Lecture 07 - Central Limit Theorem

Sim, Min Kyu, Ph.D., mksim@seoultech.ac.kr

I. 독립과 상관관계

독립 사건과 독립 변수

독립 사건 (Independent Events)

- 만약에 한 개의 사건이 다른 사건이 일어난 확률에 영향을 미치지 않는다면, 두 개의 사건이 독립이라고 한다.
- 예를들어, 동전을 2번 던진다. 첫 번째 시행에서 앞면이 나온 사건이 두 번째 시행에서 앞면이 나오는 사건에 영향을 미치지 않기 때문에 두 사건은 독립이다.

독립 변수 (Independent Random Variables)

- 만약에 한 개의 확률 변수의 값이 다른 확률 변수의 값에 영향을 미치지 않는다면, 두 확률 변수가 독립이라고 한다.
- 예를들어, 동전과 주사위를 던진다. 동전의 면과 주사위의 눈은 각각 확률 변수이며, 이들은 서로 독립이다.

독립인 두 정규분포의 합

Review

- $X \sim N(\mu, \sigma^2) \Longrightarrow aX + b \sim N(a\mu + b, a^2\sigma^2)$

Sum of two independent Normal dist.

- $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, X & Y are indep. ⊠ $X + Y \sim$
- Examples
  - $X \sim N(0, 1)$, $Y \sim N(10, 3^2) \Longrightarrow X + Y \sim$
  - $X \sim N(20, 3^2)$, $Y \sim N(100, 10^2) \Longrightarrow 2X + 3Y \sim$

- 3개의 가게를 운영하고 있다. 각 가게의 하루 매출이 각각 아래와 같다. 하루 매출의 합은? 독립을 가정하라.
  - 1번 가게: $N(100, 20^2)$
  - 2번 가게: $N(70, 10^2)$
  - 3번 가게: $N(50, 10^2)$
- 독립이 아니라면 어떻게 될까?

독립의 의미

- 2개의 가게를 운영하고 하루의 매출이 각각 $N(100, 10^2)$라고 가정하자.
  1. 가게는 각각 샌드위치 가게와 문방구이다. 매출은 독립이다. 총 매출의 분포는?
  2. 한 가게는 왼쪽 신발을 팔고 다른 가게는 오른쪽 신발을 판다. 그렇기 때문에 두 매장의 매출이 항상 같다. (왼쪽 신발은 산 사람은 건너편 다른 가게로 가서 오른쪽 신발을 사고, 오른쪽 신발은 산 사람의 경우에도 마찬가지이기 때문이다.). 총 매출은?
  3. 두 가게가 붙어있고, 한 매장은 샌드위치를 팔고 다른 매장은 커피를 판다. 많은 손님들이 두 가게를 같이 이용하는 경우가 많기에 독립도 아니고, 그렇다고 매출이 항상 같은것도 아니다. 이런 경우에는 총 매출의 표준편차가 어떻게 될까?

독립과 상관관계 - 비지니스 사례

- 사업적 집중 vs 분산/다각화
    - 데스크탑/노트북 제조사
    - 짚신/나막신 장수
- 이러한 성질은 아래등의 분야에 유용하게 적용된다.
    - Portfolio Theory
    - Risk Management
    - Insurance

상관계수 (Correlation coefficient)

- 두 변수 X와 Y가 같이 변하는 정도를 "변수 X와 Y의 상관계수"라고 하며 ρ_{XY}로 표기한다.
- ρ_{XY}의 범위는 -1과 1사이이다.
- ρ_{XY}의 범위는 아래 5가지 범주로 얘기할 수 있다.
    - If $\rho_{XY} = -1$
    - If $-1 < \rho_{XY} < 0$
    - If $\rho_{XY} = 0$
    - If $0 < \rho_{XY} < 1$
    - If $\rho_{XY} = 1$

- 두 개의 정규변수의 합은 다음과 같은 성질을 가지고 있다.

$$X \sim N(\mu_X, \sigma_X^2), Y \sim N(\mu_Y, \sigma_Y^2) \Longrightarrow X+Y \sim N(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2+2\rho_{XY}\sigma_X\sigma_Y)$$

정규분포를 따르는 변수들의 평균

어떤 제품의 무게가

- 정규 분포를 따르고
- 평균이 μ이고,
- 표준편차가 σ이다.
- 검수를 위해서 10개를 뽑았을 때에, 뽑은 10개의 제품의 무게의 합의 분포는? 평균의 분포는?

- 일반화 시켜서 $X_i \sim N(\mu, \sigma^2)$이고, 모든 $X_i, 1 \leq i \leq n$들은 서로 독립이다. 표본평균의 $\overline{X} = \left(\sum_{i=1}^{n} X_i\right)/n$분포는 무엇인가?

Distribution of Sample Mean

어떤 제품의 무게가

- 정규 분포를 따르고
- 평균이 μ 이고,
- 표준편차가 σ 이다.
- 검수를 위해서 10개를 뽑았다. 무게의 평균의 분포는?

어떤 제품의 무게가

- ~~정규 분포를 따르고~~
- 평균이 μ 이고,
- 표준편차가 σ 이다.
- 검수를 위해서 ~~10~~ 100개를 뽑았다. 무게의 평균의 분포는?

- Sample Mean의 분포는 $n > 30$ 이면 Normal을 따른다.
- 이를 **중심극한 정리**라고 한다!

중심극한정리 (Central Limit Theorem)

- 분포에 상관없이 n이 크면 표본평균은 정규분포에 근사한다.
- Regardless of the distribution, for a random variable X with mean μ and standard deviation σ, its sample mean $\overline{X} = \left(\sum_{i=1}^{n} X_i\right)/n$ follows a normal distribution as n increases.

Summary

- $X_i \sim N(\mu, \sigma^2) \Longrightarrow \overline{X} \sim N(\mu, \sigma^2/n)$
- (CLT) $X_i \sim some\_dist(\mu, \sigma^2) \Longrightarrow \overline{X} \sim N(\mu, \sigma^2/n)$ if n is large.
- In both cases, $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) = Z$

II. Estimation using CLT

모평균의 추정

- 많은 경우에 실험의 목적은 모집단의 평균(모평균, μ)을 추정하기 위함이다.
- 예를 들어 남성 100명을 뽑아서 손의 크기를 재는 것은, **전체** 남성의 손의 크기의 평균을 알고 싶기 때문이다.
- 모평균을 추정하기 위해 일반적으로 표본평균을 활용한다.

Case 1: μ를 추정하고 싶은데 σ를 안다면…

- By CLT, $\overline{X} \sim N(\mu, \sigma^2/n)$
- It follows

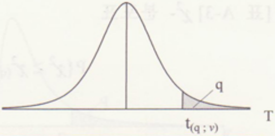
- 모집단의 표준편차가 2cm로 알려진 100명의 남성을 추출하여 손의 크기를 재니까 표본 평균이 20cm였다.
    1. 남성 손 크기의 95% 신뢰구간은?
    2. 남성 손 크기가 20.4cm가 아니다라고 할 수 있는가?

Case 2: μ를 추정하고 싶은데 σ를 모른다면 (일반적인 경우)

- By CLT, $\overline{X} \sim N(\mu, \sigma^2/n)$
- It follows

[표 A-2] t-분포표

$$P\{T \geq t_{(q\,;\,\nu)}\} = q$$



자유도 ν	꼬리확률 q 0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221

14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.792	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

- ~~모집단의 표준편차가 2cm로 알려진~~ 100명의 남성을 추출하여 손의 크기를 재니까 표본 평균이 20cm였다. 표본 표준 편차는 2cm이었다.
    1. 남성 손 크기 평균의 95% 신뢰구간은?
    2. 남성 손 크기의 평균이 cm가 아니다라고 할 수 있는가?

```r
"This note is made with Rmarkdown"
```

```
## [1] "This note is made with Rmarkdown"
```