

이따라,  
평균 표본이  
상대공간 내  
모든 원소가  
없을 수 있다.

현실 세계의 데이터는 확률변수가 가진 확률분포에 따라 실수 표본공간에서 선택된 표본이다. 이렇게 확률분포함수에 따라 표본공간의 표본이 현실 세계의 데이터로 선택되는 것을 **실현 (realization)** 혹은 **\*\*표본화(sampling)\*\***라고 한다. 표본화는 다른 의미로도 사용되는데 많은 수의 데이터 집합에서 일부 데이터만 선택하는 과정도 표본화라고 한다.

**실현**은 이상적이고(ideal), 추상적이며 (abstract), 수학적인(mathematical) 세계에서 현실로 넘어가는 과정이다. 확률변수라는 데이터 모형에 따르면 현실의 **표본 데이터**는 이러한 수학적 세계가 현실 세계에 투영된 그림자에 지나지 않는다. 따라서 "히스토그램이나 기술통계 등 분포의 특성이 깨끗한 형태를 지니지 않지만 이는 실현 혹은 표본화 과정에서 생긴 잡음일 뿐이며 그 내면에는 원래의 "수학적 특성"을 내포하고 있다."

표본의  
특성

↑ '확률변수'의 특성.

따라서 확률변수와 실제 데이터는 다음과 같은 관계가 있다.

- 확률변수로부터 데이터를 여러 번 생성하는 경우 실제 데이터값은 매번 달라질 수 있지만 확률변수 자체는 변하지 않는다.
- 확률변수의 확률분포함수는 우리가 직접 관찰할 수 없다. 다만 확률변수에서 만들어지는 실제 데이터값을 이용하여 확률분포함수가 이러한 것일 거라고 추정할 뿐이다.
- 확률변수에서 만들어지는 실제 데이터값은 확률변수가 가진 특성을 반영하고 있다. 데이터 개수가 적을수록 확률변수가 가진 특징을 정확하게 표현하지 못하지만 데이터 개수가 증가하면 보다 정확하게 확률분포함수를 묘사할 수 있게 된다.

↑ 표본의 크기가 클수록, 해당 표본이 좀더 정확하게 확률분포함수를 묘사하기 된다.