2017-07-01 (2주차-2)

1. 확률의 의미
   확률이라는 숫자가 가지는 의미는 여러가지 해석이 있을 수 있다. 대표적으로 빈도주의적인 의미와 베이지안의미가 있음.
   a. 확률의 빈도주의적 의미

      100개의 과일이 들어있는 상자가 있다고 하자. 과일은 표본이라고 보고 상자 안의 모든 과일의 집합은 표본 집합이라고 볼 수 있다. 이 상자 안의 100개 과일 중 일부는 사과이고 나머지는 오렌지다. 그러면 사과라는 과일의 집합은 하나의 사건이 된다. 이 사건의 이름을 A 라고 하자.

      이 상자안의 과일 집합(표본 공간)에서 하나의 과일(표본)을 선택하는 문제를 생각하자. 표본 집합에서 하나의 표본을 선택하는 것을 표본 추출, 또는 샘플링(sampling)이라고 한다.

      A 라는 사건이 가지는 확률이란 선택된 표본이 이 사건(부분 집합) A 의 원소가 될 경향(propensity) 또는 가능성을 말한다. 즉 100개의 과일 중에서 사과를 선택할 가능성이다.

      그렇다면 가능성이라는 것은 어떻게 정의될까? 빈도주의자(frequentist)는 "반복된 샘플링"이라는 개념을 사용하여 가능성을 정의한다.

      즉 100번 반복하여 표본을 선택하였을 때 해당 표본이 사과인 경우가 10번, 오렌지인 경우가 90번이라면 사과라는 사건의 확률은 10/100 = 0.1 로 정의할 수 있다는 것이 빈도주의적 관점에서 보는 "가능성" 즉, "확률"의 정의이다.

   b. 확률의 베이지안의미

      베이지안 관점에서 확률은 이미 발생한 사건의 진실에 대해 알고자 하는 노력이다.

      위의 과일 선택 문제에서 하나의 과일을 선택한 사람이 이 과일이 무엇인지 보여주지 않은채 "내가 선택한 과일은 사과이다"라고 주장한다고 해보자.

      이 때는 사건의 확률이 "미래에 특정한 사건에 속하는 일이 발생할 가능성"이 아니라 "이미 발생한 일이 특정한 사건에 속할 가능성"이 된다. 다른 관점에서 보면 "이미 발생한 일이 특정한 사건에 속한다는 가설(hypothesis) 혹은 주장의 신뢰도"라고도 볼 수 있다.

2. 확률의 수학적 정의
   확률을 수학적으로 정의하기 위해선 3가지 개념을 이해해야함
   - 확률 표본
     집합에서 선택된 하나의 사실을 확률 표본 또는 간단히 표본(sample)이라고 한다. 보통 ω ω(그리스 알파벳 소문자 오메가)라는 기호로 표기
     예를 들어 주사위를 던져서 2라는 숫자를 표시하는 면이 위로 나온 "사실"은 표본이 될 수 있다. 이때 주의할 점은 2라는 숫자뿐 아니라 2라는 숫자가 나온 "사실"도 표본이 될 수 있다는 점이다.
     확률표본을 숫자로 국한시키지 않고 어떤 현상, 혹은 사실로 지정하였기 때문에 하나의 표본은 숫자 하나에 해당하는 정보만 가지고 있는 것이 아니라 복수의 숫자 정보를 가질 수 있다.
     예를 들어 지구상의 살아있는 사람 중 어떤 선택된 "한 명"도 확률표본이다. 이 확률표본은 키, 몸무게, 허리둘레, 혈액형, 등 수많은 숫자 정보를 가질 수 있다.

     위키에서는 표본공간의 부분집합(사건)을 표본이라고 정의하고 있다. 표본 a, 사건 {a}의 구분은 a는 집합을 이루는 원소의 개념으로 생각하면 될듯하다.

   - 표본 공간
     표본 공간(sample space)은 선택될 수 있는 모든 표본의 집합을 말한다. 보통 Ω(그리스 알파벳 대문자 오메가)로 표기

   - 확률 사건
     확률 사건, 또는 간단히 사건(event)이라는 것은 표본 공간의 부분집합, 즉, 전체 표본 공간 중에서 우리가 관심을 가지고 있는 일부 표본의 집합을 뜻한다. 보통 A,B,C,와 같이 알파벳 대문자로 표기한다

   - 확률의 수학적 정의
     확률(probability)이란 각각의 사건에 대해 할당된, 다음과 같은 3가지 조건을 만족하는 숫자를 말한다. 보통 대문자 알파벳 P로 나타낸다. P(A) 라고 쓰면 A라는 사건에 할당된 숫자이다.

     (1) 모든 사건에 대해 확률은 실수이고 양수이다.

     $$P(A)\in,P(A)\geq 0$$

     (2) 표본공간이라는 사건에 대한 확률은 1이다.

     $$P(\Omega)=1$$

     (3) 공통 원소가 없는 두 사건의 합집합의 확률은 각각의 사건의 확률의 합이다.

     $$A\cap B= \quad \rightarrow \quad P(A\cup B)=P(A)+P(B)$$

     이 세가지를 콜모고로프의 공리(Kolmogorov's axioms)라고 한다.
     확률은 사건이라고 하는 입력을 받아서 숫자라는 출력을 내보내는 함수(function)로 볼 수 있다.

     사건→숫자

     그런데 확률은 "표본 하나 하나에 대해 정의되어 있는 숫자"가 아니다, 즉, 표본을 입력으로 가지는 함수가 아니다. 예를 들어 주사위의 면은 1/6이라는 확률을 지니고 있다고 생각한다면 다음과 같이 수식으로 쓸 수 있는데

     $$P()=1/6$$

     이러한 개념과 수식은 올바른 것이 아니다.

확률의 정의에 따라 확률은 표본이 아닌 사건에 대해 정의되는 사건을 입력으로 가지는 함수이다. 그러므로 정확한 수식은 다음과 같다.

$$P(\{\})=1/6$$

확률의 의미는 어떤 사건에 할당된 확률이라는 숫자는 그 사건(부분집합)에 해당하는 표본이 선택될 혹은 선택되었을 가능성을 뜻한다.

예들 들어 A={,,} 일 때는 주사위를 던져 나온 숫자가 3보다 클 가능성을 말한다. 혹은 뒤에서 이야기할 베이지안 확률론의 관점에서는 이러한 주장 즉, "주사위를 던져 나온 숫자가 3보다 크다"는 주장의 신뢰도를 뜻할 수도 있다.

주의해야할점은 확률의 정의를 만족하는 값이면 어떤 실수값인 상관없이 확률이라는 것이다. 주사위에서 우리가 일반적으로 생각하는 각 아토믹사건이 일어날 확률이 1/6이 아니라고 해서 잘못된 것이 아니다.

- 표본의 수가 무한한 경우
  주사위의 경우에는 표본의 수가 6개로 유한하며 표본 한개를 가지는 사건은 (2),(3)의 공리에 따라 확률을 1/6 이라고 정의할 수 있다. 하지만 표본의 수가 무한하다면 각 표본 한개를 가지는 사건은 확률이 0이되게 된다.
  따라서, 표본의 수가 무한하고 모든 표본에 대해 표본 하나만을 가진 사건의 확률이 동일하다면, 표본 하나에 대한 사건의 확률은 언제나 0이다.
  표본의 수가 무한한 경우는 구간(사건)에 대해서도 다음과 같이 확률을 정할 수 있다.
  $P(\{0 \le \theta < 30\})=1/12$ (시계의 특정구간에 초침이 있을 확률의 예)

3. 확률의 성질
   https://datascienceschool.net/view-notebook/d410ba427b5f456289ca1dd64ad2b0c5/

4. 베이즈 정리
   확률론에서는 하나의 사건(부분 집합)을 선택된 표본이 포함되어 있을 수 있는 하나의 부분 집합으로 본다. 따라서 하나의 사건(부분 집합)은 "선택된 표본이 이 사건(부분 집합) 안에 있다"라는 주장 혹은 가설이라고도 생각할 수 있다.

   따라서 사건(부분 집합)의 확률은 그 사건(부분 집합)이 선택된 표본을 포함할 가능성, 즉, 그 주장이 진실일 가능성, 다른말로 가설의 신뢰도를 뜻한다
   베이지안 확률론의 장점은 추가적인 정보가 발생하였을 때 이 추가 정보를 사용하여 기존에 가지고 있던 확률 즉, 어떤 가설에 대한 신뢰도를 좀 더 정확하게 수정할 수 있다는 점이다.

   추가적인 정보는 보통 또다른 사건의 형태로 발생한다. 즉 "어떤 또다른 사건이 진짜로 발생했다"는 말은 "실제로 발생한 표본이 확실하게 포함된 새로운 집합을 알게 되었다"는 의미이다.
   - 범인 찾기의 예

     예를 들어 살인 사건이 발생하였다고 가정하자.

     경찰은 전체 용의자 목록을 가지고 있으며 베이지안 확률론 관점에서 이 용의자 목록이 바로 표본 공간이다. 우리가 알고 싶은 것은 전체 용의자 목록(표본 공간)에서 누가 범인(실제로 발생한 표본)인가 하는 점이다.

     현재 표본 공간은 20명의 용의자로 구성되어 있으며 이 중 남자가 12 명, 여자가 8 명이라고 가정해 보자.

     만약 담당 형사가 범인은 남자라고 생각한다면, "범인이 남자이다."라는 주장은 확률론적 관점에서 남성인 용의자(표본)로만 이루어진 사건(표본 공간의 부분 집합)이 된다. 이를 사건 A 라고 하자.

     이 때 우리가 관심을 가지는 것은 "범인이 남자"라는 사건 A 의 신뢰도 즉, 사건 A 의 확률 P(A) 이다. 아무런 추가 정보가 없다면 모든 사람이 범인일 가능성이 같기 때문에 범인이 남자일 확률 P(A)는 다음과 같이 전체 용의자의 수로 남자 용의자의 수를 나눈 값이 된다.

     $P(A)=12/12+8 = 12/20 =0.6$ 이 때 새로운 사건 B가 발생하였다고 하자. 바로 범인의 머리카락이 발견된 것이다. 발견된 범인의 머리카락에서 범인은 머리가 길다는 사실을 알게되었다.
     이 새로운 사건 B 은 확률론적으로는 새로운 용의자 목록, 즉 머리카락이 긴 사람의 목록이라는 표본 공간의 새로운 부분 집합을 의미한다. 그리고 사건 B가 발생했다는 것은 이 용의자 목록에 진짜로 범인이 포함되었다는 뜻이다.

     현재 표본 공간 즉, 전체 용의자 목록에는 머리가 긴 사람이 10 명, 머리가 짧은 사람이 10 명이 있다.

     만약 이 사건이 진실이라는 보장이 없다면, 사건 B 에 대한 확률 P(B) , 즉 머리가 긴 사람이 범인이라는 주장의 신뢰도는 다음과 같다. $P(B)=10/10+10 =10/20 =0.5$
     지금까지의 상황을 요약해 보자.
     - 살인 사건 발생
     - 용의자는 20명
       - 남자 12명, 여자 8명
       - 머리가 긴 사람 10명, 머리가 짧은 사람 10명
     - 범인이 남자일 확률
       - 남자의 집합(사건) A 에 범인(선택된 표본)이 속해 있다는 주장의 신뢰도: P(A)=0.6
     - 범인이 머리가 길 확률
       - 머리가 긴 사람의 집합(사건) B에 범인(선택된 표본)이 속해 있다는 주장의 신뢰도: P(B)=0.5
     - 실제로는 범인이 머리가 길다.

     베이지안 확률론은 두 사건 A 와 B 의 관계를 알고 있다면 사건 가 발생하였다는 사실로 부터 기존에 알고 있는 사건 A 에 대한 확률 P(A) 를 좀 더 정확한 확률로 바꿀 수 있는 방법을 알려준다.

     이를 위해서는 결합 확률과 조건부 확률이라는 두 가지 개념을 정의해야 한다.
     결합 확률(joint probability)은 사건 A 와 B 가 동시에 발생할 확률이다. 다음과 같이 표기한다.

     $$P(A \cap B) \text{ 또는 } P(A,B)$$

     또한 B가 사실일 경우의 사건 A 에 대한 확률을 사건 B 에 대한 사건 A 의 조건부 확률(conditional probability)이라고 하며 다음과 같이 표기한다.

     $$P(A|B)$$

     조건부 확률은 다음과 같이 정의한다.

$$P(A|B)=P(A,B)/P(B)$$

조건부 확률이 위와 같이 정의된 근거는 다음과 같다.

a. 사건 B가 사실이므로 모든 가능한 표본은 사건 B에 포함되어야 한다. 즉, 표본 공간 Ω→B가 된다.
b. 사건 A의 원소는 모두 사건 B의 원소도 되므로 사실상 사건 A∩B 의 원소가 된다. 즉, A→A∩B 가 된다.
c. 따라서 사건 A 의 확률 즉, 신뢰도는 원래의 신뢰도(결합 확률)를 새로운 표본 공간의 신뢰도(확률)로 정규화(normalize)한 값이라고 할 수 있다.

여기서 주의할 점은 사건 A A와 사건 B B의 결합 확률의 값 P(A,B) P(A,B)은 기존의 사건 A A의 확률 P(A) P(A)나 사건 B B의 확률 P(B) P(B)와는 전혀 무관한 별개의 정보이다. 즉, 수학적으로 계산하여 구할 수 있는 값이 아니라 외부에서 주어지지 않으면 안되는 정보인 것이다.

- 독립

수학적으로는 사건 A 와 사건 B의 결합 확률의 값이 다음과 같은 관계가 성립하면 두 사건 A와 B는 서로 독립(independent)라고 정의한다.

$$P(A,B)=P(A)P(B)$$

독립인 경우 조건부 확률과 원래의 확률이 같아짐을 알 수 있다. 즉, B B 라는 사건이 발생하든 말든 사건 A A 에는 전혀 영향을 주지 않는 다는 것이다.

$$P(A|B)=P(A,B)\ P(B)\ =P(A)$$

5. 베이즈 정리

베이즈 정리는 사건B가 발생함으로써 사건 A의 확률이 어떻게 변화하는지를 표현한 정리

a. $P(A|B)=P(B|A)P(A)\ P(B)$

- P(A|B) P(A|B): 사후 확률(posterior). 사건 B가 발생한 후 갱신된 사건 A의 확률
- P(A) P(A): 사전 확률(prior). 사건 B가 발생하기 전에 가지고 있던 사건 A의 확률
- P(B|A) P(B|A): 우도(likelihood). 사건 A가 발생한 경우 사건 B의 확률
- P(B) P(B): 정규화 상수(normalizing constant): 확률의 크기 조정

b. 베이즈 정리 확장1,2
https://datascienceschool.net/view-notebook/f68d16df9ea448689ae66dc2140fe673/

6. 베이즈 정리와 분류문제

베이즈 정리는 분류문제에 사용될 수 있음
X 특성을 가진 표본을 선택하였을 때 Y일 가능성, 혹은 Y라는 주장의 신뢰도와 같이 표현할 수 있음

- ex) X = {O,A}, Y = {B,R}
  P(Y = B | X = O) => feature가 X 인 데이터가 B라는 클래스일 확률을 구할 수 있다.

7. 확률모형

기술통계등의 방법으로 자료의 분포를 기술하는 방법은 불확실하며 대략적인 정보만을 전달할 뿐이며 완벽한정보전달이 어렵다.
예를 들어 히스토그램 같은 경우 구간에 대한 정보를 더 자세하게 알고 싶을 경우

정확한 묘사를 위해 구간의 수를 증가시키면 몇 가지 문제가 발생한다.

우선 구간의 간격이 작아지면서 하나의 구간에 있는 자료의 수가 점점 적어진다. 만약 구간 수가 무한대에 가깝다면 하나의 구간 폭은 0으로 수렴하고 해당 구간의 자료 수도 0으로 수렴할 것이다. 따라서 분포의 상대적인 모양을 살펴보기 힘들어진다.

더 큰 문제는 분포를 묘사하기 위한 정보가 증가한다는 점이다. 데이터의 분포를 묘사하는 이유는 적은 갯수의 숫자를 통해 데이터의 전반적인 모습을 빠르게 파악하기 위한 것인데 묘사를 위한 정보의 양이 증가하면 원래의 목적을 잃어버린다.

이러한 문제를 해결하기 위한 만들어진 것이 확률모형이다

a. 확률모형은 확률변수를 이용하여 데이터분포를 수학적으로 정의하는 방법이다. 확률모형론에서는 데이터 그 자체에는 의미가 없으며 데이터의 분포 특성만이 중요하다고 생각
확률 모형론을 사용한다는 것은 데이터를 생성하는 가상의 주사위가 있다고 가정하는 것과 같다. 동일한 데이터가 아닌 동일한 분포를 나타내는 데이터를 생성한다는 것
b. 확률변수
이렇게 특정한 분포특성을 가지는 데이터를 만들어 내는 기계 혹은 주사위를 확률변수라고 함

확률 변수는 수학적으로 하나의 표본에 대해 하나의 실수 숫자를 대응하는 함수로 정의하며 보통 대문자 알파벳을 사용하여 표기한다.

$$X\ (\omega):\omega\rightarrow x$$

c. 확률 분포 함수(확률 밀도 함수)
확률변수의 확률적 특성을 정의 한것이 확률 분포 함수이고 함수의 계수들을 모수(parameter)라 표현
d. 샘플링 실현

확률변수를 이용하여 데이터를 생성하는 것 혹은 전통적인 관점에서 모집단에서 표본을 추출하는 것
e. 데이터 분석과정
   i. 자료를 확보한다.
   ii. 확보된 자료를 확률 변수의 표본으로 가정한다.
   iii. 확률 변수가 특정한 확률 모형을 따른다고 가정한다.
   iv. 표본에 대한 정보로부터 확률 모형의 종류나 모수를 추정한다.
   v. 구해진 확률 모형으로부터 다음에 생성될 데이터나 데이터 특성을 예측한다.

8. 확률변수

a. 정의 : w ∈Ω → 실수x ∈ 표본 공간을 정의역(domain)으로 가지고 실수를 공역(range)으로 가지는 함수
b. 확률 변수를 사용하면 모든 표본은 하나의 실수 숫자로 변하기 때문에 표본 공간을 실수의 집합 즉 수직선(number line)으로 표시할 수 있다. 일반적인 사건(event)은 이 수직선 상의 구간으로 표시된다.

예를 들어 a 보다 같거나 크고 b보다 작은 숫자의 집합인 사건 A는 다음과 같은 기호로 표시한다.

$$A=\{\omega;a\leq X\ (\omega)<b\}=\{a\leq X\ <b\}$$

c. 이산확률 변수: 확률 변수의 값이 연속적이지 않고 떨어져 있는 경우

확률 변수를 정의한다는 것은 표본(sample)이라는 추상적이고 일반적인 개념대신 숫자라는 명확한 개념을 대신 사용하겠다는 의미이다.
현실적으로도 계산이 가능한 것은 숫자 뿐이므로 데이터 분석을 수행하기 위해서는 결국 표본의 특성(feature)을 숫자로 변환하는 단계가 필요하다.

d. 연속확률 변수: 확률 변수의 값은 실수(real number) 집합처럼 연속적이고 무한개의 경우

9. 누적 분포와 함수와 확률 밀도 함수
누적 분포 함수(cumulative distribution function)와 확률 밀도 함수(probability density function)는 확률 변수의 분포 즉, 확률 분포를 수학적으로 정의하기 위한 수식