

출제에감- 03 확률분포 개념과 유형 [정규분포, Z분포, T분포, 카이제곱분포 등]

(주)씨에스리 조경미 기술사
(kmicho@cslee.co.kr)

확률분포

KeyWord

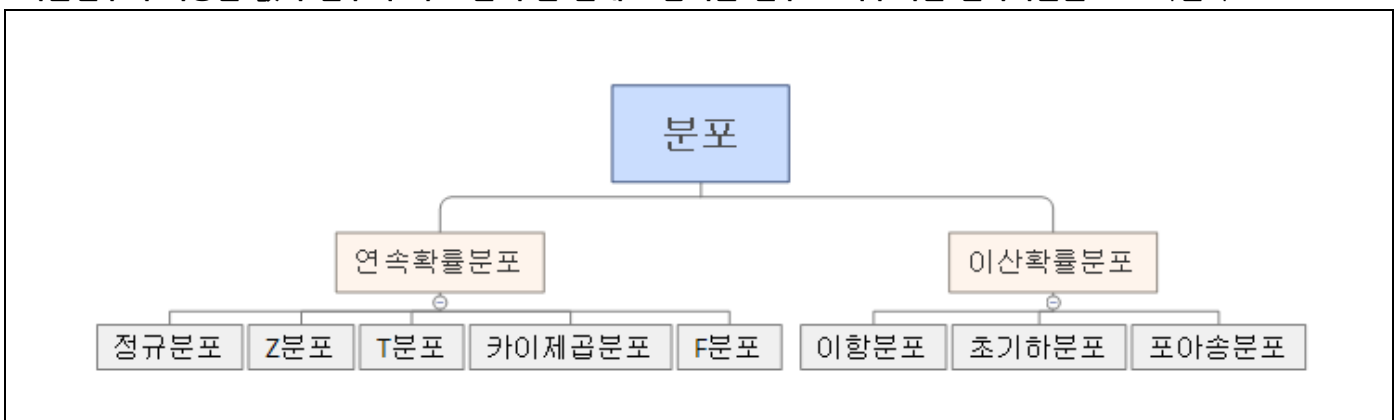
정규분포, Z분포, T분포, 카이제곱분포, F분포, 이항분포, 초기하분포, 포아송분포
연속확률분포, 이산확률분포

알파고와 이세돌로 불 붙여진 인공지능의 광풍으로 인해 IT 패러다임 변화를 더 가깝게 느끼게 된다. 90 년대의 인터넷 패러다임, 2000 년대의 모바일 패러다임, 이제 2017 년 지금의 인공지능 패러다임까지, 몇 번의 패러다임 쉬프트(Paradigm Shift)를 겪으면서 이러한 거대한 물결을 즐겁게 쉬프트(shift)하기 위해서는 그 기본이 되는 IT 원리와 기술을 이해하고, IT 기술을 융합 및 활용하는 창의력을 갖는 중요하다고 생각한다.

인공지능 패러다임으로 인해, 인공지능, 텐서플로우, 여러 가지 알고리즘, 그리고 Python 등등 IT 업계에 있는 사람이라면 최근에 누구나 관심을 갖게 되고, 나 또한 그 실체가 무엇인지 궁금해서 여러 가지를 관심있게 보게 된다. 그런데, 다시 돌아와야 하는 기본 개념은 통계인 것 같다. 고 3 막바지에 “그래, 한 문제 정도는 포기하지!”하고 시간 효율적인 선택을 했던 그 통계를 이제는 기본 원리로 꼭 알고 가야 하기에 포기하지 않고! 여러 가지 확률분포의 개념을 알아보려고 한다.

1. 여러 가지 확률 분포의 유형

- 어떤 사건들로 이루어진 모집단의 특성을 파악하여 판단하기 위해 해당 모집단의 분포를 이해하는 것이 중요하다. 분포는 확률변수가 취할 수 있는 값이 유한하거나 셀 수 있는 변수로 이루어진 이산확률분포와 확률변수의 가능한 값이 실수의 어느 한 구간 안에 포함되는 변수로 이루어진 연속확률분포로 나뉜다.



* 확률변수 : 변수가 취하는 값에 확률이 대응하고 있는 변수 (예: $P(X=1) = 0.4, P(X=2) = 0.6$)

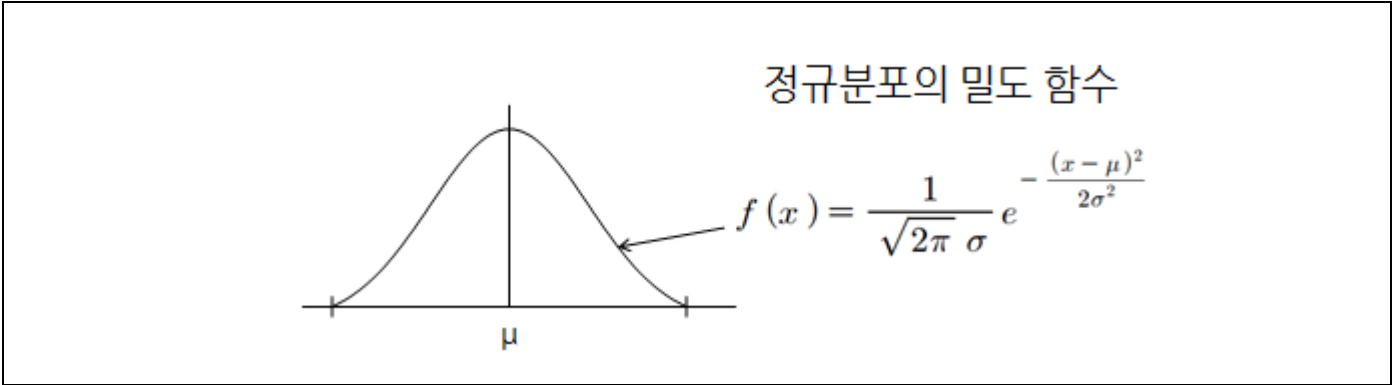
2. 연속확률분포

2.1 정규분포

가. 정규분포(Normal Distribution)의 정의

- 평균을 중심으로 좌우 대칭이면서 종모양으로 그 분포가 그려지는 선형(Linear) 분포
- 가우시안 정규분포라고도 함

나. 정규분포의 개념도 및 수식



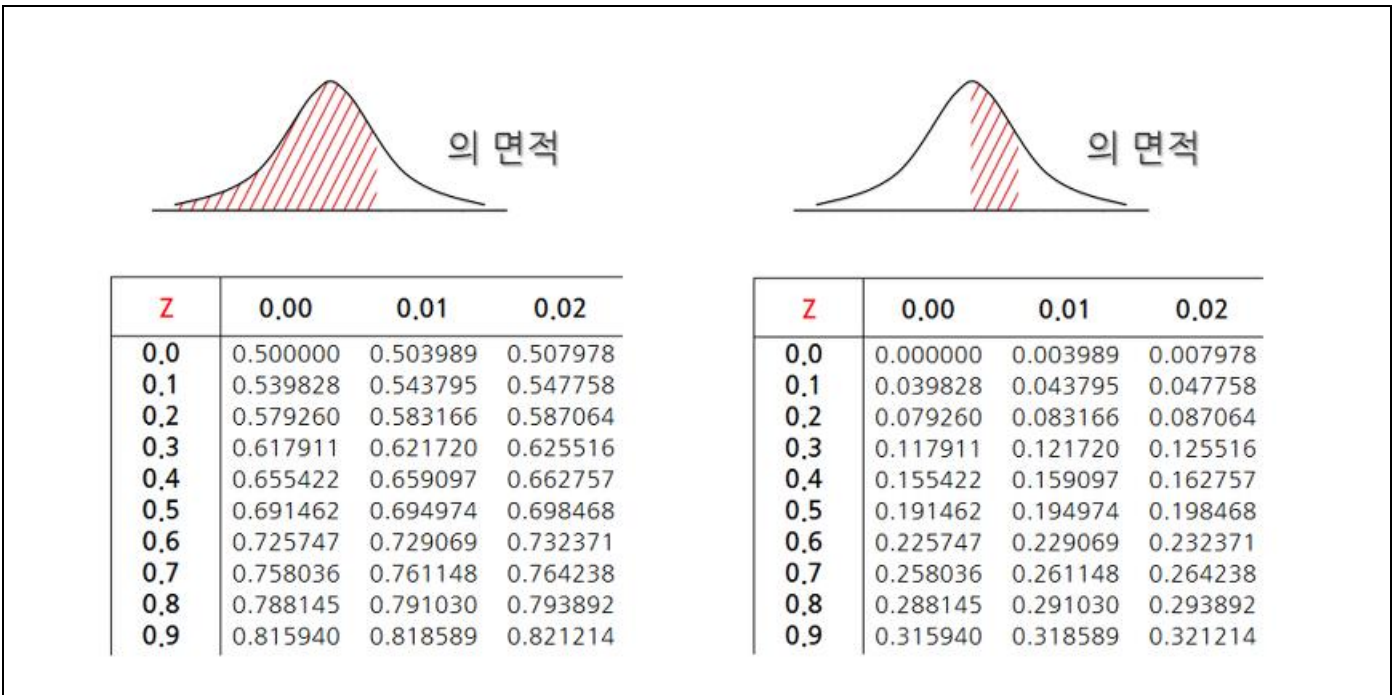
- 평균 μ 에 대하여 좌우 대칭, μ 에 따라 분포의 중심이 결정
- σ^2 에 의해 분포의 모양이 결정됨

2.2 Z 분포

가. Z 분포의 정의

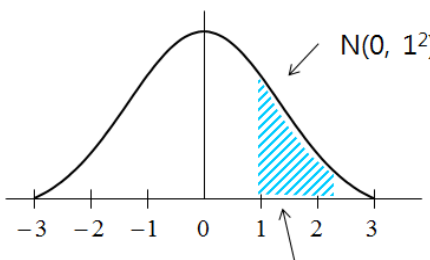
- 정규분포를 평균 0, 표준편차 1 로 표준화하여 표준정규분포가 0에서 양수 z 값의 사이에 있을 확률 분포로 표준정규분포라고도 함.

나. Z 분포 개념 이해



- Z 분포표를 이용해서 그래프의 면적을 구할 수 있으며, 이 면적은 Z 값일 때의 확률값이 된다.

나. 표준정규분포와 중심극한의 정리

<p>표준정규분포 (=Z 분포)</p>	<p>Standard Normal Distribution 평균 0, 표준편차 1 인 정규분포</p>  <p>표준 정규분포의 확률 (표)</p>
	<p>X가 평균이 μ이고 분산이 σ^2 인 정규분포 $N(\mu, \sigma^2)$ 일 때, $Z = (X - \mu) / \sigma$ 는 평균이 0, 표준편차가 1 인 표준정규분포 $N(0, 1)$을 따른다. 표준화변환 $Z = (X - \mu) / \sigma$</p>
<p>중심 극한의 정리</p>	<p>Central Limit Theorem 동일한 확률분포를 가진 독립 확률 변수 n개의 평균의 분포는 n이 충분히 크다면, 평균이 μ, 분산이 σ^2 인 정규분포에 근사함 N 이 충분히 크면 표본은 $N(\mu, \sigma^2/n)$</p>

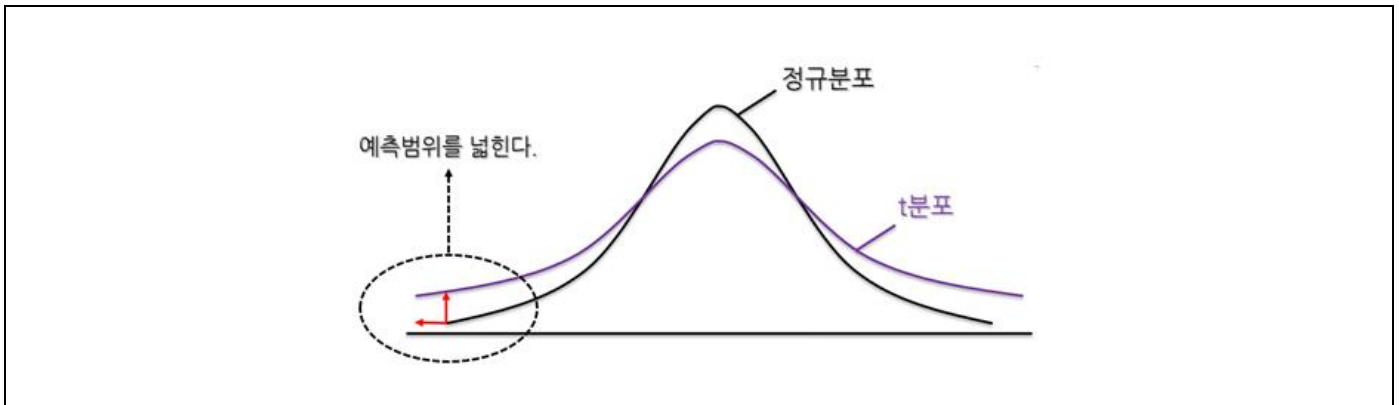
2.3 T 분포

가. T 분포의 정의

- 정규분포와 유사하나 표본의 크기가 작은 경우($n < 30$)에 기준이 되는 확률분포로 모집단의 분산을 모를 때, 모집단의 평균을 추정하는데 사용하는 분포

나. 정규분포와 T 분포의 분포비교

- T 분포는 정규분포보다 종모양의 높이가 낮고 양 옆의 높이는 더 높은 형태로 더 산개된 모양임
- 확률을 구할 때 보다, 예측범위가 보다 넓은 신뢰구간이나 가설 검정을 할 때 사용함



- 단, 표본의 수가 30 개 이상이면, 정규분포를 사용함.

2.4 카이제곱(X2)분포

가. 카이제곱분포의 정의

- 분산이 퍼져있는 모습을 만든 분포로, 모집단이 정규분포를 따르는지 검증할 때, 분산을 이용하여

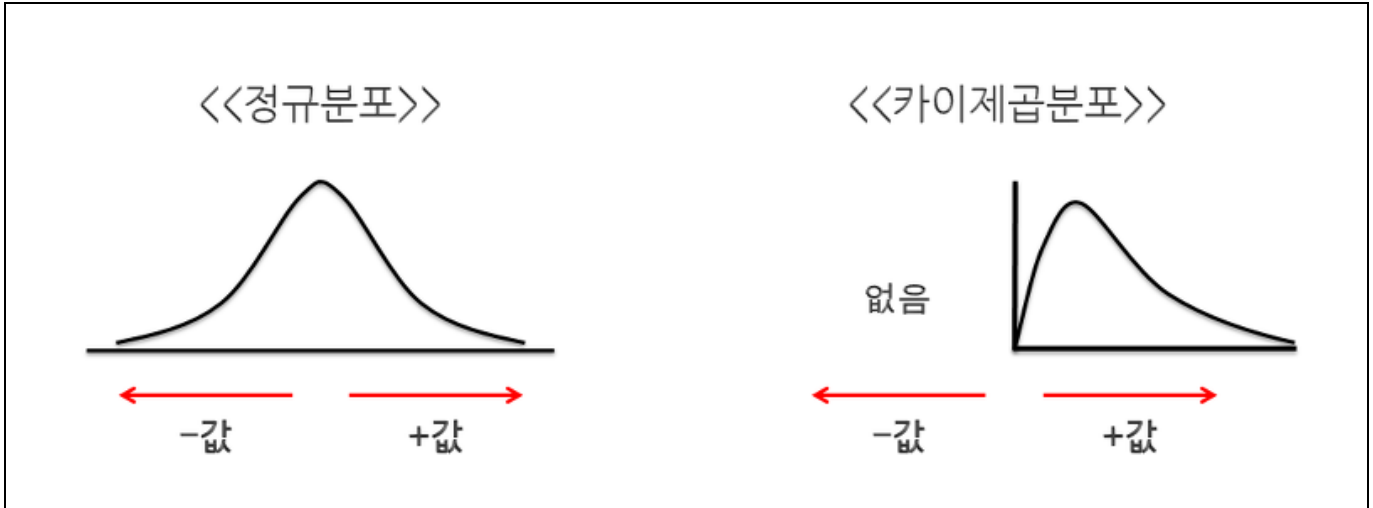
정규분포인지 아닌지 검정할 때 사용하는 분포

- 신뢰구간과 가설검정, 적합도 검정, 동질성 검정, 독립성 검정 등에 사용

나. 정규분포와 카이제곱분포의 비교도

- 카이제곱분포는 제공된 값인 분산을 이용하므로, -값은 존재하지 않고 +값만 존재함

- 0에서 가까울수록 분포가 많고, 0에서 멀어질수록 분포가 감소함



* 의사결정트리의 알고리즘 CHAID 에서 카이제곱 평가지수를 사용함

2.5 F 분포

가. F 분포의 정의

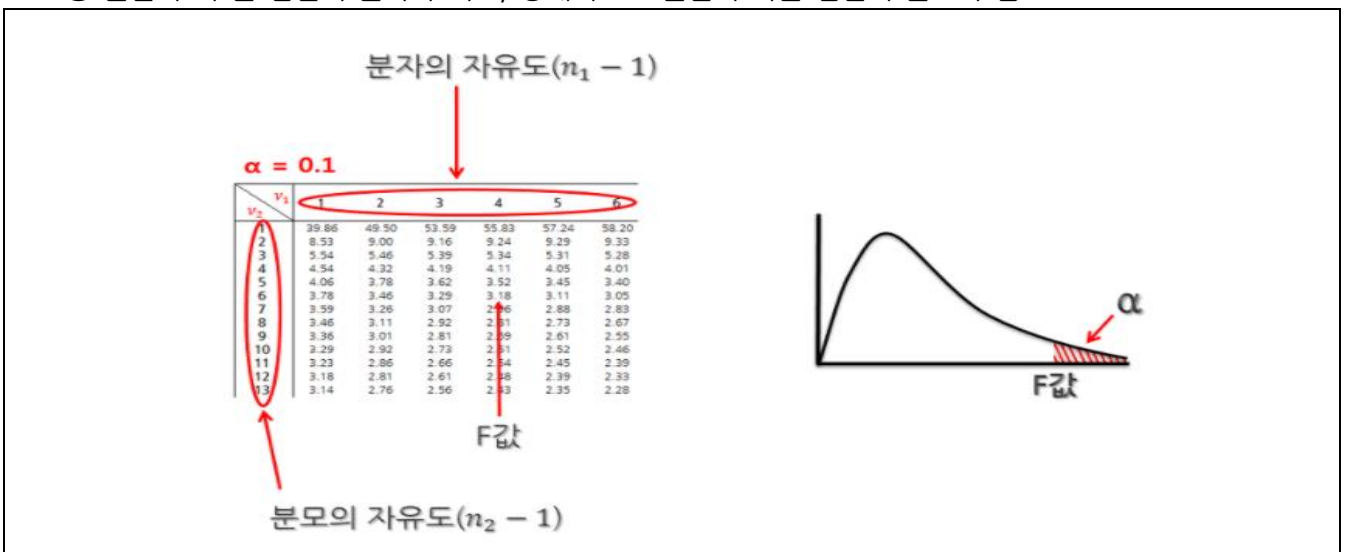
- 카이제곱분포와 같으며, 2 개이상의 모집단을 분산을 이용하여 비교(분산분석, ANOVA, Analysis of Variance)할 때 사용하는 분포

- 신뢰구간, 가설검정, 분산분석을 수행할 때 사용

나. F 분포의 분포도

- F 분포는 두 집단의 분산을 다루어 집단이 2 개이므로, 자유도도 2 개, X 축은 분자의 자유도(n_1-1), y 축은 분모의 자유도(n_2-1)임.

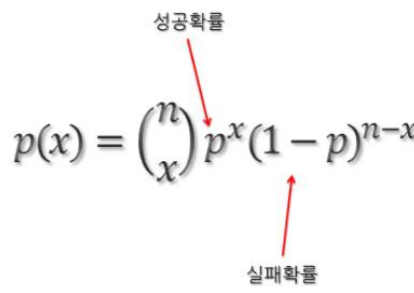
- 보통 분산이 더 큰 집단이 분자가 되고, 상대적으로 분산이 작은 집단이 분모가 됨



3. 이산확률분포

3.1 이항분포

가. 베르누이 분포의 업그레이드, 이항분포(Binomial Distribution)의 정의

정의	수식
- 베르누이 실험을 n 번 독립적으로 번갈아 실행할 때 각 시행이 확률 p를 가질 때의 이산 확률 분포	 $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$

- 베르누이 분포는 한번 하는 실험에 사용, 이항분포는 여러 번에 걸쳐 시행하는 실험에 사용

나. 베르누이 실험, 시행, 분포

베르누이 시행	확률론과 통계학에서 임의의 결과가 '성공' 또는 '실패'의 두 가지 중 하나인 실험 $P(X) = P^x (1-P)^{1-x}$ 예) 동전의 앞뒤, 여자와 남자, 소비자가 물건을 구입하였는가? 아닌가?
베르누이 실험	상호 배반인 오직 두 가지의 가능한 결과 중 하나를 갖는 실험
베르누이 분포	매 시행마다 오직 두 가지의 가능한 결과만 일어나는 분포 $P(X=0) = p, P(X=1) = 1-p$

3.2 초기하분포와 푸아송분포

가. 이항분포의 비복원추출 실험, 초기하분포(HyperGeometric Distribution)의 개념

정의	수식
- 성공과 실패 2 가지의 상황만 일어나는 실험에서 비복원추출로 실험조건이 달라지는 경우의 이산확률분포	$p(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$

나. 단위 시간의 이산확률분포, 푸아송(Poisson Distribution)의 정의

정의	수식
- 단위 시간 안에 어떤 사건이 몇 번 발생할 것인지를 표현하는 이산확률분포.	$f(n; \lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$

- 이항분포를 따르는 확률변수 X에서 n이 대단히 크고 p가 대단히 작을 때 이 확률변수 X는 $\lambda=np$ 인 푸아송 분포로 근사함

지금까지 이산확률분포와 연속확률분포로 분류되는 여러 가지 분포의 개념에 대해 알아보았다. 각각의 차이를 알고 응용하기까지는 더 깊은 이해와 응용연습이 필요할 것이다. 반복하다 보면 언젠가는 되겠지 생각하며, 일단 용어와 개념을 이렇게 반복한다^^

[참조]

<http://math7.tistory.com/> 블로그

<http://bcho.tistory.com/>

<http://ko.wikipedia.org/>

Contents connect communications!!

아이리포에 오시면 더 많은 지식을 가져가실 수 있습니다.

아이리포 IT지식창고 : <https://www.ilifo.co.kr/boards/knowledge>

아이리포 지덤시리즈 : <http://www.jidum.com>

아이리포 기술사/감리사 카페 : <http://cafe.naver.com/itlf>

서울시 마포구 상암동 1610번지, DDMC 3층 아이리포 교육센터

TEL: 02-303-9997 | MAIL: edu@ilifo.co.kr