

국내 태풍 예측

양 희 중*

*청주대학교 산업공학과

Predicting typhoons in Korea

Heejoong Yang*

*Dept. of Industrial Engineering, Cheongju University

Abstract

We develop a model to predict typhoons in Korea. We collect data for typhoons and classify those depending on the severity level. Following a Bayesian approach, we develop a model that explains the relationship between different levels of typhoons. Through the analysis of the model, we can predict the rate of typhoons, the probability of approaching Korean peninsular, and the probability of striking Korean peninsular.

We show that the uncertainty for the occurrence of various types of typhoons reduces dramatically by adaptively updating model parameters as we acquire data.

Keywords: Predicting Typhoons, Bayesian Approach, Prior Distributions, Posterior Distributions, Influence Diagrams

1. 서론

국내에는 매년 여름 수차례의 태풍이 지나가고 있다. 다행히도 태풍의 진로가 대만이나 일본으로 빗겨가는 경우가 상당 수 있어서 그 피해가 인접 국가보다 적은 편이기는 하지만 예외 없이 매년 북태평양 남서부에서 2, 30건 정도의 태풍이 발생하고 있고 이들은 우리나라를 향해 북진하고 이 중 일부는 우리나라를 근접하여 지나가거나 혹은 우리나라 내륙에 상륙하여 여전히 많은 피해를 주고 있다.

이러한 태풍은 우리가 사전에 발생을 방지할 수가 없는 자연재해이기 때문에 적어도 태풍 발생 및 경로

에 관한 지식을 습득하는 것이 관련 피해 방지를 위한 정책 마련 등에도 도움이 되는 일일 것이다.

본 논문에서는 태풍의 전체 발생 빈도를 예측하고 발생하는 태풍을 그 진로에 따라 국내에 별 영향을 미치지 않고 빗겨가는 태풍, 국내에 근접하는 태풍, 국내에 상륙하는 태풍 등 3가지 유형으로 나누어 각 각에 대한 발생빈도 및 국내근접, 국내 상륙 확률을 예측할 수 있는 모델을 개발한다. 특히 태풍 한 건 한 건에 포함되어 있는 정보를 최대한 활용하기 위하여 각기 다른 유형의 태풍 자료도 서로 연관 지어 활용하여 자료 활용성을 최대화할 수 있는 모형을 개발하고자한다.

† 이 논문은 2014-2016학년도 청주대학교 산업과학연구소가 지원한 학술연구조성비(특별연구과제)에 의해 연구되었음.

† Corresponding Author : Heejoong Yang, Industrial Engineering, Cheongju University

M·P : 010-2456-8188, E-mail: hjyang57@naver.com

Received January 20, 2015; Revision Received March 16, 2015; Accepted March 17, 2015.

2. 기존 연구 동향

그동안 태풍 및 이로 인한 피해 예측에 관한 연구는 그다지 많이 이루어지지 않고 있다. 이창수 등은[4] 보험 상품 개발을 위해 날씨 정보 확률 분석하는 방향으로 연구를 진행하였으며 서승남, 김상익은[3] 태풍으로 인한 해일 피해예측에 관한 연구를 하였다. 김백조[1]는 태풍예측 향상을 위한 전략수립에 관한 연구를 진행하였고 나인채등은[2] 온대저기압화 과정을 비교하는데 저기압위상 공간도를 활용하였다.

그러나 태풍의 발생 빈도와 진행 경로 등을 예측하는 연구는 극소수 이루어지고 있다. 국내를 강타하여 엄청난 피해를 입히는 태풍의 발생 확률은 작지만 일단 발생하면 피해가 치명적이 되는 특성이 있어 희귀 발생-심각한피해 (rare event - critical casualty) 유형의 재해로 분류될 수 있다[7], [9]. 태풍을 비롯한 희귀발생-심각한피해 유형의 재해가 어떻게 시작되고 어떤 경로를 통해 어떻게 더욱 치명적인 재해로 진행되는지를 체계적으로 설명하기위해 PRA (Probabilistic Risk Analysis) 기법이 주로 활용되고 있으며 여기서는 사건 가지(Event Tree) 와 결함나무 (fault tree) 분석을 종종 이용하고 있다[5], [8]. 그러나 이러한 분석은 사건 가지를 자료가 입수되는 대로 바로 보정할 수 없다는 단점이 있다. 또한 파라미터간의 상호 독립성 여부도 파악할 수 없다는 단점도 있다. 이에 대한 보완으로 사건 가지를 상호연관도로 변형시켜 모델 파라미터를 보정하는 방법이 연구되었다[6,10,11].

국내에 상륙하여 많은 피해를 입히는 태풍은 흔히 발생하는 것이 아니기 때문에 많은 자료의 확보가 쉽지 않다. 따라서 피해 규모는 작지만 상대적으로 빈번히 발생하고있는 다른 유형의 태풍, 즉, 국내에 상륙하지 않고 지나가는 태풍, 혹은 국내에 상륙하지는 않지만 근접하여 지나가는 태풍에 포함되어 있는 정보까지도 최대한 효율적으로 활용할 수 있어야한다. 또한 일반적인 공학지식과 전문가의견까지도 효과적으로 반영될 수 있어야한다. 그러므로 상이한 유형의 태풍 자료로부터도 정보를 발췌해낼 수 있는 모델, 전문가 의견 및 과거 기록을 포함시킬 수 있는 모델, 자료가 입수되는 대로 즉각 자료 보정이 가능한 모델이 구축되어야 한다. 본 연구에서는 이러한 모델을 구축하고 모델 파라미터를 보정하는 방법을 연구한다. 이렇게 개발된 모델에 기초하여 과거 태풍자료를 분석하고 각 유형의 태풍 발생률을 예측한다.

3. 태풍 자료 분석

본 절에서는 현재 수집되어 있는 태풍 발생 자료를 면밀히 분석하여 어느 정도의 태풍이 발생하고 이 중 국내 근접, 국내 상륙 태풍은 얼마나 되는지, 각 각이 어떤 상호관련성이 있는지를 분석한다. (출처:기상청)

<Table 1> number of typhoon occurrence per year

| year | number of typhoon occurrence | number of typhoons approaching Korean peninsular | number of typhoons striking Korean peninsular |
|------|------------------------------|--|---|
| 1951 | 21 | 3 | 1 |
| 1952 | 27 | 4 | 2 |
| 1953 | 23 | 3 | 0 |
| 1954 | 21 | 3 | 0 |
| 1955 | 28 | 3 | 0 |
| 1956 | 23 | 4 | 1 |
| 1957 | 22 | 2 | 1 |
| 1958 | 31 | 1 | 0 |
| 1959 | 23 | 7 | 1 |
| 1960 | 27 | 3 | 1 |
| 1961 | 29 | 5 | 1 |
| 1962 | 30 | 4 | 3 |
| 1963 | 24 | 3 | 1 |
| 1964 | 34 | 4 | 0 |
| 1965 | 32 | 3 | 0 |
| 1966 | 35 | 3 | 1 |
| 1967 | 39 | 1 | 0 |
| 1968 | 27 | 3 | 0 |
| 1969 | 19 | 1 | 0 |
| 1970 | 26 | 4 | 0 |
| 1971 | 36 | 3 | 2 |
| 1972 | 31 | 4 | 0 |
| 1973 | 21 | 3 | 1 |
| 1974 | 32 | 4 | 0 |
| 1975 | 21 | 2 | 0 |
| 1976 | 25 | 6 | 0 |
| 1977 | 21 | 2 | 0 |
| 1978 | 30 | 4 | 1 |
| 1979 | 24 | 2 | 2 |
| 1980 | 24 | 3 | 1 |
| 1981 | 29 | 5 | 0 |
| 1982 | 25 | 4 | 0 |
| 1983 | 23 | 1 | 0 |
| 1984 | 27 | 3 | 3 |
| 1985 | 27 | 5 | 1 |
| 1986 | 29 | 3 | 0 |

| | | | |
|------|----|---|---|
| 1987 | 23 | 3 | 1 |
| 1988 | 31 | 0 | 0 |
| 1989 | 32 | 2 | 0 |
| 1990 | 29 | 4 | 2 |
| 1991 | 29 | 5 | 1 |
| 1992 | 31 | 2 | 2 |
| 1993 | 28 | 5 | 0 |
| 1994 | 36 | 5 | 3 |
| 1995 | 23 | 3 | 2 |
| 1996 | 26 | 2 | 0 |
| 1997 | 28 | 5 | 0 |
| 1998 | 16 | 2 | 1 |
| 1999 | 22 | 5 | 4 |
| 2000 | 23 | 5 | 0 |
| 2001 | 26 | 1 | 0 |
| 2002 | 26 | 4 | 0 |
| 2003 | 21 | 4 | 1 |
| 2004 | 29 | 5 | 0 |
| 2005 | 23 | 1 | 0 |
| 2006 | 23 | 3 | 1 |
| 2007 | 24 | 3 | 1 |
| 2008 | 22 | 1 | 0 |
| 2009 | 22 | 0 | 0 |
| 2010 | 14 | 3 | 2 |
| 2011 | 21 | 3 | 0 |
| 2012 | 25 | 4 | 3 |
| 2013 | 31 | 3 | 0 |

<Table 1>에서 볼 수 있듯이 과거 60여년 간의 자료에서 해마다 큰 변화 없는 수의 태풍이 발생하고 있고 이 중 일부가 국내에 근접하거나 상륙하고 있는 것으로 볼 수 있다. 그러나 유심히 살펴보면 전체 태풍 발생 수 대비 국내 근접 태풍 수와 국내 상륙 태풍 수의 비율은 일정하지 않은 것을 알 수 있다. 이들 간의 연관성 분석을 위해 각 유형의 태풍을 아래와 같이 정의한다.

- 빗겨간 태풍 : 태풍은 발생했으나 국내에 근접하지 않아 큰 피해를 미치지 않는 태풍
- 국내근접 태풍 : 발생한 태풍이 국내에 상륙하지는 않았으나 근처로 접근하는 태풍
- 국내상륙 태풍 : 발생한 태풍이 국내에 상륙하는 태풍

이러한 서로 다른 경로의 태풍 발생 수를 아래와 같이 정의한다;

- $m_0(T)$: T 기간 동안 국내를 빗겨간 태풍 수
- $m_1(T)$: T 기간 동안 국내에 근접한 태풍 수
- $m_2(T)$: T 기간 동안 국내에 상륙한 태풍 수

이들을 상호 연관성을 고려하여 분석하기 위하여 level 0, 1, 2 태풍으로 아래와 같이 새롭게 정의한다.

- level 0 태풍 : 국내상륙태풍, 국내근접태풍, 빗겨간 태풍, 즉 모든 태풍.
- level 1 태풍 : 국내상륙태풍과 국내근접태풍
- level 2 태풍 : 국내상륙태풍

따라서 level 1 태풍은 level 0 태풍의 subset 이 되고 level 2 태풍은 level 1태풍의 subset 이 된다. 또한 각 level 의 태풍수를 아래와 같이 정의한다;

- $n_0(T)$: T 기간 동안 발생한 level 0 태풍 수
 - $n_1(T)$: T 기간 동안 발생한 level 1 태풍 수
 - $n_2(T)$: T 기간 동안 발생한 level 2 태풍 수
- 그러므로

$$n_0(T) = m_2(T) + m_1(T) + m_0(T)$$

$$n_1(T) = m_2(T) + m_1(T)$$

$$n_2(T) = m_2(T)$$

의 관계가 성립하며 표1에서 태풍발생 수는 $n_0(T)$, 국내접근 수는 $n_1(T)$, 상륙태풍 수는 $n_2(T)$ 를 나타낸다.

여기에서 level 0 태풍수와 level 1 태풍수간의 상관 분석의 결과는 상관계수가 0.126, p-value 가 0.330 으로 나타났으며, level 0 태풍수와 level 2 태풍수간의 상관분석의 결과는 상관계수가 -0.005, p-value 가 0.971 로 나타났다. 또한 level 1 태풍수와 level 2 태풍수간의 상관분석의 결과는 상관계수가 0.244, p-value 가 0.056 으로 나타났다. level 0 태풍은 level 1,2 태풍을, level 1 태풍은 level 2 태풍 수를 포함하고 있기 때문에 이들 간의 어느 정도의 상관관계는 당연히 존재할 수밖에 없다. 이를 고려하면 상관 분석의 결과는 예상보다 그 정도가 크지는 않다. 전체 태풍 발생 건수가 많은 해에는 많은 수의 태풍이 국내에 근접하거나 상륙할 것이라는 상식적인 예상을 뒤집는 결과이다.

위의 상관분석 결과를 보면 국내 근접 및 상륙 태풍 건수를 전체 태풍 발생 수의 일부 비율로서 예측할 수 없고 이들 간의 유기적인 정보를 활용할 수 있는 예측 모형 개발이 요구된다.

4. 새로운 예측모형의 개발

어떤 재해 및 사고가 발생하여 심각도가 다른 유형으로 확산될 수 있는 재해 및 사고의 분석에는 결함나무 및 사건가지를 이용하여 발생 및 확산 경위를 분석하는 방법들이 종종 사용되고 있다[7],[8],[9]. 그러나 이러한 접근 방식에는 모델 파라미터간의 상호연관성 분석이 어렵고 자료가 입수 되는대로 파라미터를 보정할 수 없는 단점이 존재한다는 것이 연구되었다[11].

따라서 본 논문에서는 태풍발생 및 국내 근접, 상륙으로의 확산 과정을 결함나무, 사건가지 분석 대신 상호연관도를 통해 도해하고 여기에 포함된 모델 파라미터를 실시간 보정하는 방법을 제시한다. 상호연관도 모형을 통해 사고 발생 및 확산과정을 명확히 이해하고 여기에 관련된 모델 파라미터의 사전 및 사후분포를 구하는데 중점을 둔다. 사전 분포 가정 후 자료가 입수 되는대로 실시간 사후분포가 구해지는 방법을 제시한다.

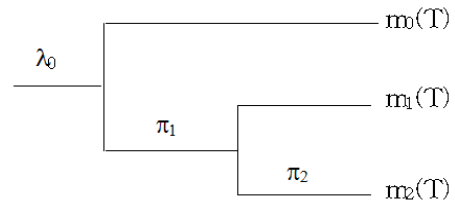
전체 태풍수의 일정 비율이 국내 근접 및 상륙하는 것은 아니지만 무시할 수 없는 상호 유기적 관계를 분석하고 태풍 발생 중 일부가 국내로 근접하고 그 중 일부가 상륙하는 현상을 반영하여 각 유형의 태풍 발생 자료를 서로 다른 유형의 태풍 발생에 활용할 수 있는 의존형 예측모형(dependent model)을 개발한다. 이를 위해서는 실시간 모델 파라미터를 보정할 수 있는 베이지안 기법을 이용하여 태풍 발생률을 예측한다.

4.1. 모형개발

태풍은 매년 북태평양 남서부에서 어떠한 발생률에 따라 발생하고 이 태풍이 어떠한 확률로 국내에 근접하며 국내근접 태풍 중 어떠한 확률로 국내에 상륙한다. 우리가 예측해야 할 이러한 태풍 발생률과 국내 근접, 혹은 상륙하여 보다 심각한 태풍으로 진행할 확률 등을 아래와 같이 정의한다;

- l0 : level 0 태풍 발생률
- l1 : level 1 태풍 발생률
- l2 : level 2 태풍 발생률
- p1 : level 0 태풍이 level 1 태풍으로 진행할 확률
- p2 : level 1 태풍이 level 2 태풍으로 진행할 확률

그러면 각 level 태풍간의 관계는 아래의 그림 1과 같은 사건가지로 표현될 수 있다.



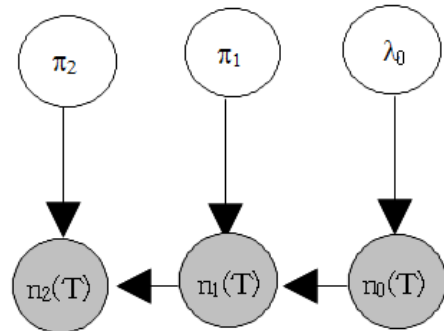
[Figure 1] An event tree showing typhoon occurrence

[Figure 1]에서 보는 것처럼 λ_0 의 발생률로 태풍이 발생하고 있으며 이중 π_1 의 확률로 국내에 접근하게 된다. 또 국내에 접근한 태풍 중 π_2 의 확률로 국내에 상륙하게 된다. 따라서 각 level 태풍의 발생률과 진행 확률 간에는 아래의 관계가 성립한다.

$$\lambda_1 = \lambda_0 \times \pi_1$$

$$\lambda_2 = \lambda_1 \times \pi_2 = \lambda_0 \times \pi_1 \times \pi_2$$

[Figure 1]과 같은 사건가지를 그리고 보면 태풍이 어떻게 발생하여 보다 심각한 태풍으로 어떻게 진행해 가는지를 명확히 볼 수 있다. 그러나 앞 절에서 지적되었던 단점을 보완하기위해 그림 2와 같이 상호연관도로서 [Figure 1]의 사건가지를 다시 표현한다.

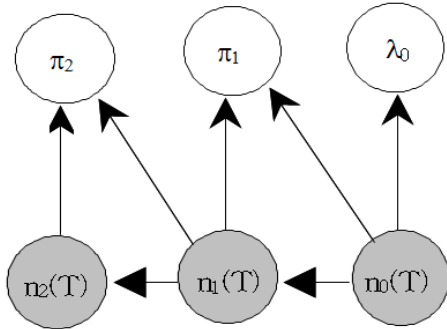


[Figure 2] An influence diagram which is statistically equivalent to figure 1

[Figure 2]의 상호 연관도에서 태풍 발생률, λ_0 , 와 보다 심각한 태풍으로 진행하는 확률, π_1 , π_2 , 간에는 서로 독립적임을 보여주고 있다. 총 태풍 수 $n_0(T)$ 는 태풍 발생률 λ_0 에만 의존하고 있으며 국내근접 태풍 수 $n_1(T)$ 는 총 태풍 수 $n_0(T)$ 와 근접확률 π_1 에만 의존하고 있다. 마찬가지로 국내 상륙 태풍 수 $n_2(T)$ 는 국내 근접 태풍 수 $n_1(T)$ 와 상륙 확률 π_2 에만 의존하고 있는 것을 보여주고 있다. 즉, 태풍 발생률 λ_0 에 따라 태풍이 발생하며 T 기간 중 발생한 태풍 수는 $n_0(T)$ 가 된다. 발생한 태풍은 π_1 의 확률로 국

내 근접 태풍으로 진행할 수 있으며 그 수는 $n_1(T)$ 가 된다. 이 중 π_2 의 확률로 국내에 상륙하게 되고 그 수는 $n_2(T)$ 가 된다.

이러한 태풍 발생 및 진행 모델이 완성 되면 다음은 데이터가 수집 되는대로 모델 파라미터를 보정하는 과정이 된다. 아래 [Figure 3]은 모델 파라미터가 보정 되는 과정을 보여주고 있다.



[Figure 3] Parameter updating process

이 [Figure 3]의 상호연관도는 이론적으로 아래 식 (1), (2), (3)의 사후분포 입수 과정을 나타낸다.

$$p(\lambda_0 | n_0(T)) = \int p(\lambda_0) p(n_0(T) | \lambda_0) d\lambda_0 \quad (1)$$

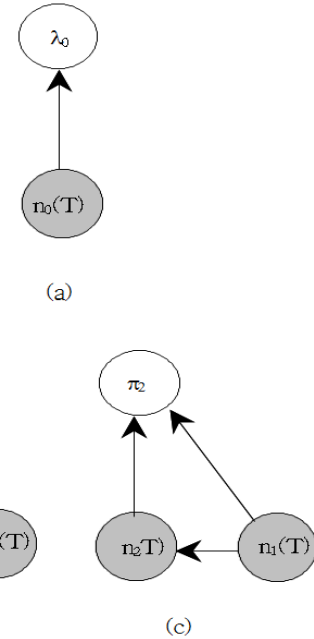
$$p(\pi_1 | n_0(T), n_1(T)) = \int p(\pi_1) p(n_1(T) | n_0(T), \pi_1) d\pi_1 \quad (2)$$

$$p(\pi_2 | n_1(T), n_2(T)) = \int p(\pi_2) p(n_2(T) | n_1(T), \pi_2) d\pi_2 \quad (3)$$

각 모델 파라미터의 사후분포는 위 식과 같이 구해지며 파라미터들의 결합확률분포는 아래 식(4)와 같이 구해진다.

$$p(\lambda_0, \pi_1, \pi_2 | n_0(T), n_1(T), n_2(T)) = \iiint p(\lambda_0 | n_0(T)) p(\pi_1 | n_0(T), n_1(T)) p(\pi_2 | n_1(T), n_2(T)) d\lambda_0 d\pi_1 d\pi_2 \quad (4)$$

위 식(1), (2), (3)에서 보듯이 사후 파라미터는 그림 4와 같이 각각 따로 따로 보정 될 수 있음을 알 수 있다.



[Figure 4] Parameter updating processes for each parameter

이는 베이زي안 접근 방식에서 나타날 수 있는 상당히 복잡한 계산 과정을 간편히 해준다. 모델 파라미터의 사후분포를 구하는 과정에서 λ_0 는 관찰한 모든 유형의 자료 중 $n_0(T)$ 에만 의존하며 π_1 은 $n_0(T)$ 와 $n_1(T)$, π_2 는 $n_1(T)$ 와 $n_2(T)$ 에만 의존한다는 것을 볼 수 있다.

λ_0, π_1, π_2 의 사후분포는 베이즈정리에 의해 사전분포와 우도를 이용해서 구해지며 이들의 결합분포는 각각 사후분포의 곱으로 구해진다.

여기서 보듯이 전체 태풍 발생률 λ_0 는 전체 태풍 수 $n_0(T)$ 에만 의존하지만 국내 근접 태풍 발생률 λ_1 은 전체 태풍 수 $n_1(T)$ 와 $n_0(T)$, 국내 상륙 태풍 발생률 λ_2 는 $n_2(T)$ 와 $n_1(T)$ 에도 의존함을 알 수 있다.

5. 자료분석 및 태풍 예측

5.1 사전, 사후, 예측 분포

위절에서 설명한 베이زي안 예측기법에 따라 본 절에서는 실제 자료를 적용하여 태풍 발생을 예측한다. 태풍 발생률 λ_0 의 사전 분포로서는 감마분포를 이용한다. 태풍 발생률의 분포는 우측으로 치우친 분포를 따를 것으로 예상할 수 있다. 감마분포는 모델 파라미터의 값에 따라 상당히 신축성 있게 모양이 변할 수 있으며 우로 치우친 분포 모양도 잘 설명할 수 있기 때문에 발생률의 사전 분포로서 아래와 같은 감마분포를 선택한다.

$$\lambda_0 \sim \Gamma(\alpha_0, \beta_0)$$

$$p(\lambda_0) = \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} \lambda_0^{\alpha_0 - 1} (1 - \lambda_0)^{\beta_0 - 1}$$

여기서 감마분포를 따르는 파라미터 λ_0 의 평균과 분산은 아래와 같다;

$$E[\lambda_0] = \frac{\alpha_0}{\beta_0}, \quad Var[\lambda_0] = \frac{\alpha_0}{\beta_0^2}$$

일정 기간 T 동안의 태풍발생 수 $n_0(T)$ 의 우도는 태풍발생률 λ_0 가 주어졌을 경우 포아송 분포를 따른다고 가정할 수 있다. 태풍은 서로 독립적으로 발생하기 때문에 포아송 분포 가정이 정당화 될 수 있다:

$$n_0(T) | \lambda_0 \sim \text{Poisson}(\lambda_0 T)$$

또한 국내 근접 태풍, 국내 상륙 태풍으로의 확산 확률 π_1, π_2 의 사전분포로는 베타분포를 이용한다. 베타분포는 0과 1사이의 거의 모든 모양의 분포를 설명할 수 있는 분포이기 때문에 확률의 사전 분포로서 적절하다:

$$\pi_1 \sim \text{Be}(a_1, b_1), \quad \pi_2 \sim \text{Be}(a_2, b_2)$$

$$p(\pi_1) = \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)\Gamma(b_1)} \pi_1^{a_1 - 1} (1 - \pi_1)^{b_1 - 1}$$

$$p(\pi_2) = \frac{\Gamma(a_2 + b_2)}{\Gamma(a_2)\Gamma(b_2)} \pi_2^{a_2 - 1} (1 - \pi_2)^{b_2 - 1}$$

여기서 베타분포를 따르는 파라미터 $\pi, i=1, 2$ 의 평균과 분산은 아래와 같다;

$$E[\pi_i] = \frac{a_i}{a_i + b_i}, \quad Var[\pi_i] = \frac{a_i b_i}{(a_i + b_i)^2 (a_i + b_i + 1)}$$

전체 태풍 수 $n_0(T)$ 와 국내 근접 태풍으로 진행할 확률 p_1 이 주어졌을 경우 국내 근접 태풍 수 $n_1(T)$ 의 우도는 이항분포로 가정한다. 마찬가지로 국내 근접 태풍 수 $n_1(T)$ 와 국내 상륙 태풍으로 진행할 확률 π_2 가 주어졌을 경우 국내 상륙 태풍 수 $n_2(T)$ 의 우도는 이항분포로 가정한다. 역시 각 태풍이 서로 독립적으로 국내 근접 및 상륙 태풍으로 진행되기 때문에 이항 분포의 가정이 타당하다고 할 수 있다:

$$n_1(T) | n_0(T), \pi_1 \sim \text{Bin}(n_0(T), \pi_1)$$

$$n_2(T) | n_1(T), \pi_2 \sim \text{Bin}(n_1(T), \pi_2)$$

λ_0 에 가정된 감마분포는 $[n_0(T) | \lambda_0]$ 에 가정된 우도, 포아송분포의 conjugate prior 가되기 때문에 자료 입수 후 식 (1)에 의해 구해지는 λ_0 의 사후분포 역시 감마분포가 된다:

$$\lambda_0' = [\lambda_0 | n_0(T)] \sim \Gamma(\alpha_0', \beta_0')$$

여기서

$$\alpha_0' = \alpha_0 + n_0(T)$$

$$\beta_0' = \beta_0 + T$$

가 된다.

그리고 level 0 태풍에서 level 1 태풍, 그리고 level 1 태풍에서 level 2 태풍으로 진행할 확률 π_1 과 π_2 의 사후분포는 식 (2) 와 식 (3) 에 의해 각각 구해진다. π_1 과 π_2 에 가정된 베타분포는 $[n_1(T) | n_0(T), \pi_1], [n_2(T) | n_1(T), \pi_2]$ 에 가정된 우도, 이항분포의 conjugate prior 가되기 때문에 자료 입수 후 식 (2) 와 식 (3) 에 의해 구해지는 p_1 과 p_2 의 사후분포 역시 베타분포가 된다:

$$\pi_1' = [\pi_1 | n_1(T), n_0(T)] \sim \text{Be}(a_1', b_1')$$

$$\pi_2' = [\pi_2 | n_2(T), n_1(T)] \sim \text{Be}(a_2', b_2')$$

여기서

$$a_1' = a_1 + n_1(T), \quad b_1' = b_1 + n_0(T) - n_1(T)$$

$$a_2' = a_2 + n_2(T), \quad b_2' = b_2 + n_1(T) - n_2(T)$$

가 된다.

5.2 과거 자료를 이용한 태풍 발생률 예측

다음으로는 태풍 발생률 λ_0 의 사전분포인 감마분포의 파라미터 a_0, b_0 의 값을 추정한다.

먼저 1951년부터 2000년까지의 자료로부터 사전 분포의 파라미터를 추정하고 매년 입수되는 자료로서 파라미터를 보정해간다. 2000년까지의 자료를 살펴보면 50년간 1343건의 태풍이 발생했다. 그 연간 평균 수는 26.860건, 분산은 23.717건²이다. 따라서 감마 분포의 파라미터 a_0, b_0 값이 위 평균과 분산을 만족시키도록 사전분포의 파라미터 값을 구한다.

$$E[\lambda_0] = \frac{\alpha_0}{\beta_0} = 26.860, \quad Var[\lambda_0] = \frac{\alpha_0}{\beta_0^2} = 23.717$$

위 식에서 $a_0 = 30.420$, $b_0 = 1.133$ 이 얻어진다.

이후 2001년부터 2013년까지의 자료는 자료가 입수 되는대로 사전 파라미터를 보정하는데 사용한다. 감마분포의 사후 파라미터는 관측기간, T, 과 그 기간 동안 추가로 발생한 총 태풍 수, $n_0(T)$, 에만 영향을 받으므로 사후분포의 파라미터는 쉽게 구해질 수 있다;

$$a_0' = a_0 + n_0(T) = 30.42 + 307 = 337.42$$

$$b_0' = b_0 + T = 1.133 + 13 = 14.133$$

따라서 연간 태풍 발생수의 연간 사후 평균과 분산은 아래와 같이 변화된다;

$$\lambda_0' = [\lambda_0 | n_0(T)] \sim \Gamma(a_0', b_0')$$

$$E[\lambda_0'] = \frac{\alpha_0'}{\beta_0'} = 23.875, \quad Var[\lambda_0'] = \frac{\alpha_0'}{(\beta_0')^2} = 1.689$$

위에서 볼 수 있듯이 연간 태풍 발생률의 사후 평균은 사전 평균과 큰 차이가 없지만 사후 분산은 사전 분산보다 현저히 줄어들고 있음을 볼 수 있다. 자료가 입수 되는대로 태풍발생률의 분포가 보정되어 불확실성이 현저히 줄고 있는 것을 보여주는 것이다.

5.3 과거 자료를 이용한 태풍 근접확률 예측

앞 절과 마찬가지로 1951년도부터 2000년도까지의 자료를 이용하여 태풍 발생 중 국내에 근접하는 확률의 사전분포 파라미터 값을 추정하고 그 이후의 자료로 실시간 파라미터를 보정하여 근접확률의 사후분포를 획득한다.

1951년부터 2000년까지 총 1343건의 발생 태풍 중 166건이 국내에 근접하였다. 각 연도별로 근접확률의 차이가 존재하여 연도별 근접확률, π_1 의 평균은 0.1265, 분산은 0.003125 로 계산된다. 이를 충족시키는 베타분포의 파라미터 a_1 , b_1 을 아래 식에서 구한다.

$$E[\pi_1] = \frac{a_1}{a_1 + b_1} = 0.1265,$$

$$Var[\pi_1] = \frac{a_1 b_1}{(a_1 + b_1)^2 (a_1 + b_1 + 1)} = 0.003125$$

위 식에서 $a_1 = 4.305$, $b_1 = 30.041$ 이 구해진다.

이후 2001년부터 2013년까지의 자료는 자료가 입수 되는대로 사전 파라미터를 보정하는데 역시 사용하며 그 결과 사후 파라미터 값은 아래와 같이 구해진다.

$$a_1' = a_1 + n_1(T) = 4.35 + 35 = 39.35$$

$$b_1' = b_1 + n(T) - n_1(T)$$

$$= 30.041 + 307 - 35 = 302.041$$

따라서 근접확률의 사후 평균 및 분산은

$$E[\pi_1'] = \frac{a_1'}{a_1' + b_1'} = 0.1153,$$

$$Var[\pi_1'] = \frac{a_1' b_1'}{(a_1' + b_1')^2 (a_1' + b_1' + 1)} = 0.000298$$

이 된다.

여기에서도 사후확률의 분산은 사전확률의 경우보다 약 1/10 로 줄어들고 있음을 알 수 있다.

5.4 과거 자료를 이용한 태풍 국내 상륙확률 예측

여기에서도 1951년도부터 2000년도까지의 자료를 이용하여 국내에 근접한 태풍 중 태풍이 국내에 상륙하는 확률의 사전분포 파라미터 값을 추정하고 그 이후의 자료로 실시간 파라미터를 보정하여 태풍 상륙확률의 사후분포를 획득한다.

<Table 1>에서 보듯이 1951년부터 2000년까지 총 1343건의 발생 태풍 중 166건이 국내에 근접하였으며 그 중 40건이 국내에 상륙하였다.

각 연도별 상륙확률 π_2 의 평균은 0.1265, 분산은 0.0943 으로 계산된다. 이를 충족시키는 베타분포의 파라미터 a_2 , b_2 , 를 아래 식에서 구한다.

$$E[\pi_2] = \frac{a_2}{a_2 + b_2} = 0.2461,$$

$$Var[\pi_2] = \frac{a_2 b_2}{(a_2 + b_2)^2 (a_2 + b_2 + 1)} = 0.0943$$

위 식에서 $a_2 = 0.2381$, $b_2 = 0.7295$ 가 구해진다.

이후 2001년부터 2013년까지의 자료는 자료가 입수 되는대로 앞의 경우와 마찬가지로 사전 파라미터를 보정하는데 사용한다. 그림 4에서 보듯이 파라미터 π_2 를

보정하는데는 $n_1(T)$, 와 $n_2(T)$ 만 있으면 되고 $n(T)$ 에는 영향을 받지 않는다. 따라서 $n_1(T)$, 와 $n_2(T)$ 의 값을 이용하여 사후 파라미터 값이 아래와 같이 구해진다;

$$a_2' = a_2 + n_2(T) = 0.2381 + 8 = 8.2381$$

$$b_2' = b_2 + n_1(T) - n_2(T)$$

$$= 0.7295 + 35 - 8 = 27.7295$$

따라서 상륙확률의 사후 평균 및 분산은

$$E[\pi_2'] = \frac{a_2'}{a_2' + b_2'} = 0.2290,$$

$$Var[\pi_2'] = \frac{a_2' b_2'}{(a_2' + b_2')^2 (a_2' + b_2' + 1)} = 0.000752$$

이 된다. 여기에서도 사후분포의 분산이 현저히 줄어들고 있음을 확인할 수 있다.

4절에서 자료가 입수 되는대로 실시간 모델 파라미터를 보정할 수 있는 예측 모형을 개발하였고 표 1의 자료를 이 모델에 적용하여 베이저안 방식으로 모델 파라미터를 보정한 결과 표2와 같은 결과가 얻어졌다. 태풍 발생률, 국내근접확률, 국내상륙확률의 사전분포가 입수된 데이터로 보정되어 사후분포가 획득되었으며 표 2에서 보듯이 사후분포의 분산은 사전분포의 분산에 비해 현저히 줄어들었음을 알 수 있다. 따라서 태풍 예측의 불확실성이 크게 줄어들었다고 할 수 있다.

<Table 2> Prior and Posterior distributions for model parameters

| | | prior distribution | posterior distribution |
|--|--------------|-------------------------|--------------------------|
| rate of occurrence | distribution | $\Gamma(30.420, 1.133)$ | $\Gamma(337.42, 14.133)$ |
| | mean | 26.860 | 23.875 |
| | variance | 23.717 | 1.689 |
| probability of approaching Korean peninsular | distribution | $Be(4.35, 30.041)$ | $Be(39.35, 302.041)$ |
| | mean | 0.1265 | 0.003125 |
| | variance | 0.1153 | 0.000298 |
| probability of striking Korean peninsular | distribution | $Be(0.2381, 0.7295)$ | $Be(8.2381, 27.7295)$ |
| | mean | 0.2461 | 0.0943 |
| | variance | 0.2290 | 0.000752 |

6. 결론

태풍발생자료를 그 경로에 따라 빗겨가는 태풍, 국내에 근접하는 태풍, 국내에 상륙하는 태풍의 3 유형으로 분류하고 각 유형의 태풍 간 상호 관계를 설명할 수 있는 모델을 구축하였다. 데이터가 입수 되는대로 모델 파라미터를 보정하여 예측 초기에 추정된 사전분포에 비해 사후파라미터는 그 불확실성이 현저히 줄어들고 있음을 알 수 있다.

이 모델을 통해 태풍 발생 수에 관한 보다 정확한 예측이 가능해짐은 물론, 이 중 국내에 상륙하여 많은 피해를 입힐 가능성이 있는 태풍 수에 관한 예측도 보다 정확히 이루어질 수 있게 되었다. 비교적 발생 빈도가 희박한 국내 상륙 태풍(level 2 태풍)을 예측하기 위해 비교적 많이 발생하는 태풍 발생(level 0 태풍)의 자료를 이용하여 국내 상륙 태풍의 확률을 보다 정확히 예측하였다.

7. References

- [1] 김백조, “태풍예측 향상을 위한 태풍관측 현황과 전략”, 한국방재학회, 제6권 3호 (2006) : 117-125
- [2] 나인채, 이우정, 강지영, 장기호, 강기룡, 신도식, “저기압공간위상도를 이용한 태풍예측 모델간의 온대저기압화 과정 비교”, 한국기상학회 학술대회 논문집 (2014)
- [3] 서승남, 김상익, “태풍 블라벤의 서해안 폭풍해일
- [4] 분석”, 한국해양과학기술지, 제26권1호 (2014) :41-48
- [5] 이창수, 권혁성, 하용준, “기온 확률 모형에 근거한날씨 보험의 가격결정에 대한 연구”, 보험금융연구, 제19권 2호 (2012) : 55-76
- [6] 한상진, 김근정, 오순미, “전통적 사고예측모형의 한계 및 개선 방안 : Hauer 사고예측모형의 소개

및 적용”, 한국도로학회 논문집, 제10권 1호 (2008) : 19-29

- [7] Aitchison, I.R., and Dunsmore, "Statistical Prediction Analysis", Cambridge University Press, (1975)
- [8] Cohan, Faisal et al. "A New Methodology for Safety Management Based on Feedback from Credible Accident-Probabilistic Fault Tree Analysis System", Journal of Hazardous Materials, 87, No.1 (2002) : 23-36
- [9] Piet,S.J. "Implication of Probabilistic Risk Assessment for Fusion Decision Making", Fusion Technology, Vol.10, pp.31-48, 1986
- [10] Bruske, S.Z. and Holland, D.F., "Risk Assessment Technology for the Evaluation of Tritium Accident Mitigation", Nuclear Technology Fusion, Vol.4, pp.539-543, 1983
- [11] Shachter, Ross D., "Evaluating Influence Diagrams" Operations Research, Vol.34, No.26 (1987) : 871-882
- [12] Yang, Heejoong, "Forecasting Accidents by Transforming Event Trees into Influence Diagrams", 산업경영시스템학회지, Vol. 29, No.1, pp.72-75, 2006

저자 소개

양희중



텍사스 주립대학에서 산업공학 석사를 취득하고 U.C. Berkeley에서 산업공학 박사학위를 취득하였다. Naval Postgraduate School과 Texas A&M University, Kingsville에서 교환교수를 역임하였으며 현재 청주대학교 산업공학과 교수로 재직 중이다.

관심분야 : 베이지안 예측이론, 시스템 안전도 분석, 품질경영 등.

주소 : 충북 청주시 상당구 내덕동 36 청주대학교 산업공학과