

불확도와 오차의 전파

Uncertainties and Error Propagation

최종일 (한남대학교 자유전공학부 물리담당 교수)

1. 계통오차와 우연오차

어떤 측정도 완전히 정확할 수는 없다. 측정의 정확도와 정밀도는 사용한 기기의 성능, 관측자의 숙련도, 그리고 실험 자체의 물리적 한계에 의해 항상 제한된다. 따라서 실험에서는 어떤 물리량의 최선의 추정값만 제시하는 것이 아니라, 가능한 참값의 범위도 함께 제시해야 한다.

같은 물리량을 여러 번 측정했을 때 서로 다른 값이 나오는 이유는 여러 가지가 있다. 여기서는 실험자의 단순한 실수는 제외하고, 측정 과정 자체에 내재한 오차만 다룬다.

측정값의 변동을 나타낼 때 이 문서에서는 불확도, 오차, 편차라는 용어를 문맥에 따라 사용한다. 그중 오차는 크게 **계통오차**와 **우연오차**의 두 종류로 나눌 수 있다.

계통오차는 측정값이 참값보다 항상 크거나 항상 작게 나오도록 만드는 오차이다. 이는 잘못 교정된 기기나 일정한 방향으로 치우친 측정 방법 때문에 생긴다. 예를 들어, 낮은 온도에서 수축한 강철 자로 길이를 측정하면 실제 길이를 과대평가할 수 있다. **계통오차는 실험을 신중하게 설계하고 필요한 보정을 적용함으로써 줄이거나 제거할 수 있다.**

반면 우연오차는 계통오차를 제거한 뒤에도 남아 있는 측정값의 불규칙한 흔들림이다. 같은 양을 반복 측정할 때 값이 조금씩 달라지는 현상이 여기에 해당한다. 이러한 **우연오차는 통계적인 방법을 사용하여 정리하고 해석할 수 있다.**

2. 우연오차의 결정

측정량의 불확도를 어떻게 추정할 수 있을까? 적용하는 경우에 따라 여러 가지 접근법을 사용할 수 있다.

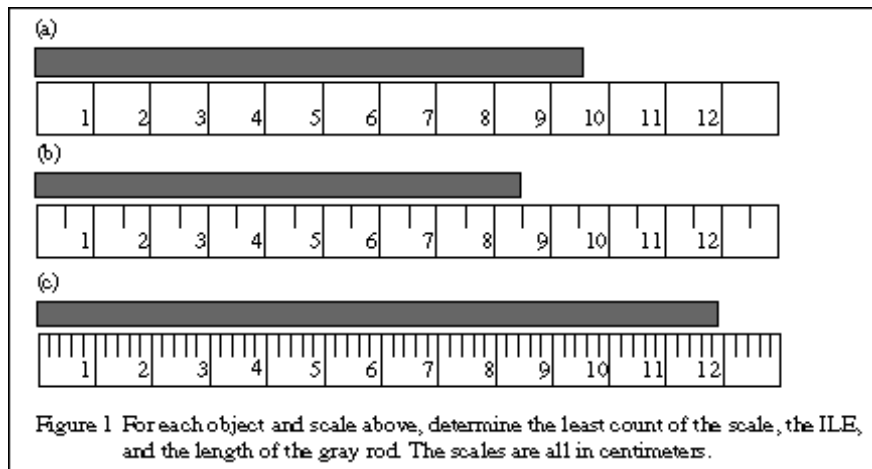
(a) 기기 한계오차(Instrument Limit of Error, ILE)와 최소눈금

최소눈금은 측정 기구에 표시된 가장 작은 눈금 간격이다. 예를 들어 미터자는 최소눈금이 **1.0 mm**이고, 디지털 초시계는 최소눈금이 **0.01 sec**일 수 있다. 기기 한계오차, 줄여서 ILE는 측정 장치를 실제로 판독할 수 있는 정밀도를 뜻하며, 항상 최소눈금과 같거나 그보다 작다. 성능이 좋은 측정 기구는 국가 표준 기관이 유지하는 기준에 맞추어 교정된다.

기기 한계오차는 일반적으로 최소눈금 자체이거나, 최소눈금의 어떤 분수 $1/2$, $1/5$, $1/10$ 등으로 잡는다. 최소눈금을 쓸지, 최소눈금의 절반을 쓸지, 아니면 다른 값을 쓸지는 고정 규칙으로 정할 수 없다. 대신 상식적인 판단이 필요하다. 눈금 사이 간격이 충분히 크면 최소눈금의 $1/5$ 또는 $1/10$ 까지 추정해서 읽을 수 있다. 눈금 사이 간격이 더 좁으면 최소눈금의 $1/2$ 정도까지만 추정할 수 있고, 눈금이 아주 촘촘하면 최소눈금까지만 읽을 수 있다.

어떤 장치에서는 ILE가 허용오차 또는 퍼센트로 주어진다. 예를 들어 어떤 저항이 허용오차 5%라고 표시되어 있다면, ILE는 그 저항값의 5%를 뜻한다.

연습문제: 다음의 각 눈금(모두 단위는 cm)에 대하여 최소눈금, ILE, 그리고 회색 막대의 길이를 구하여라.



(b) 추정 불확도

실제로는 ILE보다 더 큰 다른 불확도들이 존재하는 경우가 많다. 예를 들어, ILE가 0.01 g인 단순한 양팔저울로 질량을 맞추려고 할 수 있다. 그런데 한쪽 접시의 질량을 무려 3 g까지 바꾸어도 지침의 변화가 보이지 않을 수 있다. 이 경우에는 그 범위의 절반을 추정 불확도로 사용하여, 불확도를 ± 1.5 g로 잡는다.

또 다른 좋은 예는 렌즈의 초점거리를 구하기 위해 렌즈에서 스크린까지의 거리를 측정하는 경우이다. ILE는 0.1 cm일 수 있다. 그러나 초점 심도가 있어서, 스크린을 1.6 cm 움직여도 상이 계속 초점이 맞아 보일 수 있다. 이 경우 추정 불확도는 그 범위의 절반인 ± 0.8 cm가 된다.

연습문제: ILE가 0.5 mm인 줄자를 사용하여 서 있는 사람의 키를 측정하였다. 불확도를 추정하여라. 이때 “똑바로 서 있는지” 아니면 약간 구부정한지도 알 수 없다는 효과를 포함하여라.

(c) 평균편차: 반복 측정에 의한 추정 불확도

반복 측정을 통해 측정값과 그 불확도를 추정하는 가장 기본적인 통계적 방법은 다음과 같다. 먼저 동일한 물리량을 여러 번 측정하고, 그 측정값들의 평균을 구한다. 그런 다음 데이터가 평균값 주변에서 얼마나 퍼져 있는지를 나타내는 척도로 평균편차 또는 표준편차를 계산한다. 이 값들이 측정 결과의 불확도를 정량적으로 나타내는 기준이 된다.

어떤 물리량을 여러 번 측정하여 서로 다른 값을 얻었다고 하자. 이때 먼저 측정값들의 평균값을 계산한다. 평균값은 보통 꺾쇠괄호 기호로 나타내어 $\langle t \rangle$ 와 같이 쓰며, 이 값을 측정값의 최선의 추정값으로 사용한다. 그러나 평균값만으로는 측정 결과의 신뢰도를 판단할 수 없다. 따라서 측정값들이 평균값에서 얼마나 벗어나는지를 분석하여 불확도를 추정해야 한다.

Time, t, sec.	(t - <t>), sec	t - <t> , sec	$(t - \langle t \rangle)^2$
7.4	-0.2	0.2	0.04
8.1	0.5	0.5	0.25
7.9	0.3	0.3	0.09
7.0	-0.6	0.6	0.36
<t> = 7.6	<t - <t>> = 0.0	< t - <t> > = 0.4	$\langle (t - \langle t \rangle)^2 \rangle = 0.247$ Std. dev = 0.50

표 1. 시간 측정 데이터로부터 평균값, 평균편차, 표준편차를 구하는 과정을 보여주는 값들. 평균편차가 0이 아닌 값을 갖도록 하려면, 편차의 절댓값을 사용하여 평균을 계산해야 한다는 점에 유의하라.

예를 들어 어떤 시간 측정을 여러 번 반복하여 다음과 같은 값을 얻었다고 하자.

7.4 s, 8.1 s, 7.9 s, 7.0 s

이 값들의 산술평균은 다음과 같이 계산된다.

$$\langle t \rangle = \frac{7.4 + 8.1 + 7.9 + 7.0}{4} = 7.6 \text{ sec}$$

따라서 평균값은 7.6 s 이다. 이처럼 어떤 물리량의 평균을 나타낼 때는 보통 $\langle t \rangle$ 와 같은 표기를 사용한다.

이제 각 측정값이 평균값에서 얼마나 벗어나는지 살펴보자. 이를 위해 각 측정값에서 평균값을 뺀 값을 계산한다.

$$t - \langle t \rangle$$

이 값을 **편차**라고 한다 (표1: Column 2). 각 측정값의 편차를 모두 더해 평균을 구하면 항상 0이 된다. 편차는 평균을 기준으로 양의 값과 음의 값이 서로 상쇄되기 때문이다. 따라서 단순히 편차의 평균을 계산하는 것은 데이터의 산포를 나타내는 데 도움이 되지 않는다.

이 문제를 해결하기 위해 편차의 절댓값을 취한 뒤 그 평균을 계산한다. 이렇게 얻은 값을 **평균편차**라고 한다 (표1 Column 3). 평균편차는 측정값들이 평균값 주변에서 평균적으로 얼마나 벗어나는지를 나타내는 척도이다.

표 1의 Column 4에는 Column 2에서의 편차를 제곱한 값들이 들어 있으며, 이로 인해 모든 값이 양수가 된다. 제곱값들의 합을 3(측정 횟수보다 하나 작은 값)으로 나눈 다음, 그 결과의 제곱근을 취하면 **표준편차(standard deviation)**가 된다. 왜 N이 아니라 (N-1)로 나누는지에 대한 설명은 통계학 교재에서 확인할 수 있다.

표준편차는 평균편차와 수치적으로 약간 차이가 있지만, 두 값 모두 측정 데이터가 평균값 주변에서 얼마나 퍼져 있는지를 나타내는 척도이다. 따라서 평균편차나 표준편차 가운데 어느 것을 사용하더라도 데이터의 변동 정도를 정량적으로 표현할 수 있다.

다음은 스프레드시트(예: Excel)에서 사용하는 기본 함수와 그 의미를 설명한 것이다.

=SUM(A2:A5) 셀 A2부터 A5까지의 값들을 모두 더하여 합계를 구한다.

=COUNT(A2:A5) 셀 A2부터 A5 범위 안에 들어 있는 숫자 데이터의 개수를 센다.

=AVERAGE(A2:A5) 셀 A2부터 A5까지의 숫자들의 평균값을 계산한다.
 =AVEDEV(A2:A5) 셀 A2부터 A5까지의 데이터에 대해 평균편차(average deviation)를 계산한다.
 =STDEV(A2:A5) 셀 A2부터 A5까지의 데이터에 대해 표준편차(standard deviation)를 계산한다.

연습문제: 펜의 길이 L 에 대해 다음 다섯 번의 측정값이 있다. 평균과 평균편차를 구하여라.

12.2, 12.5, 11.9, 12.3, 12.2 cm

연습문제: 질량에 대한 다음 측정값들의 평균과 평균편차를 구하여라.

4.32, 4.35, 4.31, 4.36, 4.37, 4.34 g

(d) 위 방법들 사이의 충돌

어떤 경우에는 ILE, 추정 불확도, 평균편차를 모두 구하게 되고, 이 세 값이 서로 다를 수 있다. 이런 경우에는 보수적으로 세 값 중 가장 큰 값을 최종 불확도로 사용한다.

정답을 올바르게 쓰는 방법은 다음과 같다.

1. ILE, 추정 불확도, 평균편차 또는 표준편차 중 가장 큰 값을 고른다.
2. 불확도를 유효숫자 1자리 또는 2자리로 반올림한다.
3. 결과값도 불확도와 자리수를 맞추어 반올림한다.
4. 결과값과 불확도를 괄호 안에 쓰고, 단위와 10의 거듭제곱은 괄호 밖에 쓴다.

연습문제: 어떤 물체의 무게를 저울로 여러번 측정하였다. 저울의 ILE는 0.02 g, 평균 질량은 12.14286 g, 평균 편차는 0.07313 g 이다. 질량을 올바르게 써라. 또 다음 표기 각각의 잘못을 지적하여라.

1. 12.14286 g
2. 12.14 ± 0.02 g
3. $12.14286 \text{ g} \pm 0.07313$
4. 12.143 ± 0.073 g
5. 12.143 ± 0.073 g
6. 12.14 ± 0.07
7. 12.1 ± 0.1 g
8. $12.14 \text{ g} \pm 0.07$ g

9. 12.14 ± 0.07 g

연습문제: 최소눈금이 1 mm 인 미터자로 길이를 다섯 번 측정했더니 결과가 모두 123 mm 였다.

123, 123, 123, 123, 123 mm

평균 길이와 길이의 불확도는 얼마인가?

(e) 왜 많은 측정을 하는가? 평균의 표준오차

여러 번의 측정을 수행하면(예를 들어 4~5회 정도) 측정 대상 물리량의 평균값을 더 신뢰할 수 있게 된다는 것은 잘 알려져 있다. 그렇다면 이보다 더 많은 횟수로 측정을 반복하는 것이 과연 의미가 있을까? 통계학을 배우면, 평균값의 신뢰도는 측정 횟수에 의해 영향을 받으며 이를 나타내는 척도가 바로 평균의 표준오차라는 것을 알게 된다.

가장 단순한 경우, 평균의 표준오차는 다음과 같이 정의된다.

$$\sigma_m = \frac{\sigma}{\sqrt{N}}$$

여기서 σ 는 측정값들의 표준편차, N 은 측정 횟수, σ_m 은 평균의 표준오차를 의미한다.

이 식이 의미하는 바는 다음과 같다. 측정 횟수 N 이 증가하면 평균의 표준오차는 $\sqrt{N}$ 에 반비례하여 감소한다. 즉, 측정을 더 많이 반복할수록 평균값의 신뢰도는 점점 높아진다.

다음 예는 이 사실을 가장 단순한 형태로 보여준다. 어떤 물체의 길이를 여러 번 측정한다고 하자. 측정 횟수가 증가해도 평균값과 표준편차는 크게 변하지 않는 경우가 많다. 이는 측정값들의 분포 특성이 이미 비교적 안정적이기 때문이다. 그러나 평균의 표준오차는 측정 횟수 N 이 증가함에 따라 크게 감소한다. 다시 말해, 개별 측정값들의 퍼짐 정도(표준편차)는 거의 같더라도, 반복 측정을 많이 할수록 평균값이 실제 값에 가까울 가능성은 점점 커진다.

Finding Standard Error in the Mean

Number of Measurements, N	Average	Standard Deviation	Standard Error
5	15.52 cm	1.33 cm	0.59 cm
25	15.46 cm	1.28 cm	0.26 cm
625	15.49 cm	1.31 cm	0.05 cm
10000	15.49 cm	1.31 cm	0.013 cm

입문 과정에서는 평균의 표준오차는 크게 다루지 않고, 표준편차 또는 불확도 추정값만 사용한다.

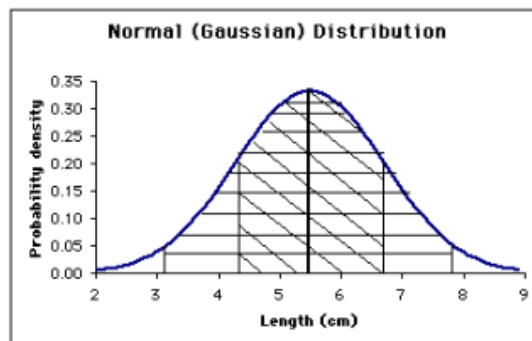
3. 가능한 값의 범위는 무엇인가?

측정 결과가 7.6 ± 0.4 s 와 같이 보고되면, 처음에는 모든 측정값이 7.2 s ($= 7.6 - 0.4$) 와 8.0 s ($= 7.6 + 0.4$) 사이에 들어 있다고 생각하기 쉽다. 그러나 표 1의 데이터를 확인해 보면 실제로는 그렇지 않다. 네 개의 측정값 가운데 이 범위 안에 들어 있는 값은 두 개뿐이다.

통계적으로 보면, 측정값의 약 68%가 $\langle x \rangle \pm Dx$ 범위 안에 들어가며, 약 95%의 값은 $\langle x \rangle \pm 2Dx$ 범위 안에 들어간다고 기대된다.

첫 번째 예에서는 모든 데이터가 6.8 s ($= 7.6 - 2 \times 0.4$)와 8.4 s ($= 7.6 + 2 \times 0.4$) 사이에 존재한다. 두 번째 예에서도 여섯 개의 측정값 가운데 다섯 개가 평균값에서 두 편차 이내에 들어 있다.

이 과목에서는 경험적인 규칙으로, 어떤 물리량의 실제 값은 보통 평균값에서 두 표준편차 이내에 존재한다고 가정한다. 통계학을 배우면 이러한 해석이 신뢰수준(confidence level)과 관련된 개념이라는 것을 자세히 다루게 된다.



정규분포(가우시안 분포)

아래 그림은 정규분포(Normal distribution) 또는 가우시안 분포(Gaussian distribution)를 나타낸 것이다. 그래프의 세로축은 특정 값 z 가 측정될 확률(또는 전체 측정에서 그 값이 차지하는 비율)을 나타낸다. 가장 가능성이 높은 값은 평균값이며, 이 예에서는 $\langle z \rangle = 5.5$ cm 이다. 표준편차는 $s = 1.2$ 이다. 그래프에서 가운데 음영으로 표시된 영역은 $-s$ 부터 $+s$ 사이에 해당하며, 전체 측정값의 약 67%가 이 범위 안에 들어간다. 더 넓게 표시된 음영 영역은 $-2s$ 부터 $+2s$ 범위이며, 전체 측정값의 약 95%가 이 범위 안에 존재한다.

통계학에서는 이러한 분포와 신뢰수준에 대해 훨씬 더 자세하게 다룬다.

4. 절대오차와 상대오차

물리량 z 에 대한 불확도가 Δz 일 때, Δz 를 **절대오차**라고 한다. 절대오차는 원래 물리량과 같은 단위를 가진다. 상대오차는 절대오차를 측정값으로 나눈 값이다.

$$\text{상대오차} = \frac{\Delta z}{z}$$

상대오차는 fractional uncertainty라고도 한다. 퍼센트오차는 상대오차에 100% 를 곱한 값이다.

실제 공학에서는 상황에 따라 절대오차 또는 퍼센트오차 중 하나의 방식으로 허용오차를 제시한다. 예를 들어 기계 가공에서 엔진 부품의 허용오차는 보통 절대오차 형태로 제시된다. 반면 전자 부품의 경우에는 대부분 퍼센트 허용오차로 표시된다.

문제: 저항값이 1200 Ω 이고 허용오차가 5% 인 저항이 있다. 이 저항의 절대오차는 얼마인가?

5. 오차의 전파, 기본 규칙

두 측정량 x 와 y 의 불확도가 각각 Δx , Δy 로 주어져 있다고 하자. 앞에서 설명한 절차에 따라 이들을 구했다면, 측정 결과는 각각 $x \pm \Delta x$, $y \pm \Delta y$ 와 같이 보고할 수 있다. 이제 이 두 측정값으로부터 새로운 물리량 z 를 계산한다고 하자. 이때 문제는, 계산으로 얻어진 z 의 불확도 Δz 를 어떻게 구할 것인가 하는 점이다.

이 과목에서는 엄밀한 통계적 처리 대신, 그것을 단순화한 방법을 사용한다. 다만 표준편차를 사용하는 보다 정식의 통계적 처리식도 함께 제시한다. 여기서 기본 원칙은 모든 경우에 대해 가장 보수적인 경우, 즉 결과가 참값에서 가장 크게 벗어날 수 있는 상황을 기준으로 생각하는 것이다. 이러한 방법의 엄밀한 이론적 배경은 통계학에서 다룬다.

이 절에 제시되는 예제들은 오차의 전파뿐 아니라, 계산 결과를 어떻게 적절히 반올림하여 보고해야 하는지도 함께 보여 준다. 반올림 규칙 자체는 6절에서 더 자세히 설명한다. 또한 여기의 예제들은 기본적으로 평균편차를 이용한 오차 전파를 기준으로 작성되어 있다.

(a) 덧셈과 뺄셈

$z = x + y$ 또는 $z = x - y$ 일 때, 단순한 방법에서는 절대오차를 더한다.

$$\Delta z = |\Delta x| + |\Delta y|$$

표준편차를 쓰면

$$\Delta z = \sqrt{(\Delta x)^2 + (\Delta y)^2}$$

를 사용한다.

예제: $w = 4.52 \pm 0.02$ cm, $x = 2.0 \pm 0.2$ cm, $y = 3.0 \pm 0.6$ cm 일 때,

$z = x + y - w = 0.5$ cm 라면

$\Delta z = 0.2 + 0.6 + 0.02 = 0.82$ cm ≈ 0.8 cm 이다.

따라서,

$$z = 0.5 \pm 0.8 \text{ cm}$$

정확한 수를 곱하는 경우에는 불확도도 같은 수만큼 곱한다.

예를 들어 반지름이 $x = 3.0 \pm 0.2$ cm 일 때, 원주의 길이는 $C = 2\pi x = 18.850$ cm 이고,

$\Delta C = 2\pi\Delta x = 1.257$ cm 이다.

따라서, $C = 18.8 \pm 1.3$ cm 이다.

(b) 곱셈과 나눗셈

$z = xy$ 또는 $z = \frac{x}{y}$ 일 때, 단순한 방법에서는 상대오차를 더한다.

$$\frac{\Delta z}{z} = \frac{\Delta x}{x} + \frac{\Delta y}{y}$$

표준편차를 쓰면

$$\frac{\Delta z}{z} = \sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2}$$

를 사용한다.

예제:

$w = 4.52 \pm 0.02$ cm, $x = 2.0 \pm 0.2$ cm 일 때, $z = wx = 9.04$ 이다.

$$\frac{\Delta z}{z} = \frac{0.02}{4.52} + \frac{0.2}{2.0} = 0.1044$$

$$\Delta z = 0.1044 \times 9.04 = 0.944 \approx 0.9$$

따라서,

$$z = 9.0 \pm 0.9$$

(c) 거듭제곱의 곱

$z = x^a y^b A^c \dots$ 와 같이 변수들이 거듭제곱으로 들어 있을 때, 단순한 방법에서는 지수만큼 상대오차에 가중치를 준다.

$$\frac{\Delta z}{z} = |a| \frac{\Delta x}{x} + |b| \frac{\Delta y}{y} + |c| \frac{\Delta A}{A} + \dots$$

예를 들어 어떤 항에 y^2 가 들어 있으면 $2 \frac{\Delta y}{y}$ 가 들어가고, $\sqrt{A}$ 가 들어 있으면 $(1/2) \frac{\Delta A}{A}$ 가 들어간다.

(d) 기타 함수의 단순 접근

$z = \sin x$ 같은 함수에는 이 자료에서 별도의 일반 공식을 바로 주지 않는다. 대신 평균값과 가장 큰 가능한 값의 차이를 사용한다.

$$\Delta(\sin x) = \sin(x + \Delta x) - \sin x$$