베이즈 법칙을 활용한 미니탭 매크로
- 한글 미니탭 Release 14를 이용 -

백호유[1], 이정미[2]

요 약

베이즈 법칙에서는 사전확률과 우도가 주어지고 어떤 실험결과가 일어났을 때 사후확률을 계산한다. 이러한 사후확률의 계산 문제를 미니탭 매크로를 이용하여 쉽게 계산할 수 있다. 또한 일련의 독립적이고 연속적인 실험결과에 따르는 사후확률도 편리하게 계산할 수 있다. 최근에는 미니탭 한글 Release 14가 출시되어 한글로 결과를 나타낼 수 있도록 매크로를 작성할 수 있다.

주요용어 : 베이즈 법칙, 사전확률, 사후확률, 미니탭 매크로

1. 서론

베이지안통계학에서의 기본 구조는 모수의 사전분포의 결정과, 자료와 사전분포를 이용한 사후분포를 계산하고, 이 사후분포를 이용한 모수를 추론하는 것이다. 베이즈 법칙에서 사전분포를 설정하고 그에 따른 실험결과(자료)를 얻어 사후분포를 계산하는 문제가 사전분포의 여러 가지 값을 갖는 형태로 주어졌을 때는 계산상 복잡하다. 또한 여러 단계를 거쳐 독립적으로 자료를 얻었을 때, 매 단계마다 사후분포를 계산하기가 매우 번거롭다. 이러한 계산문제를 컴퓨터로 처리하는 것이 매우 바람직하고 최종 사후분포로부터 의사결정에 도움을 얻는다고 할 수 있다.

본 연구에서는 통계 패키지 미니탭(Minitab)을 이용하여 사전분포를 설정하고 실험결과를 얻었을 때 그에 따르는 사후분포를 계산하는 문제를 다루고자 한다. 2절에서는 잘 알려진 베이즈 법칙과 그것에 필요한 미니탭 매크로 프로그램을 설명하고, 3절에서는 매크로 프로그램의 실행 예를 들었다.

2. 베이즈 법칙과 그것의 활용을 위한 미니탭 매크로

이 절에서는 베이즈 법칙과 그것의 일반적인 내용을 설명하고, 그것의 활용을 위한 미니탭 매크로 프로그램에 대하여 설명한다. 베이즈 법칙은 Antleman(1996)과 Lee(1997), Albert(1996), 김용대외(2003), 김병휘외(2001) 등의 일반적인 베이지안 통계학 서적에 소개되어 있다.

베이즈 법칙(Bayes' rule):

모든 가능한 모델을 $M_1, M_2, \cdots, M_m$이라고 하고, $P(M_i)$를 M_i의 사전확률이라고 하자. 자료(data)가 주어지면, M_i의 사후확률 $P(M_i \mid data)$는

1) 원광대학교 자연과학대학 수학정보통계학부, 교수, 전북 익산시 신용동 344-2
2) 원광대학교 의과대학 시간강사, 전북 익산시 신용동 344-2

$$P(M_i \mid data) = \frac{P(M_i)\,P(data \mid M_i)}{\sum_{j=1}^{m} P(M_j)\,P(data \mid M_j)}, \quad i = 1,\cdots,m,$$

을 만족한다. 여기서, $P(data|M_i)$를 우도 값(likelihood)이라고 하고, 이것은 그 모델 하에서 관측 값을 얻을 확률을 말한다. 사전확률분포는 자료를 관측하기 전에 실험결과와는 상관없이 표본공간의 사상(event)들에 대하여 과거의 실험결과 또는 이미 알려진 정보들에 의해(주관적인 정보 포함) 도출된 확률분포이다. 사후확률분포는 관측 값이 주어졌을 때, 가능한 모델 각각에 대한 조건부 확률분포를 말한다.

미니탭 매크로 베이지안 계산은 Albert(1996)와 Berry(1996)에 의해 처음 시도 되었으며, 그 당시의 미니탭 Release 10으로 매크로를 작성하였으나, 현재 미니탭 Release 14(2003)로는 매크로 실행이 불가능하다. 따라서 본 연구는 매크로를 수정하고, 한글화하여 한글자체로 자료를 읽고, 입력, 출력할 수 있도록 하였다. 베이즈 법칙을 활용하는 매크로 베이즈설정.mac은 사전확률과, 우도 값을 입력하여 작업창(worksheet)과 입출력창(session window)에 나타나게 한다. 다음으로 매크로 베이즈.mac은 실험결과회수와 실험결과를 입력하여 베이즈설정.mac에서 수정한 입력결과인 사전확률분포와 실험결과에 따라 update된 사후확률을 계산하여 나타낸다. 이때 최종적으로 나타난 사후확률분포가 의사결정에 도움을 주는 것이다.

다음에 나타난 매크로 프로그램 베이즈설정.mac로 사전분포와 우도 값을 입력하고 작업창(worksheet)에 저장한다. 그 다음으로 베이즈.mac를 이용하여 관측 값 수와 관측자료를 입력한 후 사전확률분포에 대한 사후확률분포를 연속적으로 계산하여 출력한다.

(1) 베이즈설정.mac

```
gmacro
베이즈설정
erase c1-c500
erase k1-k500
name c11 '모델' c12 '사전확률' c13 '우도값' c14 '곱' c15 '사후확률'
name c16 '모델이름'
name c1 '결과_1' c2 '결과_2' c3 '결과_3' c4 '결과_4' c5 '결과_5'
name c6 '결과_6' c7 '결과_7' c8 '결과_8' c9 '결과_9' c10 '결과_10'
name c200 '결과값' c201 '실험결과' c202 'PRIOR_C'
name c21 '모델1확률' c22 '모델2확률' c23 '모델3확률' c24 '모델4확률'
name c25 '모델5확률' c26 '모델6확률' c27 '모델7확률' c28 '모델8확률'
name c29 '모델9확률' c30 '모델10확률'
name c51 '관측값수'

note
note 모델의 수를 입력하시오:
set c200;
file "terminal";
nobs 1.
end
let k1=c200

note
note 모델의 이름을 입력하시오 (한줄에 한모델):
set '모델이름';
file "terminal";
format (a12);
nobs k1.
end

note
note 모델의 사전확률을 입력하시오:
set c200;
```

```
file "terminal";
nobs k1.
end
let '사전확률'=c200
let '사전확률'=abs('사전확률')/sum(abs('사전확률'))

note
note 가능한 실험결과의 수를 입력하시오:
set c200;
file "terminal";
nobs 1.
end
let k2=c200

note
note 각 모델의 우도값을 입력하시오:
let '모델'=1
print '모델'
read c1-ck2;
file "terminal";
nobs 1.
end
do k5=2:k1
  call 우도값
enddo

set '모델'
1:k1
end
note
note 모델과 실험결과의 확률분포표:
print '모델'  '모델이름' '사전확률' c1-ck2

endmacro

gmacro
우도값

let '모델'=k5
print '모델'
insert c1-ck2;
file "terminal";
nobs 1.
endmacro
```

(2) 베이즈.mac

```
gmacro
베이즈

note
note 관측값수를 입력하시오:
set c200;
file 'terminal';
nobs 1.
end
let k21=c200
let 'PRIOR_C'='사전확률'
set '실험결과'
0
end

note
note 관측값을 입력하시오:
insert '실험결과';
file 'terminal';
```

```
nobs k21.
end
set  '관측값수'
0 1:k21
end
let k31=21              #  사후모델확률을 저장하는 열
let k32=21+k1-1
let k41=31
let k42=31+k1-1
copy '사전확률' m1
trans m1 m1
copy m1 ck31-ck32
let k9=2
do k66=1:k21
  call update
enddo
note
note 각 모델의 사전확률과 사후확률의 요약:
print '관측값수' '실험결과' ck31-ck32
let '사전확률'='PRIOR_C'
endmacro
gmacro
update
let k5='실험결과'(k9)
let '결과값'=k5
print '결과값'
copy ck5 '우도값'
let '곱'='사전확률'*'우도값'
let '사후확률'='곱'/sum('곱')
print '모델' '모델이름' '사전확률' '우도값' '곱' '사후확률'
copy '사후확률' '사전확률'
let k9=k9+1
copy '사후확률' m1
trans m1 m1
copy m1 ck41-ck42
stack (ck31-ck32) (ck41-ck42) (ck31-ck32)
endmacro
```

3. 미니탭 매크로의 실행 예

이 절에서는 2절에서 나타난 미니탭 매크로를 이용하여 베이즈법칙을 활용하는 예를 보였다. 이 예는 김병휘 외 4인(2001, p45)에서 인용한 자료이다.

<예> 어떤 특정 희귀병에 대한 혈액반응검사 결과 양성반응(positive reaction)은 그 병에 대한 높은 가능성을 말해주고, 음성반응(negative reaction)은 그 반대의 경우로서 그 병에 대한 낮은 가능성을 말해준다. 과거의 통계치에 의하면 전체 인구 중 현재 이 병에 걸린 사람의 비율은 0.001이고, 혈액반응검사의 정확도, 즉, 실제로 그 병에 걸린 사람 중 음성반응을 나타낸 비율이 0.05, 그 병에 걸리지 않은 사람 중 양성반응을 나타낸 비율이 0.05라고 한다. 임의의 갑이 혈액반응검사결과 양성반응을 나타냈다면, 그 사람이 실제로 그 병을 지니고 있을 가능성은 얼마인가?

위의 <예> 내용을 요약하면 아래 <표 3.1>과 같다.

<표 3.1> 사전확률과 우도값

모델	사전확률	우 도	
		P(+ | 모델)	P(- | 모델)
병에 걸림(모델1)	0.001	0.95	0.05
병에 걸리지 않음(모델2)	0.999	0.05	0.95

활용방법:

앞 절에서 <표 3.1>의 내용을 입력하기 위해 먼저 '베이즈설정.mac'을 실행한 후 출력되는 명령에 따라 그 내용을 입력한다.

<그림 3.1> 미니탭 프로그램 출력결과

```
      MTB> %베이즈설정
      다음 파일에서 실행하는 중 : 베이즈설정.mac
      모델의 수를 입력하시오:
①     DATA> 2

      모델의 이름을 입력하시오(한줄에 한모델):
②     DATA> 병에 걸림
③     DATA> 병에 걸리지 않음

      모델의 사전확률을 입력하시오:
④     DATA> 0.001 0.999

      가능한 실험결과의 수를 입력하시오:
⑤     DATA> 2

      각 모델의 우도값을 입력하시오:

      데이터 표시
      모델
⑥       1
      DATA> 0.95 0.05
      1개의 행을 읽었습니다.

      데이터 표시
      모델
        2
⑦     DATA> 0.05 0.95
      1개의 행을 읽었습니다.

⑧     모델과 실험결과의 확률분포표:
      데이터 표시

      행  모델  모델이름          사전확률  결과_1  결과_2
      1     1  병에걸림            0.001    0.95    0.05
      2     2  병에걸리지않음      0.999    0.05    0.95
```

<그림 3.1>의 프로그램과 출력결과에 대한 설명은 다음과 같다. <그림 3.1>의 각 명령문에 붙여 놓은 번호를 따라 설명한다.

① 모델의 수를 입력한다. (병에걸림 :모델1, 병에걸리지않음 : 모델2 '2'개)
② ~ ③ 모델의 이름을 입력한다. ('병에걸림'과 '병에걸리지않음' 입력)
④ 모델의 사전확률을 입력한다. (병에걸림 : '0.001', 병에걸리지않음 : '0.999')
⑤ 가능한 실험결과의 수를 입력한다.(양성과 음성 '2')
⑥ ~ ⑦ 각 모델의 우도값을 차례로 입력한다.(병에 걸림 : '0.95, 0.05', 병에걸리지않음 :

'0.05, 0.95')
⑧ 이러한 입력이 끝나면 프로그램은 모델이름, 사전확률과 우도값의 정보를 요약한 출력물을 보여줌으로써 입력결과를 확인할 수 있다. (<예>를 요약한 <표 3.1>과 같다.)

베이즈법칙을 이용한 사후확률은 '베이즈설정.mac'이 실행된 후, '베이즈.mac' 프로그램을 실행함으로써 얻어진다. 다음 <그림 3.2>는 사후확률을 얻기 위한 입력과정과 출력결과이다.

<그림 3.2> 미니탭 프로그램 사후확률 출력결과

```
      MTB> %베이즈
      다음 파일에서 실행하는 중 : 베이즈.mac

      관측값수를 입력하시오:
⑨     DATA> 1

⑩     관측값을 입력하시오:
      DATA> 1

⑪     데이터 표시

      결과값
        1

⑫     데이터 표시

      행  모델  모델이름          사전확률  우도값    곱     사후확률
      1     1  병에 걸림           0.001    0.95  0.00095  0.018664
      2     2  병에 걸리지 않음     0.999    0.05  0.04995  0.981336

      각 모델의 사전확률과 사후확률의 요약:

⑬     데이터 표시

      행  관측값수  실험결과  모델1확률  모델2확률
      1         0         0  0.0010000  0.999000
      2         1         1  0.0186640  0.981336
```

<그림 3.3> 미니탭 프로그램 작업창(worksheet) 출력결과

워크시트 1 ***

↓	C1	C2	C3	C11	C12	C13	C14	C15	C16-T
	결과_1	결과_2	결과_3	모델	사전확률	우도값	곱	사후확률	모델이름
1	0.95	0.05		1	0.001	0.95	0.00095	0.018664	병에걸림
2	0.05	0.95		2	0.999	0.05	0.04995	0.981336	병에 걸리지 않음

⑨ 관측값의 수를 입력한다.
⑩ 관측값을 입력한다. 즉, 주어진 관측값은 '+'인데 우도 값들의 입력에서 '+'의 값이 첫 번째의 값으로 입력되었으므로 이를 관측값 '1'이라 하고 '1'을 입력하면 주어진 관측 값에 대한 각 모델에서의 사후확률이 사전확률, 우도 값 등과 함께 출력된다. 그리고 맨 마지막으로는 데이터가 수집되기 전의 각 모델의 사전확률이 첫 번째 줄에 나오고, 다음 관측 값이 하나씩 추가되면서 각 모델의 사후확률의 변화가 관측값의 입력된 순서대로 출력된다. 이 예에서는 관측 값이 한 개이므로 관측 값의 수가 0일때와 1개일 때의 값이 출력되었다.
⑪ 우리가 관심을 두어야 할 것은 혈액반응검사 결과 양성반응을 나타낸 갑이 실제로 그 병에 걸렸는가 하는 것이다. 갑이 혈액반응검사를 받기 전에는 그 병에 걸렸을 확률이 0.001이었는데 혈액반응검사 후 양성반응이 나타난 결과 병에 걸렸을 확률은 0.019로서 19배나 증가되었음을 알 수 있다. 그러나 0.019는 여전히 작은 값으로서 갑이 양성반응을 나타냈음에도 불구하고 실제로 그 병에 걸렸을 확률은 낮다고 할 수 있다. 따라서 갑이 그 병에 걸렸는지를 확실히 알기 위해서는 더 많은 증거를 획득해야만 한다. 즉, 혈액반응검사 외의 다른 각종 검사를 실시

함이 요구된다.

참고문헌

1. Albert, J. H. (1996), Bayesian Computation Using Minitab, Duxbury Press, Belmont, CA.

2. Anteleman, G. (1996), Elementary Bayesian Statistics, Edward Elgar Publishing, Cheltenham.

3. Berry, D. A. (1996), Statistics : A Bayesian Perspective, Duzbury Press, Belmont, CA.

4. Lee, P. M. (1997), Bayesian Statistics : an introduction, 2nd edition, John Wiley & Sons Inc., New York.

5. 김병휘 외 4인 (2001), 베이지안 통계계산, 자유아카데미.

6. 김용대 외 4인 (2003), 베이지안 통계학, 자유아카데미.