

기댓값의 벡터 해석

기하 2학기 개인주제탐구

2025-10-30

목차

I. 발상	3
II. 기댓값의 벡터 해석	3
III. 벡터 공간으로 전개	4
IV. 결론	9

I. 발상

확률과 통계에서 배우는 평균, 즉 기댓값은 변량의 합을 개수로 나누는 방법 이외에도 변량에 각각 그에 해당하는 확률을 곱하는 방법으로 구할 수 있다.

$$E = \frac{\sum_{i=1}^n x_i}{n} = \sum x_i p_i \quad (1.1)$$

또, 확률변수 X 와 $\forall a, b \in \mathbb{R}$ 에 대해 다음이 성립한다.

$$E(aX + b) = aE(X) + b \quad (1.2)$$

이로써 직관적으로 기댓값이라는 것은 선형이라는 것을 알 수 있다.

Definition 1.1 (선형성) 어떤 사상 f 에 대해 이것이 선형이라면 다음을 만족해야 한다.

1. 가산성: $f(x + y) = f(x) + f(y)$
2. 동차성: $f(ax) = af(x)$

엄밀한 증명은 식 1.1을 통해 정의 1.1을 검증하는 자명한 과정이므로 생략한다.

기댓값은 왜 선형인가에 대해 그 이유를 잠깐 생각해보았더니 기댓값의 식이 내적 꼴로 되어있기 때문이라는 것을 알았다. 변량벡터인 $\mathbf{x} = [x_1 \ \dots \ x_n]^T$ 와 확률벡터인 $\mathbf{p} = [p_1 \ \dots \ p_n]^T$ 의 내적으로 나타낼 수 있는 것이다. 따라서

$$E(X) = \sum x_i p_i = \mathbf{x} \cdot \mathbf{p} \quad (1.3)$$

II. 기댓값의 벡터 해석

벡터의 내적 꼴을 풀어써보면

$$\mathbf{a} \cdot \mathbf{x} = a_1 x_1 + a_2 x_2 + \dots + a_n x_n \quad (2.1)$$

으로, 두 벡터 중 하나를 계수의 벡터, 다른 것을 변수의 벡터로 본다면 이것은 변수 벡터의 각 항에 계수가 붙어 일정한 가중치를 주고 있다고 생각할 수 있다.

즉, 가중치 합으로서의 선형결합이다. 여기서, 평균과 기댓값이 같은 값을 가리킴에도 용어가 다른 이유를 대강 알아볼 수 있다. 평균(平均, mean, average)란 변량의 총합을 균등하게 배분한 것으로, 당장 생각나는 비유를 하자면 양동이보다 더 수북이 쌓인 모래를 삽으로 깎아 양이 적은 양동으로 옮기는 느낌이다. 반면 기댓값(expectation value)은 변량에 그 변량이 나올 확률을 곱해줌으로써 가중치를 매겨 어떤 변량이 나올 가능성이 더 큰지를 정해, 시행 전체를 대변할 수 있는 값 하나를 정하는 느낌이다.

벡터의 내적은 이러한 수식적 의미 외에도 기하학적 의미를 가지고 있는데, 한 벡터를 다른 벡터에 투영시키는 것이다. 그 말은 즉슨, 내적의 결과인 스칼라 값은 두 벡터의 방향이 일치하는 정도가 클수록 더 커진다.

$$\mathbf{u} \cdot \mathbf{v} = uv \cos \theta \quad (2.2)$$

이러한 성질을 기하학적으로 표현하면서 가중치 합의 의미까지 살리기 위해서는 기댓값을 확률벡터 \mathbf{p} 를 주어진 공간 위에서 변량벡터로 투영한 결과로 이해할 수 있다. 그러므로 기댓값이란 확률분포의 방향이 일정하게 정렬되었는가, 또는 얼마나 상관관계가 있는가 등을 나타내는 척도라고 할 수 있다.

간단히 생각해보면 기댓값은 수직선상에서 전체 변량이 균형을 이루는 점, 또는 확률 분포를 나타낸 점들의 무게중심 정도로 생각할 수 있다. 이런 결론은 평균의 원리를 체감하며 사는 우리에게는 당연하게 느껴지면서도 실망을 안겨준다. 더 깊은 의미가 있을 것이라고 생각되기 때문이었다.

III. 벡터 공간으로 전개

다시 처음으로 돌아가서 논리를 전개해보자. 어떤 이산확률변수 X 에 대해 $X \in \{x_1, \dots, x_n\}$ 이라고 하고 각 값에 해당하는 확률은 $P(X = x_i) =: p_i$ 이다. 이때 기댓값의 정의는

$$E[X] = \sum_{i=1}^n x_i p_i \quad (3.1)$$

이 식을 n 차원 유클리드 공간 \mathbb{R}^n 에서의 내적으로 표현하기 위해 변량과 확률을 각각 벡터로 정의한다.

$$\mathbf{x} := \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n, \mathbf{p} := \begin{bmatrix} p_1 \\ \vdots \\ p_i \end{bmatrix} \in \mathbb{R}^n \quad (3.2)$$

이때 확률벡터 \mathbf{p} 에 대해 $p_i \geq 0$ 이고 $\sum p_i = 1$ 이다. 정의에 따라

$$E[X] = \mathbf{x} \cdot \mathbf{p} = \mathbf{x}^T \mathbf{p} \quad (3.3)$$

기하학적 의미를 실감하기 위한 방법을 고민한 결과 \mathbb{R}^n 의 공간을 두 개의 정규직교기저를 통해 직교부분공간으로 분해해 나타내기로 했다. 먼저 평균 그 자체를 나타내는 1차원 공간 W_{mean} 을 잡을 것인데, 변량의 산술평균을 나타내는 \bar{x} 의 의미를 기하학적으로 대변한다고 수 있다.

$$W_{\text{mean}} = \left\{ k \cdot \mathbf{1} \mid k \in \mathbb{R}, \mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^n \right\} \quad (3.4)$$

이 공간의 기저인 \mathbf{u}_1 은 $\mathbf{1}$ 을 정규화한 것으로

$$\mathbf{u}_1 = \frac{\mathbf{1}}{|\mathbf{1}|} = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad (3.5)$$

자명하게

$$W_{\text{mean}} = \text{Span}\{\mathbf{u}_1\} \quad (3.6)$$

이 공간에 속한 벡터는 (k, \dots, k) 의 형태로 모두 균등한 값을 가진다. 따라서 이 공간에 변량 벡터를 정사영하면

$$\mathbf{x}_{\parallel} = \bar{x} \cdot \mathbf{1} = \begin{bmatrix} \bar{x} \\ \vdots \\ \bar{x} \end{bmatrix} \quad (3.7)$$

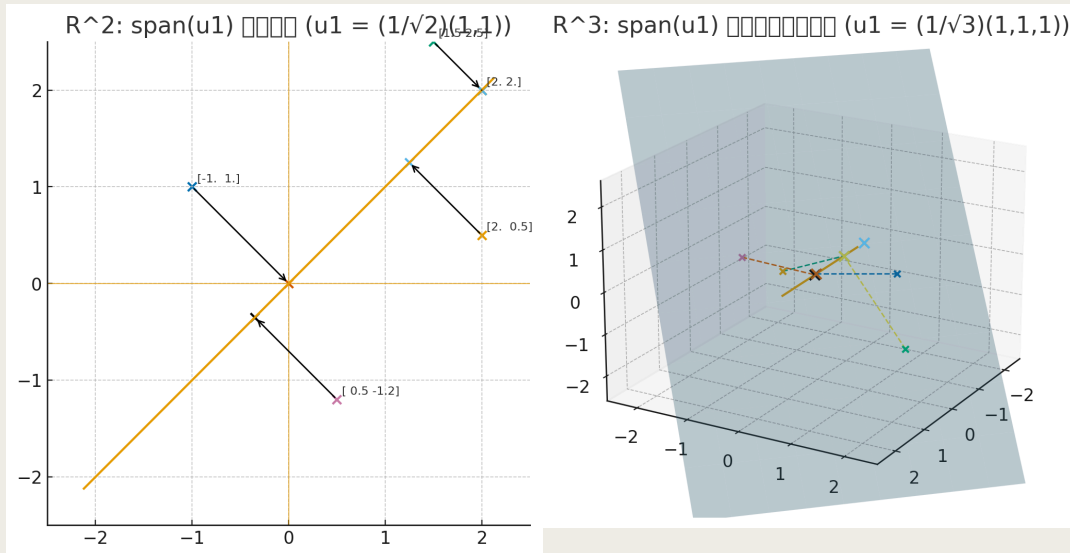
또 이 공간에 확률벡터를 정사영하면

$$\mathbf{p}_{\parallel} = \frac{1}{n} \mathbf{1} = \begin{bmatrix} 1/n \\ \vdots \\ 1/n \end{bmatrix} \quad (3.8)$$

이 공간에서 둘을 내적하면 이때 \mathbf{p}_{\parallel} 이 $\frac{1}{n}$ 으로 균등하므로 자명하게

$$\mathbf{x}_{\parallel} \cdot \mathbf{p}_{\parallel} = \bar{x} \quad (3.9)$$

Figure 3.1 — matplotlib 로 시각화한 W_{mean} 공간



즉 W_{mean} 에 어떤 \mathbb{R}^n 상의 벡터를 정사영하면 벡터의 원소의 평균을 모든 좌표로 가지는 벡터로 나타나게 된다.

이제 $n - 1$ 차원 부분공간 W_{dev} 를 생각하자. W_{dev} 는 W_{mean} 에 직교하며, 이 공간 상의 벡터 \mathbf{v} 는 $\mathbf{v}^T \mathbf{u}_1 = 0$, 즉 $\sum v_i = 0$ 을 만족한다. 이 공간의 기저는 $\{\mathbf{u}_2, \dots, \mathbf{u}_n\}$ 으로 \mathbf{u}_1 , 즉 W_{mean} 에 직교한다는 것에 초점을 둔다.

공간 $W_{\text{dev}} = \text{Span}\{\mathbf{u}_1\}^\perp$ 이 W_{mean} 에 직교하므로 내적이 0이라는 것을 강조한다.

$$\begin{aligned} W_{\text{dev}} \perp W_{\text{mean}} &\Rightarrow \forall \mathbf{v} \in W_{\text{dev}} \quad \mathbf{v} \cdot \mathbf{u}_1 = 0 = \sum v_i \cdot 1 \\ &\therefore \sum v_i = 0 \end{aligned} \quad (3.10)$$

이 공간 상의 모든 벡터는 원소의 총합이 0이 되는데, 이것은 편차의 성질이다. 따라서 명명도 deviation을 따 W_{dev} 라고 하였고, 이 공간을 편차의 의미를 대변한다고 하자.

이제 변량벡터 \mathbf{x} 와 확률벡터 \mathbf{p} 를 위에서 정의한 두 직교 부분공간 W_{mean} 과 W_{dev} 로 분해하여 $E[X]$ 를 다시 계산하자. 이때, 벡터는 두 부분공간의 벡터합이다.

$$\begin{aligned} \mathbf{x} &= \mathbf{x}_{\parallel} + \mathbf{x}_{\perp} \\ \mathbf{p} &= \mathbf{p}_{\parallel} + \mathbf{p}_{\perp} \end{aligned} \quad (3.11)$$

먼저 변량벡터를 분해하자. 위에서는 직관적으로 보았지만, 수식을 전개함으로써 좀 더 엄밀히 살펴보면

$$\mathbf{x}_{\parallel} = (\mathbf{x}^T \mathbf{u}_1) \mathbf{u}_1 = \left(\frac{1}{\sqrt{n}} \sum x_i \right) \left(\frac{1}{\sqrt{n}} \mathbf{1} \right) = \left(\frac{1}{n} \sum x_i \right) \mathbf{1} \quad (3.12)$$

이때 변량의 평균 \bar{x} 는 $\frac{1}{n} \sum x_i$ 이므로

$$\mathbf{x}_{\parallel} = \bar{x} \cdot \mathbf{1} = \begin{bmatrix} \bar{x} \\ \vdots \\ \bar{x} \end{bmatrix} \quad (3.13)$$

또,

$$\mathbf{x}_{\perp} = \mathbf{x} - \mathbf{x}_{\parallel} = \begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix} \quad (3.14)$$

가 되어 W_{dev} 위의 \mathbf{x}_{\perp} 는 실제로 편차임을 확인할 수 있다.

확률벡터 \mathbf{p} 도 똑같이 하자.

$$\begin{aligned} \mathbf{p}_{\parallel} &= (\mathbf{p}^T \mathbf{u}_1) \mathbf{u}_1 = \left(\frac{1}{\sqrt{n}} \sum p_i \right) \left(\frac{1}{\sqrt{n}} \mathbf{1} \right) \\ &= \left(\frac{1}{\sqrt{n}} \cdot 1 \right) \left(\frac{1}{\sqrt{n}} \mathbf{1} \right) = \frac{1}{n} \mathbf{1} = \begin{bmatrix} 1/n \\ \vdots \\ 1/n \end{bmatrix} \\ &\because \sum p_i = 1 \end{aligned} \quad (3.15)$$

또,

$$\mathbf{p}_{\perp} = \mathbf{p} - \mathbf{p}_{\parallel} = \begin{bmatrix} p_1 - 1/n \\ \vdots \\ p_n - 1/n \end{bmatrix} \quad (3.16)$$

이는 확률분포가 균일한 분포에서 얼마나 벗어났는지를 나타내게 된다.

이제 기댓값은

$$E[X] = (\mathbf{x}_{\parallel} + \mathbf{x}_{\perp}) \cdot (\mathbf{p}_{\parallel} + \mathbf{p}_{\perp}) = \mathbf{x}_{\parallel} \cdot \mathbf{p}_{\parallel} + \mathbf{x}_{\perp} \cdot \mathbf{p}_{\parallel} + \mathbf{x}_{\parallel} \cdot \mathbf{p}_{\perp} + \mathbf{x}_{\perp} \cdot \mathbf{p}_{\perp} \quad (3.17)$$

최종적으로

$$E[X] = \mathbf{x}_{\parallel} \cdot \mathbf{x}_{\perp} + \mathbf{x}_{\perp} \cdot \mathbf{p}_{\perp} \quad (3.18)$$

첫 항은 \bar{x} 로, 만일 $\mathbf{p} = \mathbf{p}_{\parallel}$ 였다면 $E[X] = \bar{x}$ 일 것이다. 둘째 항은 아래와 같다.

$$\mathbf{x}_{\perp} \cdot \mathbf{p}_{\perp} = \sum_{i=1}^n (x_i - \bar{x}) \left(p_i - \frac{1}{n} \right) \quad (3.19)$$

이것이 기댓값의 기하학적 의미를 잘 나타내는 것으로, 변량 편차와 확률 편차가 얼마나 일치, 상관하는가를 나타낸다.

만약 $E[X] > 0$ 이면 $\mathbf{x}_{\perp} \cdot \mathbf{p}_{\perp} > 0$ 으로, 두 편차 벡터가 W_{dev} 에서 비슷한 방향을 보고 있다는 뜻이다. 엄밀히는, $\theta < 90^\circ$ 이다. 즉, \bar{x} 보다 큰 x_i 값이 더 큰 p_i 값을 가져서, 큰 값에 큰 가중치가 부여된다는 뜻이다.

반대로 $E[X] < 0$ 이면 $\theta > 90^\circ$ 로, 작은 값에 큰 가중치가 부여된다는 뜻이다.

$E[X] = \bar{x}$ 라면 두 편차 벡터가 직교한다는 뜻으로, $\mathbf{x}_{\perp} \neq \mathbf{0} \wedge \mathbf{p}_{\perp} \neq \mathbf{0}$ 이라면 변량의 크기와 확률의 크기에 아무런 상관관계가 없음을 말한다.

III.1. 분산에서의 적용

분산은 편차의 제곱의 기댓값이기 때문에 같은 논리를 적용해볼 수 있다. 먼저, $\mu := E[X]$ 라고 났을 때 분산의 정의는

$$V(X) = E[(X - \mu)^2] = E[X^2] - (E[X])^2 \quad (3.20)$$

$Y = (X - \mu)^2$ 으로 놓으면 $V(X) = E[Y]$ 이다. 편차의 제곱의 변량벡터를 \mathbf{v} 로 놓자.

$$\mathbf{v} = \begin{bmatrix} (x_1 - \mu)^2 \\ \vdots \\ (x_n - \mu)^2 \end{bmatrix} \in \mathbb{R}^n \quad (3.21)$$

이제 분산은

$$V(X) = \sum_{i=1}^n (x_i - \mu)^2 p_i = \mathbf{v} \cdot \mathbf{p} \quad (3.22)$$

기댓값에서 얻은 결론과 상통하게

$$V(X) = \mathbf{v}_{\parallel} \cdot \mathbf{p}_{\parallel} + \mathbf{v}_{\perp} \cdot \mathbf{p}_{\perp} \quad (3.23)$$

여기서 \mathbf{v}_{\parallel} 의 모든 성분은 편차의 제곱의 산술평균으로 $\frac{1}{n} \sum (x_i - \mu)^2$ 인데, 이 값을 \bar{v} 라고 하겠다. 그러면

$$\mathbf{v}_{\perp} = \begin{bmatrix} (x_1 - \mu)^2 - \bar{v} \\ \vdots \\ (x_n - \mu)^2 - \bar{v} \end{bmatrix} \quad (3.24)$$

가 된다.

이제 분해한 성분들을 위에서 구해놓았던 기댓값 분해 식에 대입하여

$$V(X) = \bar{v} + \sum_{i=1}^n \left[((x_i - \mu)^2 - \bar{v}) \left(p_i - \frac{1}{n} \right) \right] \quad (3.25)$$

둘째항은 기댓값에서와 같이, 확률 가중치 \mathbf{p}_{\perp} 가 흩어진 정도 \mathbf{v}_{\perp} 를 증폭하는지 감쇄하는지를 나타낸다.

IV. 결론

$$E[X] = \mathbf{x} \cdot \mathbf{p} = \bar{x} + \sum_{i=1}^n (x_i - \bar{x}) \left(p_i - \frac{1}{n} \right) \quad (4.1)$$

벡터로 꾸역꾸역 해석해서 얻어낸 결론이란, 기댓값은 변량의 가중치 선형 합으로 변량에 확률을 내적하여 그 확률이 각 변량의 크기의 효과가 최종 계산 결과에 미치는 영향을 증폭하거나 감쇄하는 쪽으로 보정하여 만들어낸, 산술평균에 확률이 적용된 확률분포의 대푯값이라는 것이었다. 사실 계산법이 간단하고 만들어진 의도가 분명한만큼 특출난 다른 기하학적 의미가 있진 않겠지만 그래도 특출난 것을 기대하면서 여러 방법으로 고민한 결실로 만족스럽지는 못하다. 다만 깨달은 점도 분명 있다.

물론 계산하는 대상이 무엇인가에 따라 다르겠지만, 이 “기댓값”이라는 용어의 의미가 단순한 “평균”과는 무엇이 다른가 대해 직관적으로 와닿게 실감할 수 있었다는 의미가 있었다. 또, 원래 기댓값이란 변량에 확률이 가중치합으로 작용하는 것이라고 이해하고 있었는데, 이번 탐구를 통해 이것이 틀린 말은 아니지만, 더 의미를 잘 나타내는 서술로는 변량 편차에 확률 편차가 가중치로 작용한다는 것이 있다는 것도 새로 알게 되었다.