Central Limit Theorem

2016314726 정영준

Central Limit Therorem (중심 극한 정리)는 동일한 확률분포를 가진 독립 확률 변수 n개의 평균 분포는 n이 크다면 정규분포에 가까워진다는 정리입니다. 중심 극한 정리에 대해 설명하기 전에 먼저 i.i.d (Independency and identically Distributed)에 대해 설명하겠습니다. Iid는 확률 변수가 여러 개 있을 때 이들이 상호 독립적이며 모두 동일한 확률 분포를 가졌을 때를 의미합니다. 독립 항등 분포라고도 불립니다. 확률 변수들 사이의 모종의 상관관계가 존재하는 다중 공선성등은 iid를 해치는 주 요인입니다. 위의 중심 극한 정리의 정의에서 “동일한 확률분포를 가진 독립 확률 변수 n개”는 iid를 의미합니다. 동전 던지기의 경우 앞면과 뒷면의 볼록 오목한 정도에 따라 확률이 조금씩 달라져 정확히 1/2가 앞면, 1/2가 뒷면이 나올 확률이 되지 않는다고 합니다. 그러므로 앞면, 뒷면만이 나오고 동전이 서는 등 예외가 없고 그 확률이 1/2로 동일한 가상의 동전던지기를 생각해보면 여기서 동전을 100번 던져 앞면의 개수를 구하는 경우를 생각해보면 이 100번 던지는 세트를 많이 시행하면 그 히스토그램 그래프는 정규분포와 비슷하게 나타난다는 것입니다. 즉 표본 평균의 분포가 정규분포에 가까워진다는 것을 의미합니다. 이번엔 연속형 확률 변수에 대한 예시를 들어보겠습니다. 특정 집단의 남성의 키를 조사하고자 합니다. 모집단의 분포는 정규분포가 아닌 160cm 이하 20명, 161cm ~ 170cm : 500명 , 170cm ~ 180cm : 200명 , 180cm ~ 190cm : 300명 이라고 가정합니다. 여기에서 50명을 표본으로 추출하여 평균을 구하고 이를 충분히 큰 n번 시행하면 추출한 표본의 평균의 분포는 정규분포에 똑같이 가까워집니다. 모집단이 정규분포가 아니라도 이와 같은 결과가 나오게 됩니다. 중심 극한 정리에서 말하는 표본 평균 분포란 Sampling distribution of sample mean으로 쉽게 설명하면 모집단에서 표본크기가 n인 표본을 여러 번 반복 추출하였을 때 각각의 표본의 평균이 이루는 분포를 말합니다. 중심 극한 정리는 이 표본 평균 분포가 정규 분포에 가까워진다는 것을 의미합니다. 확률 밀도 함수에서 확률 밀도 함수가 중요하고 필요한 이유에 대해 알아보았습니다. 중심 극한 정리도 마찬가지로 그 필요와 중요성에 대해 알아보겠습니다. 중심 극한 정리가 필요한 이유는 모집단의 분포와 관계없이 표본 평균 분포가 추출 횟수가 충분히 클 때 정규분포를 이루는 데 있습니다. 모집단의 분포와 관계없이 표본의 크기가 충분히 크다면 표본 평균들의 분포는 모집단의 모수를 기반으로 한 정규분포를 이루는 것을 이용하여 특정 사건이 일어날 확률 값을 계산할 수 있기 때문입니다. 표본 평균이 이루는 표본 분포와 모집단 사이의 관계를 증명함으로써, 수집한 표본의 통계량을 토대로 모집단의 모수를 추정할 수 있는 근거가 되는 것 입니다. 이처럼 유용한 중심 극한 정리는 이용할 때 주의점이 있습니다. 표본이 충분하지 않은데 이를 정규분포로 가정하면 안됩니다. 이는 정규분포의 특성에 따라 양 극에 위치한 표본평균의 확률값을 낮게 추정하는데 표본이 충분하지 않을 때 이를 정규분포로 가정하면 이것이 큰 오류를 초래할 수 있기 때문입니다.