11장. 보험료에 추세를 반영하려면?

◈ Introduction

최근 보험회사들은 물가나 이자율이 변동할 리스크를 줄이기 위해 갱신형 상품을 늘리는 추세이다. 갱신형 상품은 짧게는 매년, 길게는 10년~15년 주기로 보험료를 갱신하는데, 이때 보험료는 어떤 근거로 바뀌게 될까? 암 최초 진단 시 보험금을 주는 보험상품을 생각해보자. 보험회사는 3년전, 그 당시의 정보를 통해 암진단율을 예측하여 보험료를 정했다. 과연 그 회사는 현재의 보험료를 정할 때 여전히 3년 전의 암진단율을 사용해도 괜찮을까? 기존의 암진단율은 최근 3년의 데이터를 포함하지 않기 때문에 보험회사는 최신화의 필요성을 느낄 것이다. 그렇다면 최근 3년간의 데이터로만 산출한 암진단율을 사용하는 것은 괜찮을까? 이는 3년 전보다 더 오래된 과거의 데이터를 전혀 고려하지 않는다. 실제 암진단율 예측에는 과거의 암진단율과 최신 데이터로 구한 암진단율의 가중평균을 사용하고, 이를 참고하여 보험료를 수정하는 것이다. 이번 시간에는 과거의 위험률과 경험 위험률에 대한 가중치인 신뢰도(Credibility)에 대해 배울 것이다. 신뢰도를 구하는 대표적인 방법으로는 Partial Credibility, Greatest Accuracy Credibility, 그리고 Bühlmann Credibility가 있다.

1. Partial Credibility

Partial Credibility는 데이터의 양으로만 신뢰도를 구하는 방법으로 다른 방법들에 비해 간단하나 데이터의 질을 고려하지 않는 단점이 있다. 식은 다음과 같다.

$$Z(\text{신뢰도}) = \sqrt{\frac{\text{실제 데이터 개수}}{\text{기준 데이터 개수}}}$$

실제 데이터 개수: 앞선 암 진단 보험의 예시에서 볼 때 최근 3년간의 암 진단 건수를 말한다. 즉, 특정 경험 기간 동안에 총 데이터 개수를 의미한다.

기준 데이터 개수: 경험 데이터에 대해 100% 신뢰하여 최신 암진단율(위험율)을 산출할 때 경험 데이터만 사용해도 무방한 수준의 데이터 개수를 의미한다.

$Pr(-k\times E[X]\le X-E[X]\le k\times E[X])=P$ 식을 만족하는 n이 기준 데이터 개수이다. k는 오차범위를 의미하고, P는 변동에 대한 확률수준을 의미한다. 이 식을 구체적으로 설명하기엔 아직 이르니 궁금하면 나중에 책을 참고하도록 하자.

이제 데이터를 통해 계산한 신뢰도를 사용하여 위험률을 최신화 할 수 있다.

최신화된 위험률 = Z× 경험 위험률+(1-Z)× 과거에 산출한 위험률

즉, 계산한 최근 3년의 데이터와 기존 위험률에 각각 Z와 1-Z를 곱하여 가중평균한다. 이는 실제 데이터 개수가 높을수록 경험위험률에 의존하고 반대로 기준 데이터 개수가 높을수록 과거 위험률에 의존한다는 것을 알 수 있다. Partial Credibility 방법은 데이터의 개수와 평균값만 알고 있으면 가중치를 쉽게 구할 수 있어 실무적으로 편리하다. 실제로 보험개발원에서 공표한 2009년 6월 "위험률산출 및 적용에 관한 모범규준"을 보면, 보험회사는 Partial Credibility 방식으로 위험률을 갱신해야 한다.

2. Greatest Accuracy Credibility

Greatest Accuracy Credibility 방법부터는 다소 어려운 Bayesian Statistics의 개념을 사용하기 때문에 대략적으로 흐름만 살펴보도록 하겠다. 지급건수가 포아송분포를 따르고 모수인 λ가 특정 prior distribution을 따른다고 가정하자. 우리가 관측한 3년 데이터를 통해 모수에 대한 믿음이 prior distribution에서 posterior distribution로 바뀐다.

Definition

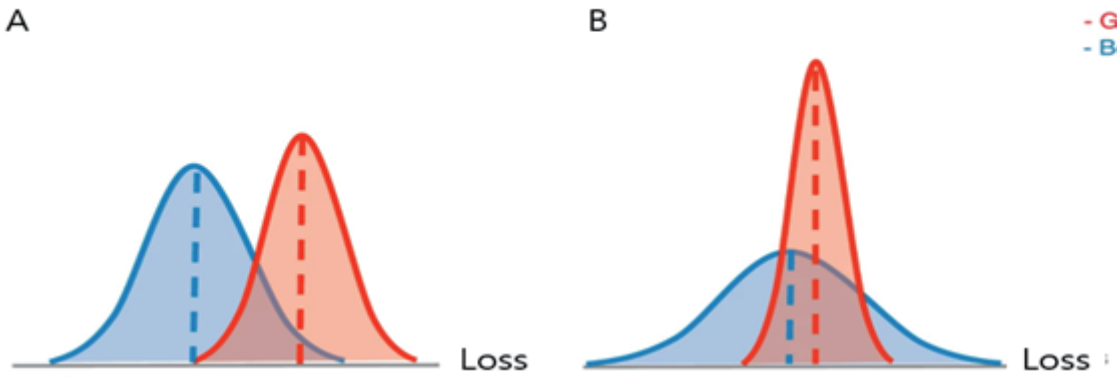
Prior distribution을 $\pi(\theta)$, posterior distribution을 $\pi_{\Theta|X}(\theta|\boldsymbol{X})$, 그리고 지급건수에 대한 model distribution을 $f_{X|\Theta}(\boldsymbol{x}|\theta)$이라고 할 때 posterior distribution은 다음과 같이 표현된다.

$$\pi_{\Theta|X}(\theta|\boldsymbol{X}) = \frac{f_{X|\Theta}(\boldsymbol{X}|\theta)\pi(\theta)}{\int f_{X|\Theta}(\boldsymbol{X}|\theta)\pi(\theta)\,d\theta} \dots (1)$$

또한 조건부 기대값을 통해 다음 예상건수를 수리적으로 표현하면 posterior distribution의 기대값이 된다는 사실을 확인할 수가 있으며 이는 zx+(1-z)μ로 표현된다. (자세한 설명은 Loss Models 책을 참고하기를 바란다.) 즉, 최신화한 위험률이 산출된다. 하지만 이 방법에는 실무적인 문제가 발생한다. 위의 식(1)은 어떤 prior distribution를 선택하느냐에 따라 계산이 복잡해지기도 한다. 첫째는 적절한 prior distribution 선택의 문제, 둘째는 선택한 prior distribution에 따라 posterior distribution을 컴퓨터로 계산하지 못하는 문제가 발생한다.

3. Bühlmann Credibility

앞선 exact accuracy credibility가 갖고 있는 실무적인 문제를 보완하기 위해서 나온 이론이 Bühlmann Credibility이다. 즉, Bühlmann Credibility는 Exact Accuracy Credibility의 Z를 계산하는 것이 어렵기 때문에 Z를 근사한다. Bühlmann Credibility에서 신뢰도 Z는 관측된 데이터의 혼잡도가 얼마나 작은지를 나타낸다. 혼잡도가 정확히 무엇을 뜻하는지 그림 1을 통해 살펴보자.



<그림 1>

보험회사가 보험계약자 갑의 성별을 모른다고 가정해보자. 만약 보험계약자 성별 손실금액 분포가 A 혹은 B와 같이 주어질 때 보험계약자 갑의 손실 금액 데이터로부터 어떻게 성별을 알아낼 수가 있을까? 먼저 남녀의 손실금액 분포들이 A를 따를 때 남성의 손실금액 분포의 평균과 여성의 손실금액의 평균이 육안상으로 구별된다. 또한 각 분포들의 분산도 비교적 작아서 두 분포들이 겹치는 부분이 많지 않아 보인다. 그러므로 갑의 손실 금액 데이터를 통해 어떤 성별인지를 쉽게 유추할 수 있다. 즉, 혼잡도가 낮다. 이와 반대로 만약 계약자들의 성별 손실 금액 분포들이 B와 같다고 가정해보자. 성별 손실 금액 분포의 평균이 육안상으로 구별하기 힘들고 남성 손실 금액 분포의 분산이 커서 두 분포가 겹치는 구간이 많다. 즉, 혼잡도가 높다. 이는 육안상으로만 내린 결론이기에 수리적으로 살펴보면 다음과 같다.

$$\text{EPV(Expected Process Variance)} = E[Var(X|\Theta)]$$

$$\text{VHM(Variance of the Hypothetical Means)} = Var[E(X|\Theta)]$$

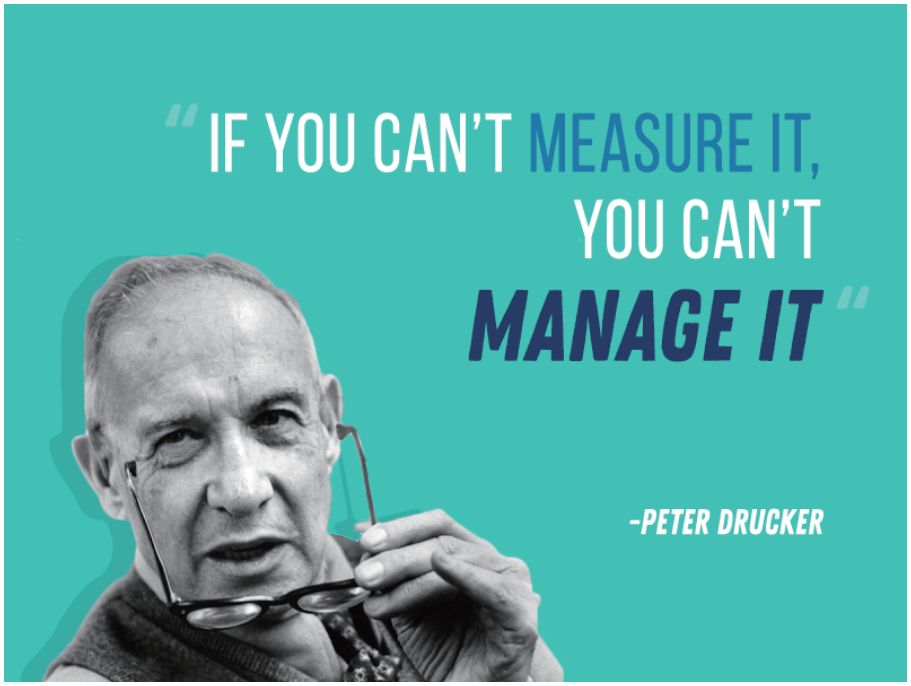
$$k = \frac{\text{EPV}}{\text{VHM}} \qquad Z = \frac{n}{n+k}$$

EPV(Expected Process Variance)는 각 분포의 분산을 평균한 값으로, 즉 분포들이 겹치는 정도를 나타낸다. VHM(Variance of the Hypothetical Means)는 각 분포 평균의 분산으로, 분포들의 사이가 얼마나 떨어져 있는지를 나타낸다.

EPV가 클수록, VHM이 작을수록 혼잡도(k)는 커지고 이에 따라 신뢰도(Z)는 작아진다. Bühlmann Credibility는 Z 근사를 통해 Exact Accuracy Credibility의 계산 문제를 해결해준다. (첫번째 문제는 계리사의 직관의 영역이기 때문에 마땅한 해결책이 없다.) 그래서 Bühlmann Credibility는 실무적으로 Exact Accuracy Credibility 보다 큰 장점을 가진다.

다양한 방법을 통해 산출한 신뢰도는 결국 정형화된 식에 불과하기 때문에 미래를 정확하게 대변해주지 않는다. 때로는 계리사의 직관이 정형화된 식보다 더 중요할 수도 있다. 예를 들면, 경험 데이터의 양이 부족해 추세에 가중치를 덜 부여했지만 나중에 보면 엄청난 변화를 암시하는 데이터였을 수도 있다. 계리사는 외부환경을 항상 염두에 두며 최근 급격히 변하는 상황을 잘 반영하고 과거에 쌓인 데이터에 너무 의존하지 않는 합리적인 보험료를 산출해야한다.

★ 쉬어가는 페이지 ★



"위험을 측정할수 없다면 관리할 수도 없다"

경영학자 피터 드러커는 위와 같은 말을 남겼다. 물론 이 말은 경영 관리 측면에서 해주는 조언이지만 Actuarial Science 와도 맞닿아있다. 계리 측면으로는 It을 Risk로 해석한다. 위험을 제대로 측정하지 못한다면, 미래 변화에 맞춰 위험을 관리할 수 없다는 것이다. 신뢰도도 결국 위험에 대응하기 위한 하나의 방법이다. 신뢰도를 결정할 때는 식만에 의존해서는 위험을 관리할 수 없다. 현재와 같이 COVID-19 상황이 지속된다면 과거 데이터가 아무리 많고 안정적이라도 더 이상 도움이 되진 않는다. COVID-19으로 인해 전과 다른 새로운 추세가 시작됐기 때문이다. 진정한 계리사는 미래에 다가올 위험을 관리하기 위해 항상 외부상황에 주의를 기울이며 위험에 미치는 요인들을 측정할 수 있어야 할 것이다.