3/6

2019년 3월 6일 수요일   11:57

(i) 주사위 1, 2, 3, 4, 5, 6

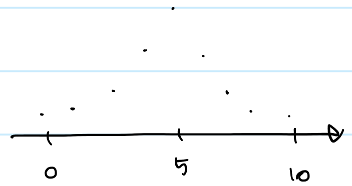
Ω, S → 표본공간 ; Sample space.

$P(X=x)$ → 확률분포함수 ; discrete → 확률질량함수 (p.m.f)

$P(X=x) = \frac{1}{6}$ → 균등분포.

(ii) 동전 10번 던져서 앞면 몇번? $\Omega = \{0, \cdots, 10\}$ → 이산사건

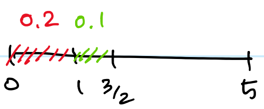
$$P(X=x) = {}_{10}C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{10-x}$$



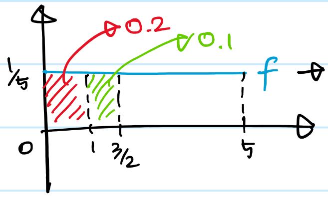
→ 이항분포

(iii) [0, 5] 에서 실수 택하기 $\Omega = \{x \mid 0 \le x \le 5\}$. → 연속사건

$P(X=x) = 0$.



→



f → 확률밀도함수 (p.d.f); 가능도 (likelihood).

$$\to f(x) = \begin{cases} 1/5 & \text{if } 0 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}$$: 균등분포.

cf p.d.f :

(i) $f(x) \ge 0$ (ii) $\int_\Omega f(x)dx = 1$.

(i) unfair coin 의 앞면이 나올 확률 p?

: 1000 번 던져서 뒷면 600, 앞면 400

$$L(p) = {}_{1000}C_{400}\, p^{400} (1-p)^{600}$$

$$L(p) = {}_{1000}C_{400}\, p^{400}(1-p)^{600}$$

$$L'(p) = c\left(400p^{399} - 600(1-p)^{599}\right) = 0 \;\rightarrow\; p = 0.4$$

→ MLE, Max. Likelihood Estimation

(ii) 178, 179, 180, 181, 182 ⋯ μ?

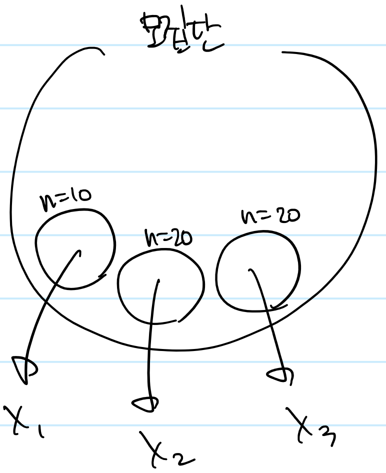
: 정규분포를 따른다고 가정.

$$\mu,\ \sigma^2 \rightarrow \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

$$f(178)f(179)f(180)f(181)f(182)$$

$$\rightarrow (x-178)^2+(x-179)^2+(x-180)^2+(x-181)^2+(x-182)^2 \quad \text{min.}$$

$$\rightarrow x = 180$$



'19 신입생

(i) 수학성적 → 평균 ; μ.

(ii) 정시 지원 → 비율 ; p.

공장 생산제품

(iii) 강도 → 평균 ; μ

(iv) 안정성 → 분산 ; σ

μ, p, σ : 모수

$x_1, x_2, x_3 \cdots \dfrac{x_1+x_2+x_3}{3} = \bar{X} \rightarrow$ 표본평균

$\bar{X}(x_1, x_2, x_3)$?

$\rightarrow x_1 = 1$

$\rightarrow x_2 = 1$ $\cdots$ $\dfrac{x_1+x_2+x_3}{3} = \hat{p} \rightarrow$ 표본비율

$\rightarrow x_3 = 0$

모수	표본 통계량	
μ	$\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$	: 표본평균
p	$\hat{p} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$	: 표본비율

σ^2 $\qquad$ $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$: 표본분산.

??

3/11

2019년 3월 11일 월요일 12:01

< Axioms of Probability >

$P : \Omega \to \mathbb{R}$

$A \mapsto P(A)$

s.t.

(i) $P(\Omega) = 1$

(ii) $P(A) \geq 0$

(iii) $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$ if $A_i \cap A_j = \emptyset$, $i \neq j$.

Thm

(1) $P(A^c) = 1 - P(A)$

pf. $P(\Omega) = P(A \cup A^c)$ $(\because A \cap A^c = \emptyset)$.

$\Rightarrow 1 = P(A) + P(A^c)$

$\Rightarrow P(A^c) = 1 - P(A)$.

(2) $0 \leq P(A) \leq 1$

pf. $P(A) + P(A^c) = 1$ $(\because (1))$.

$\Rightarrow P(A) = 1 - P(A^c) \leq 1$.

(3) $P(\emptyset) = 0$

pf. $P(\Omega) = P(\Omega \cup \emptyset)$

$= P(\Omega) + P(\emptyset)$

$\Rightarrow P(\emptyset) = 0$.

< Conditional Probability >.

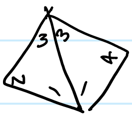
Def

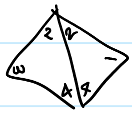
$$P(A|B) = \frac{P(A \cap B)}{P(B)} \; ; \; P(B) > 0.$$

Remark

$$P(A\cap B) = P(B)P(A|B).$$
$$= P(A)P(B|A).$$

EX 1

X : 

Y : 

Y \ X	1	2	3	4
1	×			
2		×	×	×
3		×		
4		×		

× $B : \min(X, Y) = 2$

$\rightarrow P(B) = \frac{5}{16}$.

× $M : \max(X, Y) = 1$.

$\rightarrow P(M) = \frac{1}{16}$.

$P(M=1|B) = 0$

$P(M=2|B) = 1/5$

EX 2.

100개 중 10개 불량품

$P(정상_1 \cap 정상_2) \rightsquigarrow \frac{90}{100} \cdot \frac{89}{99}$

$= P(정상_1)\, P(정상_2|정상_1)$

EX 3. Russian Roulette

$P(A_1) = \frac{1}{6}$; $P(A_2) = \frac{1}{6}\cdot 0 + \frac{5}{6}\cdot\frac{1}{5} \rightarrow \frac{1}{6}$

$P(A_1) = \frac{1}{6}$; $P(A_2) = \frac{1}{6} \cdot 0 + \frac{5}{6} \cdot \frac{1}{5} \rightarrow \frac{1}{6}$

$P(A_3) = \frac{5}{6} \cdot \frac{4}{5} \cdot \frac{1}{4} + \sim \rightarrow \frac{1}{6}$

Remark

(1) $P(A \cap B) = P(A)P(B|A)$.

(2) $P(A \cap B \cap C) = P(A \cap B)P(C|(A \cap B))$

$= P(A)P(B|A)P(C|B \cap A)$.

EX 4.

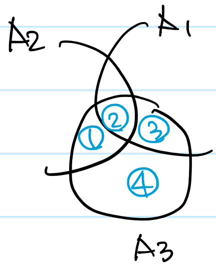
1000 → 3장 당첨 쿠폰

$P(A_1) = \frac{3}{1000}$

$P(A_2) = P(A_1)P(A_2|A_1) + P(A_1^c)P(A_2|A_1^c)$.

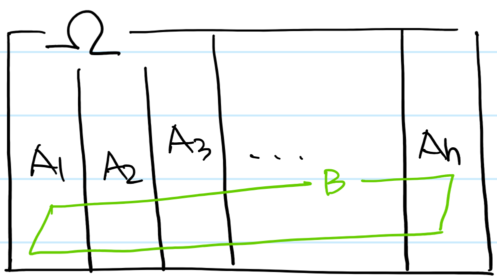
$= \frac{2}{999} \cdot P(A_1) + \frac{3}{999} P(A_1^c)$

$P(A_3) = P(A_1 \cap A_2 \cap A_3) + P(A_1 \cap A_2^c \cap A_3)$
$+ P(A_1^c \cap A_2 \cap A_3) + P(A_1^c \cap A_2^c \cap A_3)$.



→ partition.

Thm 전확률 정리



$P(B) = P\left[\left(\bigcup_{i=1}^{n} A_i\right) \cap B\right]$

$= P\left[\bigcup_{i=1}^{n} (A_i \cap B)\right]$

$= \sum_{i=1}^{n} P(A_i \cap B)$

$= \sum_{i=1}^{n} P(A_i)P(B|A_i)$

Thm Bayes' Rule

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)\,P(B \mid A_i)}{P(B)}.$$

$$= \frac{P(A_i)\,P(B \mid A_i)}{\sum_{i=1}^{n} P(A_i)\,P(B \mid A_i)}.$$

EX 5. 고혈압 키트 O, X

① 전국민 → 0.03 $= P(A^c)$ A : 고혈압

② 환자대상 O → 0.96 $= P(O \mid A)$

③ 정상대상 X → 0.98 $= P(X \mid A^c)$.

Q : O 인데 실제로 고혈압일 확률?

$$= P(A \mid O)$$

$$= \frac{P(A \cap O)}{P(O)} = \frac{P(A)\,P(O \mid A)}{P(O)}.$$

$$= \frac{P(A)\,P(O \mid A)}{P(A)\,P(O \mid A) + P(A^c)\,P(O \mid A^c)}$$

Def Independence of two events

If $P(A \mid B) = P(A)$

or $P(B \mid A) = P(B)$

EX 6.

② \ ①	H	T
H	○×	△○
T	△×	

× A : ① = H

○ B : ② = H

△ C : ① ≠ ②

$P(A \cap B) = 1/4 = P(A)P(B)$ → A, B is independent

$P(B \cap C) = 1/4 = P(B)P(C)$ → B, C 〃

$P(A \cap C) = 1/4 = P(A)P(C)$ → A, C 〃

$P(B\cap C) = 1/4 = P(B)P(C) \rightarrow B, C$ "

$P(A\cap C) = 1/4 = P(A)P(C) \rightarrow A, C$ "

A, B, C is independent? $\rightarrow$ false!

3/13

2019년 3월 13일 수요일 11:59

X : Random Variable. 확률변수.

$X : \Omega \to \mathbb{R}$

$: \omega \mapsto$ ___ an assignment of a value (number).

e.g. 2 coins

X : 앞면의 개수 ~ 0, 1, 2

1 coin

$X : H, T \to X(H) = 0,\ X(T) = 1.$

Cumulative Distribution Function : CDF 누적분포함수.

$$F(x) = P((-\infty, x])$$
$$= P(X \le x).$$

(i) $F(x) \ge 0$

(ii) right-continuous

(iii) $F(\infty) = 1,\quad F(-\infty) = 0,\quad 0 \le F(x) \le 1.$

(iv) $P(a < X \le b) = F(b) - F(a).$

- $P(a \le X \le b) = F(b) - F(a) + P(X = a)$
- cf) $P(a < X < b) = P(a \le X < b)$

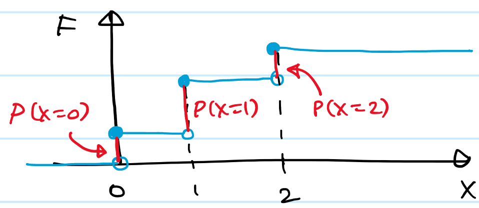
$= P(a < X \le b) = P(a \le X \le b)$ if X is continuous.

(v) $P(X = a) = F(a) - F(a^-)$

$$= \lim_{h \to 0^+} P(a - h < X \le a).$$

EX 1. $X = 0, 1, 2$

X	0	1	2
P(X=x)	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$



$\{ \ 0 \quad x < 0$

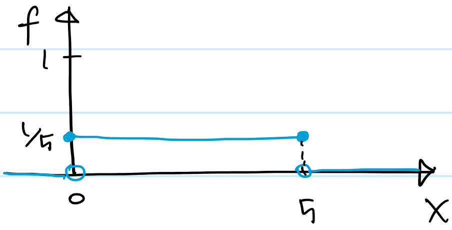
$P(X=x) \mid \frac{1}{4} \quad \frac{1}{2} \quad \frac{1}{4}$

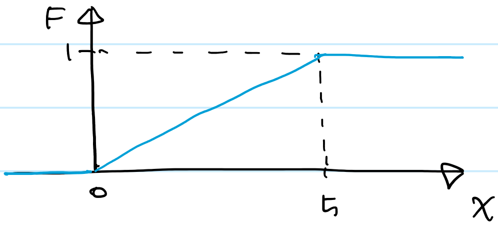
$$\rightsquigarrow f(x) = \begin{cases} 1/4 & x = 0, 2 \\ 1/2 & x = 1 \\ 0 & \text{else} \end{cases}, \quad F(x) = \begin{cases} 0 & x < 0 \\ 1/4 & 0 \le x < 1 \\ 3/4 & 1 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

EX 2. $[0, 5]$

continuous

$$\rightsquigarrow f(x) = \begin{cases} 1/5 & x \in [0, 5] \\ 0 & \text{else} \end{cases}, \quad F(x) = \begin{cases} 0 & x < 0 \\ x/5 & 0 \le x < 5 \\ 1 & 5 \le x \end{cases}$$





EX 3

x	0	1	2	3
$y = x^2$	0	1	4	9
$z = (x - 3/2)^2$	7/4	1/4	1/4	7/4
$P(X=x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
$P(Y=y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
$P(Z=z)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

① $P(Y=0) = \frac{1}{8}$

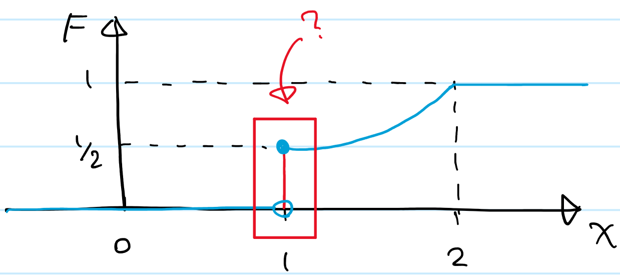
② $P\left(Z = \frac{7}{4}\right) = \frac{1}{4}$.

$\rightarrow$

z	7/4	1/4
$P(Z=z)$	$\frac{1}{4}$	$\frac{3}{4}$

☆ EX 4

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{x^2 - 2x + 2}{2} & 1 \le x \le 2 \\ 1 & 2 \le x \end{cases}$$



$$f(x) = \begin{cases} 0 & x<1 \\ 1/2 & x=1 \\ x-1 & 1<x\le 2 \\ 0 & 2<x \end{cases}$$

"혼합형"

Discrete

$\leftarrow F(x) = \int_{-\infty}^{x} f(t)\,dt$

$F'(x) = f(x)$

Q. $0 \le X \le 1,\ 0 \le Y \le 1$

$P\left(0 \le X \le \frac{1}{2},\ 0 \le Y \le \frac{1}{2}\right) \sim \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} \cdots\, dx\,dy$

$f(x, y) = 1.$

3/18

2019년 3월 18일 월요일 12:03

X : 동전 3번 → 앞면

X	0	1	2	3
$P(X=x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Expectation

$$E(X) = \sum_{x} x P(X=x) = \int_{\Omega} x f(x) dx.$$

Remark

$$E(aX+b) = a\int_{\Omega} x f(x)\, dx + b\int_{\Omega} f(x) dx = aE(X)+b.$$

$$V(X) = E(X^2) - E(X)^2$$

$$V(aX+b) = E(a^2X^2 + 2abX + b^2) - E(aX+b)^2$$
$$= a^2(E(X^2) - E(X)^2)$$
$$= a^2 V(X).$$

Examples

1. Uniform on $0, 1, 2, 3, \cdots, n$

$$\to \quad f(x) = \frac{1}{n+1}, \quad x = 0, 1, \cdots, n$$

(1) $E(X) = \dfrac{1}{n+1}(0+1+\cdots+n) = \dfrac{n}{2}.$

$$V(X) = E(X^2) - \left(\frac{n}{2}\right)^2$$

$$= \frac{1}{n+1}(0^2+1^2+\cdots+n^2) - \left(\frac{n}{2}\right)^2 = \frac{n^2-2n}{12}$$

2. Uniform on $[a, b]$

$$\rightarrow f(x) = \frac{1}{b-a}, \quad a \le x \le b.$$

(1) $E(X) = \int_a^b \frac{x}{b-a}\,dx = \frac{a+b}{2}.$

(2) $V(X) = E(X^2) - \left(\frac{a+b}{2}\right)^2$

$$= \int_a^b \frac{x^2}{b-a}\,dx - \left(\frac{a+b}{2}\right)^2 = \frac{b-a}{12}$$

(3) $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$

$$= \begin{cases} \int_a^x \frac{1}{b-a}\,dt = \frac{x-a}{b-a}. & (a \le x \le b), \\ 0 & (x < a) \\ 1 & (b < x). \end{cases}$$

3. Standard Normal : $N(0,1)$.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty$$

(1) $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx.$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2}\,dx$$

$$= \frac{1}{\sqrt{2\pi}} \left(\lim_{t\to\infty} \int_0^t x e^{-x^2/2}\,dx + \lim_{t\to\infty} \int_{-t}^0 x e^{-x^2/2}\,dx \right)$$

$$= \sqrt{\frac{2}{\pi}} \lim_{t\to\infty} \int_0^t x e^{-x^2/2}\,dx$$

$$= \sqrt{\frac{2}{\pi}} \lim_{t\to\infty} \left(1 - e^{-t^2/2}\right) = 0.$$

(2) $V(X) = E(X^2) - E(X)^2 = E(X^2)$

$$= \int_{-\infty}^{\infty} x^2 f(x)\,dx$$

$$= \int_{-\infty}^{\infty} x^2 f(x)\,dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}\, dx.$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{x}_{f} \cdot \underbrace{x e^{-x^2/2}}_{g'}\, dx.$$

$$\int x \cdot x e^{-x^2/2}\, dx$$

$$= \underbrace{x}_{f} \underbrace{\left(-e^{-x^2/2}\right)}_{g} + \int \underbrace{1}_{f'} \cdot \underbrace{e^{-x^2/2}}_{g}\, dx$$

$$= \frac{1}{\sqrt{2\pi}}\left(-\lim_{x\to+\infty} x e^{-x^2/2} + \lim_{x\to-\infty} x e^{-x^2/2} + \int_{-\infty}^{\infty} e^{-x^2/2}\, dx \right)$$

$$= 0 + 0 + \int_{-\infty}^{\infty} f(x)\, dx. = 1.$$

< Joint CDF >

$$F : \Omega_1 \times \Omega_2 \to \mathbb{R}$$

$$(x, y) \mapsto F(x,y) = P(X \le x, Y \le y).$$

(i) $0 \le F(x,y) \le 1.$

$F(-\infty, y) = 0,\quad F(x, -\infty) = 0.\quad F(\infty, \infty) = 1.$

(ii) $x_1 \le x_2,\ y_1 \le y_2$

$\Rightarrow F(x_1, y_1) \le F(x_2, y_1) \le F(x_2, y_2)$

(iii) $\lim_{x\to a^+} F(x, y) = F(a, y),\quad \lim_{y\to b^+} F(x,y) = F(x, b)$

(iv) $P(x_1 < X \le x_2, Y \le y) = F(x_2, y) - F(x_1, y)$

$P(X \le x, y_1 < Y \le y_2) = F(x, y_2) - F(x, y_1)$

(v) $P(x_1 < X \le x_2, y_1 < Y \le y_2)$

$= P(x_2, y_2) - P(x_1, y_2) - P(x_2, y_1) + P(x_1, y_1).$

(vi) $F_X(x) = F(x, \infty).$

$F_Y(y) = F(\infty, y).$

< Joint PMF >

X, Y : Discrete $\mathbb{R}$

$$f(x, y) = p(x, y) = P(X = x, Y = y)$$

(i) $0 \le p(x, y) \le 1$.

(ii) $\displaystyle\sum_{x \le \infty} \sum_{y \le \infty} p(x, y) = 1$.

(iii) $\displaystyle\sum_{x \le a} \sum_{x \le b} p(x, y) = F(a, b)$

(iv) $\displaystyle P_X(x) = \sum_{y \le \infty} p(x, y), \quad P_Y(y) = \sum_{x \le \infty} p(x, y)$.

: marginal pmf.

(v) $P_{X|Y}(x|y) = P(X = x \mid Y = y)$

$$= \frac{p(x, y)}{P_Y(y)}, \quad \sum_{x \le \infty} p(x|y) = 1.$$

< Joint PDF >

X, Y : Continuous $\mathbb{R}$.

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, dv\, du$$

or $\displaystyle f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$.

(i) $0 \le f(x, y)$

(ii) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$.

(iii) $P(x_1 < X \le x_2,\ y_1 < Y \le y_2) = \int_{y_1}^{y_2}\int_{x_1}^{x_2} f(x, y)\, dx\, dy$.

(iv) $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$,

$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$: marginal pdf.

(v) $\displaystyle f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}, \quad \int_{-\infty}^{\infty} f_{X|Y}(x|y)\, dx = 1$.

← y is fixed. : conditional pdf of X

← y is fixed. : conditional pdf of X

given $Y=y$.

3/20

2019년 3월 20일 수요일 11:55

$$F(x, y) = P(X \le x, Y \le y)$$
$$f(x, y) = P(X = x, Y = y)$$

EX 1 동전 3개 던지기

X : 1^{st} coin ~ $H=0$, $T=1$.

Y : $n(H)$ ~ $0 \cdots 3$

	$Y=0$	$Y=1$	$Y=2$	$Y=3$	
$X=0$	0	$1/8$	$2/8$	$1/8$	$f_X(0) = 4/8$
$X=1$	$1/8$	$2/8$	$1/8$	0	$f_X(1) = 4/8$
	$f_Y(0)=1/8$	$f_Y(1)=3/8$	$f_Y(2)=3/8$	$f_Y(3)=1/8$	

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}.$$

$$\to f(x|y=1) = \frac{f(x, 1)}{f_Y(1)} = \frac{8}{3} f(x, 1)$$

$$= \begin{cases} 1/3 & x = 0 \\ 2/3 & x = 1. \end{cases}$$

EX 2.

X: coin toss. $0, 1$

Y: die. $1, 2, 3, 4$.

	$Y=1$	$Y=2$	$Y=3$	$Y=4$	
$X=0$		$\left(\frac{1}{8}\right)$			$f_X(x) = \frac{1}{2}$
$X=1$					
		$f_Y(y) = \frac{1}{4}$			

$$f(x|Y=3) = \frac{1}{2} \to = f_X(x).$$

$$f(x \mid Y=3) = \frac{1}{2} \rightsquigarrow = f_X(x).$$

~ 독립사건

EX 3.

$\Omega_1 \times \Omega_2 \in (X, Y)$

$$f(x, y) = x e^{-y/2}, \quad 0 \le x \le 1, \quad y \geq 0.$$

$$\iint_{\Omega_1 \times \Omega_2} x e^{-y/2}\, dA = \int_0^\infty \int_0^1 x e^{-y/2}\, dx\, dy. = 1.$$

(1) $$f_X(x) = \int_{\Omega_2} x e^{-y/2}\, dy, \quad 0 < y < \infty$$

$$= -2x e^{-y/2} \Big|_{y=0}^{y=\infty} = 2x.$$

$$f_Y(y) = \int_{\Omega_1} x e^{-y/2}\, dx$$

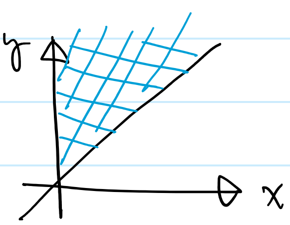
$$= \frac{1}{2} x^2 e^{-y/2} \Big|_{x=0}^{x=1} = \frac{1}{2} e^{-y/2}.$$

(2) $$f(y \mid x) = \frac{f(x, y)}{f_X(x)}.$$

$$= \frac{x \cdot \exp(-y/2)}{2x} = \frac{1}{2} e^{-y/2}, \quad y > 0.$$

EX 4. $$f(x, y) = k e^{-x-y}, \quad 0 < x < y < \infty.$$

(1)



$$\int_0^\infty \int_0^y f(x, y)\, dx\, dy$$

$$= k \int_0^\infty \int_0^y e^{-x} e^{-y}\, dx\, dy.$$

$$= k \int_0^\infty e^{-y} \left[-e^{-x} \Big|_0^y \right] dy = \frac{k}{2} = 1 \quad \therefore k = 2.$$

(2) $f_X(x) = \int_x^{\infty} f(x,y)\,dy$

$= 2e^{-x} \int_x^{\infty} e^{-y}\,dy.$

$= 2e^{-x} \left(-e^{-y}\right)\Big|_x^{\infty} = \underline{2e^{-2x}}. \quad \underline{0 < x}.$

$\rightarrow \quad f(y|x) = \dfrac{f(x,y)}{f_X(x)} = \dfrac{2e^{-x}e^{-y}}{2e^{-2x}} = \underline{e^{x-y}}. \quad \underline{x < y}$

3/25

2019년 3월 25일 월요일 11:58

Note 주어진 확률변수 X의 분포를 사용해 g(X)의 확률밀도함수 구하기

X: a r.v. with $F(x)$ and $f(x)$.

$Y = g(X)$.

assume g^{-1} exists.

$$F(y) = P(Y \le y)$$
$$= P(g(X) \le y).$$
$$= P(X \le g^{-1}(y))$$
$$= \int_{-\infty}^{g^{-1}(y)} f(x)dx.$$

$$\frac{d}{dy} F(y) = \frac{d}{dy} \int_{-\infty}^{g^{-1}(y)} f(x)dx$$

$$\to f(y) = f(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y)$$

* g가 1-1이 아니면 유한개의 1-1 함수로 나누어 결과를 더해 얻을수 있다.

예1)

$X \sim N(0,1) \to f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad -\infty < x < \infty.$

일 때 $Y = X^2$ 의 확률밀도함수

If $x > 0 \to$

$$f(y) = f(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \cdot \frac{1}{2\sqrt{y}}.$$

$$\therefore f(y) = 2\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \frac{1}{2\sqrt{y}}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \cdot \frac{1}{\sqrt{y}}.$$

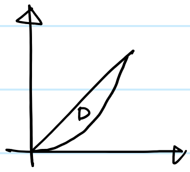
$(X, Y) \in \Omega_1, \Omega_2$

$$F(x, y) = P(X \le x, Y \le y)$$
$$= \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\, dv\, du.$$

where $\iint_{\Omega_1\times\Omega_2} f(x,y)\,dx\,dy = 1$, $f(x,y) \geq 0$.

EX.

$$f(x,y) = \begin{cases} 15y, & x^2 \leq y \leq x. \\ 0, & \text{o.w.} \end{cases}$$



$$\iint_D f(x,y)\,dA = \int_0^1 \int_y^{\sqrt{y}} 15y\,dx\,dy$$

$$= \int_0^1 15y(\sqrt{y}-y)\,dy$$

$$= 15\left(\frac{2}{5}y^{\frac{5}{2}} - \frac{1}{3}y^3\right)\Big|_0^1 = 1.$$

(1) $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$

$$= \int_{x^2}^{x} 15y\,dy = \frac{15}{2}(x^2 - x^4).$$

(2) $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$

$$= \int_y^{\sqrt{y}} 15y\,dx = 15y(\sqrt{y}-y).$$

(3) $f(y|X=\frac{1}{2}) = \dfrac{f(x,y)}{f_X(\frac{1}{2})} = \dfrac{32}{45}f(x,y) = \dfrac{32}{3}y.$ $\quad \frac{1}{4} \leq y \leq \frac{1}{2}.$

Def X, Y are independent r.vs.

$\Leftrightarrow$ $f_Y(y) = f(y|X=x)$ or

$f_X(x) = f(x|Y=y)$.

Remark X, Y are independent.

$\rightarrow$ the followings are equivalent.

(1) $f(x,y) = f_X(x)f_Y(y)$

(2) $f(x,y) = g(x)h(y)$

(3) $F(x,y) = F_X(x)F_Y(y)$

(4) $P(a<X<b,\ c<Y<d) = P(a<X<b)P(c<Y<d)$.

(3) $F(x,y) = F_X(x)F_Y(y)$

(4) $P(a<X<b,\ c<Y<d) = P(a<X<b)P(c<Y<d)$.

Def Expectation

$(X,Y) \in \Omega_1 \times \Omega_2$. $f(x,y)$: pdf of X,Y.

(1) $E(X) = \iint x f(x,y)\, dA$.

(2) $E(g(x,y)) = \iint g(x,y) f(x,y)\, dA$.

(3) $Var(g(x,y)) = E\left[(g(x,y) - E(g(x,y)))^2\right]$

Remarks

(1) $E(aX+b)$

$= \iint (ax+b) f(x,y)\, dx\, dy = aE(X)+b$.

(2) $E(X+Y)$

$= \iint (x+y) f(x,y)\, dx\, dy = E(X)+E(Y)$.

(3) $E(XY)$

$= \iint xy\, f(x,y)\, dx\, dy$

$= E(X)E(Y)$ if X,Y is independent.

(4) $Var(aX+b) = a^2 Var(X)$

(5) $Var(X+Y)$

$= E\left[((x+y) - E(X+Y))^2\right]$

$= \iint \left[(x+y)^2 + (E(X)+E(Y))^2 - 2(x+y)(E(X)+E(Y))\right] f(x,y)\, dx\, dy$.

$\vdots$

$= Var(X) + Var(Y) + 2\underbrace{(E(XY) - E(X)E(Y))}_{\text{Covariance}}$.

$\to Var(X)+Var(Y)$ if X,Y is independent.

(3/27)

2019년 4월 3일 수요일 12:01

$\boxed{1}$ Bernoulli Trial

$X \sim Ber(p)$, $f(x) = p(1-p)^{x-1}$, $x = 0, 1$

$E(X) = p$, $Var(X) = p(1-p)$.

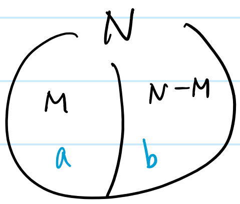
4/1

2019년 4월 1일 월요일 12:00

- 지난 수업 결석함. 복습 필요

이산 : 이항, poisson
연속 : 지수, γ

복원추출	비복원추출
① Bernoulli	Hypergeometric
② Binomial	
③ Geometric	



3개?

$\boxed{2}$

$$P(X=0) = \frac{N-M}{N}\cdot\frac{N-M-1}{N-1}\cdot\frac{N-M-2}{N-2} = \frac{\binom{N-M}{3}}{\binom{N}{3}}$$

$$P(X=1) = \frac{M}{N}\cdot\frac{N-M}{N-1}\cdot\frac{N-M-1}{N-2} = \frac{\binom{M}{1}\binom{N-M}{2}}{\binom{N}{3}}$$

$$\cdots \rightarrow f(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$\rightarrow X \sim H(N, M, n)$: Hypergeometric Trial.

$$(1)\ E(X) = \sum_{x=0}^{n} x f(x)$$

$$= 0 + \sum_{x=1}^{n} x \cdot \frac{\frac{(N-M)!}{(n-x)!(N-M-n+x)!}\cdot\frac{M!}{x!(M-x)!}}{\frac{N!}{n!(N-n)!}}$$

$$= \sum_{x=1}^{n} \frac{\frac{[(N-1)-(M-1)]!}{(n-x)![(N-M)-(n-x)]!}\cdot\frac{M\cdot(M-1)!}{(x-1)![(M-1)-(x-1)]!}}{\frac{N\cdot(N-1)!}{n(n-1)!(N-n)!}}$$

$$= n\cdot\frac{M}{N}\sum_{x=1}^{n}\frac{\binom{M-1}{x-1}\binom{N-M}{(n-1)-(x-1)}}{\binom{N-1}{n-1}}$$

$$\cdots = n\frac{M}{N}$$

$E(X) = E(X_1) + E(X_2) + \cdots$

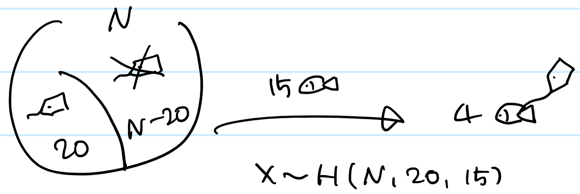
$E(X_1) = \frac{M}{N}$

$E(X_2) = \frac{M}{N}\frac{M-1}{N-1} \cdot \frac{M}{N}\frac{M}{N-1} = \frac{M}{N}$

$\vdots$

$\therefore E(X_i) = \frac{M}{N} \rightarrow E(X) = n\frac{M}{N}.$

e.g.



$X \sim H(N, 20, 15)$

$$\max_N \frac{\binom{20}{4}\binom{N-20}{11}}{\binom{N}{15}} = ?$$

(2) $Var(X) = E(X^2) - E(X)^2$

$= E(X(X-1)) + E(X) - E(X)^2.$

$= n \cdot \frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-M}{N-1}$

$N \to \infty$: $Var(X) \to 1 \sim$ Binomial ?

theorically incorrect ; show with pmf.

4/3

2019년 4월 3일 수요일 12:00

3 Binomial Distribution

X: Bernoulli 시행 n번 → 성공횟수

$X = X_1 + X_2 + \cdots + X_n$. $X_i \sim Ber(p)$.

$E(X) = np$. $Var(X) = np(1-p)$

$f(x) = P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$.

Example

(1) 항암제 치료율 50% → 15명에게 투약

(i) 8명 완치 확률 $P(X=8) = f(8)$

(ii) 10명 이상 $P(X \geq 10) = 1 - P(X \leq 9) \sim$

(2) 새로운 항암제 : 12명 완치 (15명 중)

4 Geometric Distribution

X: 첫 번째 성공까지의 시행횟수.

$X = 1, 2, \cdots$

$f(x) = P(X=x) = (1-p)^{x-1} \cdot p$, $x = 1, 2, \cdots$

$E(X) = \sum_{x=1}^{\infty} x f(x) = \sum_{x=1}^{\infty} x(1-p)^{x-1} p$.

$\frac{d}{dp}(1-p)^x = -x(1-p)^{x-1}$

$= -p \sum_{x=1}^{\infty} \frac{d}{dp}(1-p)^x = -p \cdot \frac{d}{dp} \sum_{x=1}^{\infty} (1-p)^x$.

$= \frac{1}{p}$.

$Var(X) = E(X(X-1)) + E(X) - E(X)^2$.

$\therefore = \frac{1-p}{p^2}$.

4/8

2019년 4월 8일 월요일 11:54

Remark "Memorylessness" 무기억성

e.g. (1) 기하분포.

$$P(X>a+b \mid X>a) = P(X>b).$$

$$\frac{P(X>a+b)}{P(X>a)} = \frac{\sum_{x=a+b+1}^{\infty} p(1-p)^{x-1}}{\sum_{x=a+1}^{\infty} p(1-p)^{x-1}} = (1-p)^b$$

$$= \sum_{x=b+1}^{\infty} p(1-p)^{x-1}.$$

5 (2) 음이항 분포 : Negative Binomial

~ 같은 시행 반복, r 번째 성공할 확률

$x = r, r+1, r+2, \cdots$

$$f(x) = P(X=x) = \binom{x-1}{r-1} p(1-p)^{x-r}$$

6 Poisson Distribution $n>50,\ p<0.1 \Rightarrow B(n,p) \approx P(np)$

단위 구간 (연속적)에서 발생 횟수(비율)가 λ 인 확률실험에 대해

- 단위 구간 $[0,1]$을 n등분하되, 각 소구간에 사건이 한 번 발생하거나 발생하지 않도록 분할하여
- 사건의 성공확률 $\lambda/n = p$ 이고 n 번 시행하는 이항분포로 접근하면
- n 번 시행할 때 성공 횟수를 확률변수 X라 하면

$$P(X=x) = \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} \to \frac{\lambda^x}{x!}e^{-\lambda} \quad \text{as } x \to \infty$$

이 때 $f(x) = \dfrac{\lambda^x}{x!}e^{-\lambda},\ x = 0, 1, 2, \cdots$ 는 하나의 pdf.

즉 단위 구간에서 발생 횟수(비율)이 λ일 때, 실제 발생 횟수 X에 대한 분포를 모수 λ인 포아송 분포라 하고, pdf는 $f(x) = \dfrac{\lambda^x}{x!}e^{-\lambda},\ x = 0, 1, 2, \cdots$

이 때 $E(X) = \lambda$, $Var(X) = \lambda$ 임을 확인할 수 있다.

Remark 단위시간 안에 발생 확률이 희박한 경우 $(p \ll)$

Remark 단위시간 안에 발생 확률이 희박한 경우 $(p \ll)$

X_i) i 번째에서 성공

X : 단위시간 당 발생횟수가 λ 일 때
특정 구간에서 발생할 사건의 발생횟수.

$$P(X=x) = \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}.$$

$$= \frac{n(n-1)\cdots(n-x+1)}{x!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}.$$

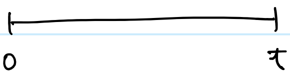
$$= \frac{\lambda^x}{x!}\underbrace{\left(\frac{n}{n}\cdot\frac{n-1}{n}\cdot\cdots\cdot\frac{n-x+1}{n}\right)}_{\downarrow\ 1}\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\downarrow\ e^{-\lambda}}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-x}}_{\downarrow\ 1}$$

$$\to \frac{1}{x!}\cdot\lambda^x\cdot e^{-\lambda}. \quad as\ n\to\infty.$$

$$\therefore f(x) = \frac{1}{x!}\lambda^x e^{-\lambda}, \quad E(X) = \sum x f(x) = \lambda,$$

$$Var(X) = \lambda.$$

Remark 단위시간 $[0, t]$ 에서 λ 번 사건 발생



주어진 구간을 n등분 ∴ 각 소구간의 사건의 수는 1개여야 1

사건이 발생할 확률을 $\lambda/n = p$ 라고 하면 특정 구간에서 발생할 사건의

수를 X 라고 할 때

$$P(X=x) = \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} \quad x=0,1,\cdots,n.$$

$$\therefore P(X=x) \longrightarrow \frac{1}{x!}\lambda^x e^{-\lambda} \quad as \quad n\to\infty.$$

$$X \sim P(\lambda).$$

< Poisson Assumption >

4/10

2019년 4월 10일 수요일 11:59

$B(10000, 0.0001)$

$$\sim P(X \geq 4) = \sum_{x=4}^{10000} \binom{10000}{x} 0.0001^{x}\, 0.9999^{10000-x}$$

$$\approx 0.018982$$

$P(1)$

$$\sim P(X \geq 4) = \sum_{x=4}^{\infty} \frac{1}{x!} e^{-1}$$

$$\approx 0.018988$$

차이가 작다.

$B(n, p) \longrightarrow P(np)$.

$\begin{pmatrix} n > 50 \\ p < 0.1 \end{pmatrix}$

$\boxed{7}$ Exponential Distribution ~ 무기억성

X : 첫 사건이 발생할 때까지의 시간. $x \geq 0$.

X(t): 크기가 t인 구간에서 발생하는 사건의 수.

$F(t) = P(T \leq t) = 1 - P(T > t)$.

→ 시간이 t 될 때까지는 사건이 발생하지 않았다!

$$P(T > t) = P(X(t) = 0) = e^{-\lambda t}$$

$$E(X) = \int_0^{\infty} x \cdot \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}.$$

$$Var(X) = \int_0^{\infty} x^2 \lambda e^{-\lambda x}\, dx - E(X). = \frac{1}{\lambda^2}.$$

e.g. 말기암 환자 "평균 50일" :

X : ________

$P(X \geq 30) =$

$$\updownarrow \quad \int_{30}^{\infty} \frac{1}{50} x e^{-\frac{x}{50}}\, dx$$

$P(X(30) = 0) \quad X(30) \sim Poi\left(\frac{30}{50}\right)$.

Note 지수분포의 모수 λ는 E(X)의 역수이다.

Note $P(X \geq a+b \mid X \geq a) = P(X \geq b)$.

Remark Gamma Function $\Gamma(\alpha)$, $\alpha > 0$.

Def: $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$. $\alpha > 0$.

(1) $\Gamma(1) = 1$; $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$

(2) $\alpha \in \mathbb{N} \Rightarrow \Gamma(\alpha+1) = \alpha!$

(3) $\Gamma\left(\frac{1}{2}\right) = \int_0^\infty t^{-\frac{1}{2}} e^{-t}\, dt$

$$= \int_0^\infty \frac{e^{-u^2}}{u} 2u\, du = \int_{-\infty}^\infty e^{-u^2}\, du = \sqrt{\pi}.$$

(4) $1 = \int_0^\infty \frac{t^{\alpha-1} e^{-t}}{\Gamma(\alpha)}\, dt$ ... pdf?

$\Downarrow t = \frac{x}{\beta}$

$$= \int_0^\infty \frac{1}{\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\frac{x}{\beta}} \frac{dx}{\beta}.$$

$$= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}\, dx.$$

$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$, $x \geq 0$ ← pdf. of Γ-dist

(4-a) $E(X) = \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^\alpha e^{-\frac{x}{\beta}}\, dx$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha}\left[\cancel{x^\alpha(-\beta e^{-\frac{x}{\beta}})\Big|_0^\infty} + \alpha\beta \int_0^\infty x^{\alpha-1} e^{-\frac{x}{\beta}}\, dx \right].$$

$$= \alpha\beta \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}\, dx = \alpha\beta$$

(4-b) $Var(X) = \alpha\beta^2$.

4/17

2019년 4월 15일 월요일 12:02

8 Gamma Distribution

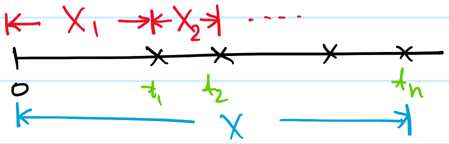
$$\Gamma(\alpha,\beta) \sim, \quad f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}}$$

$$E(X) = \alpha\beta$$

$$Var(x) = \alpha\beta^2.$$

Remark

(1)



X_i : $i-1$ 번째 사건 발생후 첫사건까지의 시간

$X_1 \sim Exp(\lambda)$ $\quad X_2 \sim Exp(\lambda)$ $\cdots$

$\underline{X_1 + X_2 + \cdots + X_n}$

: $E(X) = \frac{n}{\lambda}$, $Var(X) = \frac{n}{\lambda^2}$.

(2) $f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}}$

$\downarrow$ $\alpha=1$ $\beta\lambda=1$

$f(x) = \lambda e^{-\lambda x}$. $\therefore \Gamma(1, \frac{1}{\lambda}) = Exp(\lambda)$.

(3) $X = X_1 + X_2 + \cdots + X_n$

$\sim \Gamma(n, \frac{1}{\lambda})$.

Example

시스템에 도달되는 신호 평균 2초

- X : 첫 도달까지의 시각

  $\rightarrow X \sim Exp(0.5)$

  $X(t) \sim Poi(\frac{1}{2}t)$

- $P(X(10) = 2)$

$$= \frac{(10/2)^2}{2!} e^{-\frac{10}{2}}$$

- 부신 신호가 2건이 되는데 걸리는 평균시간

$X \sim \Gamma(2,2)$

$\rightarrow E(X) = 4.$

9 Uniform Distribution

$$f(x) = \frac{1}{b-a}, \quad a < x < b$$

$E(X) =$

$Var(X) =$

10 Normal Distribution

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

$\sim N(\mu, \sigma^2)$

$E(X) = \mu, \quad Var(X) = \sigma^2.$

$Y = aX + b$

$$\rightarrow F(Y) = P(Y \le y)$$
$$= P(aX + b \le y)$$
$$= P\left(X \le \frac{y-b}{a}\right)$$
$$= \int_{-\infty}^{\frac{y-b}{a}} f(x)\,dx.$$

$$\therefore f(y) = \frac{1}{a} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(y-(b+\mu a))^2}{2\sigma^2 a^2}},$$

$Y \sim N(a\mu + b, a^2\sigma^2).$

Remark X: random variable, $x \ge 0$, $x = 0, 1, 2, \cdots$

(1) Markov's Inequality

$$P(X \ge a) \le \frac{E(X)}{a}, \quad a > 0$$

a

(2) Chebyshev: Inequality

$P(|X-\mu| > \varepsilon\sigma) < 1/\varepsilon^2$, $\varepsilon > 0$.

5/8

2019년 5월 8일 수요일 12:05

- 확률표본 (random sample)

$$\bar{X} = \frac{1}{n} \sum X_i$$ ← 독립 : 표본평균 (sample mean) → 통계량.

$$S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$: 표본분산 (sample variance)

(1) $E(\bar{X}) = \frac{1}{n} \sum E(X_i) = \mu.$

$Var(\bar{X}) = \frac{1}{n^2} \sum Var(X_i) = \frac{\sigma^2}{n}.$

(2) $E(S^2) = \frac{1}{n-1} \sum E[(X_i - \bar{X})^2]$ …??

- 추정량
- 점추정량 (point estimator).

모수 θ를 추정하기 위해 선정된 통계량 $\hat{\theta}$

모집단의 통계량 $\hat{\theta}$은 θ의 (점) 추정량.

현실적으로 점추정에 의해 모수 θ를 추정하는 것은 어려우므로 점추정량 $\hat{\theta}$의 분포를 구하여 θ를 **확률적으로 구간추정** 하는 것이 일반적.

$\boxed{1}$ $\bar{X}$의 확률분포 ~ 표본분포.

(i) $X_i \overset{iid}{\sim} N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

(ii) $X_i \overset{iid}{\sim} ??(\mu, \sigma^2) \Rightarrow \bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$

any random sample

∵ Central Limit Thm.

Remark $X_i \overset{iid}{\sim} Ber(p).$ 이항분포의 정규분포근사

Remark $X_i \overset{iid}{\sim} Ber(p)$. 이항분포의 정규분포근사

$\rightarrow \hat{p} = \bar{X} = \dfrac{X_1 + \cdots + X_n}{n} \approx N\left(p, \dfrac{p(1-p)}{n}\right)$

한편 $X = \sum X_i \rightarrow X \sim B(n, p)$

그런데 $\hat{p} \approx N\left(p, \dfrac{p(1-p)}{n}\right)$ 이므로,

$X \approx N(np, np(1-p))$.

$\downarrow$ normalize

즉 $\dfrac{X - np}{\sqrt{np(1-p)}} \approx N(0, 1)$.

eg. $X \sim B(30, 0.2) \Rightarrow X' \sim N(0.2, 0.16/30)$

(1) $P(X=4) = \binom{30}{4} 0.2^4 0.8^{26} \sim 0.1325$

$P(3.5 \le X' < 4.5) \simeq P\left(\dfrac{3.5-6}{\sqrt{4.8}} < z < \dfrac{4.5-6}{\sqrt{4.8}}\right)$

$\simeq 0.1212$ (approximation)

(2) $P(13 \le X < 17)$ $\quad P(13 < X \le 17)$? $\quad P(13 < X < 17)$? $\quad P(13 \le X < 17)$?

$\times$ $P(12.5 \le X < \cancel{17.5})$

✓ $P(12.5 \le X < 16.5)$ 이항분포의 연속성 수정.

$\boxed{2}$ S^2 의 확률분포

Assume $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$.

(1) $\dfrac{X_i - \mu}{\sigma} \sim N(0, 1)$

(2) $\left(\dfrac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(1)$.

(3) $S^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

$\Rightarrow \underbrace{\sum_{i=1}^{n} \left(\dfrac{X_i - \mu}{\sigma}\right)^2}_{\chi^2(n)}$

$= \dfrac{1}{\ } \sum^{n} (X_i - \mu)^2$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

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

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left[(X_i-\bar{X})^2+\underbrace{2\overbrace{(X_i-\bar{X})}^{\text{편차}}\overbrace{(\bar{X}-\mu)}^{\text{const}}}_{\therefore\ =0}+\overbrace{(\bar{X}-\mu)^2}^{\text{const}}\right].$$

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

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

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

$$\to\ \frac{n-1}{\sigma^2}\,\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \sim \chi^2(n-1).$$

점추정량의 조건

→ 불편성, 유효성,

어떤 모수 θ를 추정하기 위한 점추정량 $\hat{\theta}$ 이라 하면

$E(\hat{\theta})-\theta$; 편의 (bias)

$E(\hat{\theta})-\theta=0$: 불편성 (unbiased)

$\bar{X}_1$: $\frac{1}{n}(X_1+\cdots+X_n)$

$\bar{X}_2$: $\frac{1}{n+1}(X_1+\cdots+X_{n+1})$

$\bar{X}_3$: $\frac{1}{n}(2X_1+\cdots+X_n)$

→ $E(\bar{X}_1)=E(\bar{X}_2)=\mu$, $E(\bar{X}_3)=\frac{n+1}{n}\mu$.

$\therefore$ $E(\bar{X}_1)$, $E(\bar{X}_2)$ → unbiased

$E(\bar{X}_3)$ → biased.

$Var(\bar{X}_1)=\sigma^2/n$, $Var(\bar{X}_2)=\sigma^2/n+1$.

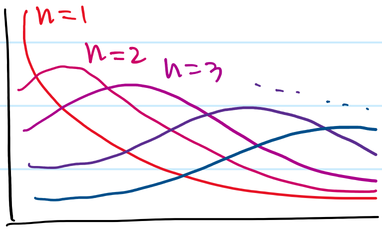
5/15

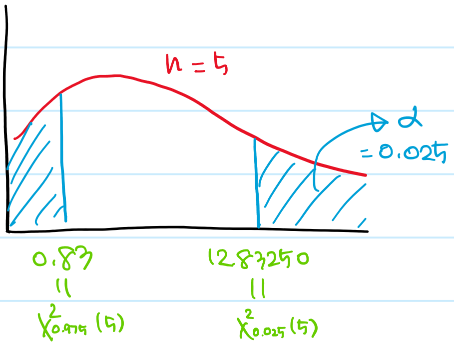
2019년 5월 13일 월요일 11:59

$\boxed{1}$ χ^2-distribution.

$X : \chi^2(n) = \Gamma(\frac{n}{2}, 2)$.

$E(X) = n$, $Var(X) = 2n$.





	⋯	0.025
⋮		
5		12.83
⋮		

Remark σ^2의 95% 신뢰구간?

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

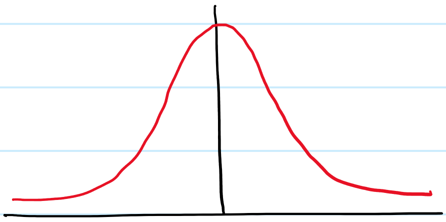
$$P\left(\chi^2_{0.975}(n-1) < \frac{n-1}{\sigma^2} S^2 < \chi^2_{0.025}(n-1)\right) = 0.95$$

$$\Rightarrow P\left(\frac{S^2(n-1)}{\chi^2_{0.025}(n-1)} < \sigma^2 < \frac{S^2(n-1)}{\chi^2_{0.975}(n-1)}\right) = 0.95.$$

$\boxed{2}$ Student's t-distribution

$$T = \frac{Z}{\sqrt{V/r}}.$$

$E(T) = 0$, $Var(T) = \frac{r}{r-2}$.



$$\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} = \frac{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{\frac{n-1}{\sigma^2}S^2}{n-1}}} \sim t(n-1).$$

$$P\left(-t_{0.025}(n-1) < \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} < t_{0.025}(n-1)\right) = 0.95$$

cf) normal dist. :

$$P(\bar{X}-1.96\sigma/\sqrt{n} < \mu < \bar{X}+1.96\sigma/\sqrt{n}).$$

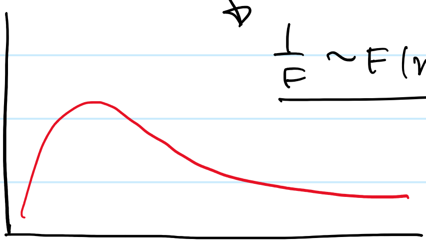
Remark 95% 신뢰구간

$P(169 < \mu < 173) \leftarrow \times$

$\boxed{3}$ F-distribution.

$$F = \frac{U/m}{V/n} \sim F(m,n).$$

$$\Rightarrow \frac{1}{F} \sim F(n,m).$$



$$P(F > f_\alpha(m,n)) = \alpha.$$

$$\Rightarrow P\left(\frac{1}{F} < \frac{1}{f_\alpha(m,n)}\right) = \alpha.$$

$$\therefore f_{1-\alpha}(n,m) = \frac{1}{f_\alpha(m,n)}.$$

$$E(F) = \frac{n}{n-2}, \quad Var(F) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$$

5/20

2019년 5월 20일 월요일 12:01

(1) $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \Rightarrow \bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right).$

(2) $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \Rightarrow \frac{n-1}{\sigma^2}S^2 \approx \chi^2(n-1).$

(3) $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i \Rightarrow \hat{p} \approx N\left(p, \frac{p(1-p)}{n}\right).$

where $X_i \sim Ber(p)$.

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1).$$

$$P\left(-Z_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z_{\frac{\alpha}{2}}\right) \cdots P\left(L < \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} < U\right)$$

(1) σ known : $\Rightarrow P(L < \mu < U)$.

(2) σ unknown : $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$.

(4) $f(x) = \frac{2}{\theta^2}(\theta - x), \quad 0 < x < \theta, \quad (\theta > 0)$

$E(X) = \frac{\theta}{3}, \quad Var(X) = \frac{\theta^2}{18}$

$\Rightarrow \theta \leftarrow \hat{\theta} = 3\sum_{i=1}^{n}\frac{X_i}{n}.$

(ㄱ) $E(\hat{\theta}) = E\left(3 \cdot \frac{1}{n}\sum_{i=1}^{n} X_i\right)$

$= 3 \cdot \frac{1}{n} \cdot \frac{\theta}{3} \cdot n = n\theta$

(ㄴ) $\hat{\theta}$ 의 분포 : $3\bar{X} \approx N\left(3 \cdot \frac{\theta}{3}, \; 9 \cdot \frac{\theta^2}{18n}\right)$

$= N\left(\theta, \frac{\theta^2}{2n}\right).$

i.e. $\frac{\hat{\theta}-\theta}{\cdots} \sim N(0,1)$

i.e. $\dfrac{\hat{\theta}-\theta}{\theta/\sqrt{2n}} \sim N(0,1)$

(9) $N(\mu_1, \sigma_1^2)$ $\qquad$ $N(\mu_2, \sigma_2^2)$

$X_1, X_2, \cdots, X_n \qquad Y_1, Y_2, \cdots, Y_n.$

$\bar{X} \sim N\left(\mu_1, \dfrac{\sigma_1^2}{n}\right), \quad \bar{Y} \sim N\left(\mu_2, \dfrac{\sigma_2^2}{n}\right).$

$X \sim N(\mu_1, \sigma_1^2)$
$Y \sim N(\mu_2, \sigma_2^2)$
$\Rightarrow X-Y \sim N(\mu_1-\mu_2, \sigma_1^2+\sigma_2^2).$

$\mu_1 - \mu_2 \leftarrow \bar{X} - \bar{Y}.$

$\bar{X}-\bar{Y}$ 의 분포 :

$$\bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2+\sigma_2^2}{n}\right).$$

(i) $\sigma = \sigma_1 = \sigma_2$ known

$$\frac{\bar{X}_n - \bar{Y}_m - (\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim N(0,1).$$

(ii) $\sigma = \sigma_1 = \sigma_2$ unknown

$$\frac{\bar{X}_n - \bar{Y}_m - (\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} = \frac{\overbrace{\left(\dfrac{\bar{X}_n - \bar{Y}_n - (\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}\right)}^{N(0,1)}}{\sqrt{\dfrac{\frac{(n+m-2)}{\sigma^2}S_p^2}{n+m-2}}} \sim t(n+m-2).$$

$$\therefore P\left(-t_{\frac{\alpha}{2}}(n+m-2) < \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_p\sqrt{n^{-1}+m^{-1}}} < t_{\frac{\alpha}{2}}(n+m-2)\right) = 1-\alpha.$$

(iii) $\sigma_1 \neq \sigma_2$, unknown.

$$\frac{\bar{X}-\bar{Y}-(\mu_2-\mu_1)}{\sqrt{\frac{S_1^2}{\quad}+\frac{S_2^2}{\quad}}} \sim t(v).$$

$$\frac{\quad}{\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}} \sim t(\nu).$$

$$\nu = \frac{\left(\frac{S_1^2}{n}+\frac{S_2^2}{m}\right)^2}{\frac{\left(\frac{S_1^2}{n}\right)^2}{n-1}+\frac{\left(\frac{S_2^2}{m}\right)^2}{m-1}}$$

5/22

2019년 5월 22일 수요일 12:45

(3) $$\frac{(n-1)s_1^2}{\sigma_1^2} \sim \chi^2(n-1),$$

$$\frac{(n-1)s_2^2}{\sigma_2^2} \sim \chi^2(m-1)$$

$\cancel{\sigma_1^2 - \sigma_2^2 \cdots ?}$

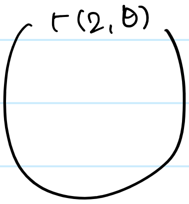
$\frac{\sigma_1^2}{\sigma_2^2} \cdots ?$

(4) $$\frac{\frac{s_1^2}{s_2^2}}{\frac{\sigma_1^2}{\sigma_2^2}} = \frac{\frac{(n-1)s_1^2}{\sigma_1^2} \cdot \frac{1}{n-1}}{\frac{(m-1)s_2^2}{\sigma_2^2} \cdot \frac{1}{m-1}} \sim F(n-1, m-1).$$

$$P\left(f_{1-\frac{\alpha}{2}}(n-1, m-1) < \frac{\frac{s_1^2}{s_2^2}}{\frac{\sigma_1^2}{\sigma_2^2}} < f_{\frac{\alpha}{2}}(n-1, m-1) \right) = 1-\alpha.$$

Examples

1. $\Gamma(2, \theta)$



$E(X) = 2\theta$

$Var(X) = 2\theta^2$

θ의 95% confidence

(i) $\hat{\theta} =$ ~~~

(ii) $\bar{X} \sim N(2\theta, \frac{2\theta^2}{n})$.

$$P\left(-1.96 < \frac{\bar{X} - 2\theta}{\frac{\sqrt{2}\theta}{\sqrt{n}}} < 1.96\right)$$

$\theta \leftarrow \hat{\theta} = \frac{1}{2}\overline{X_{50}}$: unbiased.

$E(\hat{\theta}) = \frac{1}{2}E(\overline{X_{50}}) = \frac{1}{2} \cdot 2\theta = \theta$.

$\rightsquigarrow \overline{X_{50}} \approx N\left(2\theta, \dfrac{2\theta^2}{50}\right)$ 이므로,

(1) $\hat{\theta} \approx N\left(\theta, \dfrac{\theta^2}{100}\right)$.

$$\therefore \frac{\hat{\theta}-\theta}{\theta/10} \approx N(0,1).$$

$$P\left(-1.96 < \frac{\hat{\theta}-\theta}{\theta/10} < 1.96\right)$$

$$= P\left(-1.96 < \frac{10}{\theta}(\hat{\theta}-\theta) < 1.96\right)$$

$$= P\left(-1.96 < \frac{10\hat{\theta}}{\theta} - 10 < 1.96\right).$$

$$= P\left(\frac{8.04}{10\hat{\theta}} < \frac{1}{\theta} < \frac{11.96}{10\hat{\theta}}\right)$$

$$= P\left(\frac{10\hat{\theta}}{11.96} < \theta < \frac{10\hat{\theta}}{8.04}\right).$$

$\rightarrow$ 95% conf.

(2) $\dfrac{\overline{X_{50}}-2\theta}{\theta/5} \sim N(0,1)$.

$\downarrow$

지역 1	12일간	$\overline{X_{12}} = 4.76$	$S_1^2 = 3.17^2$
지역 2	10일간	$\overline{X_{10}} = 2.58$	$S_2^2 = 0.92^2$

$\underline{X \sim N.}$ σ_1? σ_2? $\sigma_1 = \sigma_2$???

$$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n-1, m-1).$$

$\mu_1 - \mu_2$

$$\rightarrow \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{12}+\frac{\sigma_2^2}{10}}} \sim N(0,1)$$

(1) $\sigma_1 = \sigma_2 \rightarrow$

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t(n+m-2).$$

$$P\left(-t_{\frac{\alpha}{2}}(\ ) < \frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} < t_{\frac{\alpha}{2}}(n+m-2)\right).$$

$= 0.95.$

$$P\left(-t_{\frac{\alpha}{2}}(\ ) < \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} < t_{\frac{\alpha}{2}}(n+m-2)\right).$$

$$=0.95.$$

(2) $\sigma_1 \neq \sigma_2$

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}} \approx t(u).$$

$$\frac{\frac{S_1^2}{S_2^2}}{\frac{\sigma_1^2}{\sigma_2^2}} \sim F(n-1, m-1).$$

$$\rightarrow \frac{3.17^2/0.98^2}{\sigma_1^2/\sigma_2^2} \sim F(11,9).$$

$$\therefore P\left(f_{0.975}(11,9) < \frac{3.17/0.98^2}{\sigma_1^2/\sigma_2^2} < f_{0.025}(11,9)\right) = 0.95$$

$$f_{0.975}(11,9) = \frac{1}{f_{0.025}(9,11)}.$$

$$P\left(\underset{3.5}{L} < \frac{\sigma_1^2}{\sigma_2^2} < \underset{42.6}{U}\right) = 0.95 \rightarrow (\sigma_1 \neq \sigma_2)!$$

6/3 - 1

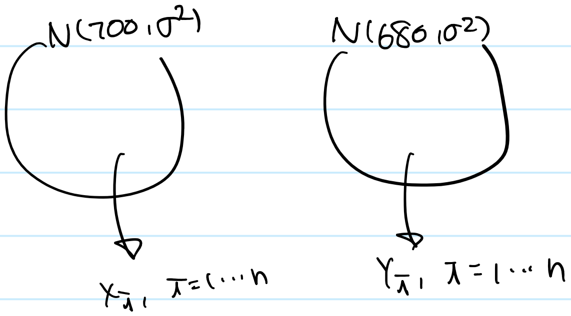
2019년 6월 3일 월요일 12:03

$\bar{X}-\bar{Y}$ 의 분포

$P(\bar{X}-\bar{Y} > t_0) = 0.05$

A gr.: 17개 $\bar{X} = 704$ $s_X^2 = 39.25^2$

B gr.: 10개 $\bar{Y} = 675$ $s_Y^2 = 43.75^2$



$\rightarrow$ $\bar{X}-\bar{Y} \sim N(20, \sigma^2/n^2)$

$$\frac{\bar{X}-\bar{Y}}{\sqrt{1/17+1/10}} \sim t(17+10-2)$$

6/3 - 2

2019년 6월 3일 월요일 12:20

통계적 추론
- 추정. 점추정. 구간추정.
- 가설검정

(i) 모집단의 모수에 대한 추측, 주장에 대한 가설 확립

(ii) 표본이 나타내는 정보를 근거로

(iii) 귀무가설 하에서 주어진 표본의 자료가 비정상인지 정상적인지를 기준을 가지고 판단하는 것.

H_0 귀무가설 null hypothesis

H_1 대립가설 alternative hypothesis

(iv) 귀무가설을 채택 혹은 기각한다.

대립가설 입장에서는 '보류'.

㉾ 45 kcal

16개 $x_1, \cdots, x_{16}$ → $\bar{x} = 42$ kcal

→ $H_1 : \mu < 45$

$H_0 : \mu \geq 45$: 채택하거나 기각

e.g. (1) H_1 : '신제품의 kcal 이 기존보다 낮다'

→ $H_1 : \mu < 45$

$H_0 : \mu \geq 45$.

$\bar{x} = \frac{1}{16} \Sigma x_i = 42$

'비정상적' 자료! 귀무가설의 입장에서 45보다 작을수록

(2) 아들 57, 딸 43

$H_1 = \hat{p} \neq 0.5$ / $\hat{p} > 0.5$ $H_0 : \hat{p} = 0.5$ / $\hat{p} \leq 0.5$

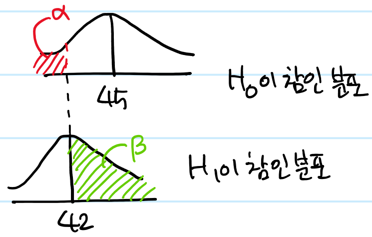
H_0 : 반드시 (=) 이 포함되어야 함

H_1 : (=) 이 있으면 안됨

$H_0^c = H_1$, $H_1^c = H_0$

e.g. $H_0: \mu \geq 45$

$H_1: \mu < 45$.



$\alpha = P(H_0$ 기각 $| H_0$ 참$)$.

: 제 1종 오류.

$\beta = P(H_0$ 채택 $| H_1$ 참$)$.

: 제 2종 오류.

$\alpha\downarrow \rightarrow \beta\uparrow$, $\alpha\uparrow \rightarrow \beta\downarrow$

: α와 β를 모두 작게 할 수는 없으므로,

α를 고정시킨 후 β가 최소가 되도록

검정통계량을 정한다.

$1-\beta$: 검정력.

$$\alpha = 0.05 = H_0\left(\frac{\bar{X}-45}{8/4} < -1.643\right)$$

$$\beta = H_1\left(\frac{\bar{X}-42}{8/4} > z_n\right).$$

6/5

2019년 6월 5일 수요일 11:59

Example (π) 16개의 kcal 값 $x_1 \cdots x_{16}$ ← $N(\mu, 8^2)$

→ $\overline{x_{16}} = 42.7$

$H_1 : \mu < 45$

$H_0 : \mu \geq 45$

- H_0가 참이라는 전제로 주어진 표본정보가 얼마나 비정상인지 판단

  기준에 근거 표본이 비정상이면 H_0 기각 (H_1 채택)

  표본이 정상이면 H_0 채택 (H_1 보류)

예를 들어 $H_0 : \mu = 45$

$H_1 : \mu = 42$ 인 경우

$\alpha = 0.05$ 라고 하면

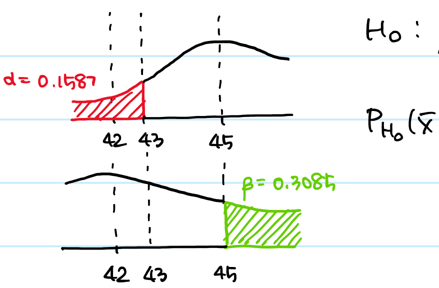
$$P_{H_0}(\bar{X} < 43) = P\left(\frac{\bar{X}-45}{8/\sqrt{16}} < \frac{43-45}{2}\right)$$

$$= P(Z < -1) = 0.1587$$

$$P_{H_1}(\bar{X} > 43) = P\left(\frac{\bar{X}-42}{8/\sqrt{16}} < \frac{43-42}{2}\right)$$

$$= P(Z > 0.5) = 0.3085$$

(1)



검정력 = 1 − 0.3085
= 0.6925

(2)

$H_0 : \mu = 46$

$$P_{H_0}(\bar{X} < 43) = P\left(\frac{\bar{X}-46}{8/\sqrt{16}} < \frac{43-46}{2}\right)$$

$$= P(Z < -1.5)$$

$$= 0.0668$$

- 단일모집단의 모수에 대한 가설검정

  → 정규모집단 여부?

§9.2 1~7
§9.3 1~3
§9.4 1~2

(1) 모수 μ 에 대한 추측.

(i) $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ σ known

(가) $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ $\quad \sigma$ unknown.

관측값 $z_0 = \dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$.

e.g. 대졸취업률 $= 54.5\ \%$

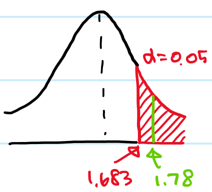
컴공 42개대 $x_1 \dots x_{42}$: $\bar{x}_{42} = 58.77\ \%$ $(s^2 = 15.54^2)$

$H_1 : \mu > 54.5$ $\quad H_0 : \mu \le 54.5$

$(\mu = 54.5)$

$$T = \frac{\bar{X} - 54.5}{15.54/\sqrt{42}} \sim t(41).$$

$$t_0 = \frac{58.77 - 54.5}{15.54/\sqrt{42}} = 1.78$$



정규모집단이 아니면? $\bar{X} \approx N(\mu, \sigma^2/n)$.

(대표본)

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \approx N(0,1)$$

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \approx \cancel{t(n-1)}.\ N(0,1).$$

(2) 모분산 σ^2 에 대한 추측.

(나) $\dfrac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$.

정규성을 가정할 수 없을 때

(1) $\mu \to \bar{X} \approx N\left(\mu, \dfrac{\sigma^2}{n}\right)$ by CLT.

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \approx N(0,1), \quad \frac{\bar{X}-\mu}{S/\sqrt{n}} \approx N(0,1).$$

(2) $p \to \hat{p} \approx N\left(p, \dfrac{p(1-p)}{n}\right)$. by CLT.

$$\frac{p-\hat{p}}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1).$$

$$\frac{p-\hat{p}}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1).$$

Example 보건복지부 − 20.5%

500 명 중 음주경험 120 명 − 24%

$\hat{p} = 0.24$, $\hat{p} \approx N\left(p, \frac{p(1-p)}{n}\right)$. $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$.

$H_0: p = 0.205 \longleftarrow H_1: p \neq 0.205$. 양측검정.

$H_0: p \leq 0.205 \longleftarrow H_1: p > 0.205$. 단측검정

$(=)$

$$z_0 = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = 1.94$$

Example 제품의 강도의 분산

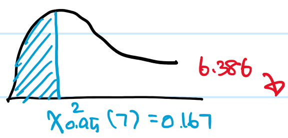
$s^2 = 3.65$, 8개 _ _ _ _ . .

$\sigma < 2 \to$ 안정적이다 : H_1

$\sigma \geq 2$: H_0.

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

$x_0 = \frac{n \times 3.65}{2^2} = 6.386$



6/10

2019년 6월 10일 월요일 13:06

(1) 모집단 2개 (정규성 가정).

$N(\mu_1, \sigma_1^2)$

$N(\mu_2, \sigma_2^2)$.

① $\mu_1 - \mu_2 \leftarrow \bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$.

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \sim N(0,1). \qquad \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t(n+m-2).$$

② $\frac{\sigma_2}{\sigma_1} \leftarrow \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n-1, m-1)$.

(2) 정규성 가정이 없고 대표본 → CLT에 의해 근사.

cf) 정규성 가정이 없고 소표본?

Example $\bar{x} = 69.8$: 5개 $s_1^2 = 13.7$

$\bar{y} = 48.75$: 4개 $s_2^2 = 9.58$.

$H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 > 0$.

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_p\sqrt{1/5+1/4}} \sim t(5+4-2).$$