Journal of the Korean Data & Information Science Society
2016, **27**(1), 9–18

http://dx.doi.org/10.7465/jkdi.2016.27.1.9
한국데이터정보과학회지

베이지안 접근법과 모수불확실성을 반영한 보험위험 측정 모형

조재린[1] · 지혜수[2] · 이항석[3]

[1]보험연구원 금융전략실 · [2]KB생명보험 계리부 · [3]성균관대학교 보험계리학과/수학과

접수 2015년 9월 22일, 수정 2015년 10월 29일, 게재확정 2016년 1월 14일

요 약

모수불확실성을 반영하는 손실모형으로는 Heckman과 Meyers가 제안한 모형이 주로 인용되고 있다. 이 모형은 모수 자체가 어떤 확률분포를 따른다는 가정을 하고 있으며 IAA, Swiss Solvency Test, EU Solvency II 등에서 참고하고 있다. 반면 베이지안 기법을 이용한 연구는 모수에 대한 선험적 정보 즉, 사전분포를 이용하여 모수불확실성을 반영한다. 그러나 현실에서는 두 가지 방법을 동시에 고려해야 하는 상황이 빈번히 발생한다. 이에 본 연구는 Heckman-Meyers의 모형과 베이지안 접근법을 동시에 고려한 베이지안 H-M CRM모형을 제안하고 그 특성을 분석하였다.

주요용어: 모수불확실성 리스크, 베이지안, 보험리스크, 손실모형.

1. 서론

보험회사는 보험계약자에게 장래의 보험금 지급의무를 이행하기 위해 책임준비금을 적립하지만 보험료 산출시 기대한 것과 달리 책임준비금만으로는 보험금을 지급하지 못할 위험이 있다. 이러한 위험을 대비하여 예상치 못한 손실이 발생하더라도 보험계약자에게 보험금을 지급할 수 있는 충분한 자산을 보유하도록 하고 있다. 보험회사가 노출된 위험은 보험리스크, 금리리스크, 시장리스크, 신용리스크, 운영리스크 등으로 구분하며 이 중 큰 부분을 차지하는 것이 보험리스크이다. 보험리스크란 손해율 또는 사망률의 악화 등으로 보험료 산출시 예상한 것과 다른 현금흐름으로 인하여 발생하는 위험이다.

리스크는 변동성 리스크 (volatility risk)와 모수불확실성 리스크 (parameter uncertainty risk)로 구분할 수 있다. 변동성 리스크는 모수가 변하지 않는다는 가정 하에 미래에 발생할 보험금지급액이 기댓값으로부터 벗어날 가능성을 말한다. 따라서 계약건수가 충분하다면 대수의 법칙 (law of large number)에 따라 변동성 리스크는 줄어들 것이다. 그러나 실제로 계약건수가 많은 자동차보험이나 사망보험의 경우에도 여전히 큰 변동성이 존재하는데 이러한 이유는 모수불확실성 (parameter uncertainty)이 존재하기 때문이다.

반면 모수불확실성 리스크는 표본분포와 실제 모집단의 차이에서 오는 표본위험 (sampling risk)과 표본을 추출할 당시와 리스크를 측정하는 기간 사이에 일어난 제도 및 사회경제 환경의 변화 등으로 인한 데이터 편의 (data bias) 등 계약건수가 충분히 크더라도 나타날 수 있어서 리스크의 크기를 측정할 때 주요 관심대상이 된다.

[1] (07328) 서울특별시 영등포구 국제금융로 6길 38, 보험연구원 금융전략실, 연구위원.
[2] (07325) 서울특별시 영등포구 국제금융로 2길 28, KB생명보험 계리부, 사원.
[3] 교신저자: (03063) 서울특별시 종로구 성균관로 25-2, 성균관대학교 보험계리학과/수학과, 교수.
E-mail: hangsuck@skku.edu

모수불확실성을 반영하는 손실모형에 대한 연구로는 frequentist 접근법인 Heckman과 Meyers (1983)의 연구가 주로 인용되고 있으며, IAA (Wason, 2004), Swiss Solvency Test (Luder, 2005), EU Solvency II 등에서 참고하고 있다. 다른 한편으로는 Pai (1997), Migon과 Moura (2005), Migon과 Penna (2006) 등 Bayesian 기법을 CRM (collective risk model)에 적용한 연구가 있다. Pai (1997) 등이 제안한 베이지안 방법론을 이용하여 모수불확실성을 반영하는 모형은 기본적으로 CRM 모형의 모수가 어떤 사전분포 (prior distribution)를 갖는다고 가정하고, 예측분포 (predictive distribution)를 통해 모수불확실성이 나타나도록 하고 있다.

두 부류의 연구에서 나타나는 근본적인 차이는 Heckman과 Meyers는 모수 자체를 확률변수로 보는 반면 Bayesian 접근법은 모수에 대한 선험적 정보 즉, 사전분포를 이용하여 모수불확실성을 반영한다는 점이다. 이러한 차이는 두 가지 방법을 동시에 고려한 연구를 찾아보기 어렵게 하는 이유인 것으로 보인다. 그러나 현실에서는 두 가지 방법을 동시에 고려해야 하는 상황이 빈번히 발생한다. 이에 본 연구는 Heckman과 Meyers와 Bayesian 접근법을 동시에 반영하는 모형을 제안하고 각각의 모형이 가지는 특성을 분석하고자 한다.

본 연구의 구성은 다음과 같다. 2절에서는 기존의 총손실분포 모형을 소개하고 3절에서 두 모형을 모두 고려한 모형을 제안한다. 그리고 4절에서는 각 모형별 위험계수를 비교하고, 5절에서는 결론을 내린다.

2. 보험리스크 측정 모형

2.1. CRM (collective risk model) 모형

CRM은 빈도와 심도로 총 손실액을 표현하는 방법으로 개별보험사고가 독립적이고 동질적으로 분포되어 있다는 가정 하에 총손실을 개별 손실액의 합으로 나타낸다. 또한 사고의 빈도도 미래에 사고가 얼마나 일어날지 모르기 때문에 확률변수이다. 즉 전통적 CRM은 총손실액을

$$S = \sum_{i=1}^{N} X_i \tag{2.1}$$

으로 정의한다. 여기서 N은 빈도를 나타내는 확률변수이며, 심도 $X_i,\ i = 1, \cdots, N$ 는 건당 손해액의 크기를 나타내는 확률변수로서 서로 독립적이고 동일한 분포 (i.i.d.)를 따른다. 또한 N과 $X_i,\ i = 1, \cdots, N$는 서로 독립이라고 가정한다. 따라서 총손실액 S의 기댓값은

$$E[S] = E[N]E[X]$$

이고 분산은

$$Var[S] = E[N]Var[X] + Var[N]E[X]^2$$

이므로 총손실액 의 변동계수 CV (coefficient of variation)의 제곱은

$$CV^2[S] = \frac{Var[S]}{E^2[S]} = \frac{Var[X]}{E[N]E^2[X]} + \frac{Var[N]}{E^2[N]}$$

이다.

한편, 일반적으로 빈도의 기댓값 대비 분산의 비율 $CM = Var[N]/E[N]$은 상수라고 가정한다. 예를 들어 빈도가 포아송분포를 따른다면 $CM = 1$이고, 모수가 r, β인 음이항분포를 따른다면 $CM = 1 + \beta$이다. 따라서

$$CV^2[S] = \frac{CV^2[X]}{E[N]} + \frac{CM}{E[N]} \tag{2.2}$$

로 표현할 수 있고, $E[N]$ (즉, 보험계약집단의 크기)이 커질수록 변동성 리스크는 작아진다는 것을 알 수 있다. 예를 들어 빈도 N은 기댓값이 λ인 포아송분포를 따르고 심도 X 는 건당 손해액의 크기를 나타내는 확률변수로 기댓값이 v이고 분산은 r^2인 로그정규분포를 따른다고 가정하면, 식 (2.2)에 의해

$$CV^2[S] = \frac{1}{\lambda}\left(\frac{\tau^2}{v^2}+1\right) \tag{2.3}$$

이므로 $\lambda(=E[N])$가 커짐에 따라 $CV[S]$는 0으로 수렴한다.

2.2. H-M CRM 모형

전통적 CRM 모형은 보험계약집단의 크기가 커짐에 따라 변동계수가 0에 가까워진다. 그러나 현실은 계약건수가 많은 자동차보험이나 사망보험의 경우에도 여전히 큰 변동성이 존재한다. 이러한 이유는 모수자체의 변화가능성인 데이터 편의가 존재하기 때문이라고 생각할 수 있다. 예를 들어 실손의료보험의 빈도와 심도를 나타내는 모수는 의료물가상승률, 국가의료보험제도의 변화 등 외부사건의 영향에 의해서 변화할 수 있다.

Heckman과 Meyers (1983)는 전통적 CRM 모형에 빈도와 심도의 모수불확실성을 반영하는 모형 (H-M CRM 모형)을 제시하였다. 이 모형은 빈도의 모수가 어떤 분포를 따른다고 가정하고, 총손실은 개별 손실액에 심도의 모수불확실성을 나타내는 확률변수를 곱한 값들의 합으로 아래와 같이 정의한다.

$$S_H|(K,Y) = Y\sum_{i=1}^{N|K} X_i \tag{2.4}$$

여기서 빈도의 모수불확실성을 반영하기 위해 K의 변화에 따라 $N|K$의 모수가 달라진다고 가정한다. 그리고 확률변수 Y는 심도의 모수불확실성을 나타내고 개별 손실액에 곱해짐으로 심도의 모수 또한 Y에 의해 변할 수 있음을 의미한다.

예를 들어 N은 $K=\kappa$인 경우 기댓값이 $\kappa\lambda$인 포아송분포를 따르고, K는 기댓값이 1이고 분산이 c인 감마분포를 따른다고 가정하자. 그러면 $K\sim\Gamma(1/c,c)$는 모수불확실성을 반영하기 위한 확률변수이고, $N|(K=\kappa)\sim Pois(\kappa\lambda)$는 모수불확실성이 반영된 보험사고건수를 나타낸다. 또한 심도는 평균이 $E[X]=v$ 이고 분산은 $Var(X)=r^2$인 로그정규분포를 따르며, 심도의 모수불확실성을 나타내는 확률변수 $Y\sim\Gamma(1/b,b)$는 기댓값이 1이고 분산이 b인 감마분포를 따른다고 가정하면,

$$E[S_H] = \lambda v \tag{2.5}$$

이고 분산은

$$Var[S_H] = (1+b)[\lambda\tau^2 + v^2(\lambda + c\lambda^2)] + bv^2\lambda^2 \tag{2.6}$$

이다. 따라서 식 (2.5)와 식 (2.6)에 의해, H-M CRM의 총손실 S_H 변동계수는 다음과 같다.

$$\begin{aligned} CV^2[S_H] &= \frac{(1+b)[\lambda\tau^2 + v^2(\lambda + c\lambda^2)] + bv^2\lambda^2}{\lambda^2 v^2} \\ &= (1+b)\left(\frac{\tau^2+v^2}{\lambda v^2}\right) + c(1+b) + b \end{aligned} \tag{2.7}$$

위의 식 (2.7)에서 알 수 있듯이 H-M CRM에서 변동계수는 λ가 커짐에 따라 변동계수는 작아지기는 하지만 전통적인 CRM과는 다르게 0에 수렴하지 않고 $c(1+b)+b$에 가까워진다. 모수불확실성을 반영하기 위해 도입한 확률변수 K와 B의 분산인 c와 b는 보험계약집단의 크기에 영향을 받지 않기 때문이다.

2.3. 베이지안 접근법

CRM 모형을 이용하여 보험위험을 측정하기 위해서는 주어진 표본 (sample)을 이용하여 모수를 추정하여야 한다. 그러나 표본의 분포는 모집단의 분포와 다르기 때문에 점추정 값은 모집단의 모수와 다르다. 즉 표본위험이 존재한다. 특히 표본의 크기가 작을 때에는 표본위험이 상당히 크다고 할 수 있다. 이러한 위험을 반영하는 방법으로, 데이터가 가지는 정보의 질을 예측분포에 반영하는 베이지안 방법론은 매우 유용하다.

이제 베이지안 방법론을 CRM 모형에 적용하기 위해 빈도의 모수 벡터 θ_1과 심도의 모수 벡터 θ_2로 이루어진 벡터 $\theta = (\theta_1, \theta_2)$는 사전분포 $\pi(\theta)$ 를 따르고, 주어진 θ에 대해

$$M|\Theta = (N, X_1, X_2, \ldots, X_N)|\Theta$$

는 확률밀도함수가 $f_{M|\Theta}(n, x_1, x_2, \ldots, x_n|\theta)$인 모델분포 (model distribution)를 따른다고 가정하자. 그러면 주어진 데이터 d에 대해 θ의 사후분포 (posterior distribution)는

$$\pi_{\Theta|D}(\theta|d) = \frac{f_{M|\Theta}(d|\theta)\pi(\theta)}{\int f_{M|\Theta}(d|\theta)\pi(\theta)d\theta}$$

이고, 총손실 예측분포의 확률밀도함수는 다음과 같다.

$$f_{S_B}(s) = f_{S|D}(s|d) = \int f_{S|\Theta}(s|\theta)\pi_{\Theta|D}(\theta|d)d\theta$$

여기서 $d = ((n_1, x_{1,1}, \ldots, x_{1,n_1}), (n_2, x_{2,1}, \ldots, x_{2,n_2}), \ldots, (n_t, x_{t,1}, \ldots, x_{t,n_T}))$이다.

예를 들어, 빈도 모수의 표본위험을 반영하기 위하여 빈도 모수의 사전분포는 모수가 α,θ 인 감마분포를 따르고, 빈도는 포아송분포 심도는 로그정규분포를 따르는 전통적인 CRM 모형을 따른다고 가정하자. 즉 사전분포는

$$\Lambda \sim \Gamma(\alpha, \theta)$$

이고, 총손실분포는

$$S = \sum_{i=1}^{N} X_i$$

라고 가정하자. 여기서 빈도는

$$N|\Lambda = \lambda \sim Poisson(\lambda)$$

이고, 심도 $X_i,\ i = 1, \cdots, N$ 는 기댓값이 v이고 분산은 τ^2인 로그정규분포를 따른다. 이때 t년간의 빈도 데이터가 $N_1, N_2, N_3, \cdots, N_t$로 주어지면, 사후분포는 감마분포를 따르고 $(t+1)$년 빈도의 예측분포는 음이항분포를 따르며,

$$E[N_{t+1}|N_1, N_2, \ldots, N_t] = \frac{(t\overline{N} + \alpha)\theta}{1 + t\theta} \tag{2.8}$$

이고

$$Var[N_{t+1}|N_1, N_2, \ldots, N_t] = (t\overline{N} + \alpha)\frac{\theta}{1 + t\theta}\left(1 + \frac{\theta}{1 + t\theta}\right) \tag{2.9}$$

이다. 여기서 $\overline{N} = (N_1 + N_2 + \cdots + N_t)/t$이다.

이제 $(t+1)$년의 예상 총손실 S_B는 $N_{t+1}|N_1, N_2, \ldots, N_t$개의 개별손실액의 합으로 나타낼 수 있으며 다음과 같다.

$$S_B = S|N_1, \ldots, N_t = \sum_{i=1}^{N_{t+1}|N_1,\ldots,N_t} X_i.$$

따라서 식 (2.8)에 의해, S_B의 기댓값은

$$E[S_B] = \frac{(t\overline{N} + \alpha)\theta}{1 + t\theta} v \tag{2.10}$$

이고 식 (2.9)에 의해, 분산은

$$Var[S_B] = \frac{(t\overline{N} + \alpha)\theta}{1 + t\theta} \tau^2 + v^2 (t\overline{N} + \alpha) \frac{\theta}{1 + t\theta} \left(1 + \frac{\theta}{1 + t\theta}\right) \tag{2.11}$$

이다. 그러므로 식 (2.10)과 식 (2.11)에 의해, 변동계수는

$$\begin{aligned} CV^2[S_B] &= \frac{\frac{(t\overline{N}+\alpha)\theta}{1+t\theta}\tau^2 + v^2(t\overline{N} + \alpha)\frac{\theta}{1+t\theta}\left(1 + \frac{\theta}{1+t\theta}\right)}{\left(\frac{(t\overline{N}+\alpha)\theta}{1+t\theta}v\right)^2} \\ &= \frac{\tau^2(1 + t\theta) + v^2(1 + (t + 1)\theta)}{(t\overline{N} + \alpha)\theta v^2} \end{aligned} \tag{2.12}$$

으로 표현될 수 있다.

위의 식 (2.12)를 식 (2.3)과 비교해보면 빈도 모수에 사전분포를 적용함으로써 표본위험이 반영된 위험계수가 산출되었음을 알 수 있다. 또한 t값이 커짐에 대수의 법칙에 의해 $\overline{N}$는 λ로 수렴하므로, $CV^2[S_B]$은 $CV^2[S]$로 수렴한다는 것을 확인할 수 있다. 즉, 표본의 크기가 커질수록 표본위험은 작아진다는 것을 알 수 있다.

3. 베이지안 H-M CRM 모형

2절에서 살펴보았듯이 H-M CRM 모형과 베이지안 방법론 모두 모수불확실성을 반영한 총손실모형이다. 그러나 상대적으로 H-M CRM 모형은 표본을 추출할 당시와 리스크를 측정하는 기간 사이에 일어난 제도 및 사회경제 환경의 변화 등으로 인한 데이터 편의를 반영하기에 적합한 모형이고, 반면 베이지안 방법론은 표본 분포와 실제 모집단의 차이에서 오는 표본위험을 반영하기에 적합한 모형이다.

한편 현실에서는 데이터 편의와 표본위험이 동시에 존재하는 경우가 많다. 예를 들어 단체 실손의료보험의 경우 빈도와 심도 모수가 의료물가상승률, 국가의료보험제도의 변화 등 외부사건의 영향에 의해서 변화할 수 있는 등 데이터 편의가 발생할 가능성이 높을 뿐 아니라, 가입단체의 규모가 작거나 가입기간이 짧아서 생기는 데이터의 부족이 야기하는 표본위험이 동시에 존재한다. 따라서 두 가지 방법론을 동시에 적용하는 것이 보다 합리적인 방안이라고 할 수 있다. 다시 말해 데이터의 부족으로 인한 표본위험은 사전분포를 이용하여 반영하고, 환경 변화에 의한 데이터 편의는 H-M CRM을 이용하여 반영하는 것이 적절한 모형이라 할 수 있다. 이제 H-M CRM에 베이지안 방법론을 적용한 모형을 베이지안 H-M CRM 모형이라 하자. 이 장에서는 베이지안 H-M CRM 모형에 대해 설명하고, 그 예를 보이고자 한다.

3.1. 베이지안 H-M CRM 모형

베이지안 H-M CRM 모형은 표본위험을 반영하는 단계와 데이터 편의를 반영하는 두 단계로 이루어져 있다. 먼저 표본위험을 반영하기 위해 빈도 N과 심도 X의 모수들로 이루어진 벡터 θ는 사전분포 $\pi(\theta)$를 따르고, 주어진 θ에 대해

$$M|\Theta = (N, X_1, X_2, \ldots, X_N)|\Theta$$

는 확률밀도함수가 $f_{M|\Theta}(n, x_1, x_2, \ldots, x_n|\theta)$인 모델분포를 따른다고 가정하자. 그러면 주어진 데이터 d에 대해 θ의 사후분포는 다음과 같다.

$$\pi_{\Theta|D}(\theta|d) = \frac{f_{H|\Theta}(d|\theta)\pi(\theta)}{\int f_{H|\Theta}(d|\theta)\pi(\theta)d\theta}$$

여기서 $d = ((n_1, x_{1,1}, \ldots, x_{1,n_1}), (n_2, x_{2,1}, \ldots, x_{2,n_2}), \ldots, (n_t, x_{t,1}, \ldots, x_{t,n_T}))$이다.

다음으로 데이터 편의를 반영하기 총손실은 2.2절에 소개한 H-M CRM 모형을 따른다고 가정한다. 즉, 총손실은

$$S_H|(K,Y) = Y \sum_{i=1}^{N|K} X_i$$

이고, 의 K변화에 따라 $N|K$의 모수가 달라지며, 확률변수 Y가 개별 손실액에 곱해짐으로 심도의 모수 또한 변한다고 가정한다. 그러면 총손실 예측분포의 확률밀도함수는 다음과 같다.

$$f_{S_{BH}}(s) = f_{S_H|D}(s|d) = \int f_{S_H|\Theta(s|\theta)\pi_{\Theta|D}(\theta|d)d\theta}$$

여기서 주의할 점은 총손실 모형이 H-M CRM 모형을 따르므로 현실적으로 주어지는 데이터는 위에서 사후분포를 산출하기 위해 사용한 d 형태로 주어지는 경우 보다는

$$d_{BH} = ((l_1, z_{1,1}, \ldots, z_{1,l_1}), (l_2, z_{2,1}, \ldots, z_{2,l_2}), \ldots, (l_t, z_{t,1}, \ldots, z_{t,l_T}))$$

의 형태로 주어지는 경우가 많다는 점이다. 여기서 $L = N|K$이고 $Z = YX$이다. 따라서 실제 적용에서는 d_{BH}을 d로 전환 또는 보정하는 과정이 필요하다.

3.2. 베이지안 H-M CRM의 예

빈도 모수 λ의 표본위험을 반영하기 위해 λ의 사전분포는 모수가 α, θ인 감마분포를 따른다고 가정하고, 빈도와 심도모수의 데이터 편의를 반영하기 위해 총손실 분포는 H-M CRM 모형을 따른다고 가정하자. 즉 빈도모수의 사전분포는

$$\Lambda \sim \Gamma(\alpha, \theta)$$

이고, 총손실분포는

$$S_H|(K,Y) = Y \sum_{i=1}^{N|K} X_i$$

라고 가정하자. 이때 N은 $K = \kappa$인 경우 기댓값이 $\kappa\lambda$인 포아송분포를 따르고, K는 기댓값이 1이고 분산이 c인 감마분포를 따르며, 심도는 평균이 $E[X] = v$이고 분산은 $Var(X) = r^2$인 로그정규분포를 따르고, 심도의 모수불확실성을 나타내는 확률변수 Y는 기댓값이 1이고 분산이 b인 감마분포를 따른다.

$$N|K \sim Poisson(\kappa\lambda),\ K \sim \Gamma(\frac{1}{c}, c)$$

위의 가정 하에서 보험사고 발생건수에 대한 데이터가 $N_1, N_2, N_3, \ldots, N_n$인 경우 빈도모수의 사후분포는 다음과 같은 감마분포를 따른다.

$$\Lambda|N_1, N_2, N_3, \ldots, N_n \sim \Gamma\left(n\overline{N} + \alpha, \frac{\theta}{1+n\theta}\right)$$

$\Lambda|N_1, N_2, N_3, \ldots, N_n$가 감마분포를 따르는 이유는 다음과 같다. 빈도 N은 $K = \kappa$인 경우 평균이 $\kappa\lambda$인 포아송분포를 따르므로 N은 기댓값이 λ이고 분산이 $\lambda + c\lambda^2$인 음이항분포를 따른다. 따라서

$N_1, N_2, N_3, \ldots, N_n$는 음이항분포를 따르는 데이터이다. 그러나 이 단계에서는 빈도 모수 λ의 표본위험만을 반영하는 단계이므로 데이터 편의를 반영하는 확률변수 K의 영향은 제거해야 한다. 따라서 K의 영향이 없을 때, N은 모수가 λ인 포아송분포를 따른다고 할 수 있다. 한편 사전분포가 감마분포이고 모델분포는 포아송분포이므로 사후분포 $\Lambda|N_1, N_2, N_3, \ldots, N_n$은 감마분포를 따른다. 따라서

$$E[\Lambda_p] = E[\Lambda|N_1, N_2, N_3, \ldots, N_n] = \frac{(n\overline{N}+\alpha)\theta}{1+n\theta},$$

$$E[\Lambda_p^2] = \frac{(n\overline{N}+\alpha)(n\overline{N}+\alpha+1)\theta^2}{(1+n\theta)^2}, \quad Var[\Lambda_p] = \frac{(n\overline{N}+\alpha)\theta^2}{(1+n\theta)^2}$$

이다. 여기서 $\Lambda_p = \lambda|N_1, N_2, N_3, \ldots, N_n$이다.

이제 식 (2.5)와 식 (2.6)에 의해,

$$E[S_H|\Lambda_p = \lambda] = \lambda v$$

이고 분산은

$$Var[S_H|\Lambda_p = \lambda] = (1+b)[\lambda\tau^2 + v^2(\lambda + c\lambda^2)] + bv^2\lambda^2$$

이므로,

베이지안 H-M CRM 모형에서 예측 총손실은

$$E[S_{BH}] = E[S_H|N_1, \ldots, N_n] = E[E[S_H|\Lambda_p]] = E[\Lambda_p]v = \frac{(n\overline{N}+\alpha)\theta v}{(1+n\theta)} \tag{3.1}$$

이고,

$$\begin{aligned} Var[S_{BH}] &= Var[S_H|N_1, N_2, \ldots, N_n] = E[Var[S_H|\Lambda_p]] + Var[E[S_H|\Lambda_p]] \\ &= (1+b)[E[\Lambda_p]\tau^2 + v^2(E[\Lambda_p] + cE[\Lambda_p^2])] + bv^2E[\Lambda_p^2] + v^2Var[\Lambda_p] \\ &= (1+b)\left[\frac{(n\overline{N}+\alpha)\theta}{1+n\theta}\tau^2 + v^2\left(\frac{(n\overline{N}+\alpha)\theta}{1+n\theta}\left(1 + c\frac{(n\overline{N}+\alpha+1)\theta}{1+n\theta}\right)\right)\right] \\ &\quad + bv^2\left(\frac{(n\overline{N}+\alpha)\theta}{1+n\theta}\right)^2 + v^2\frac{(n\overline{N}+\alpha)\theta^2}{(1+n\theta)^2} \end{aligned} \tag{3.2}$$

이다.

그러므로 베이지안 H-M CRM 모형의 변동계수를 구하면 식 (3.1)과 식 (3.2)에 의해, 다음과 같은 결과를 얻을 수 있다.

$$CV^2[S_{BH}] = (1+b)\frac{(\tau^2+v^2)(1+n\theta)}{(n\overline{N}+\alpha)\theta v^2} + \frac{c+bc+1}{n\overline{N}+\alpha} + c(1+b) + b \tag{3.3}$$

위의 식 (3.3)를 보면 베이지안 H-M CRM 모형의 위험계수는 기본적으로 표본위험을 반영한 부분인

$$\frac{(\tau^2+v^2)(1+n\theta)}{(n\overline{N}+\alpha)\theta v^2}$$

과 데이터 편의를 반영하는 부분인

$$c(1+b) + b$$

으로 이루어져 있음을 알 수 있다. 또한 2.3절의 베이지안 접근법과는 다르게 보험계약집단이 커짐에 따라 즉, $\overline{N}$가 커짐에 따라 위험계수는 0이 아닌 $c(1+b)+b$로 수렴한다.

4. 시뮬레이션

A 보험회사 단체 실손의료보험의 지난 5년간 사고 빈도는 다음과 같다고 하자.

$$N_1 = 62, N_2 = 42, N_3 = 56, N_4 = 51, N_5 = 39$$

그리고 심도는 기댓값 $v = 100$ 이고 표준편차 $r = 25$ 인 로그정규분포를 따른다고 하자. 이때 빈도의 기댓값 λ의 추정값은 아래와 같이 빈도 데이터의 평균을 이용하는 것이 가장 자연스러울 것이다.

$$\hat{\lambda} = \overline{N} = 50$$

다음으로 H-M CRM의 총손실 분포를 구하기 위해서는 심도와 빈도의 모수불확실성을 반영하기 위해 b와 c를 추정하여야 한다. b와 c를 추정하는 방법에 대해서는 Cho (2013)에서 자세히 다루고 있으므로 본 연구에서는 b와 c를 추정하는 방법에 대한 논의는 생략하고

$$\hat{b} = 0.01,\ \hat{c} = 0.0025 \tag{4.1}$$

라고 하자. 이것은 빈도 모수가 기댓값은 λ이고 분산은 0.01λ인 분포를 따른다. 즉 표준편차가 기댓값의 10%인 분포를 따른다는 의미이다. 마찬가지로 심도모수는 표준편차가 기댓값의 5% 인 분포를 따른다는 의미로 해석할 수 있다.

베이지안 방식을 적용하기 위해서는 사전분포를 정의해야 한다. 일반적으로 사전분포는 선험적인 정보를 이용하거나 업계 전체의 자료를 분석하여 추정한다. 이 연구에서는 빈도 모수의 사전분포는 $\alpha = 75$, $\theta = 2$인 감마 분포를 따른다고 하자. 다시 말해 기댓값은 150이고 분산은 300인 분포를 따른다고 하자. 마지막으로 베이지안 H-M CRM의 총손실 분포는 위에서 언급한 모든 정보를 이용하여 구한다.

총손실 분포는 일반적으로 정규분포를 따르지 않으며 수리적으로 구하기 어려운 경우가 많다. 따라서 이 논문에서는 각 모형별로 임의표본 (random sample) 100,000개씩을 각각 발생시켜 총손실 분포를 구하였다. Figure 4.1은 시뮬레이션을 통해 얻은 분포를 히스토그램으로 표현하였으며, Table 4.1은 각 모형별 기댓값, 분산, 변동계수를 정리한 표이다.

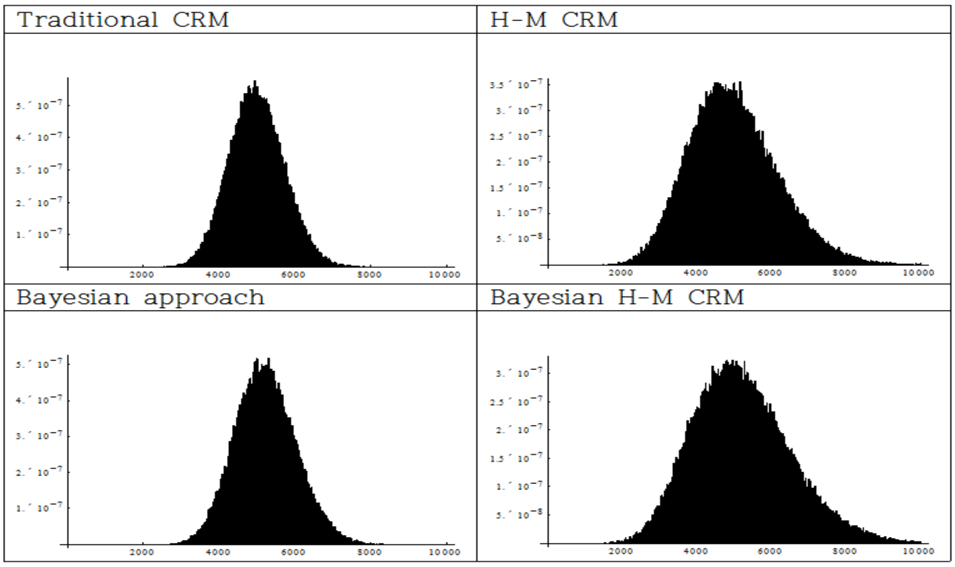

Figure 4.1 Aggregate loss distributions

Table 4.1 Expectation, Variance and CV's

	Expectation	Expectation Variance	CV
Traditional CRM	4998.08	$0.53 * 10^6$	0.14567
H-M CRM	4996.29	$1.42 * 10^6$	0.23842
Bayesian approach	5229.64	$0.65 * 10^6$	0.15418
Bayesian H-M CRM	5229.68	$1.74 * 10^6$	0.25226

Table 4.1을 보면 전통적 CRM에 비해 데이터 편의를 고려한 H-M CRM과 표본위험을 고려한 베이지안 모형의 변동성이 크고, 두 가지 리스크를 모두 반영한 베이지안 H-M CRM의 변동성이 가장 큰 것을 알 수 있다.

5. 결론

모수불확실성 리스크는 표본분포와 실제 모집단의 차이에서 오는 표본위험, 표본을 추출할 당시와 리스크를 측정하는 기간 사이에 일어난 제도 및 사회경제 환경의 변화 등으로 인한 데이터 편의 등 계약건수가 충분히 크더라도 나타날 수 있어서 리스크의 크기를 측정할 때 주요 관심대상이 된다.

모수불확실성을 반영하는 손실모형으로는 모수자체가 확률변수라는 아이디어를 사용한 H-M CRM과 베이지안 접근법을 이용한 모형이 있다. 그리고 2절에서 살펴보았듯이 H-M CRM 모형과 베이지안 방법론 모두 모수불확실성을 반영한 모형이지만 상대적으로 H-M CRM 모형은 표본을 추출할 당시와 리스크를 측정하는 기간 사이에 일어난 변화 등으로 인한 데이터 편의를 반영하기에 적합한 모형이고, 반면 베이지안 방법론은 표본 분포와 실제 모집단의 차이에서 오는 표본위험을 반영하기에 적합한 모형이다. 한편 현실에서는 단체 실손의료보험의 경우처럼 데이터 편의와 표본위험이 동시에 존재하는 경우가 상당히 많다. 따라서 두 가지 방법론을 동시에 적용하는 것이 보다 합리적인 방안이라고 할 수 있다.

이에 본 연구는 베이지안 접근법을 이용하여 표본위험를 반영하는 단계와 Heckman과 Meyers의 방법론을 이용하여 데이터 편의를 반영하는 단계로 이루어진 베이지안 H-M CRM 모형을 제안하였다. 아울러 이 모형은 리스크의 측정 뿐 아니라 데이터의 수가 적고 외부 환경의 영향을 크게 받는 보험의 보험료 산출에도 적용할 수 있다.

References

Cho, Y. and Cho, J. (2013). Measuring insurance risk of health insurance using a collective risk model. *Journal of Insurance and Finance*, **24**, 3-35

Heckman, P. E. and Meyers, G. G. (1983). The calculation of aggregate loss distributions form claim severity and claim count distributions. *Proceedings of the Casualty Actuarial Society*, **LXX**, 22-61.

International Actuarial Association. (2004). *A global framework for insurer solvency assessment*, IAA Insurer Solvency Assessment Working Party Research Report, Availabe from `http://www.actuaries.org/LIBRARY/Papers/Global_Framework_Insurer_Solvency_Assessment-public.pdf`.

Luder T. (2005). *Swiss solvency test in non-life insurance*, Federal Office of Private Insurance, 36th ASTIN Colloquium.

Migon, H. S., Edison M. O. and Penna (2006). Bayesian analysis of a health insurance Model. *Journal of Actuarial Practice*, **13**, 61-80.

Migon, H. S. and Moura, F. A. S. (2005). Hierarchical Bayesian collective risk model: An application to health insurance. *Insurance: Mathematics and Economics*, **36**, 119-135.

Pai, J. S. (1997). Bayesian analysis of compound loss distributions. *Journal of Econometrics*, **79**, 129-146.

Journal of the Korean Data & Information Science Society
2016, **27**(1), 9–18

http://dx.doi.org/10.7465/jkdi.2016.27.1.9
한국데이터정보과학회지

Bayesian analysis of insurance risk model with parameter uncertainty

Jaerin Cho[1] · Hyesu Ji[2] · Hangsuck Lee[3]

[1]Department of Financial and Strategy Research, Korea Insurance Research Institute
[2]Actuarial Department, KB Life Insurance
[3]Department of Actuarial Science/Mathematics, Sungkyunkwan University

Received 22 September 2015, revised 29 October 2015, accepted 14 January 2016

Abstract

In the Heckman-Meyers model, which is frequently referred by IAA, Swiss Solvency Test, EU Solvency II, the assumption of parameter distribution is key factor. While in theory Bayesian analysis somewhat reflects parameter uncertainty using prior distribution, it is often the case where both Heckman-Meyers and Bayesian are necessary to better manage the parameter uncertainty. Therefore, this paper proposes the use of Bayesian H-M CRM, a combination of Heckman-Meyers model and Bayesian, and analyzes its efficiency.

Keywords: Bayesian, CRM, insurance risk, parameter uncertainty risk.

[1] Research fellow, Department of Financial and Strategy Research, Korea Insurance Research Institute, Seoul 07328, Korea.
[2] Assistant, Actuarial Department, KB Life Insurance, Seoul 07325, Korea.
[3] Corresponding author: Professor, Department of Actuarial Science/Mathematics, Sungkyunkwan University, Seoul 03063, Korea. E-mail: hangsuck@skku.edu