

# Chapter 2 확률(Probability) vs 가능도(Likelihood)

## 2.1 시작하면서

본 챕터에서는 **가능도(Likelihood)**가 무엇인지 직관적으로 이해하는 것을 목표로 한다. 가능도는 정규분포부터 회귀분석과 최신 인공지능 알고리즘에 이르기까지 통계학의 모든 부분에서 빠질 수 없는 개념인데, 이상하게도 의학 또는 보건학을 다루는 통계학 책에서는 이 개념을 잘 설명하지 않는다. 물론 통계 비전공자에게 설명하기 까다로운 개념임을 인정하며, 이번 기회에 **확률(Probability)**과 비교를 통해 엄밀한 정의는 아니더라도 대략적인 느낌은 파악하도록 하자. 마지막에는 **최대 가능도 추정량(Maximum Likelihood Estimator)**이 나오는데, 통계적 추론에서 중심극한정리와 더불어 가장 중요한 개념 중 하나이므로 반드시 이해하고 넘어가도록 하자.

## 2.2 확률

주사위를 예를 들어 보자. 주사위를 던져서 나올 수 있는 숫자는 1,2,3,4,5,6이고 각 숫자가 나올 확률은  $\frac{1}{6}$ 로 모두 같다. 동전 10번 던져서 앞면은 0~10번 나올 수 있으며 각각의 확률은 계산해 보면 각각 0.001, 0.01, 0.044, 0.117, 0.205, 0.246, 0.205, 0.117, 0.044, 0.01, 0.001 이다. 두 경우 모두 일어날 수 있는 사건이 6개, 11개로 정해져 있으며 각각에 대한 확률을 구할 수 있고 확률의 합은 1이 된다.



## 2.3 연속사건의 확률

### 2.3.1 특정 사건의 확률은 모두 0

이번에는 1에서 6사이의 숫자 중 랜덤으로 아무 숫자나 뽑는다고 하자. 이 때 정확히 5가 뽑힐 확률은 얼마일까? 1과 6사이에는 무한개의 숫자가 있으니 정확히 5가 뽑힐 확률은  $\frac{1}{\infty} = 0$ 이다. 2가 뽑힐 확률,  $\pi$ 가 뽑힐 확률도 마찬가지로 0일 뿐만 아니라 어떤 특정 숫자가 뽑힐 확률은 전부 0이다. 이는 연속된 숫자 사이에서 뽑을 수 있는 숫자의 갯수가 무한하기 때문이다(1과 6사이에 놓 무한히 많은 숫자가 있다). 따라서, 이런 연속사건인 경우 특정 숫자가 나올 확률을 말하는 것은 의미가 없어 다른 방법을 생각해야 하는데, 숫자가 특정 구간에 속할 확률을 말하는 것이 그 대안이다.

### 2.3.2 특정 구간에 속할 확률: 확률밀도함수(Probability Density Function, PDF)

아까의 1에서 6사이의 숫자를 뽑는 상황을 다시 생각해 보자. 1에서 6사이의 숫자 중 정확히 5가 뽑힐 확률은 0지만 4에서 5사이의 숫자가 뽑힐 확률은 20%이다. 전체 구간의 길이는  $6-1=5$ 이고 4에서 5사이 구간의 길이는 1이기 때문이다. 마찬가지로 논리로 2에서 4사이의 숫자가 뽑힐 확률은  $\frac{2}{5} = 40\%$ 가 된다. 이처럼 우리는 특정 사건에 대한 확률 대신 특정 구간에 속할 확률을 구함으로써 간접적으로 특정 사건의 확률에 대한 감을 잡을 수 있다. 이것을 설명하는 곡선이 바로 고등학교 때 배운 **확률밀도함수(Probability Density Function: PDF)**이다. PDF는 특정 구간에 속할 확률을 계산하기 위한 함수이며 함수를 나타내는 <그래프에서 특정 구간에 속한 넓이=특정 구간에 속할 확률>이 되게끔 정한 함수이다. 아래 그림을 예를 들어 살펴보자. 왼쪽의 그림에서 PDF의 값은 1에서 6사이에서는 전부 0.2이고 나머지 구간에서는 전부 0인데, 이는 1에서 6사이의 숫자를 뽑는 상황을 그림으로 나타낸 것이다. 1보다 작거나 6보다 큰 숫자를 뽑을 수는 없으므로

이에 해당하는 확률밀도함수의 함수의  $y$ 값은 전부 0이고, 1~6사이에서는 무작위로 숫자를 뽑으므로  $y$ 값은 전부 같다. 전체 확률은 1이므로 그림의 직사각형의 넓이는 1이 되고  $y$ 값은 전부 0.2가 되며, 이를 바탕으로 2에서 4사이의 숫자가 뽑힐 확률을 계산하면  $2 \times 0.2 = 0.4$ 로 40%가 된다. 오른쪽의 그림은 정확한 의미는 잘 몰랐더라도 모양은 많이 봤을 **정규분포(Normal distribution)**이며, 그 중에서도 가장 흔히 쓰이는 평균 0, 분산 1인 **표준정규분포(Standard normal distribution)**를 나타내고 있다. 표준정규분포의 PDF는 다들 알고 있는대로(?)  $\frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ 로 표현되며 그림에서 보듯이  $z$ 가 -1.96~1.96에 안에 있을 확률이 95%임이 잘 알려져 있다.

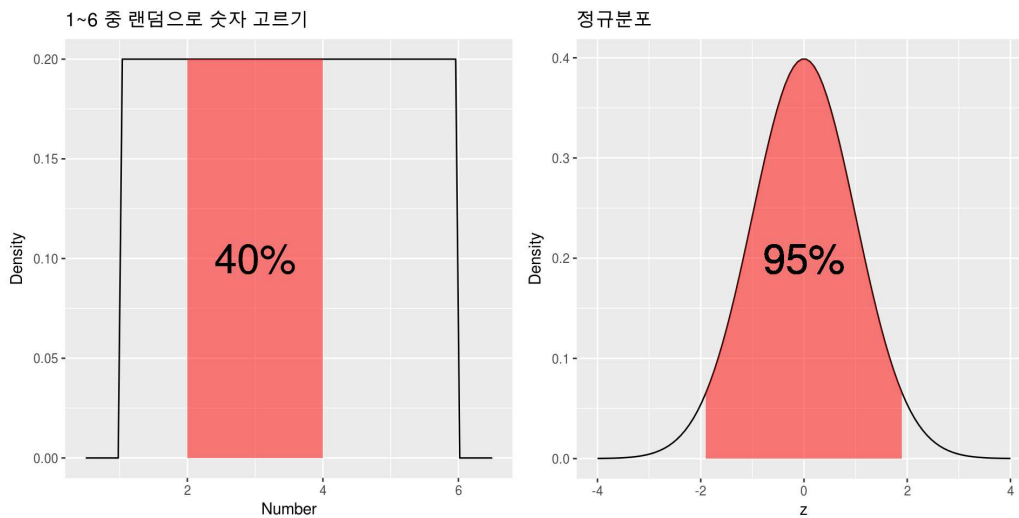


Figure 2.2: 연속 사건의 확률 예시

정리하자면 연속사건의 경우에는 특정 사건이 일어날 확률은 모두 0이며, 어떤 구간에 속할 확률은 PDF를 이용해서 구할 수 있다고 할 수 있다. 그러면 특정 사건에 대한 해석은 할 수 없는 것인가? 단순히 할 수 없다고 말하기는 아쉽다. 위의 정규분포의 경우를 보면 0이 나올 확률도 0, 1이 나올 확률도 0, 999가 나올 확률도 0으로 모두 같으므로 0, 1, 999가 나올 가능성은 전혀 차이가 없다고 말해야 한다. 그러나 우리 모두는 정

규분포의 그림을 보고 직관적으로 느끼고 있다. 가장 위로 솟아올라 있는 0 근처가 나올 가능성이 가장 높고, 1 근처가 나올 가능성은 그보다 낮으며, 999같이 큰 수가 나올 가능성은 거의 없다는 것을... 그러나 확률이 라는 지표로는 이런 연속사건간의 가능성 차이를 표시할 수가 없다는 문제가 있다.

## 2.4 특정 사건이 일어날 가능성을 비교할 수는 없을까?: 가능도(Likelihood)

방금 설명한 대로 연속사건에서는 특정 사건이 일어날 확률이 전부 0으로 계산되기 때문에 사건들이 일어날 가능성을 비교하는 것이 불가능하며, **가능도**라는 개념을 적용해야 이를 비교할 수 있다. 그러나 지금 가능도의 엄밀한 정의를 설명하는 것은 이해를 돕는데 도움이 안될 것이며, 직관적인 설명을 이용할 것인데, 쉽게 말하자면 **위에 있는 그래프들에서  $y$ 값을 가능도로 생각하면 된다.** 즉,  **$y$ 값이 높을수록 일어날 가능성이 높은 사건**이라는 것이다. 주사위나 동전을 던지는 경우는  $y$ 값이 각 사건이 일어날 확률을 나타내었으므로 가능도 = 확률이 되어, 확률이 높을수록 일어날 가능성이 높은 사건이 된다. 한편 정규분포같이 연속사건인 경우는 PDF의 값이 바로  $y$ 가 되며 0에 해당하는 PDF값이 0.4로 1에 해당하는 PDF값인 0.24보다 높아 0 근처의 숫자가 나올 가능성이 1 근처의 숫자가 나올 가능성보다 높다고 할 수 있으며, 0이 나올 확률과 1이 나올 확률이 모두 0인 것과는 대조적이다. 이를 정리하면 가능도의 직관적인 정의는 다음과 같다.

- 가능도의 직관적인 정의 : 확률분포함수의  $y$ 값
  - 셀 수 있는 사건: **가능도 = 확률**
  - 연속 사건: **가능도  $\neq$  확률, 가능도 = PDF값**

## 2.5 사건이 여러 번 일어날 경우에서의 가능도

이번에는 사건이 여러 번 일어날 경우를 생각해 보자. 먼저 아래의 두 문제를 풀어보자.

1. 주사위를 3번 던져 각각 1,3,6이 나올 확률은 얼마일까?

2. 동전을 10번 던지는 일을 3회 시행하여 앞면이 각각 2,5,7번 나올 확률은 얼마일까?

1번의 경우 주사위를 던져 1,3,6이 나올 확률은 전부  $\frac{1}{6}$ 이므로 정답은  $\frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{216}$ 이고, 2번의 경우 동전을 10번 던져 앞면이 2,5,7번 나올 확률은 앞에서와 같이 각각 0.044, 0.246, 0.117이므로 정답은  $0.044 \times 0.246 \times 0.117 = 0.001$ 이다. 가능성도 마찬가지이다. 앞서 셀 수 있는 사건에서는 확률과 가능성이 같다고 했으므로 주사위를 3번 던져 각각 1,3,6 이 나올 가능성을 나타내는 가능성도는  $\frac{1}{216}$ 이 되고, 동전을 던지는 경우의 가능성도도 마찬가지로 확률과 같은 0.001이 된다. 이제 연속사건이 여러 번 일어날 경우를 살펴보자. 앞서 언급한 평균 0, 분산 1인 정규분포에서 숫자를 3번 뽑았을 때 차례대로 -1,0,1이 나올 확률은 각각의 사건이 일어날 확률이 모두 0이므로 결국 0이 된다. 그러나 가능성도의 경우 -1,1이 나올 가능성도는 0.24, 0이 나올 가능성도는 0.4이므로 -1,0,1이 나올 가능성도는  $0.24 \times 0.4 \times 0.24 = 0.02$ 가 되어 확률과는 다른 값으로 나타나게 된다.

## 2.6 진실을 찾는 방법: 최대가능도 추정량(Maximum Likelihood Estimator, MLE)

가능도 관련 마지막 주제로 최대 가능도 추정량(이하 MLE)에 대해 알아보겠다. MLE는 기초적인 통계분석에서 회귀분석에 이르기까지 거의 모든 통계분석에서 참값을 추정하는 원리이지만, 설명의 어려움 때문인지 거의 모든 기초통계 교과서에서 설명이 빠져 있다. 여기서는 두 가지 예를 가지고 MLE의 개념에 대해 설명할 것인데 먼저 모양이 변형된 동전을 생각해 보자.

### 2.6.1 예1: 모양이 일그러진 동전

지금까지와는 다르게 이 동전은 모양이 많이 일그러져서 앞이 나올 확률이 0.5라고 말할 수가 없고, 실제로 던져봐도 그 확률을 알 수 있을 것 같다. 실제로 1000번을 던져봤더니 앞이 400번, 뒤가 600번 나왔다면 우리는 동전을 던져 앞이 나올 확률  $p$ 가 대략 얼마 정도라고 생각할까? 아마 대부분은 0.4정도라고 생각할 것

이며 이것은  $p$ 의 MLE값과 일치한다. 풀어서 설명하면 동전을 1000번 던져서 앞이 400번 나올 가능성을 최대로 하는  $p$ 는 0.4라는 뜻이며 수식을 이용한 엄밀한 증명은 다음과 같다.

1. 앞면이 나올 확률이  $p$ 라면 1000번을 던져 앞이 400번, 뒤가 600번 나올 가능도(=확률)  
 $L = {}_{1000}C_{400} p^{400} (1-p)^{600}$ 이다. 수식이 싫으면 이  $L$ 이 최대값을 가지는  $p$ 를 계산해보면 0.4가 나온다는 것을 인정하고 그냥 넘어가자.
2.  $L$ 이 언제 최대가 되는지 살펴보자.  $L$ 이 최대가 되는 것은  $p^{400}(1-p)^{600}$ 가 최대가 될 때인데 산술-기하 평균 부등식에 의하여  $600 = \frac{3}{2}p \times 400 + (1-p) \times 600 \geq 1000 \times \{(\frac{3}{2}p)^{400} (1-p)^{600}\}^{\frac{1}{1000}}$ 이 된다.
3. 따라서  $p^{400}(1-p)^{600}$ 의 최대값은  $(\frac{600}{1000})^{1000} \times (\frac{2}{3})^{400}$ 이 되며,
4.  $L$ 이 최대값이 될 때는 앞의 산술-기하평균 부등식의 등호조건이 성립할 때이므로  $\frac{3}{2}p = 1-p$  즉,  $p = 0.4$ 가 된다.

더 직관적으로 이해해보기 위해 앞면이 나올 확률  $p$ 에 따른 가능도  $L$ 의 값을 그래프로 그려보면 아래와 같다.

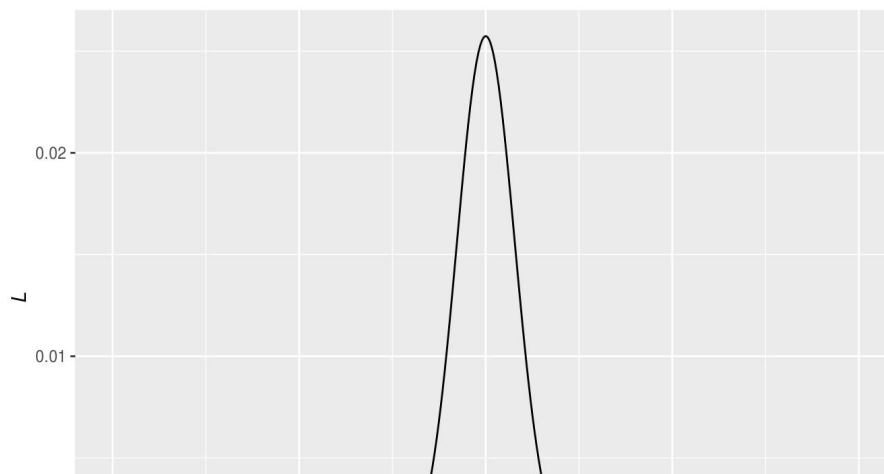


Figure 2.3: 앞면이 나올 확률  $p$ 에 따른 가능도  $L$ 의 값

그림으로 살펴봐도  $p$ 가 0.4일 때  $L$ 의 값이 최대임을 알 수 있다. 지금까지 이야기를 다시 한 번 정리 해 보자. 동전을 1000번 던져 앞이 400번, 뒤가 600번 나왔다면 우리는 직관적으로 앞이 나올 확률  $p$ 는 0.4 정도 라고 생각할 것이며, 실제 이런 일이 발생할 가능성을 최대로 하는  $p$ 를 계산하면 0.4가 된다. 이를 간략히  $p$ 의 MLE는 0.4라고 표현한다. 즉, MLE는 우리의 가능성과 확률에 대한 직관을 수리적으로 표현한 것에 불과 하며 어렵게 생각할 필요가 없다고 말하고 싶다.

## 2.6.2 예2: 나의 실제 키

새로운 예시로 나의 키를 재는 상황을 생각해 보자. 키는 재는 시간, 방법에 따라서 실제 키와 조금씩 오차가 있을 수 밖에 없다. 만약 키를 5번 측정해서 178,179,180,181,182(cm)가 나왔다면 나의 실제 키는 얼마일 까? 물론 많은 사람들이 가장 높게 나온 182cm을 자기 실제 키라고 생각할 수 있겠지만... 합리적(?)인 사람 이라면 아마 다섯 번의 평균인 180cm을 실제 자신의 키라고 생각할 것이다. 그렇다면 키의 MLE 값은 얼마 일까? 즉, 키를 5번 측정했을 때 178,179,180,181,182cm이 나올 가능성이 최대가 되는 나의 키는 얼마일 까? 일그러진 동전 때와 마찬가지로 직관과 MLE는 일치할까? 이제부터 나의 키의 MLE값을 계산할 것인데 그 전에 한 가지 가정을 하려고 한다. 그것은 바로 키의 측정값은 참값을 평균으로 하는 정규분포를 따른다 는 것이다. 키의 측정값은 어떤 값이든 갖을 수 있으며 오차가 적은 값이 오차가 큰 값보다 나올 가능성이 높 다는 점으로 생각해 볼 때, 정규분포를 따른다는 가정은 큰 무리가 없을 것 같다. 이 가정 하에서 키의 MLE를 구하는 과정은 다음과 같다. 아까와 마찬가지로 수식이 싫은 사람은 바로 그림으로 내려가 직관적으로 이해 하길 바란다.

1. 키의 참값이  $\mu$ 일 때 측정값은 평균  $\mu$ , 분산  $\sigma^2$ 인 정규분포를 따른다.

2. 키의 측정값이  $x$ 일 때의 가능도, 즉 정규분포의  $y$ 값은  $\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ 이다.

3. 5번 측정한 키가 178,179,180,181,182가 나올 가능도  $L$ 은 각각의 가능도의 곱인

$$\frac{1}{\sqrt{2\pi}\sigma^2}e^{-\frac{(178-\mu)^2}{2\sigma^2}} \times \frac{1}{\sqrt{2\pi}\sigma^2}e^{-\frac{(179-\mu)^2}{2\sigma^2}} \times \frac{1}{\sqrt{2\pi}\sigma^2}e^{-\frac{(180-\mu)^2}{2\sigma^2}} \times \frac{1}{\sqrt{2\pi}\sigma^2}e^{-\frac{(181-\mu)^2}{2\sigma^2}} \times \frac{1}{\sqrt{2\pi}\sigma^2}e^{-\frac{(182-\mu)^2}{2\sigma^2}}$$

4.  $L$ 이 최대가 된다는 것은  $e^{-\frac{(178-\mu)^2}{2\sigma^2}} \times e^{-\frac{(179-\mu)^2}{2\sigma^2}} \times e^{-\frac{(180-\mu)^2}{2\sigma^2}} \times e^{-\frac{(181-\mu)^2}{2\sigma^2}} \times e^{-\frac{(182-\mu)^2}{2\sigma^2}} = e^{-\frac{(178-\mu)^2+(179-\mu)^2+(180-\mu)^2+(181-\mu)^2+(182-\mu)^2}{2\sigma^2}}$ 가 최대가 된다는 뜻이고, 이는 다시

$(178-\mu)^2 + (179-\mu)^2 + (180-\mu)^2 + (181-\mu)^2 + (182-\mu)^2$ 가 최소가 되는 것으로 해석 할 수 있다.

5.  $(178-\mu)^2 + (179-\mu)^2 + (180-\mu)^2 + (181-\mu)^2 + (182-\mu)^2$ 은  $\mu = 180$ 에서 최소값을 가 짐을 쉽게 알 수 있고, 따라서  $\mu$ 의 MLE는 180이다.

이번에도 직관적인 이해를 위해  $\sigma^2 = 1$ 로 가정한 후, 실제 키인  $\mu$ 와 그에 해당하는 가능도  $L$ 의 그래프를 살펴보자.

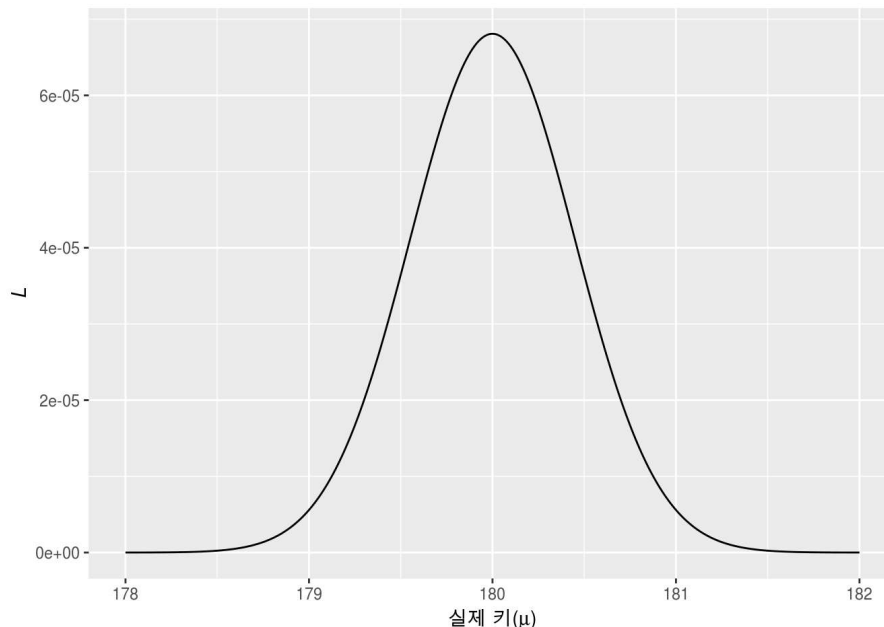


Figure 2.4: 실제 키에 따른 가능도  $L$ 의 값

그림에서도  $\mu$ 가 180일 때  $L$ 의 값이 가장 큰 것을 쉽게 확인할 수 있으며, 앞서 일그러진 동전 때와 마찬가지로 실제 값은 측정값의 평균 정도일 것이라는 우리의 직관과 실제로 계산한 MLE값이 일치하는 것으로 나타났다.

## 2.7 마치며

본 챕터에서는 통계학에서 필수적인 개념인 가능도와 MLE를 이해하기 위해 확률과의 비유를 통해 가능도와 MLE의 개념을 순서대로 설명하였으며, 각 순서마다 이산사건(셀 수 있는 사건)과 연속사건의 2가지 예를 제시하였다. 다음 챕터에서는 정규분포부터 시작해서 최소한으로 꼭 알아야 할 분포를 이야기 할 예정이며, 실제 키의 MLE를 구하는 과정이 거꾸로 적용될 것이니 잘 기억하도록 하자.