



## 왜 평균절대편차 대신 표준편차를 쓸까? 왜 중앙값이 아니라 평균인가?

주어진 값들의 퍼짐 정도를 나타내는 분산variance은 다음과 같이 구한다.

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

분산은 평균과 각 값의 차이를 제곱한 값들의 평균이다. 처음 분산을 배웠을 때, 왜 제곱을 해야 하는지 의아했을 것이다. 물론 편차( $x_i - \bar{x}$ ; 각 값에서 평균을 뺀 값)를 모두 더하면 항상 0이 되므로, 편차의 평균은 항상 0이다.<sup>9)</sup> 그렇기에 편차의 제곱의 평균을 구한다고 설명할 수도 있지만 어딘지 꺼림직하다. 편차의 절대값(절대편차absolute deviation)을 평균할 수도 있지 않은가? 각 값과 평균의 편차의 절대값을 평균한 값을 평균절대편차average absolute deviation라 하고 식은 다음과 같다.

---

9)  $\sum_{i=1}^n (x_i - \bar{x}) = 0$ . 평균의 정의에 의해  $\frac{\sum_{i=1}^n x_i}{n} = \bar{x}$ 이므로, 양변에  $n$ 를 곱하면,  $\sum_{i=1}^n x_i = n\bar{x}$ , 평균을 왼쪽으로 옮기면  $\sum_{i=1}^n x_i - n\bar{x} = 0$ 이 되는데,  $n\bar{x}$ 는  $\sum_{i=1}^n \bar{x}$ 과 같으므로( $\bar{x}$ 가  $i$ 에 따라 변하는 값이 아니므로,  $\sum_{i=1}^n \bar{x}$ 는  $\bar{x}$ 를  $n$ 번 더한 값과 같다.)  $\sum_{i=1}^n x_i - n\bar{x} = \sum_{i=1}^n (x_i - \bar{x}) = 0$ 이 성립한다.



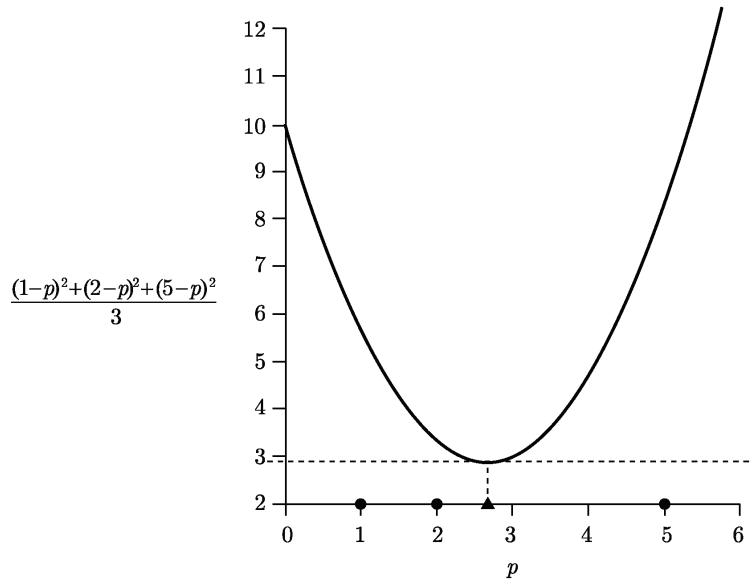
$$AAD = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} \quad (AAD: \text{Average Absolute Deviation})$$

왜 분산을 사용하는지에 대해 알아보기 전에 분산과 절대편차를 구하는 식을 좀더 자세히 살펴보자. 위의 분산을 구하는 식에서 평균  $\bar{x}$ 는  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$  이지 만, 다른 값을 넣어볼 수도 있지 않을까?

$\bar{x}$ 를 변수  $p$ 로 대치한 다음의 식  $\sigma_p^2 = \frac{\sum_{i=1}^n (x_i - p)^2}{n}$ 를  $p$ 를 중심으로 한 분산이라고 정의하자.  $x_i$ 값이 주어졌을 때,  $p$ 값에 따른  $\sigma_p^2$ 의 변화를 살펴보면 재미있는 현상을 볼 수 있다.

구체적으로  $x_1 = 1, x_2 = 2, x_3 = 5$ 인 경우를 생각해보자.

$\sigma_p^2 = \frac{(1-p)^2 + (2-p)^2 + (5-p)^2}{3}$ 이고, 이것을  $p$ -좌표 위의 그래프로 그려 보면 다음과 같다.

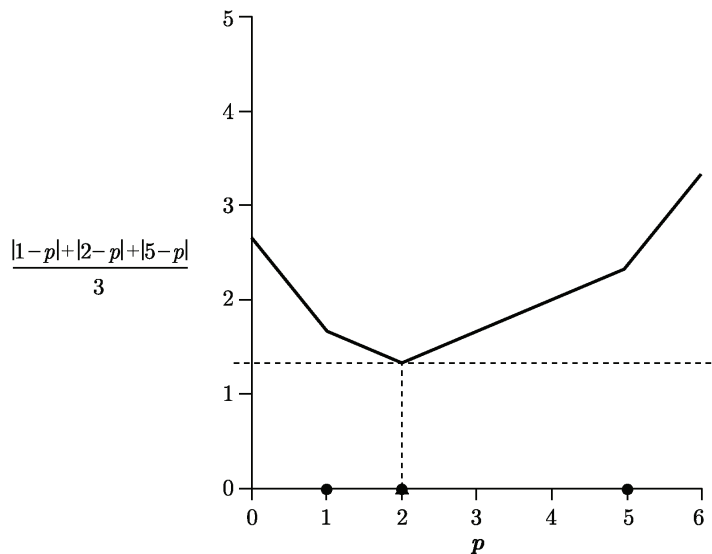


자료의 수가 많아지더라도  $\sigma_p^2$ 은 항상  $p$ 의 2차 함수 꼴로 나타난다. 그리고  $\sigma_p^2$ 의 최소값은  $p$ 가  $x_i$ 의 평균일 때 나타난다. 평균을 특정한 관찰값들의 대푯값이라 생각하고, 분산을 대푯값으로부터 각 관측값들이 얼마나 넓게 퍼져 있는지를 정량화한 값이라고 생각한다면, 대푯값은 분산(자료들이 퍼져 있는 정도)을 최소로 해야 할 것이다. 따라서 분산을 최소화하는 평균을 대푯값으로 정하는 것이 자연스러워 보인다.

하지만 주어진 값들이 어떤 값을 중심으로 얼마나 퍼져 있는지를 정량화하는 방법은 분산만 있는 것이 아니다. 평균절대편차를 쓰면 어떨까? 변수  $p$ 를 중심으로 평균절대편차를 구하는 식은 다음과 같다.

$$D_p = \frac{\sum_{i=1}^n |x_i - p|}{n}$$

앞에서와 마찬가지로  $p$ 값의 변화에 따른  $D_p$ 값을 그래프로 그려보면 다음과 같다.





평균절대편차의 경우에도 주어진 값의 수( $n$ )가 그래프의 모양에 큰 영향을 미치지 못한다. 함수의 모양은 항상 주어진 값  $x_i$ 에서 꺾어진 직선의 모양으로 나타난다. 그리고 최소값은 항상 자료의 중앙값에서 나타난다. 따라서 평균절대편차를 가지고 주어진 값들이 얼마나 퍼져 있는지를 측정할 때, 대푯값을 중앙값으로 쓰는 것이 자연스럽다.

그렇다면 왜 거의 대부분의 경우 평균절대편차 대신 분산을 쓰는 것일까?

그것을 설명하기 위해 먼저 최대 가능도 방법(maximum likelihood method(혹은 최대우도 방법이라고도 한다))을 알 필요가 있다. 최대 가능도 방법이란 모수를 추정함에 있어, 주어진 값이 나올 확률이 가장 큰 모수를 사용하는 것이다.

예를 들어서 어떤 값을 측정할 확률분포가 아래의 정규분포를 따른다고 해보자. 그리고 우리에게 주어진 값들은 이 확률분포에서 독립적으로 측정된 것이라 생각하자.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

위의 정규분포는 두 개의 모수  $\mu$ ,  $\sigma$ 가 있는데, 이 두 수가 결정되면 확률분포  $f(x)$ 를 결정할 수 있다. 이 두 개의 모수  $\mu$ ,  $\sigma$ 를 어떻게 추정할 것인가? 최대 가능도 방법은 주어진 값이 나타날 확률을 가장 크게 하는 모수를 찾아서 그 값을 사용한다.

$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$  을 따르는  $x$  값( $x_1, x_2, \dots, x_n$ )을 실제로 관찰하였다. 이 값을 가지고 모수  $\mu$ ,  $\sigma$ 를 추정할 때,  $x_1, x_2, \dots, x_n$ 을 관찰할 확률(확률밀도)를 최대가 되게 하는  $\mu$ ,  $\sigma$ 를 구하는 방법은 다음과 같다.

관찰값( $x_1, x_2, \dots, x_n$ )이 주어졌고, 그 값들이 서로 독립적으로 얻어졌을 때, 그 값들을 동시에 얻을 확률밀도인  $f(x_1)f(x_2)\dots f(x_n)$ 을 최대로 하는  $\mu$ ,  $\sigma$ 를 구하기 위해서는  $f(x_1)f(x_2)\dots f(x_n)$ 를 먼저 구해야 한다.

$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$  이므로  $f(x_1)f(x_2)\cdots f(x_n)$ 는 다음과 같이 전개할 수 있다.

$$f(x_1)f(x_2)\cdots f(x_n) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_1-\mu}{\sigma}\right)^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_2-\mu}{\sigma}\right)^2} \cdots \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2}$$

(지수 법칙( $e^a e^b = e^{a+b}$ )에 따라 공통된 지수의 밑( $e$ )를 합친다.)

$$= \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2}\left(\frac{x_1-\mu}{\sigma}\right)^2 - \frac{1}{2}\left(\frac{x_2-\mu}{\sigma}\right)^2 - \cdots - \frac{1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2}$$

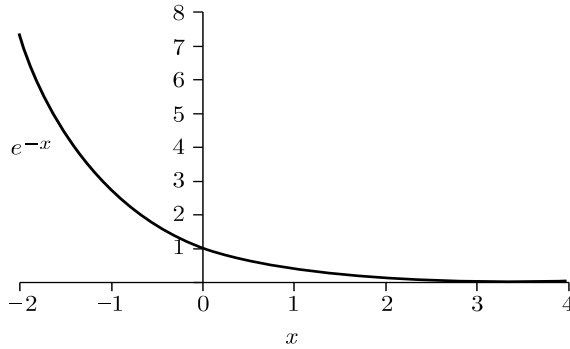
( $e$ 의 지수에 반복적으로 나타나는  $-\frac{1}{2}$ 을 합친다.)

$$= \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2}\left[\left(\frac{x_1-\mu}{\sigma}\right)^2 + \left(\frac{x_2-\mu}{\sigma}\right)^2 + \cdots + \left(\frac{x_n-\mu}{\sigma}\right)^2\right]}$$

( $e$ 의 지수에 반복적으로 나타나는  $\frac{1}{\sigma^2}$ 을 합친다.)

$$= \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2\sigma^2}[(x_1-\mu)^2 + (x_2-\mu)^2 + \cdots + (x_n-\mu)^2]}$$

$e^{-x}$  함수는 다음과 같이 나타난다.



따라서  $e^{-x}$  값이 최대가 되기 위해서는  $x$  값이 최소가 되어야 한다. 즉,  
 $f(x_1)f(x_2)\cdots f(x_n) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{2\sigma^2}[(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_n-\mu)^2]}$ 의 값이 최대  
 가 되기 위해서는  $\frac{1}{2\sigma^2}[(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_n-\mu)^2]$ 가 최소가 되어야  
 하는데, 그것은 앞서 봤듯이  $\sigma$ 의 값에 상관없이  $\mu = \frac{\sum_{i=1}^n x_i}{n}$ 일 때이다.

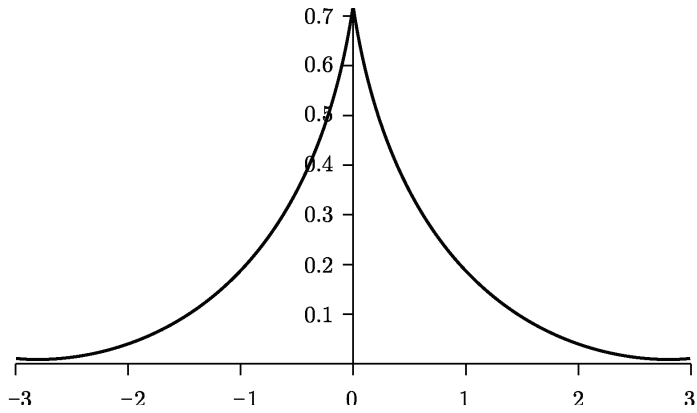
결국 정규분포일때, 모수  $\mu$ 의 최대 가능도 추정치는 표본평균이다!

그렇다면 중앙값은 분포가 어떤 경우에 최대 가능도 추정치일까?

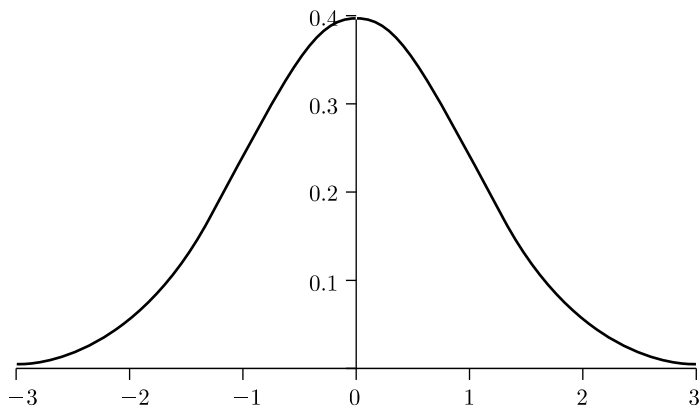
만약 확률분포함수가 다음과 같다면 중앙값이 최대 가능도 추정치가 된다.  
 정규분포일 때와 같이  $f(x_1)f(x_2)\cdots f(x_n)$ 를 전개한 후, 최대값이 되기 위한  $\mu$   
 값을 구하면 된다.

$$f(x) = \frac{1}{\sqrt{2}\sigma} e^{-\sqrt{2}\left|\frac{x-\mu}{\sigma}\right|}$$

이 확률밀도의 그래프는 다음과 같은 모양이다.



이에 반해 정규분포의 모양은 다음과 같다.



어떤 모양의 확률분포가 더 흔하게 발견될까?

 요약

평균은 분산과 표준편차(표준편차는 분산의 양의 제곱근이다)을 최소로 만드는 대푯값이고, 중앙값은 평균절대편차를 최소로 만드는 대푯값이다.





정규분포에서는 분산을 최소화 하는 값(평균)이 모평균의 최대 우도 추정치MLE 이고,  $f(x) = \frac{1}{\sqrt{2}\sigma} e^{-\sqrt{2}\left|\frac{x-\mu}{\sigma}\right|}$ 를 따르는 확률분포(이런 확률분포를 라플라스 분포Laplace distribution 혹은 이중지수분포double exponential distribution라고 부른다.)에서는 평균절대편차를 최소화 하는 값(중앙값)이 모평균의 최대 우도 추정치이다.