

7장. 사고 발생 횟수를 어떻게 모델링할까?

◆ Introduction

정부가 내년 COVID-19으로 인한 의료비용을 추정하길 원한다고 생각해보자. 일반적으로 계리사는 과거 COVID-19 환자의 1인당 의료비와 발병 환자 수를 참고할 것이다. 당신이 정부로부터 환자 수에 대한 확률분포를 모델링해달라는 요청을 받는다면 어떠한 방식으로 모델링을 해야 할까? 사고의 정도를 나타내는 의료비와 달리, 사고 발생 횟수를 나타내는 발병 환자 수는 0을 포함하는 양의 정수로 정확히 떨어진다. 그러므로 당신은 이산형 확률분포 중에서 후보를 찾아야 할 것이다. 수많은 분포들 중 최적의 분포를 선택하는 일은 많은 고민을 요하지만 보통 $(a,b,0)$ class와 $(a,b,1)$ class의 분포들은 좋은 답이 될 수 있다. 이 장에서 $(a,b,0)$ class와 $(a,b,1)$ class 분포가 무엇인지 배우고 이들의 특징을 알아보자.

Definition. $(a,b,0)$ class

Let p_k be the probability function of a discrete random variable. It is a member of the $(a,b,0)$ class of distributions provided that there exist constants a and b such that

$$\text{Let } p_k = \left(a + \frac{b}{k}\right) p_{k-1}, \quad k = 1, 2, 3, \dots \quad (1)$$

(1)의 조건을 충족시키는 분포를 $(a,b,0)$ class 라고 정의한다. 대표적인 $(a,b,0)$ class 분포는 포아송분포, 이항분포 그리고 음이항분포가 있다. $(a,b,0)$ class 분포의 어떠한 특징들이 COVID-19 발병 환자 수의 모델 분포로 사용하기 적합하게 해주는지 분포별로 살펴보자.

확률분포	확률질량함수	모수	평균	분산	a	b	p_0
Poisson Distribution	$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$	λ	λ	λ	0	λ	$e^{-\lambda}$
Binomial Distribution	$p(x) = \binom{m}{x} q^x (1-p)^{m-x}$	m, q	mq	$mq(1-q)$	$-\frac{q}{1-q}$	$(m+1)\frac{q}{1-q}$	$(1-q)^m$
Negative Binomial Distribution	$p(x) = \binom{x+r-1}{x} \left(\frac{1}{1+\beta}\right)^r \left(\frac{\beta}{1+\beta}\right)^x$	r, β	$r\beta$	$r\beta(1+\beta)$	$\frac{\beta}{1+\beta}$	$(r-1)\frac{\beta}{1+\beta}$	$(1+\beta)^{-r}$

7장. 사고 발생 횟수를 어떻게 모델링할까?

1. Poisson Distribution

포아송분포의 경우 평균과 분산이 같다는 첫번째 특징이 있다. 만약 발병 환자 수의 평균과 분산이 비슷하다면 포아송 분포가 모델 분포로 적합하다는 근거가 될 수 있다. 두번째 특징은 독립된 포아송 분포를 따르는 확률변수들의 합으로 이루어진 확률변수 또한 포아송 분포를 따른다는 사실이다. 만약 1년 동안의 발병 환자 수의 분포가 포아송 분포라고 가정할 경우, 두번째 특징으로 인해 계리사는 리스크를 세분화하는데 실무적 편리함의 장점을 가질 수가 있다. 예를 들어 만약 1년 동안의 발병 환자 수들이 의료비의 크기에 따라 세분화할 수 있다면 각 세분화된 셀의 발병 환자 수들 또한 포아송 분포를 따른다. 만약 정부가 의료비 예산을 줄이기 위해서 특정 의료비 미만은 보장해주지 않는다면 해당 발병 환자 수를 빼더라도 남은 발병 환자 수들은 여전히 포아송 분포를 따른다. 여기서 주의해야하는 것은 포아송 분포의 모수는 바뀐다는 점이다. 만약 의료비의 크기에 따라 10만원 이하, 10만 이상 50만원 이하 그리고 50만원 이상인 발병 환자 수가 각각 총 발병 환자 수의 30%, 20%, 그리고 50%를 차지한다고 하면 각 영역의 포아송 분포의 모수는 원래의 평균인 λ 값에 0.3, 0.2 그리고 0.5를 곱한 값이 되겠다.

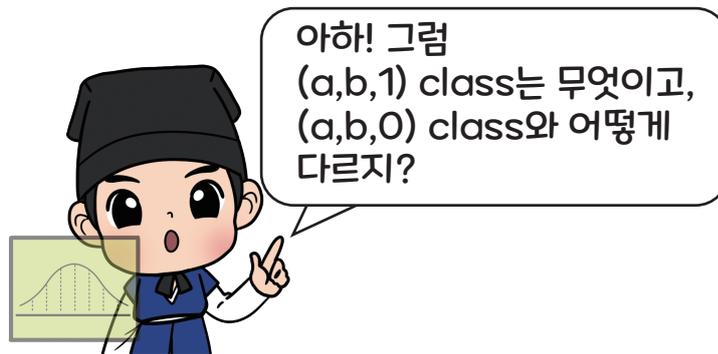
2. Binomial Distribution

이항분포의 분산은 평균보다 작기 때문에 발병 환자 수 분포의 분산이 평균보다 작게 관측될 경우에 적합하다. 또한 같은 리스크를 공유하는 집단을 대상으로 유용하게 사용할 수 있다. 예를 들어 같은 연령에서 비슷한 건강상태를 유지하는 개인들을 모아둔 집단 생명보험을 생각해보자. 같은 리스크를 공유하는 집단이기에 개개인들이 1년 안에 사망할 확률은 비슷할 것이고 계약자들의 수를 사전에 알고 알 수 있기 때문에 1년 동안의 보험청구건수가 이항분포를 따를 것이라는 근거가 된다. 마지막 특징은 이항분포의 모수를 (n, q) 라고 가정을 했을 때, n 을 초과하는 청구건수의 값을 관측하지 못한다는 사실이다. 사전에 모델 선택자가 청구 건수가 제한되어 있는 보험 상품을 모델링 한다면 이항분포의 사용을 고려할 수 있다.

7장. 사고 발생 횟수를 어떻게 모델링할까?

3. Negative Binomial Distribution

음이항분포의 분산은 평균보다 크다. 그러므로 발병 환자 수 분포의 분산이 평균보다 크게 관측될 경우에 음이항분포가 유용하게 쓰일 수가 있다. 음이항분포는 확률변수가 포아송 분포를 따르고 분포의 모수가 감마분포를 따르는 mixing된 분포이기도 하다. 즉, 모델 선택자가 발병 환자 수의 분포를 포아송 분포로 사용하고 싶지만 모수에 대한 확신이 없을 경우 음이항분포를 사용할 수 있다.



Definition. (a,b,1) class

Let p_k be the pf of a discrete random variable. It is a member of the (a,b,1) class of distributions provided that there exist constants a and b such that

$$\text{Let } p_k = \left(a + \frac{b}{k}\right) p_{k-1}, \quad k = 2, 3, \dots \quad (2)$$

식 (1)은 $k=1$ 부터 시작하지만, 식 (2)는 $k=2$ 부터 시작한다. 즉, (a,b,0) class 분포들은 p_0 부터 관계식이 성립하나 (a,b,1) class 분포들은 p_1 부터 관계식이 성립한다는 말인데, (a,b,0) class에서 p_0 값만 인위적으로 바꾸고 바뀐 값만큼은 다른 p_k 값들에 비율적으로 할당한 분포가 (a,b,1) class 분포인 것이다. 항공보험을 예로 들어보자. 비행기 사고는 매우 드물게 발생하기 때문에 보험금 청구건수가 0일 확률이 높은 편이다. p_0 는 보험금 청구건수가 0건일 확률을 나타내기 때문에, 사고 발생 확률이 낮지만 매우 높은 위험을 담보하는 위 예시와 같은 상품에서는 p_0 의 설정이 보험료에 큰 영향을 미친다. 그러므로 회사의 경험데이터나 직관을 통해 p_0 값을 변경해야 할 때 (a,b,1) class 분포를 사용하게 된다.



★ 쉬어가는 페이지 ★



사진출처 : 한국공제신문

“모 아니면 도, P. 가 높은 보험”
 (높은 보험료 수익) (막대한 보험금)

‘도쿄올림픽 연기, 보험사 영향은?’

코로나19 여파로 도쿄올림픽이 1년 연기되면서 올림픽 상품을 다루는 보험회사들이 감당해야 할 손해액은 얼마인지 관심이 쏠리고 있다. 로이터통신에 따르면 올림픽 연기로 인한 보험손해는 런던 로이드, 뮤닉리, 스위스리 등이 감당할 것으로 보인다. 또한 8억 달러 상당의 보험을 들어놓은 IOC는 거의 손해를 보지 않을 전망이다. 올림픽이 연기된 것은 124년 역사상 처음이다. 이와 관련, 글로벌 투자 은행 제프리스(Jefferies)의 분석가들은 TV방송권과 기업 후원을 포함해 보험사들이 20억 달러(약 2조4180억원)의 보험손실과 6억 달러(약 7255억원)의 관광산업손해가 발생할 것으로 추정했다.

IOC, 주최자, 스폰서 및 관광업계 등이 올림픽 연기로 인한 손해를 보험회사에 청구할 수 있지만 실제 이 과정은 복잡해 보인다. 올림픽 전면 취소가 아닌 만큼, 내년에 다시 개최하고 나면 생기는 수익금과 올림픽 지연에 따른 손해를 따져야 하기 때문이다. 보험업계에 따르면 연기에 불구하고 회사들이 올림픽에서 돈을 벌 수 있기 때문에 올림픽이 끝날 즈음엔 보험금이 줄어들지만, 이미 지출한 자금 이외에 행사 재정비를 위한 추가비용이 발생할 수 있다는 것이다. 원수보험사와 보험위험을 공유하는 전업 재보험사 중 뮤닉리의 관계자는 도쿄올림픽에 5억 달러 정도의 담보를 제공하고 있다고 밝혔으나 회사 측은 공식 논평을 거부했다.

출처 : 한국공제신문 박형재 기자