가 도출된다. 본 연구에서는 베이즈 위험에 대한 논의가 주요 대상인 바, 추정량의 정확도에 대한 지표(index)는 동일 모형에 대하여 Kim과 Kim (2014)에 의해 제안된 다음의 지표를 준용한다.

$$\text{Index} = p\left[E\left(\bar{\theta}|\boldsymbol{x}\right) - m^{-1}\sum_{i=1}^{m} t_i(\boldsymbol{x})\right]^2 + (1-p)\left|\left[E\left[\sum_{i=1}^{m}\left(\theta_i - \bar{\theta}\right)^2\middle|\boldsymbol{x}\right] - \sum_{i=1}^{m}\left[t_i(\boldsymbol{x}) - \bar{t}(\boldsymbol{x})\right]^2\right]\right|,$$

궁극적으로 모집단의 분포에 일치성을 추구한다는 목적에 부합하도록 일차적률과 이차적률의 각 항에 가중치를 동일하게 배분($p = 0.5$)하는 것으로 한다.

Table 3.2를 통하여 본 연구에서 고려하는 추정량들과 손실함수의 특성을 쉽게 확인할 수 있다. 가중치가 0일 경우, 제곱평균 손실함수는 적합도를 추가적으로 고려하는 균형 손실함수와 동일한 형태이기 때문에 손실함수별 해당 추정치들이 일치됨을 확인할 수 있다. 또한 CB는 위치 모수 뿐만 아니라 모수의 분포를 근사시키고자 이차 적률을 고려하기 때문에 새롭게 제안된 지표에서 기존의 베이즈 통계량들보다 우수함을 나타낸다. 이러한 추정량의 특성과 손실함수의 특성을 바탕으로 고려한 베이즈 위험의 결과들은 Table 3.3과 Table 3.4와 같으며, 균형 손실함수에서 정도와 적합도에 부여하는 가중치를 변경하여 도출한 결과이다.

적합도의 가중치를 일부($w = 0.2$) 부여한 결과, 동일한 경험적 베이즈 추정량에 대하여 균형 손실함수 하에서의 베이즈 위험이 감소되는 현상이 발생함을 확인할 수 있다. 이는 적합도, 즉 데이터가 가지는 정보의 적정성을 고려하는 손실함수라는 특성이 반영된 결과이다. 또한 모형의 모수에 대한 정보를 가정한 베이즈 추정량들에 대하여서는 CB의 위험의 감소가 발생하지 않는 현상이 나타나는데, 이는 오차를 과대축소(overshrink) 하는 베이즈 추정량의 단점을 보완한 추정량의 특성으로 이해할 수 있다. 참고적으로 제곱평균 손실함수와 균형 손실함수하에서의 지표가 동일하게 나타나는 것은 CB가 가지는 로버스트한 특성으로 손실함수의 종류와 상관없이 동일한 추정량이 도출되기 때문이다.

Table 3.3. Bayes risks when $w = 0.2$ and Index

Bayes risks	Group				Index
	Small size car	Medium size car	Large size car	SUV & Van	
$\hat{\boldsymbol{\theta}}^{B}$	0.997716	0.998580	0.997437	0.998604	1.512762
$\hat{\boldsymbol{\theta}}^{\mathrm{EB}}$	0.999429	0.999645	0.999359	0.999651	1.515160
$\hat{\boldsymbol{\theta}}^{\mathrm{CB}}$	0.998038	0.998779	0.997797	0.998801	1.512208
$\hat{\boldsymbol{\theta}}^{\mathrm{CEB}}$	0.999001	0.999379	0.998878	0.999389	1.479300
$\hat{\boldsymbol{\theta}}^{\mathrm{NB}}$	1.090541	0.980709	1.126093	0.977559	1.529947
$\hat{\boldsymbol{\theta}}^{\mathrm{NEB}}$	0.872635	0.845177	0.881523	0.844390	1.462077
$\hat{\boldsymbol{\theta}}^{\mathrm{NCB}}$	1.041362	0.950110	1.070903	0.947493	1.512208
$\hat{\boldsymbol{\theta}}^{\mathrm{NCEB}}$	0.927641	0.879378	0.943267	0.977994	1.479300

Table 3.4. Bayes risks when $w = 0.5$ and Index

Bayes risks	Group				Index
	Small size car	Medium size car	Large size car	SUV & Van	
$\hat{\boldsymbol{\theta}}^{B}$	0.997716	0.998580	0.997437	0.998604	1.512762
$\hat{\boldsymbol{\theta}}^{\mathrm{EB}}$	0.999429	0.999645	0.999359	0.999651	1.515160
$\hat{\boldsymbol{\theta}}^{\mathrm{CB}}$	0.998038	0.998779	0.997797	0.998801	1.514384
$\hat{\boldsymbol{\theta}}^{\mathrm{CEB}}$	0.999001	0.999379	0.998878	0.999389	1.424525
$\hat{\boldsymbol{\theta}}^{\mathrm{NB}}$	0.784588	0.677006	0.819411	0.673921	1.410819
$\hat{\boldsymbol{\theta}}^{\mathrm{NEB}}$	0.571147	0.544251	0.579853	0.543480	1.335659
$\hat{\boldsymbol{\theta}}^{\mathrm{NCB}}$	1.105870	0.876806	1.180024	0.870238	1.514384
$\hat{\boldsymbol{\theta}}^{\mathrm{NCEB}}$	0.820771	0.699445	0.859972	0.695968	1.424525

Table 3.4는 정도와 적합도의 비중을 균등하게 고려한 실증 결과이다. EB에 대한 베이즈 위험이 적합도에 대한 가중치를 20% 부여한 Table 3.2의 결과보다 급격하게 감소함을 확인할 수 있다. 이는 데이터가 가지는 정보에 보다 많은 신뢰를 부여함으로써 해당 오차를 감소시킬 수 있음을 보여주는 것이다. 따라서 보험업계에서는 다년간 축척해 온 자료의 신뢰성, 즉 모수에 대한 신뢰성과 당해년도에 발생한 사고 정보에 대한 신뢰성을 고려하여 가중치를 조정함으로써 해당 추정량의 오차를 감소시킬 수 있음을 의미한다. 대부분의 경우 다년간의 경험으로 가정한 모수에 대한 신뢰도를 많이 고려하기는 하나, 제도적인 변화, 특정한 이벤트의 발생 등으로 인하여 당해년도의 정보를 보다 많이 활용하여야 하는 경우에는 본 연구에서 제시한 결과를 토대로 하여 다양한 선택이 가능할 것으로 판단된다. 또한 베이즈 위험은 각 그룹별 오차의 평균적 개념이나, 보험사 입장에서는 전체적인 오차보다는 각 그룹별 오차에 더 많은 관심이 있는 점을 고려하여 베이즈 위험을 분리하여 제시하였음을 밝혀두는 바이다.

4. 결론

기존의 베이지안 추정과 관련한 연구는 모수 또는 모수의 분포에 대하여 보다 정확하고 적합한 대안을 제시하고자 하는 주제가 주요 대상이었다. 그러나 추정량이 갖는 장점에 비해 추정량에 따르는 오차의 크기가 간과되게되면 해당 추정량의 의미가 퇴색될 개연성이 존재한다. 또한 실제 현업에서 손익의 결과와 직결되는 경우, 오차에 대한 고려가 매우 조심스러울 수 밖에 없다. 따라서 본 연구에서는 다양한 베이즈 추정량들이 손실함수의 종류에 따라 갖게 되는 베이즈 위험의 특성을 실증 자료 분석을 통하여 확인해 보았다. 기존에 증명된 추정량의 장점과 주어진 상황에 따라 선택 가능한 복수의 손실함수를 고

려함으로써 다양한 선택적 상황을 제시하였으며, 이에 대한 베이즈 위험들의 추정과 실제 자료의 분석 결과의 비교를 통하여 해당 특징을 증명하고 방안을 제시하였다. 이러한 분석 결과들을 토대로 보험업계에서 복잡 다양해진 사회적 현상을 해석하고 예측하는데 있어 보다 폭넓은 선택을 할 수 있을 것으로 기대한다.

References

Kim, Y., Kim. M. and Kim, M. (2011). Estimating the auto insurance premium based on credibilities, *The Korean Journal of Applied Statistics*, **24(2)**, 279–292.

Ghosh, M. (1992). Constrained Bayes estimation with applications, *Journal of the American Statistical Association*,

Ghosh, M. and Kim, D. (2002). Multivariate constrained Bayes estimation, *Pakistan Journal of Statistics*, **18(2)**, 143–148.**87**, 533–540.

Ghosh, M., Kim, D. and Kim, M. (2004). Asymptotic mean squared error of constrained James-Stein estimators, *Journal of Statistical Planning and Inference*, **126**, 107–118.

Ghosh, M., Kim, M. and Kim, D. (2008). Constrained Bayes and empirical Bayes estimation under random effects normal ANOVA model with balanced loss function, *Journal of Statistical Planning and Inference*, **138**, 2017–2028.

Kim, M. and Kim, Y. (2005). Asymptotic mean squared error of positive part James-Stein estimators, *Journal of the Korean Statistics Society*, **34(2)**, 99–107.

Kim, M. and Kim, Y. (2013). Constrained Bayes and empirical Bayes estimator application in insurance pricing, *Communication for Statistical Applications and Methods*, **20(4)**, 321–327.

Kim, M. and Kim, Y. (2014). Application of Constrained Bayes Estimation under Balanced Loss Function in Insurance Pricing, *Communication for Statistical Applications and Methods*, **21(3)**, 235–243.

Louis, T.A. (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods, *Journal of the American Statistical Association*, **79**, 393–398.

Zellner, A. (1988). Bayesian analysis in econometrics, *Journal of Econometrics*, **37**, 27–50.

Zellner, A. (1992). Bayesian and non-Bayesian estimation using balanced loss functions, *Statistical Decision Theory and Related Topics V*, Springer-Verlag, New York, 377–390.

손해보험 위험도 추정에 대한 베이즈 위험 비교 연구

김명준[a] · 우호영[b] · 김영화[c,1]

[a]한남대학교 비즈니스 통계학과; [b]중앙대학교 대학원 통계학과; [c]중앙대학교 응용통계학과

(2014년 8월 12일 접수, 2014년 10월 14일 수정, 2014년 10월 15일 채택)

요 약

잘 알려져 있는 것처럼 일반적인 베이즈 추정량(Bayes estimator)과 경험적 베이즈 추정량(empirical Bayes estimator)은 모수를 추정하는데 있어서 오차를 과다축소하는 단점을 가지고 있다. 따라서 이러한 단점을 극복하기 위하여 constrained 베이즈 추정량이 일차 적률과 이차 적률을 일치시키는 성질을 만족시키며 제안되었다. 또한 평균제곱오차 함수와 같은 전통적인 손실함수에서는 추정의 정확성만을 고려하는 특징을 가지고 있기 때문에, 추정의 정확성과 정합성을 동시에 고려하는 균형 손실함수가 제안되었다. 이러한 이유로 인하여 균형손실 함수하에서의 제한적 베이즈 추정량의 활용이 손해 보험의 가격 산출에 제안되는 것은 타당하다. 그러나 대부분의 연구는 추정의 문제에만 집중하는 경향이 있으며. 이는 새롭게 제안되는 특정 손실함수하에서의 constrained 베이즈 추정량과 constrained empirical 베이즈 추정량의 베이즈 위험의 계산이 어렵다는 점에서 기인한다. 본 연구에서는 다양한 베이즈 추정량들에 대한 베이즈 위험을 서로 다른 두 손실함수하에서 비교하였으며, 그 대상은 자동차 보험 산업에서의 위험도 추정 분야이다. 또한 자동차 보험 산업의 실제 사고 데이터를 이용하여 새롭게 제안된 베이즈 추정량의 베이즈 위험을 비교함으로써 그 효용성을 입증하였다.

주요용어: 베이즈 위험, 보험위험, 균형 손실함수, constrained 베이즈 추정량.

[1]교신저자: (156-756) 서울시 동작구 흑석동 221, 중앙대학교 응용통계학과. E-mail: gogators@cau.ac.kr