Seoul National University

- 일반통계학 -
제 4장
표본분포

개요

1. 통계적 추론의 정확도를 나타내는 중요한 도구
 - 추정량의 확률분포
2. 추정량의 확률분포를 결정 짓는 중요한 요소
 - 모집단의 분포
 - 표본 추출 방법
3. 모집단의 대표적 분포로 베르누이분포와 정규분포를 소개
4. 통계량에 대한 표본 분포를 유도

1 베르누이 분포와 정규분포

1.1 베르누이 시행

-어떤 실험의 결과를 오직 두 가지 중의 하나로 생각하는 시행

(1) 베르누이 시행의 표본공간 : S={s ('성공'), f ('실패') }

(2) 성공확률 p=P({s}), 실패확률 q=1-p

1.2 베르누이 확률변수

- 베르누이 표본공간 S에서 실수로 가는 함수 X, X(s)=1, X(f)=0

1.3 베르누이 분포

베르누이 확률변수 X의 확률분포

(1) 특성값이 이원적인 모집단의 분포를 나타냄

X	0	1
P(X=x)	1-p	p

(2) 기호 X~B(1,p) or Ber(P)

(3) 평균과 분산

1.4 베르누이 시행이 독립적으로 반복되는 실험

-기호 : X_1, …, X_n ~ B(1,p) iid (Independent identically distributed)
베르누이 시행 X_i 각각은 서로 독립(independent)이고 성공의 확률이 p로 동일하다(identically distributed)는 것을 의미.

-(예) 10개 중 3개의 당첨제비가 있을 때, 2개의 제비를 단순랜덤**복원**추출하는 경우
X_1=‘첫번째 시행의 결과’, X_2 =‘두번째 시행의 결과’

※n 개의 제비를 단순랜덤**복원**추출하는 경우에 X_1, …, X_n을 각 시행의 결과라고 하면, X_1, …, X_n은 서로 독립이며 동일한 분포를 갖는다.

1 베르누이 분포와 정규분포

1.5 정규분포(normal distribution) - 가우스 분포

- 특성값이 연속적인 무한모집단의 분포로서 가장 대표적인 분포.

(1) 평균 μ, 표준편차 σ인 정규분포의 확률밀도 함수

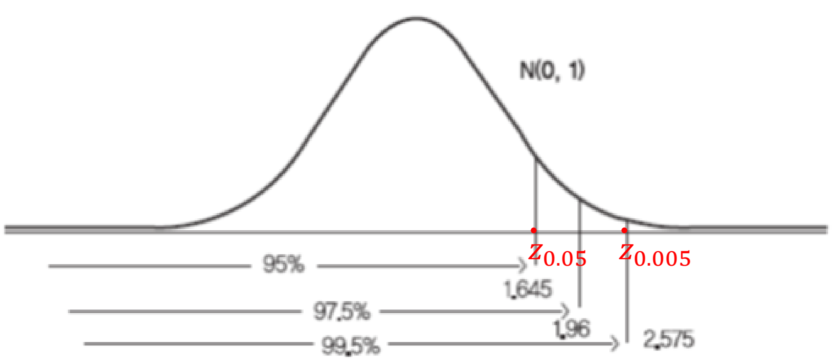
$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ -\infty < x < \infty$$

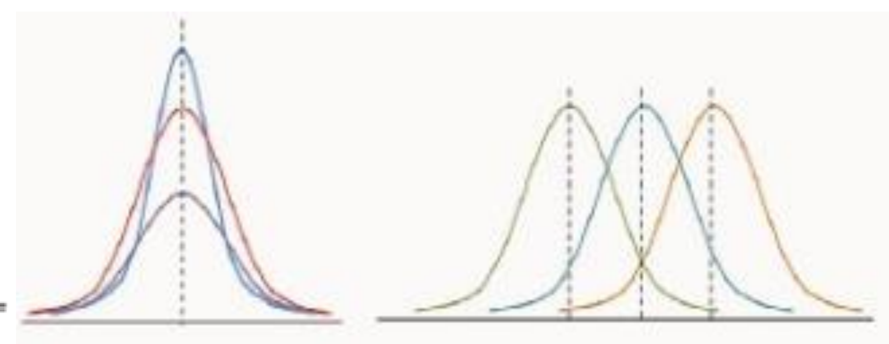
(2) 평균 μ를 중심으로 대칭

(3) 표준정규분포 : Z~N(0,1), X~N(μ, σ^2) 일 때, Z=$\frac{X-\mu}{\sigma}$~N(0,1)

(4) z_α : P($Z \geq z_\alpha$) = α 인 값, 즉 100(1- α)% 백분위수

(5) $P(a \leq X \leq b) = P(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma})$





평균은 같으나 표준
편차가 다른 경우

표준편차는 같으나
평균이 다른 경우

- 표준정규분포의 확률 계산
  - 예) Z가 0과 1.21 사이에 포함될 확률

z	.00	.01	.02
⋮		⋮	
.9	.3159	.3186	.3212
1.0	**.3413**	.3438	.3461
1.1	.3643	.3665	.3686
1.2	**.3849**	.3869	.3888
⋮		⋮	

$P(0 \le Z \le z)$

TABLE 1 Normal Curve Areas

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.00	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.10	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.20	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.30	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.40	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.50	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.60	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.70	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.80	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.90	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.00	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.10	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.20	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.30	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.40	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.50	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.60	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545

  - 예) Z가 구간 -1.00과 0.00사이에 있을 확률

$$P(-1.00 \le Z \le 0) = P(0 \le Z \le 1.00) = 0.3413$$

⇒ 표준정규확률밀도함수는 0을 중심으로 대칭

(예) 우체국에서 소포 무게의 상한선을 설정하고자 하여 기존 고객이 부치는 짐의 무거운 5% 정도를 제한하고자 한다. 만약 기존 고객의 소포 무게의 분포가 평균 5kg, 표준편차 1kg인 정규분포를 따른다면 상한선은 얼마로 해야 할까?

1.6 정규분포의 성질

(1) X~N(μ, σ^2) 일 때, 임의의 상수a, b에 대하여 aX+b ~N(aμ+b, $a^2\sigma^2$)
(2) X_1~N(μ_1, $\sigma_1{}^2$) , X_2~N(μ_2, $\sigma_2{}^2$)이고 X_1과 X_2가 서로 독립일 때,
$a_1X_1+a_2X_2$~N(a_1μ$_1$+a_2Xμ$_2$,$a_1{}^2\sigma_1{}^2+a_2{}^2\sigma_2{}^2$)

(예) 한 가정에서 아침식사로 빵과 우유 한 잔을 먹는다고 하자. 빵에서 섭취하는 칼로리~N(200, 225), 우유에서 섭취하는 칼로리~N(80, 25)일 때, 300칼로리 이상을 섭취할 확률은 얼마인가?

2.1 용어정리

• 모수 (parameter) - 모집단의 특성을 결정하는 상수

• 통계량 (statistic)

-표본(X_1, …, Xn)으로부터 계산 가능한 표본의 특성을 나타내는 함수

-$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (통계량 o), $\sum_{i=1}^{n} (X_i - \bar{X})^2$ (통계량 o), $\sum_{i=1}^{n}(X_i - \mu)$ (단, μ를 모름) (통계량 X)

•추정량 (estimator) - 모수의 추정을 위한 통계량 예) $\hat{\mu} = \bar{X}$, $\widehat{\sigma^2} = S^2$

•표본분포 – 통계량의 확률분포

•표본분포는 추정의 정확도를 나타내는 중요한 도구

(Why? 참값인 모수를 모를 때, 표본으로부터의 추정값이 모수를 얼마나 정확하게 추정한 것인지 알 수가 없다. 표본분포를 이용해서 정확도에 대한 확률적 접근이 가능해짐)

Q)통계량은 확률변수인가? YES!

(예) 유한모집단 {2,2,4,5}에서 크기 2인 표본을 단순랜덤비복원추출하였을 때, 표본평균의 표본분포를 구하여라.

2 표본분포

- 통계량의 확률분포(표본분포)는 **모집단의 분포**와 **표본의 추출방법**에 의하여 결정된다.
- 랜덤표본 (Random sample)

\- 유한 모집단인 경우 단순랜덤비복원추출로 뽑은 확률변수의 모임 $\{X_1, X_2 \ldots, X_n\}$

\- 무한모집단인 경우 다음의 두 조건을 만족하는 확률변수의 모임 $\{X_1, X_2 \ldots, X_n\}$

(i) $X_1, X_2 \ldots, X_n$은 서로 독립(independent)

(ii) $X_1, X_2 \ldots, X_n$ 각각의 확률분포는 모집단 분포와 동일 (identically distributed)

-> $X_i \sim i.i.d. f(\theta)$, $i = 1, \cdots, n$

❖ 유한모집단에서의 랜덤표본인 경우 표본분포의 유도는 매우 복잡하고 어렵다. 따라서 모집단의 크기가 큰 경우 흔히 무한모집단에서의 랜덤표본으로 간주하여 표본분포를 구한 다음 이를 실제 표본분포의 근사분포로 사용한다.

이후의 모든 추정과 검정에서는 무한모집단에서의 랜덤표본을 가정한다.

3 초기하분포와 이항분포

3.1 초기하분포 (hypergeometric distribution)

- 두가지 특성값({s,f})만 가지는 **유한** 모집단의 모비율을 추정하기 위해 표본비율을 사용할 때 표본비율의 확률분포를 나타내기 위해 사용한다.
- 특성값 1의 개수가 D, 0의 개수가 N - D인 크기 N 의 유한모집단에서 크기 n인 랜덤표본을 뽑을 때, 표본 1의 개수의 확률분포를 초기하 분포라고 한다.

$$p(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}},\ x = 0,1,\dots,n \quad (\text{단},\ n \le D,\ \text{n} \le N - D\ )$$

- 평균$=np$, 단, $p = \frac{D}{N}$
- 분산$=np(1-p) \cdot \frac{N-n}{N-1}$

(예) 신세대 의식구조 조사에서 모집단 크기가 10이며, 이 중 찬성자의 수가 6명, 반대자의 수가 4명이라고 하자. 크기 3인 랜덤표본에서 찬성자수의 평균과 분산을 구하여라.

(풀이) X=찬성자의 수, $p(x) = \frac{\binom{6}{x}\binom{10-6}{3-x}}{\binom{10}{3}}$: x =0,1,2,3

3.2 이항분포 (binomial distribution)

이제, 특성값이 1또는 0과 같이 이원적으로 분류되고 1의 비가 p (0< p<1)인 **무한** 모집단을 생각해보자. 랜덤표본 X_1, X_2..., X_n은 서로 독립이며 각각 1의 확률이 p인 베르누이 분포를 따른다. 1을 성공, 0을 실패로 생각하면

X_1, X_2..., X_n은 성공률 p인 베르누이 시행을 n번 반복 시행할 때, 시행의 결과이며

X = 표본에서 1의 개수 =성공의 수=$X_1+X_2+\ldots+X_n \sim$B(n, p)

- **이항분포**

시행 횟수가 n이고 성공의 확률이 p일 때 성공의 수에 대한 확률분포

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}\ , \ x = 0,1,2,\ldots,\mathrm{n}$$

- 평균$=np$

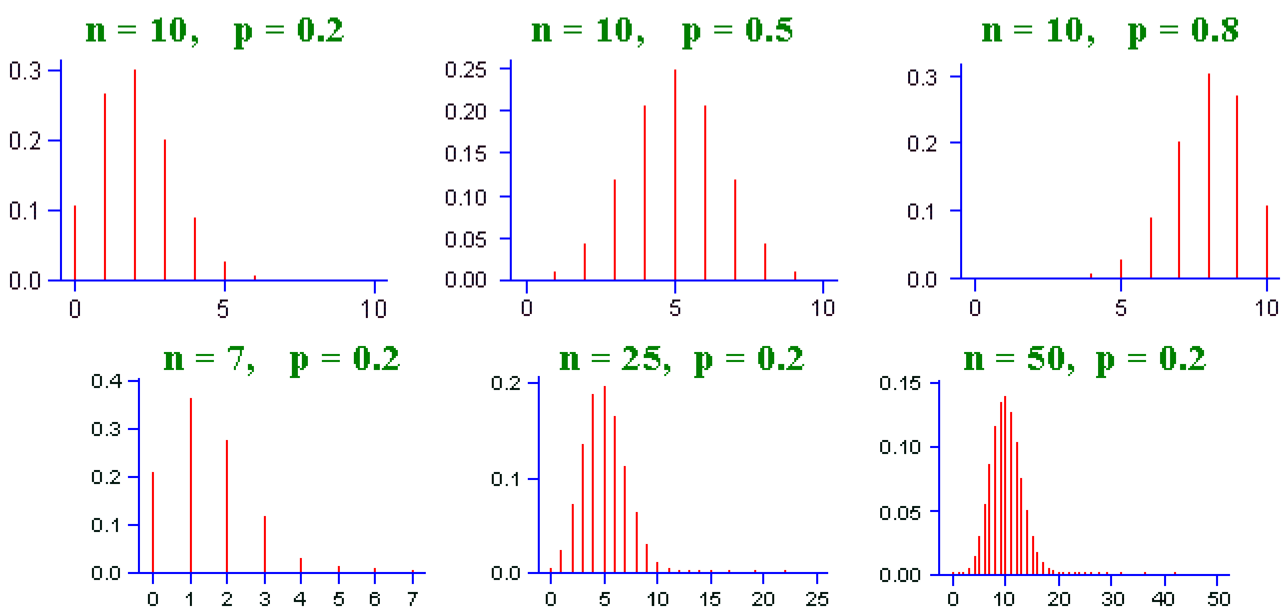
E(X)=E($X_1+X_2+\ldots+X_n$)=$\sum_{i=1}^{n} E(X_i) = np$ ($\because$ $X_i \sim$B(1, p), i= 1,2,...,n)

- 분산$=np(1-p)$

Var(X)=Var($X_1+X_2+\ldots+X_n$)=$\sum_{i=1}^{n} Var(X_i) = np(1-p)$

($\because$ $X_i \sim$B(1, p), i= 1,2,...,n 이고 독립이므로)

3 초기하분포와 이항분포



(예) 5개 중 하나를 택하는 선다형 문제가 20문항 있는 시험에서 랜덤하게 답을 써 넣는 경우에 다음 물음에 답하여라.

(a) 정답이 하나도 없을 확률은 얼마인가?
(b) 8개 이상의 정답을 맞힐 확률은 얼마인가?
(c) 4개부터 6개 사이의 정답을 맞힐 확률은 얼마인가?

3 초기하분포와 이항분포

1의 비 $D/\ N$ 이 p에 무한히 가까워지고 $N \to \infty$일 때, 초기하분포는 이항분포 B(n, p)에 근사한다.

- 초기하분포-일정한 시행에서 성공의 횟수에 대한 분포 (단, 비복원추출)
- 이항분포 -일정한 시행에서 성공의 횟수에 대한 분포 (단, 복원추출)

(예) 어떤 제품을 생산하는 공정의 불량률은 5%로 알려져 있다. 오늘 생산한 10000개의 제품 중에서 20개를 단순랜덤추출하여 조사할 때 표본의 불량률이 10%이상일 확률을 구하여라.

(풀이) X=20개 중 불량품의 수, X는 초기하 분포를 따르며 근사적으로 B(20, 0.05)를 따른다.

4 표본평균의 분포

- X_1, X_2..., X_n이 모평균 μ, 모분산 σ^2인 무한모집단으로부터의 랜덤표본인 경우 표본평균 $\bar{X}=\frac{1}{n}(X_1+X_2+\ldots+X_n)$ 의 기대값과 분산은

-E($\bar{X}$)=

-Var($\bar{X}$)=

-sd($\bar{X}$)=

- 모집단의 분포가 정규분포 N(μ, σ^2)일 때, 표본평균 $\bar{X}$는 정규분포 N(μ, $\frac{\sigma^2}{n}$)을 따른다.

중심극한정리

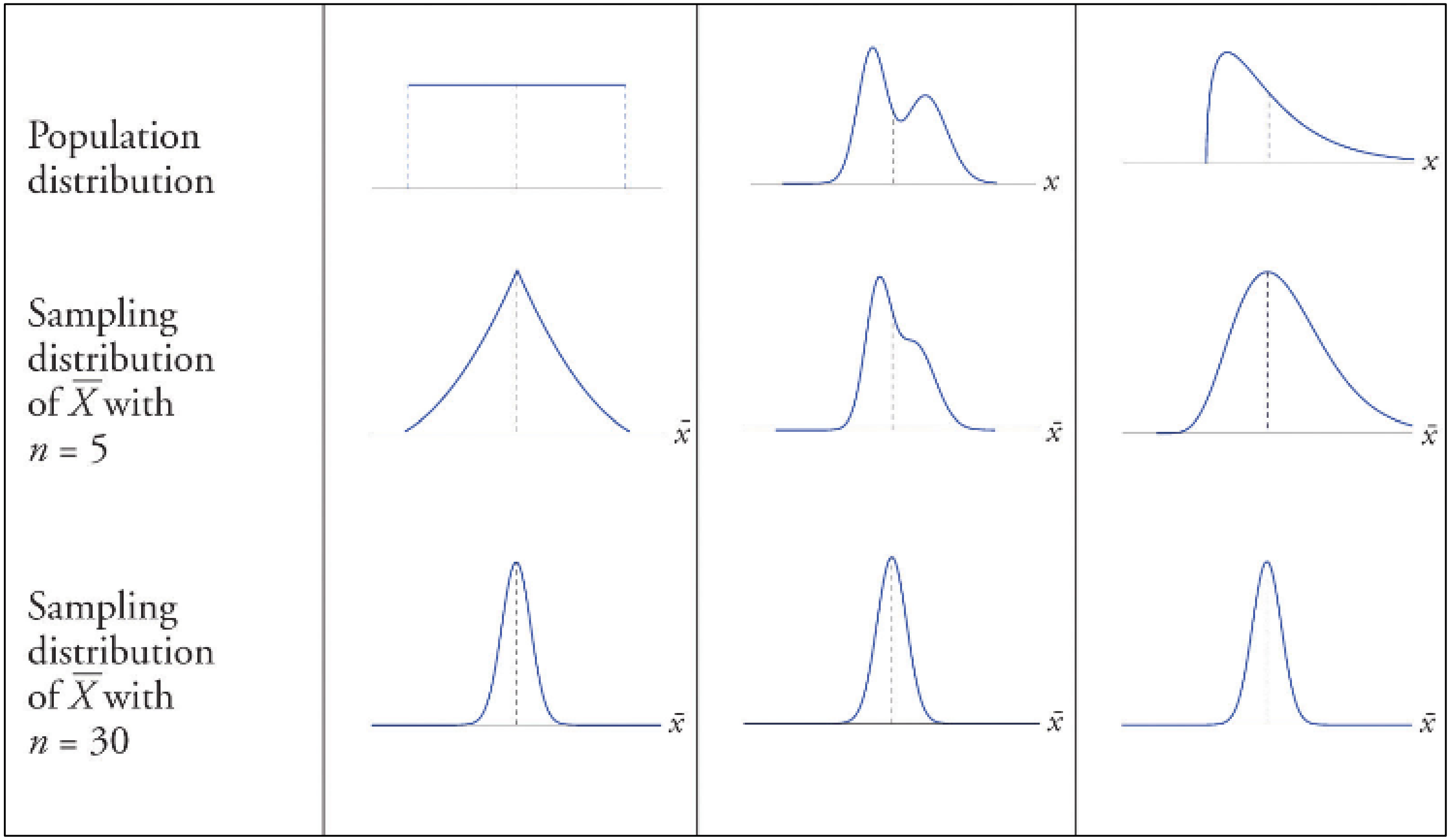
$X_1, X_2 \dots, X_n$이 모평균 μ, 모분산 σ^2인 무한모집단으로부터의 랜덤표본인 경우, n 이 충분히 크면 (일반적으로 30이상)

$$\bar{X} \approx N(\mu, \frac{\sigma^2}{n})$$

$$\text{즉, } \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \approx N(0,1)$$

(예) 대학 신입생 신장의 평균이 168cm이고 표준편차가 6cm라고 알려져 있다. 100명의 신입생을 단순랜덤추출하는 경우 표본평균이 167cm이상 169cm이하일 확률을 구하여라.

4 표본평균의 분포

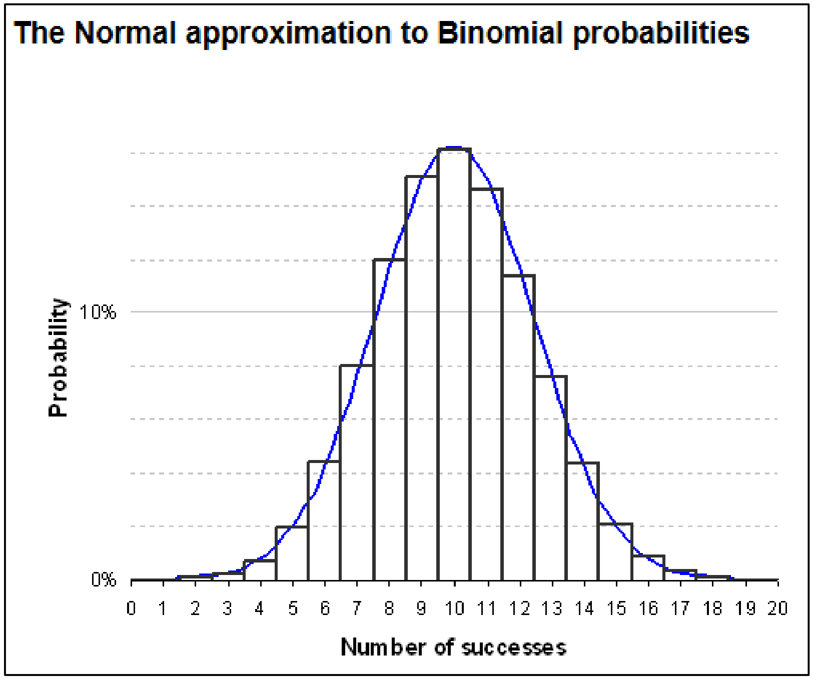

Population distribution
Sampling distribution of $\overline{X}$ with $n = 5$
Sampling distribution of $\overline{X}$ with $n = 30$
x
$\bar{x}$

4 표본평균의 분포

- **이항분포의 정규근사**

X~B(n, p)이고 n이 충분히 클 때 (혹은 np>5, $n(1-p)$>5), X는 근사적으로 정규분포 N($np, np(1-p)$) 를 따른다. 즉, $\frac{X-np}{\sqrt{np(1-p)}} = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \approx$ N(0,1) (단, $\hat{p}=X/n$)

- **연속성 수정계수**



X~B(n, p)일 때,

P(X<5) =P(X≤ 4) $\approx \Phi(\frac{4.5-np}{\sqrt{np(1-p)}})$

P(X=5) =$\approx \Phi(\frac{5.5-np}{\sqrt{np(1-p)}})$- $\Phi(\frac{4.5-np}{\sqrt{np(1-p)}})$

P(X≥5) $\approx 1 - \Phi(\frac{4.5-np}{\sqrt{np(1-p)}})$

P(X≤5) $\approx \Phi(\frac{5.5-np}{\sqrt{np(1-p)}})$

$\Phi(x)=\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$

(표준정규분포에 대한 누적확률분포함수값)

(예) 어떤 공장의 생산품 중 10%가 불량품이라고 한다. 이 공장의 생산품에서 단순랜덤복원추출로 100개의 상품을 꺼냈을 때, 그 중 불량품이 12개를 넘을 확률을 구하여라.

(풀이)

1. 이항분포 사용

X~B(100, 0.1) 이므로 $P(X\geq 13) = 1-\sum_{x=1}^{12}\binom{100}{x}0.1^{x}0.9^{100-x} = 0.1982$

2. 근사분포 사용

X~B(100, 0.1) 이고 $np=10>5$, $n(1-p)=90>5$ 이므로 정규분포근사를 이용할 수 있다.

$P(X\geq 13) = 1-P(X\leq 12) \approx 1-\Phi(\frac{12-10}{3}) = 0.2524$

3. 연속성 수정계수 사용

$P(X\geq 13) = 1-P(X\leq 12) \approx 1-\Phi(\frac{12.5-10}{3}) = 0.2023$

근사값의 정확도를 개선하기 위해서 연속성 수정계수를 사용하여 보정을 하게 된다.

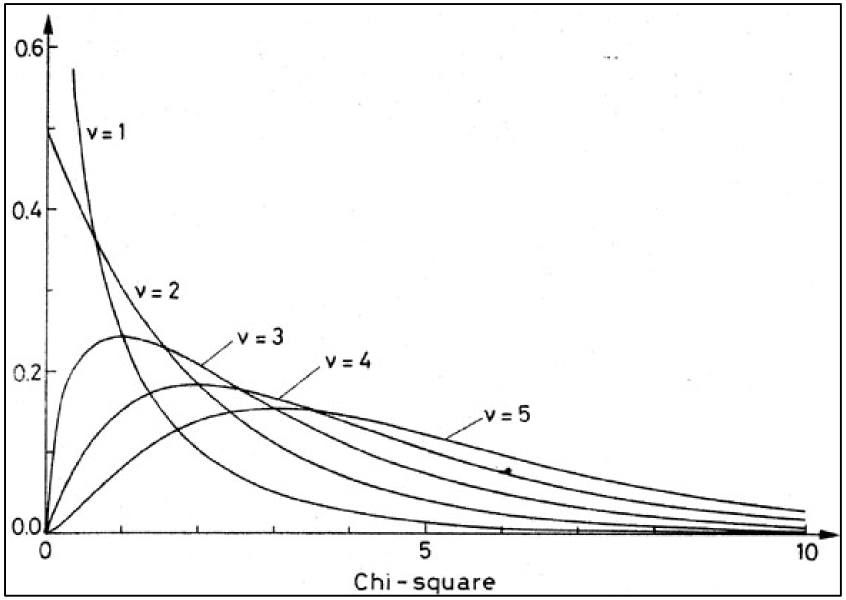
5 카이제곱, t, F 분포 (필기노트)

- **카이제곱분포**

$Z_1, Z_2, \dots Z_v$가 N(0,1)로부터의 랜덤표본일 때, $\sum_{i=1}^{v} {Z_i}^2 \sim \chi^2(v)$.
즉, 자유도(degrees of freedom) v인 카이제곱분포를 따른다.

- **카이제곱분포의 가법성**

$X_1 \sim \chi^2(v_1)$, $X_2 \sim \chi^2(v_2)$이**고** X_1과 X_2가 서로 독립이면 $X_1 + X_2 \sim \chi^2(v_1 + v_2)$



- **정규모집단에서의 표본분산의 분포**

X_1, X_2..., X_n이 $\mathrm{N}(\mu, \sigma^2)$으로부터의 랜덤표본이라 할 때,

표본분산 $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$ 에 대하여 다음이 성립한다.

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

(∵)

σ^2으로 양변을 나누면

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 + \left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2$$

$$\chi^2(n) \quad = \quad \chi^2(n-1) \quad + \quad \chi^2(1)$$

$$\therefore \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

- **분산이 동일한 정규모집단에서의 표본분산의 분포**

$$\frac{(n_1+n_2-2){S_p}^2}{\sigma^2} \sim \chi^2(n_1+n_2-2), \quad \text{단, } {S_p}^2 = \frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}$$

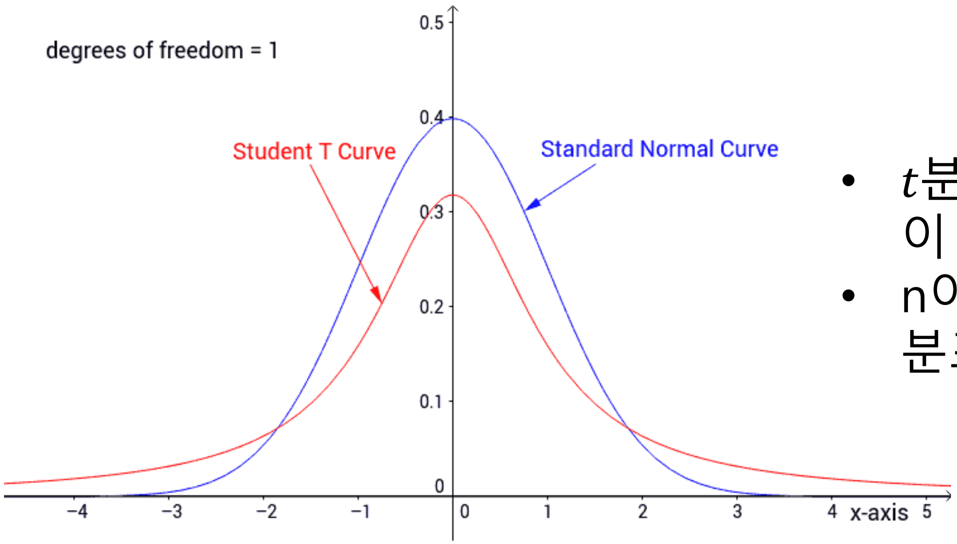
(∵) $X_{11}, X_{12}\ldots, X_{1n_1} \sim \mathrm{N}(\mu_1, \sigma^2)$ 이고 $X_{21}, X_{22}\ldots, X_{2n_2} \sim \mathrm{N}(\mu_2, \sigma^2)$ 이고 각 랜덤표본이 독립일 때, $\frac{(n_1-1){S_1}^2}{\sigma^2} \sim \chi^2(n_1-1)$, $\frac{(n_2-1){S_2}^2}{\sigma^2} \sim \chi^2(n_2-1)$이고 서로 독립이 된다. 카이제곱분포의 가법성에 가 성립한다.

$({S_1}^2 = \sum_{i=1}^{n}(X_{1i}-\overline{X_1})^2/(n_1-1),\ {S_2}^2 = \sum_{i=1}^{n}(X_{2i}-\overline{X_2})^2/(n_2-1))$

- **t 분포**

$Z \sim \mathrm{N}(0,1)$ 와 $V \sim \chi^2(k)$가 독립일 때, $T = \frac{Z}{\sqrt{\frac{V}{k}}} \sim t(k)$. 즉, 자유도 k인 t분포를 따른다.

5 카이제곱, t, F 분포 (필기노트)



- t분포는 표준정규분포보다 꼬리부분이 더 두껍다.
- n이 충분히 클 때, t분포는 표준정규분포에 근사한다. (보통, n은 30이상)

- **스튜던트화된 표본평균의 분포**

$X_1, X_2 \ldots, X_n$이 $\mathrm{N}(\mu, \sigma^2)$으로부터의 랜덤표본이라 할 때, $T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$.

($\because$) $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathrm{N}(0,1)$, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ 이고, $\bar{X}$ 과 S^2 이 독립이므로, t분포의 정의에 의해

5 카이제곱, t, F 분포 (필기노트)

- **분산이 동일한 두 정규모집단에서의 t분포**

$X_{11}, X_{12}\ldots, X_{1n_1} \sim \mathrm{N}(\mu_1, \sigma^2)$, $X_{21}, X_{22}\ldots, X_{2n_2} \sim \mathrm{N}(\mu_2, \sigma^2)$인 독립인 랜덤표본일 때, $\overline{X_1} - \overline{X_2}$도 정규분포를 따르므로, $\frac{\overline{X_1}-\overline{X_2}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim \mathrm{N}(0,1)$이고 카이제곱의 가법성에 의해 $\frac{(n_1+n_2-2){S_p}^2}{\sigma^2} \sim \chi^2(n_1+n_2-2)$이므로

$$\frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \Bigg/ \sqrt{\frac{\frac{(n_1+n_2-2){S_p}^2}{\sigma^2}}{(n_1+n_2-2)}} = \frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

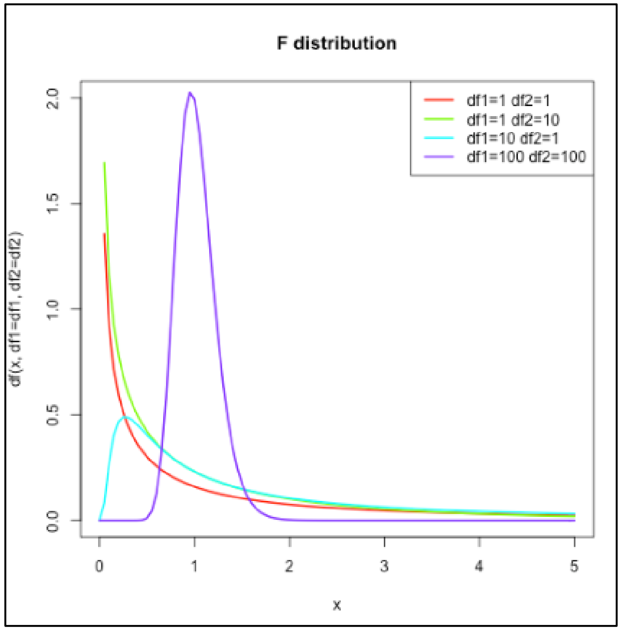
- **분산이 동일하지 않은 두 정규모집단에서의 t분포**

$$\frac{\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} \approx t(v)$$

단, $v = \frac{\left(\frac{{s_1}^2}{n_1}+\frac{{s_2}^2}{n_2}\right)^2}{({s_1}^2/n_1)^2/(n_1-1)+({s_2}^2/n_2)^2/(n_2-1)}$

- **F 분포**

$V_1 \sim \chi^2(k_1)$ 과 $V_2 \sim \chi^2(k_2)$가 독립일 때, $F = \dfrac{V_1/k_1}{V_2/k_2} \sim F(k_1,\ k_2)$. 즉, 자유도 $(k_1,\ k_2)$ 인 F분포를 따른다.



- **F 분포의 성질**

$\alpha = P(F = \frac{V_2/k_2}{V_1/k_1} > f_\alpha(k_2,\ k_1)) = P(\frac{1}{F} = \frac{V_1/k_1}{V_2/k_2} < f_{1-\alpha}(k_1,\ k_2)) = P(\frac{1}{\frac{V_2/k_2}{V_1/k_1}} < f_{1-\alpha}(k_1,\ k_2))$

$= P(\frac{V_2/k_2}{V_1/k_1} > \frac{1}{f_{1-\alpha}(k_1,\ k_2)})$ $\quad \therefore \boldsymbol{f_{1-\alpha}(k_1,\ k_2) = \frac{1}{f_\alpha(k_2,\ k_1)}}$

$f_{0.95}(k_1,\ k_2) = \frac{1}{f_{0.05}(k_2,\ k_1)}$ 를 이용하여 $f_{0.95}(k_1,\ k_2)$를 구함.

- **F 분포와 t분포의 관계**

$V_1 \sim \chi^2(k_1)$ 과 $V_2 \sim \chi^2(k_2)$가 독립일 때, $F = \frac{V_1/k_1}{V_2/k_2} \sim F(k_1,\ k_2)$이다.

$Z \sim \mathrm{N}(0, 1)$ 와 $V \sim \chi^2(k)$가 독립일 때, $\mathrm{T} = \frac{Z}{\sqrt{\frac{V}{k}}} \sim t(k)$이고, $Z^2 \sim \chi^2(1)$ 이므로

$\boldsymbol{T^2 = \frac{Z^2}{\frac{V}{k}} \sim F(1,\ k)}$이다.

- **두 정규모집단에서의 표본분산의 비에 대한 분포**

$X_{11}, X_{12}\ldots, X_{1n_1} \sim \mathrm{N}(\mu_1, \sigma_1{}^2)$, $X_{21}, X_{22}\ldots, X_{2n_2} \sim \mathrm{N}(\mu_2, \sigma_2{}^2)$인 독립인 랜덤표본일 때,
$\frac{(n_1-1)S_1{}^2}{\sigma_1{}^2} \sim \chi^2(n_1 - 1)$, $\frac{(n_2-1)S_2{}^2}{\sigma_2{}^2} \sim \chi^2(n_2 - 1)$이고 서로 독립이 된다.
F분포의 정의에 의해,

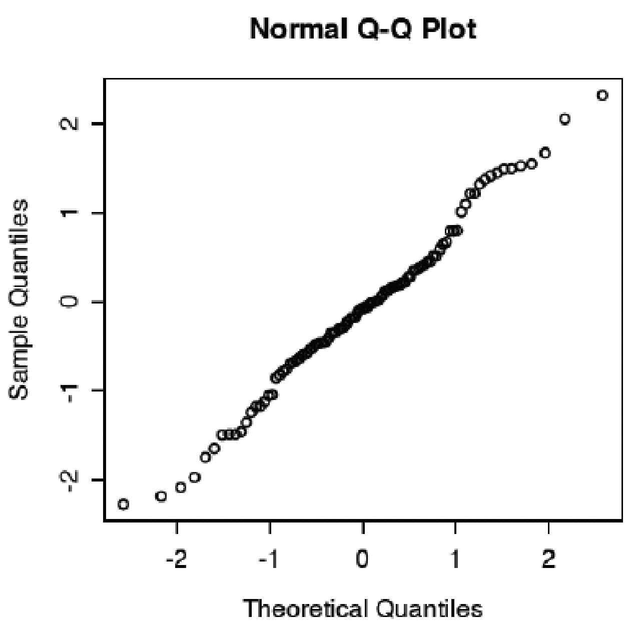
6 정규분포 분위수 대조도

- **정규분포 분위수 대조도**

- 모집단의 분포가 정규분포라는 가정을 검토하기 위한 방법

- 정규분포의 분위수와 이에 대응하는 자료 분포의 분위수를 좌표평면에서 각각 수평축과 수직축의 좌표로 하여 점을 찍은 것

- 정규분포에 가까우면 직선의 모양

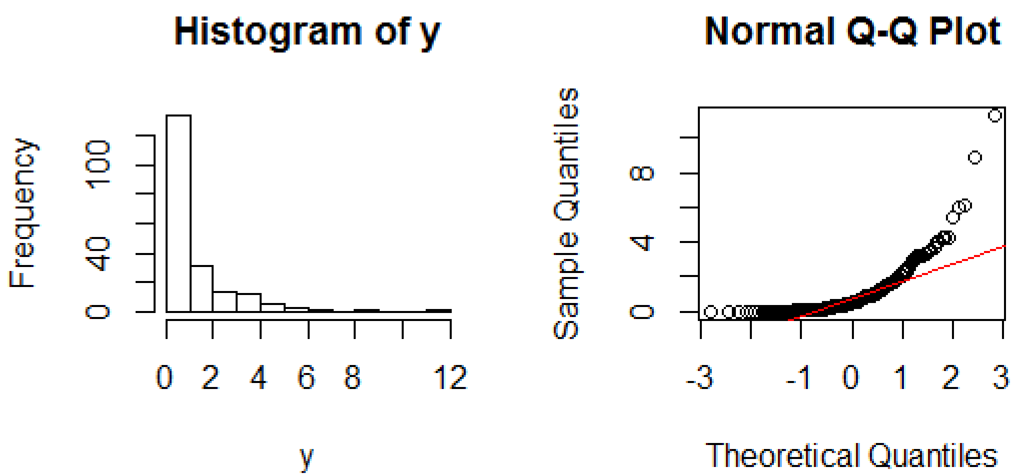


$Q.$ $x_1, x_2 ..., x_{100}$이 $\mathrm{N}(2, 3^2)$을 따르는 모집단으로부터의 표본인가?

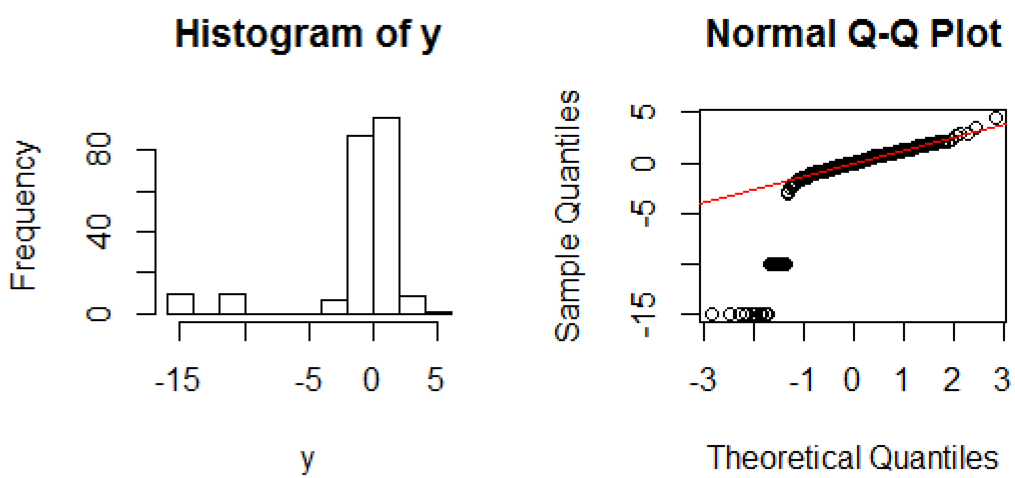
자료를 크기 순서로 나열한 $x_{(1)} < x_{(2)} < ... < x_{(100)}$에서 $x_{(i)}$를 대략 이 자료분포의 $(\frac{i}{100})$분위수라고 생각할 수 있다.

$x_{(1)}, x_{(2)}$ ... 에 대응하는 $\mathrm{N}(2, 3^2)$의 분위수는 $3\times z_{0.99}+2$, $3\times z_{0.98}+2$...이고 만약 실제 모집단이 $\mathrm{N}(2, 3^2)$인 경우에는 서로 대응하는 분위수가 가깝게 나타날 것이다. 즉, $(z_{0.99}, x_{(1)})$, $(z_{0.98}, x_{(2)})$...이 직선주위에 밀집하여 나타날 것이다.

6 정규분포 분위수 대조도



한쪽으로 치우친 분포



꼬리가 긴 분포

결론

- 5장부터는 랜덤표본은 독립이고 분포가 동일한 확률변수의 모임 $X_1, X_2..., X_n$ 을 의미한다.
- 랜덤표본으로 만들어진 함수를 통계량이라고 한다.
- 특별히 모수를 추정하기 위한 통계량을 추정량이라고 한다.
- 통계량도 확률변수이므로 확률분포를 가진다.
- 랜덤표본으로 구성된 통계량의 분포인 표본분포를 통해 모수에 대한 추론이 가능해진다.