

## Contents

1 確率空間	2
2 中心極限定理・母分散既知の場合の信頼区間	4
3 仮設検定	6
3.1 $z$ 検定	6
3.2 母比率の検定・母平均の差の検定	7
3.3 母分散の検定	8
3.4 母分散の検定 (母平均未知)	10
3.5 $t$ 検定	13
3.6 母平均の差の検定 (母分散未知)	15
4 適合度の検定・独立性の検定	16
5 一元配置の分散分析	18
6 第一種誤り・第二種誤り・検定力	21
7 回帰分析	23
8 重回帰分析	25

# 1 確率空間

## Definition 1.1. $\sigma$ 代数

集合  $\Omega$  の部分集合族  $\mathcal{F}$  が次を満たすとき、 $\mathcal{F}$  を  $\Omega$  上の  $\sigma$  代数という：

- $\Omega \in \mathcal{F}$
- $A \in \mathcal{F} \Rightarrow A^c = \Omega \setminus A \in \mathcal{F}$
- $A_n \in \mathcal{F} (n \in \mathbb{N}) \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

## Definition 1.2. 確率

$\sigma$  代数  $\mathcal{F}$  に対し、写像  $P : \mathcal{F} \rightarrow [0, 1]$  が次を満たすとき確率という：

- $P(\Omega) = 1$
- 互いに素な  $A_1, A_2, \dots \in \mathcal{F}$  に対して

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \quad (1)$$

## Definition 1.3. 確率空間

集合  $\Omega$ 、その上の  $\sigma$  代数  $\mathcal{F}$ 、及び  $\mathcal{F}$  上の確率  $P$  の組  $(\Omega, \mathcal{F}, P)$  を確率空間という

## Proposition 1.1. Markov の不等式

$X$  を非負の確率変数とする。このとき

$$P(X \geq a) \leq \frac{E[X]}{a} \quad (2)$$

Proof.

$$E[X] = \int_{-\infty}^{\infty} xf(x) dx \quad (3)$$

$$= \int_0^{\infty} xf(x) dx \quad (4)$$

$$\geq \int_a^{\infty} xf(x) dx \quad (5)$$

$$\geq \int_a^{\infty} af(x) dx \quad (6)$$

$$= a \int_a^{\infty} f(x) dx \quad (7)$$

$$= aP(X \geq a) \quad (8)$$

□

**Corollary 1.1.**  $\phi$  を非負の単調増加関数とし、その値も非負であるとする。このとき、 $X$  を確率変数とすると、 $a \geq 0$  かつ  $\phi(a) > 0$  なら

$$P(|X| \geq a) \leq \frac{E(\phi(|X|))}{\phi(a)} \quad (9)$$

## Proposition 1.2. Chebyshev の不等式

$X$  を期待値が  $\mu$  で分散が  $\sigma^2$  の確率変数とする。このとき、任意の  $k > 0$  に対して

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad (10)$$

ただし、 $k > 1$  の場合にだけ意味がある

Proof. Markov の不等式  $\phi(|Y| > a) \leq \frac{E[|Y|]}{a}$  を  $Y = (X - \mu)^2, a = (k\sigma)^2$  と適用すればいい  $\square$

**Proposition 1.3. Hölder の不等式**

確率変数  $X, Y$  に対して,  $1 < p, q < \infty$  かつ  $\frac{1}{p} + \frac{1}{q} = 1$  のとき

$$E|XY| \leq (E|X|^p)^{\frac{1}{p}} (E|Y|^q)^{\frac{1}{q}} \quad (11)$$

となる. また,  $0 < r < s$  なる  $r, s$  に対して

$$E|X|^r \leq (E|X|^s)^{\frac{r}{s}} \quad (12)$$

**Proposition 1.4. Jensen の不等式**

$p(x)$  が確率密度関数の場合,  $E[f(y)] \geq f(E[y])$

## 2 中心極限定理・母分散既知の場合の信頼区間

### Definition 2.1. 独立同分布

確率変数列  $\{X_n\}_{n \geq 1}$  が次の条件を満たすとき、独立同分布である

- 各  $X_n$  は同じ分布に従う 言い換えれば、 $\forall x \in \mathbb{R}, P(X_n \leq x)$  は  $n$  に依らない
- $X_1, X_2, \dots$  は互いに独立である

### Theorem 2.1. 中心極限定理

確率変数列  $\{X_n\}_{n \geq 1}$  が独立同分布で、 $m = \mathbb{E}[X_1], 0 < V[X_1] = \sigma^2 < \infty$  とする。標本平均

$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$  に対して、 $\forall a < b$  で

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}} \leq b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad (13)$$

### Definition 2.2. 標準正規分布の $\alpha$ 点

$z(\alpha)$  は次の条件を満たす実数とする

- $P(-z(\alpha) \leq Z \leq z(\alpha)) = 1 - \alpha$
- $Z \sim N(0, 1)$

### Definition 2.3. 母集団・母数・母平均・母分散

- 確率変数の全体は母集団であり、そこから得られた有限個のデータ列は標本である
- 母集団を特徴付ける未知の値は母数であり、理論的に一定だが直接観測できないため、標本から推定される
- 母集団から取り出した確率変数  $X$  が同一の期待値・分散をもつとすると、母平均と母分散はそれぞれ  $\mu = \mathbb{E}[X], \sigma^2 = V[X]$  である

### Definition 2.4. 標本平均・標本分散

標本  $(X_1, \dots, X_n)$  に対して、標本平均と標本分散はそれぞれ次で定義される

- $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$
- $S_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2$

### Definition 2.5. 信頼係数 $1 - \alpha$

母分散  $\sigma^2$  が既知のとき、母平均  $\mu$  の信頼係数  $1 - \alpha\%$  の信頼区間は

$$\left[ \bar{X}_n - z(\alpha) \frac{\sigma}{\sqrt{n}}, \bar{X}_n + z(\alpha) \frac{\sigma}{\sqrt{n}} \right] \quad (14)$$

ここで、 $z(\alpha)$  は標準正規分布  $Z \sim N(0, 1)$  に対し

$$P(-z(\alpha) \leq Z \leq z(\alpha)) = 1 - \alpha \quad (15)$$

例：95% の信頼区間に対し、 $\alpha = 0.05$  であり、 $z(0.05) \approx 1.96$  である

### Theorem 2.2. 大数の法則

- 大数の弱法則：独立同分布に従う可積分な確率変数の無限列  $X_1, X_2, \dots$  が与えられたとき、その平均を  $\mu$  とおく。標本平均  $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$  のとる値が平均  $\mu$  の近傍から外れる確率は、十分大きな  $n$  を取れば、いくらでも小さくできる。すなわち、 $\forall \epsilon > 0$

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \epsilon) = 0 \quad (16)$$

- 大数の強法則：同じ仮定の下で  $n \rightarrow \infty$  とするとき、 $\bar{X}_n$  は  $\mu$  にほとんど確実に収束する。すなわち、 $P\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1$

Proof. 大数の弱法則は Chebyshev の不等式を用いて示せる。確率変数列は独立同分布に従っているので、確率変数  $\bar{X}_n$  の平均と分散はそれぞれ  $\mu$  と  $\frac{\sigma^2}{n}$  である。したがって、Chebyshev の不等式より

$$0 \leq P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{1}{\epsilon^2} V[\bar{X}_n] = \frac{1}{\epsilon^2} \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0 \quad (17)$$

□

### 3 仮説検定

#### 3.1 $z$ 検定

##### Definition 3.1. 帰無仮説・対立仮説・検定統計量・判定

- 帰無仮説  $H_0$  : 差がない・効果がないなどの基準仮説
- 対立仮説  $H_1$  : 差がある・効果がある (両側・片側)
- 検定統計量 : 標本データから計算する指標
- 判定 : 有意水準  $\alpha$  を固定し, 臨界値または  $p$  値で判断

詳しく言うと

- 帰無仮説  $H_0$ 
  - 「差がない」「効果がない」など, まず前提として置く仮説
  - 例 : 母平均  $\mu = \mu_0$
  - 検定では,  $H_0$  を全体にデータを評価する
- 対立仮説  $H_1$ 
  - 帰無仮説が誤りであるときに成り立つ仮説
  - 「差がある」「効果がある」などを主張する
  - 例 :  $\mu \neq \mu_0, \mu > \mu_0, \mu < \mu_0$

##### Definition 3.2. 両側 $z$ 検定の $p$ 値

標本から得られた検定統計量を  $Z$  とする. 両側検定における  $p$  値とは

$$p = P(|Z^*| \geq |Z|) \quad (18)$$

ここで,  $Z^*$  は標準正規分布に従う確率変数である

詳しく言うと

- $H_0$  が正しいと仮定しときに,  $Z$  以上となる確率
- $p$  が小さいほど, データは  $H_0$  のもとでは起こりにくい  $\implies H_0$  に反する証拠が強い
- $p$  値は, 片側検定か両側検定かで値が異なる. 片側の場合,  $p = P(Z^* \geq |Z|)$  などを使う

### 3.2 母比率の検定・母平均の差の検定

標本  $X_1, \dots, X_n$  について,  $X = X_1 + \dots + X_n$  は二項分布に従う. そこで,  $\frac{X}{n}$  を標本比率である

- $X \sim B(n, p)$  のとき,  $E[X] = np, V[X] = np(1-p)$
- 標本比率  $\hat{p} = \frac{X}{n}$  の平均・分散はそれぞれ  $E[\hat{p}] = p, V[\hat{p}] = \frac{p(1-p)}{n}$  である

#### Definition 3.3. 母比率の検定

$X \sim B(n, p)$  とし, 標本比率  $\hat{p} = \frac{X}{n}$  を観測  
 帰無仮説  $H_0: p = p_0$ , 対立仮説  $H_1: p \neq p_0$  (両側) を検定する  
 $n$  が十分大きいとき, 検定統計量  $Z$  は

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad (19)$$

で,  $H_0$  が正しければ  $Z \approx N(0, 1)$ . 目安:  $np_0 \geq 5, n(1-p_0) \geq 5$

二つの母集団から, 別々に標本を取り出す状況を考えると

- $X_1, \dots, X_n$  は独立同分布,  $E[X_i] = \mu_X, V[X_i] = \sigma_X^2$
- $Y_1, \dots, Y_m$  は独立同分布,  $E[Y_j] = \mu_Y, V[Y_j] = \sigma_Y^2$
- $\{X_i\}$  と  $\{Y_j\}$  は互いに独立
- $H_0: \mu_X - \mu_Y = \Delta_0$  (通常  $\Delta_0 = 0$ )

標本平均  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \bar{Y} = \frac{1}{m} \sum_{j=1}^m Y_j$  について

$$E[\bar{X} - \bar{Y}] = \mu_X - \mu_Y \quad V[\bar{X} - \bar{Y}] = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m} \quad (20)$$

#### Definition 3.4. 2標本の母平均差の $z$ 検定

独立な二つの母集団  $X_1, \dots, X_n$  と  $Y_1, \dots, Y_m$  とする. 母平均を  $\mu_X, \mu_Y$  で, 既知分散 (または  $m, n$  が十分大きい) を仮定. 検定したい仮説:  $H_0: \mu_X - \mu_Y = \Delta_0$ . 検定統計量

$$Z = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \approx N(0, 1) \quad (21)$$

### 3.3 母分散の検定

#### Definition 3.5. $\Gamma$ 関数

$a \in \mathbb{R}_{>0}$  とすると

$$\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx \quad (22)$$

$$\bullet a = 1, \Gamma(1) = \int_0^{\infty} e^{-x} dx = 1$$

$$\bullet a \in \mathbb{N}$$

$$\Gamma(a+1) = \int_0^{\infty} x^a e^{-x} dx \quad (23)$$

$$= [-x^a e^{-x}]_0^{\infty} + a \int_0^{\infty} x^{a-1} e^{-x} dx \quad (24)$$

$$= a\Gamma(a) \quad (25)$$

$$\bullet \forall n \in \mathbb{N}, \Gamma(n) = (n-1)!$$

$$\bullet \Gamma\left(\frac{1}{2}\right) = 2 \int_0^{\infty} e^{-t^2} dt = \int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}$$

$$\bullet \forall a, b > 0, \int_0^{\infty} w^{a-1} e^{-bw} dw = b^{-a} \int_0^{\infty} u^{a-1} e^{-u} du = \frac{\Gamma(a)}{b^a}$$

#### Definition 3.6. $\chi^2$ 分布

自由度  $\nu > 0$  に対し、確率変数  $W$  が  $\chi^2$  分布  $\chi^2_{\nu}$  に従うとは、その確率密度関数が

$$f_{\nu}(w) = \begin{cases} \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} w^{\frac{\nu}{2}-1} e^{-\frac{w}{2}} & w > 0 \\ 0 & w \leq 0 \end{cases} \quad (26)$$

で与えられることをいう

$$\bullet \text{期待値 } E[W] = \nu$$

$$\bullet \text{分散 } V[W] = 2\nu$$

$$\bullet \text{確率母関数 } M_W(t) = E[e^{tW}] = (1-2t)^{-\frac{\nu}{2}} \quad \left(t < \frac{1}{2}\right)$$

**Proposition 3.1.**  $\nu \in \mathbb{N}$  とする. 独立な  $Z_1, \dots, Z_{\nu} \sim N(0, 1)$  に対し

$$W = \sum_{i=1}^{\nu} Z_i^2 \sim \chi^2_{\nu} \quad (27)$$

Proof.  $Z \sim N(0, 1)$  について

$$E[e^{tZ^2}] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(\frac{1}{2}-t)z^2} dz = (1-2t)^{-\frac{1}{2}} \quad \left(t < \frac{1}{2}\right) \quad (28)$$

独立性より

$$M_W(t) = \prod_{i=1}^{\nu} E[e^{tZ_i^2}] = \left(E[e^{tZ^2}]\right)^{\nu} = (1-2t)^{-\frac{\nu}{2}} \quad (29)$$

一方、 $\chi^2_{\nu}$  の確率母関数も  $(1-2t)^{-\frac{\nu}{2}}$  である. したがって、 $t = 0$  近傍で一致し、一意性より  $W \sim \chi^2_{\nu}$  である  $\square$

**Proposition 3.2.** 独立な  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$  のとき,  $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi_n^2$

Proof.  $Z_i := \frac{X_i - \mu}{\sigma}$  とすると,  $Z_i \sim N(0, 1)$  かつ  $\sum_{i=1}^n Z_i^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2$  で, 標準正規分布の二乗和の性質より,  $\sum_{i=1}^n Z_i^2 \sim \chi_n^2$  である □

### 3.4 母分散の検定 (母平均未知)

#### Theorem 3.1. Fisher-Cochran の定理

$$\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\sigma^2} \sim \chi_{n-1}^2 \quad (30)$$

さらに  $\bar{X}_n$  と  $\sum_{i=1}^n (X_i - \bar{X}_n)^2$  は独立である。これによって、母平均を使わずに母分散の検定が可能になる

#### Definition 3.7. 多次元正規分布

確率ベクトルを  $X = (X_1, \dots, X_n)$  とし、平均ベクトル  $\mu \in \mathbb{R}^n$ 、共分散行列  $\Sigma \in \mathbb{R}^{n \times n}$  は対称で狭義正定値とする。  $X$  が多次元正規分布  $N_n(\mu, \Sigma)$  に従うとは、次のような同時確率密度関数をもつことをいう

$$f(x) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det \Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \langle x - \mu, \Sigma^{-1}(x - \mu) \rangle\right) \quad (31)$$

さらに

- $E[X] = \mu = (\mu_1, \dots, \mu_n)$
- $\Sigma = (\Sigma_{ij})_{1 \leq i, j \leq n}$
- $\Sigma_{ij} = \text{Cov}(X_i, X_j)$
- $\{\text{Cov}(X_i, X_j) = 0, i \neq j\} \iff \{X_1, \dots, X_n \text{ は互いに独立}\}$
- $t \in \mathbb{R}^n, M_X(t) = E[e^{\langle t, X \rangle}] = \exp\left(\langle t, \mu \rangle + \frac{1}{2} \langle t, \Sigma t \rangle\right)$

#### Lemma 3.1. 2次元正規分布の線型写像の像

$X = {}^t(X_1, \dots, X_n) \sim N_n(m, \sigma^2 I_n)$  とする,  $Y = \sum_{i=1}^n b_i X_i, Z = \sum_{i=1}^n c_i X_i$  とおく。このとき,

$A = \begin{pmatrix} b_1 & b_2 & b_3 & \cdots & b_n \\ c_1 & c_2 & c_3 & \cdots & c_n \end{pmatrix}$  とし,  $\text{rank} A = 2$  も仮定する

- ${}^t(Y, Z) = AX$  は二次元正規分布に従う:  ${}^t(Y, Z) = N_2(Am, \sigma^2 A^t A)$
- $Y$  と  $Z$  が独立であるための必要十分条件は,  $\text{Cov}(Y, Z) = \sigma^2 \sum_{i=1}^n b_i c_i = \sigma^2 \langle b, c \rangle = 0$

Proof. •  $X = {}^t(X_1, X_2, \dots, X_n) \sim N_n(m, \sigma^2 I_n), {}^t(Y, Z) = AX$  で

$$M_{(Y, Z)}(t) = E[\exp(\langle t, {}^t(Y, Z) \rangle)] \quad (32)$$

$$= E[\exp(\langle t, AX \rangle)] \quad (33)$$

$$= E[\exp(\langle {}^t A t, X \rangle)] \quad (34)$$

$X \sim N_n(m, \sigma^2 I_n)$  の MGF を用いると

$$M_{(Y, Z)}(t) = \exp\left(\langle {}^t A t, m \rangle + \frac{1}{2} \sigma^2 \langle {}^t A t, {}^t A t \rangle\right) \quad (35)$$

$$= \exp\left(\langle t, Am \rangle + \frac{1}{2} \sigma^2 \langle t, A^t A t \rangle\right) \quad (36)$$

したがって,  ${}^t(Y, Z) \sim N_2(Am, \sigma^2 A^t A)$ . また,  $\forall u \in \mathbb{R}^2 \setminus \{0\}$

$${}^t u (\sigma^2 A^t A) u = \sigma^2 \|{}^t A u\|^2 > 0 \quad (37)$$

$$\text{Cov}(Y, Z) = \text{Cov}\left(\sum_{i=1}^n b_i X_i, \sum_{j=1}^n c_j X_j\right) \quad (38)$$

$$= \sum_{i=1}^n \sum_{j=1}^n b_i c_j \text{Cov}(X_i, X_j) \quad (39)$$

$$= \sigma^2 \sum_{i=1}^n b_i c_i = \sigma^2 \langle b, c \rangle \quad (40)$$

ここで共分散の双線形性と  $\text{Cov}(X_i, X_j) = \sigma^2 \delta_{ij}$  を使っている. 非退化な二次元正規分布では  $\text{Cov}(Y, Z) = 0 \iff Y, Z$  は独立であるから,  $\text{Cov}(Y, Z) = \sigma^2 \langle b, c \rangle = 0 \iff Y$  と  $Z$  が独立

□

### Definition 3.8. 共分散・双線形性

$E[|XY|] < \infty$  とする

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \quad (41)$$

双線形性:

$$\text{Cov}(aX_1 + bX_2, Y) = a\text{Cov}(X_1, Y) + b\text{Cov}(X_2, Y) \quad (42)$$

$$\text{Cov}(X, aY_1 + bY_2) = a\text{Cov}(X, Y_1) + b\text{Cov}(X, Y_2) \quad (43)$$

照明は第一式だけ:

$$\text{Cov}(aX_1 + bX_2, Y) = E[(aX_1 + bX_2 - E[aX_1 + bX_2])(Y - E[Y])] \quad (44)$$

$$= E[(a(X_1 - E[X_1]) + b(X_2 - E[X_2]))(Y - E[Y])] \quad (45)$$

$$= aE[(X_1 - E[X_1])(Y - E[Y])] + bE[(X_2 - E[X_2])(Y - E[Y])] \quad (46)$$

$$= a\text{Cov}(X_1, Y) + b\text{Cov}(X_2, Y) \quad (47)$$

### Lemma 3.2. $\chi^2$ の加法性

独立に  $U \sim \chi_{\nu_1}^2, V \sim \chi_{\nu_2}^2$  なら,  $U + V \sim \chi_{\nu_1 + \nu_2}^2$

Proof. 独立に  $U \sim \chi_{\nu_1}^2, V \sim \chi_{\nu_2}^2$  とする

$$M_{\chi_{\nu}^2}(t) = E[e^{tW}] = (1 - 2t)^{-\frac{\nu}{2}} \quad \left(t < \frac{1}{2}\right) \quad (48)$$

独立性より

$$M_{U+V}(t) = E[e^{t(U+V)}] = E[e^{tU}] E[e^{tV}] \quad (49)$$

$$= (1 - 2t)^{-\frac{\nu_1}{2}} (1 - 2t)^{-\frac{\nu_2}{2}} = (1 - 2t)^{-\frac{\nu_1 + \nu_2}{2}} \quad \left(t < \frac{1}{2}\right) \quad (50)$$

MGF の一意性より,  $U + V \sim \chi_{\nu_1 + \nu_2}^2$

□

Fisher-Cochran の定理の証明.

### Lemma 3.3.

$$\sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2 \sim \chi_{n-1}^2 \quad (51)$$

標準化して  $W_i = \frac{X_i - \mu}{\sigma}$  とおく. このとき  $\sum_{i=1}^n W_i^2 \sim \chi_n^2$

また,  $T_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0, 1) \implies T_n^2 \sim \chi_1^2$

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 + n(\bar{X}_n - \mu)^2 \quad (52)$$

より,  $\sum_{i=1}^n W_i^2 = \sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2 + T_n^2$ . ここで,  $U := \sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2$

$Y = \bar{X}_n = \sum_{k=1}^n \frac{1}{n} X_k$ ,  $Z_1 = X_1 - \bar{X}_n = \sum_{k=1}^n \left(\delta_{1k} - \frac{1}{n}\right) X_k$  とし

$$A := \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ 1 - \frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \end{pmatrix} \quad (53)$$

とおけば,  $(Y, Z_1) = AX$  であり, 明らかに  $\text{rank}A = 2$  となり, 共分散は

$$\text{Cov}(Y, Z_1) = \sigma^2 \sum_{k=1}^n \frac{1}{n} \left(\delta_{1k} - \frac{1}{n}\right) = \sigma^2 \left(\frac{1}{n} - \frac{1}{n}\right) = 0 \quad (54)$$

$\implies (Y, Z_1)$  は二次元正規で無相関  $\implies$  独立

同じようにして,  $Y$  と  $Z_i = X_i - \bar{X}_n$  は独立であるから, よって,  $Y = \bar{X}_n$  と  $U = \sum_{i=1}^n \frac{(X_i - \bar{X}_n)^2}{\sigma^2}$

が独立. よって,  $T_n^2 = \left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}\right)^2$  と  $U$  は独立. よって,  $U = \sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2 \sim \chi_{n-1}^2$

二乗和の等式の証明.

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n [(X_i - \bar{X}_n) + (\bar{X}_n - \mu)]^2 \quad (55)$$

$$= \sum_{i=1}^n (X_i - \bar{X}_n)^2 + 2(\bar{X}_n - \mu) \sum_{i=1}^n (X_i - \bar{X}_n) + n(\bar{X}_n - \mu)^2 \quad (56)$$

$$= \sum_{i=1}^n (X_i - \bar{X}_n)^2 + n(\bar{X}_n - \mu)^2 \quad (57)$$

□

□

### 3.5 $t$ 検定

#### Definition 3.9. $t$ 分布

自由度  $\nu > 0$  に対し, 確率変数  $T$  が  $t$  分布  $t_\nu$  に従うとは, その確率密度関数が次のように与えられることをいう

$$f_\nu(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad (58)$$

- 期待値  $E[T] = 0$  ( $\nu > 1$ )
- 分散  $V[T] = \frac{\nu}{\nu-2}$  ( $\nu > 2$ )
- 分位点: 右側  $\alpha$  点を  $t_{\nu,\alpha}$  と書く ( $P(T \geq t_{\nu,\alpha}) = \alpha$ ) 両側  $\alpha$  点は  $t_{\nu,\frac{\alpha}{2}}$  と書く

#### Theorem 3.2. 正規分布と $\chi^2$ の比による $t$ 分布

$Z \sim N(0, 1), U \sim \chi_\nu^2$  が独立ならば

$$T = \frac{Z}{\sqrt{\frac{U}{\nu}}} \sim t_\nu \quad (59)$$

#### Theorem 3.3.

独立に  $X_1, \dots, X_n$  が  $N(\mu, \sigma^2)$  に従うとする. このとき標本平均・不偏分散を

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (60)$$

とおく. このとき

$$T = \frac{\bar{X}_n - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1} \quad (61)$$

$X_1, \dots, X_n$  は独立同分布で, 正規  $N(\mu, \sigma^2)$  に従うとする. 母分散  $\sigma^2$  が未知, 仮説  $H_0: \mu = \mu_0$  (対立は両側・片側). 検定統計量

$$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (62)$$

$H_0$  の下で  $T \sim t_{n-1}$

- 両側  $\alpha$ :  $|T| \geq t_{n-1, \frac{\alpha}{2}}$  で棄却
- 右側  $\alpha$ :  $T \geq t_{n-1, \alpha}$  で棄却
- 左側  $\alpha$ :  $T \leq -t_{n-1, \alpha}$  で棄却

独立な確率変数  $X, Y$  を  $X \sim N(0, 1), Y \sim \chi_n^2$  とする. さらに  $T = \frac{X}{\sqrt{\frac{Y}{n}}}$  を定める. 確率

密度関数を  $P_T(t)$  と書く.  $\forall a < b, P(a \leq T \leq b) = \int_a^b P_T(t) dt$  を計算する

定義より,  $a \leq T \leq b \iff \sqrt{\frac{Y}{n}}a \leq X \leq \sqrt{\frac{Y}{n}}b$  で

$$D := \left\{ (x, y) : y > 0, \sqrt{\frac{y}{n}}a \leq x \leq \sqrt{\frac{y}{n}}b \right\} \quad (63)$$

を用いて

$$P(a \leq T \leq b) = \iint_D P_{X,Y}(x,y) dx dy \quad (64)$$

$$= \int_0^\infty \int_{\sqrt{\frac{y}{n}}a}^{\sqrt{\frac{y}{n}}b} P_{X,Y}(x,y) dx dy \quad (65)$$

独立性より  $P_{X,Y}(x,y) = P_X(x)P_Y(y)$

$$P_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, P_Y(y) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} \quad (y > 0) \text{ となり, 従って}$$

$$P(a \leq T \leq b) = \int_0^\infty \int_{\sqrt{\frac{y}{n}}a}^{\sqrt{\frac{y}{n}}b} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} dx dy \quad (66)$$

ここで,  $s = \sqrt{\frac{n}{y}}x$  とし,  $dx = \sqrt{\frac{y}{n}}ds$  になる

$$P(a \leq T \leq b) = \frac{1}{\sqrt{2\pi} 2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \int_0^\infty y^{\frac{n-1}{2}} e^{-\frac{y}{2}} \left( \int_a^b e^{-\frac{ys^2}{2n}} ds \right) dy \quad (67)$$

$$= \frac{1}{\sqrt{2\pi} 2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \int_a^b \left\{ \int_0^\infty y^{\frac{n-1}{2}} \exp\left(-\frac{1}{2} \left(1 + \frac{s^2}{n}\right) y\right) dy \right\} ds \quad (68)$$

$$z = \frac{1}{2} \left(1 + \frac{s^2}{n}\right) y \implies dy = \frac{2}{1 + \frac{s^2}{n}} dz$$

$$\int_0^\infty y^{\frac{n-1}{2}} \exp\left(-\frac{1}{2} \left(1 + \frac{s^2}{n}\right) y\right) dy = \left(\frac{2}{1 + \frac{s^2}{n}}\right)^{\frac{n+1}{2}} \int_0^\infty z^{\frac{n-1}{2}} e^{-z} dz \quad (69)$$

$$= \left(\frac{2}{1 + \frac{s^2}{n}}\right)^{\frac{n+1}{2}} \Gamma\left(\frac{n+1}{2}\right) \quad (70)$$

よって

$$P(a \leq T \leq b) = \int_a^b \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \left(1 + \frac{s^2}{n}\right)^{-\frac{n+1}{2}} ds \quad (71)$$

で, すなわち,  $T$  の確率密度関数は

$$P_T(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} = \frac{1}{\sqrt{n} B(\frac{n}{2}, \frac{1}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \quad (72)$$

すなわち,  $T \sim t_n$

### 3.6 母平均の差の検定 (母分散未知)

#### Definition 3.10.

独立かつ互いに独立に  $X_1, \dots, X_{n_1} \sim N(\mu_X, \sigma^2)$ ,  $Y_1, \dots, Y_{n_2} \sim N(\mu_Y, \sigma^2)$  とする. 標本平均と二乗和を

$$\bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i \quad \bar{Y} = \frac{1}{n_2} \sum_{j=1}^{n_2} Y_j \quad (73)$$

$$S_{XX} = \sum_{i=1}^{n_1} (X_i - \bar{X})^2 \quad S_{YY} = \sum_{j=1}^{n_2} (Y_j - \bar{Y})^2 \quad (74)$$

$$u_X^2 = \frac{S_{XX}}{n_1 - 1} \quad u_Y^2 = \frac{S_{YY}}{n_2 - 1} \quad (75)$$

とおく. 合計二乗和は

$$S = S_{XX} + S_{YY} = (n_1 - 1)u_X^2 + (n_2 - 1)u_Y^2 \quad (76)$$

#### Definition 3.11. 二標本の二乗和と $t$ 統計量

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim N(0, 1) \quad (77)$$

とする. Fisher-Cochran の定理より,  $\frac{S_{XX}}{\sigma^2} \sim \chi_{n_1-1}^2$ ,  $\frac{S_{YY}}{\sigma^2} \sim \chi_{n_2-1}^2$  かつ互いに独立. 従って, その和は

$$\frac{S}{\sigma^2} = \frac{S_{XX}}{\sigma^2} + \frac{S_{YY}}{\sigma^2} \sim \chi_{n_1+n_2-2}^2 \quad (78)$$

$t$  分布の定理より

$$t = \frac{Z}{\sqrt{\frac{S}{(n_1+n_2-2)\sigma^2}}} = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2}}{\sqrt{\left( \frac{1}{n_1} + \frac{1}{n_2} \right) \frac{S}{n_1+n_2-2}}} \sim t_{n_1+n_2-2} \quad (79)$$

## 4 適合度の検定・独立性の検定

1 回の試行の結果が  $k$  種類  $A_1, \dots, A_k$  のいずれかになるとする。これを独立に  $n$  回繰り返し、 $i$  番目のカテゴリの観測度数（実際に観測された回数）を  $X_i$  と書く。検定したい仮説（適合度の検定）：

$$\begin{aligned} H_0 &: \text{与えられた理論分布に等しい} \\ H_1 &: \text{与えられた理論分布と異なっている} \end{aligned}$$

ここで  $p_i$  を  $A_i$  が起こる確率とする。  $\sum_{i=1}^k p_i = 1$  とする。同じことを  $n$  回繰り返して、 $k$  個のカテゴリに分類され、その分類されたものの個数を数えるので二項分布の一般化が必要となる

### Theorem 4.1. 多項定理

$\forall k \geq 2, \forall n \in \mathbb{N}$  と変数  $y_1, \dots, y_k$  に対して

$$(y_1 + \dots + y_k)^n = \sum_{\substack{r_1, \dots, r_k \in \mathbb{Z}_{\geq 0} \\ r_1 + \dots + r_k = n}} \frac{n!}{r_1! \dots r_k!} y_1^{r_1} \dots y_k^{r_k} \quad (80)$$

- $\frac{n!}{r_1! \dots r_k!}$  :  $r_1$  個の  $y_1$ ,  $r_2$  個の  $y_2$ ,  $\dots$ ,  $r_k$  個の  $y_k$  の並べ方の総数を表す
- $k = 2$  で二項定理に一致
- 数学的帰納法で証明できる

### Definition 4.1. 多項分布の定義

$k, n \in \mathbb{N}$  と確率  $p_1, \dots, p_k$  を与える  $\left( p_i > 0, \sum_{i=1}^k p_i = 1 \right)$ .  $k$  次元確率変数  $X = (X_1, \dots, X_k)$  が多項分布に従うとは、その分布が次のように与えられることである：

$$r_1, \dots, r_k \in \{0, 1, 2, \dots\} \quad r_1 + \dots + r_k = n \quad (81)$$

$$P(X_1 = r_1, \dots, X_k = r_k) = \frac{n!}{r_1! \dots r_k!} p_1^{r_1} \dots p_k^{r_k} \quad (82)$$

- 二項分布の多数版であり、各試行において、 $k$  個の異なる結果があり、それを独立に  $n$  回繰り返したときに、 $k$  個の結果がそれぞれ起こる回数をあらかず分布である
- 多項定理により確率の総和は 1 である
- $E[X_i] = np_i$
- $V[X_i] = np_i(1 - p_i)$
- $\text{Cov}(X_i, X_j) = -np_i p_j \quad (i \neq j)$

### Definition 4.2. 観測度数と期待度数の表

観測度数（実際に観測された回数）を  $x_i$  とする。このとき、帰無仮説  $H_0$  の下での期待度数（ $H_0$  から計算される理論的な回数）は

$$E_i := np_i^{(0)} \quad (83)$$

次のような表を作る：

カテゴリ	$A_1$	$A_2$	$\dots$	$A_k$
観測度数（実測）	$x_1$	$x_2$	$\dots$	$x_k$
期待度数（理論； $H_0$ 下）	$E_1$	$E_2$	$\dots$	$E_k$

**Definition 4.3. 多項分布に基づく  $\chi^2$  統計量**

多項分布のもとで、次の量を統計量と定義する

$$\chi^2 = \sum_{i=1}^k \frac{(X_i - np_i)^2}{np_i} \tag{84}$$

$n$  が十分大きいとき、 $\chi^2$  は自由度  $k - 1$  の  $\chi^2$  に従うとみなしてよいことが知られている。  
 $k = 2$  の場合は、 $X_2 = n - X_1, p_2 = 1 - p_1$  であるから

$$\chi^2 = \frac{(X_1 - np_1)^2}{np_1} + \frac{(n - X_1 - n(1 - p_1))^2}{n(1 - p_1)} \tag{85}$$

$$= \frac{(X_1 - np_1)^2}{np_1(1 - p_1)} \tag{86}$$

ドモアブルプラス（または中心極限定理）より  $n$  が大きいときに右辺が正規分布の 2 乗で近似され、自由度 1 の  $\chi^2$  分布に従うことが直観的にわかる

**Definition 4.4. 棄却域と判断基準**

- 有意水準  $\alpha$  に対し、上側臨界値  $\chi_{k-1}^2(1 - \alpha)$  を用いると、 $P(\chi^2 \geq \chi_{k-1}^2(1 - \alpha)) = \alpha$
- 観測された値を  $x_i$  とするとき、 $\sum_{i=1}^k \frac{(x_i - np_i)^2}{np_i} \geq \chi_{k-1}^2(1 - \alpha)$  であれば、観測値と期待値の差が大きいと判断し、帰無仮説  $H_0$  を棄却する
- これまでの検定と異なり、検定量  $\chi^2$  が小さくても棄却しない。理由は、小さい場合は、 $x_i$  と  $x_{pi}$  が近いということであり、観測度数と期待度数が近いことを意味するからである

母集団の各要素が二つの属性  $A, B$  を持つとする。属性  $A$  と  $B$  に対応する事象族  $A_1, \dots, A_M \subset \Omega, B_1, \dots, B_N \subset \Omega$  を次のように仮定する：

$$A_i \cap A_{i'} = \emptyset \quad (i \neq i') \quad \bigcup_{i=1}^M A_i = \Omega \tag{87}$$

$$B_j \cap B_{j'} = \emptyset \quad (j \neq j') \quad \bigcup_{j=1}^N B_j = \Omega \tag{88}$$

このとき、各組合せ事象  $A_i \cap B_j$  の確率を  $p_{ij} = P(A_i \cap B_j), p_i = P(A_i), q_j = P(B_j)$  とおく。独立であるとは  $p_{ij} = p_i q_j$  が成り立つことである

**Definition 4.5. 観測度数表**

属性  $A$  (行) と属性  $B$  (列) によるクロス集計表 (分割表) を作る

	$B_1$	$B_2$	$B_3$	$B_4$	計
$A_1$	$s_{11}$	$s_{12}$	$s_{13}$	$s_{14}$	$s_{1\cdot}$
$A_2$	$s_{21}$	$s_{22}$	$s_{23}$	$s_{24}$	$s_{2\cdot}$
計	$s_{\cdot 1}$	$s_{\cdot 2}$	$s_{\cdot 3}$	$s_{\cdot 4}$	$n$

ここで、 $s_{ij}$  は「属性  $A_i$  かつ属性  $B_j$  に属する個体の数 (観測度数)」である  
 帰無仮説  $H_0$  : 属性  $A$  と  $B$  は独立であり、 $H_0 : p_{ij} = p_i q_j$ . このとき、 $A_i \cap B_j$  に属する期待度数は  $E_{ij} = np_i q_j$ . 推定値を用いると、 $\hat{E}_{ij} = n\hat{p}_i \hat{q}_j = \frac{R_i C_j}{n}$  である。観測度数  $s_{ij}$  と期待度数  $E_{ij}$  を用いて、検定統計量を次で定義する

$$\chi^2 = \sum_{i=1}^M \sum_{j=1}^N \frac{(s_{ij} - E_{ij})^2}{E_{ij}} \tag{89}$$

$n$  が十分大きいとき、この  $\chi^2$  は自由度  $\nu = (M - 1)(N - 1)$  の  $\chi^2$  分布に従うとみなしてよい

## 5 一元配置の分散分析

- 母集団： $k$  個の正規母集団 ( $k$  個の級 (クラス)) を  $\Pi_i : N(m_i, \sigma^2)$  とし、母分散は未知でもいいが、全ての母集団で等しいとする
- 各母集団からの標本： $X_{i1}, X_{i2}, \dots, X_{in_i}$
- 標本平均： $\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$
- 検定する仮説： $H : m_1 = m_2 = \dots = m_k = m$  で、つまり級が異なるとき、母集団に差があるかを調べたい

### Definition 5.1. $F$ 分布

$n_1, n_2 > 0$  とする. 確率変数  $X$  が自由度  $n_1, n_2$  の  $F$  分布に従うとは、その確率密度関数が次のように与えられることである

$$f_{n_1, n_2}(x) = \begin{cases} \frac{1}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1} \left(1 + \frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (90)$$

ここで、 $B(u, v) = \int_0^1 t^{u-1} (1-t)^{v-1} dt$  はベータ関数である. このとき、 $X \sim F_{n_1, n_2}$  と書く

- 期待値  $E[F] = \frac{n_2}{n_2 - 2}$  ( $n_2 > 2$ )
- 分散  $V[F] = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}$  ( $n_2 > 4$ )
- 分位点： $X \sim F_{n_1, n_2}$  とする.  $0 < \alpha < 1$  に対し、 $P(X \geq F_{n_1, n_2}(\alpha)) = \alpha$  をみたす実数  $F_{n_1, n_2}(\alpha)$  を自由度  $n_1, n_2$  の  $F$  分布の  $\alpha$  分位点 (上側  $\alpha$  点) という

### Theorem 5.1. $\chi^2$ の比による $F$ 分布

$\chi_1^2, \chi_2^2$  をそれぞれ自由度  $n_1, n_2$  の  $\chi^2$  分布に従う独立な確率変数とし、これらが独立であるとす. このとき、 $F = \frac{\chi_1^2/n_1}{\chi_2^2/n_2}$  は自由度  $n_1, n_2$  の  $F$  分布に従う (すなわち  $F \sim F_{n_1, n_2}$ )

### Theorem 5.2.

$X_1, \dots, X_{n_1}$  を正規母集団  $N(m_1, \sigma_1^2)$  からの無作為標本、 $Y_1, \dots, Y_{n_2}$  を正規母集団  $N(m_2, \sigma_2^2)$  からの無作為標本とする. 不偏分散

$$u_X^2 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (X_j - \bar{X})^2 \quad (91)$$

$$u_Y^2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (Y_j - \bar{Y})^2 \quad (92)$$

として、 $F = \frac{u_X^2/\sigma_1^2}{u_Y^2/\sigma_2^2}$  は自由度  $n_1 - 1, n_2 - 1$  の  $F$  分布に従う (すなわち  $F \sim F_{n_1-1, n_2-1}$ )

### Definition 5.2. 分散分析

仮説  $H$  が正しいとすると、 $X_{ij}, (i = 1, \dots, k, j = 1, \dots, n_i)$  はいずれも同一の正規母集団  $N(m, \sigma^2)$  からの無作為標本とみなすことができる. このとき正規分布と  $\chi^2$  分布の関係より

$$\frac{1}{\sigma^2} \sum_{i=1}^k \sum_{j=1}^{n_i} [X_{ij} - \bar{X}]^2 \sim \chi_{n-1}^2 \quad (93)$$

ただし,  $n = n_1 + n_2 + \dots + n_k$  かつ標本平均

$$\bar{X} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij} = \frac{1}{n} \sum_{i=1}^k n_i \bar{X}_i \quad (94)$$

**Definition 5.3.** 全変動・級間変動・級内変動

- 全変動 (全平方和) :  $S = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$
- 級間変動 (級間平方和) :  $S_1 = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2$
- 級内変動 (級内平方和) :  $S_2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$

ここで,  $\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$ ,  $\bar{X} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}$ ,  $n = \sum_{i=1}^k n_i$

Fisher-Cochran の定理より

$$\frac{S}{\sigma^2} = \sum_{i=1}^k \sum_{j=1}^{n_i} \left( \frac{X_{ij} - \bar{X}}{\sigma} \right)^2 \sim \chi_{n-1}^2 \quad (95)$$

同じく Fisher-Cochran より  $\sum_{j=1}^{n_i} \left( \frac{X_{ij} - \bar{X}}{\sigma} \right)^2$  は  $\chi_{n_i-1}^2$  に従うが, クラスが異なる分布は独立なので,  $\chi^2$  分布の性質より

$$\frac{S_2}{\sigma^2} = \sum_{i=1}^k \sum_{j=1}^{n_i} \left( \frac{X_{ij} - \bar{X}_i}{\sigma} \right)^2 \sim \chi_{n-k}^2 \quad (96)$$

$S = S_1 + S_2$  が成立するから,  $\frac{S}{\sigma^2} = \frac{S_1}{\sigma^2} + \frac{S_2}{\sigma^2}$  も成立する. Fisher-Cochran より,  $X_i - \bar{X}_i$  と  $\bar{X}_i$  は独立であり,  $\bar{X}$  も  $\bar{X}_i$  の一次結合で書けるから,  $X_i - \bar{X}_i$  と  $\bar{X} - \bar{X}_i$  も独立. よって,  $S_1$  と  $S_2$  は独立.  $\frac{S}{\sigma^2} \sim \chi_{n-1}^2$ ,  $\frac{S_2}{\sigma^2} \sim \chi_{n-k}^2$  より,  $\frac{S_1}{\sigma^2} \sim \chi_{k-1}^2$

**Definition 5.4.** 検定統計量と棄却規準

検定統計量  $F$  を  $F = \frac{S_1/(k-1)}{S_2/(n-k)}$  で定めると,  $F \sim F_{k-1, n-k}$  が成り立つ. 帰無仮説が正しくないとき (クラス間で差がある) とき, 級間変動  $S_1$  が大きくなるが  $S_2$  は変化しない. 従って,  $F$  の値が大きいつきに  $H_0$  を棄却するのが妥当である  
正規母集団  $N(m_i, \sigma^2)$  の母平均に関する帰無仮説  $H_0 : m_1 = m_2 = \dots = m_k$  の危険率  $\alpha$  の検定は, 次の判定による:

$$\begin{aligned} F &\geq F_{k-1, n-k}(\alpha) \text{ なら } H_0 \text{ を棄却} \\ F &< F_{k-1, n-k}(\alpha) \text{ なら } H_0 \text{ を採択} \end{aligned}$$

確率変数  $X$  と  $Y$  がそれぞれ自由度  $m, n$  の  $\chi^2$  分布に従うとする,  $X \sim \chi_m^2, Y \sim \chi_n^2$ . 確率変数  $Z = \frac{X/m}{Y/n}$  と定義する.  $Z$  の確率密度関数を  $P_Z(z)$  とすると

$$P(a \leq Z \leq b) = \int_a^b P_Z(z) dz \quad (97)$$

定義より,  $a \leq Z \leq b \iff \frac{1}{b} \frac{n}{m} X \leq Y \leq \frac{1}{a} \frac{n}{m} X$ . 従って,  $P(a \leq Z \leq b) = P\left(\frac{1}{b} \frac{n}{m} X \leq Y \leq \frac{1}{a} \frac{n}{m} X\right)$ .  
この領域を  $D$  とすると

$$P(a \leq Z \leq b) = \iint_D P_{X,Y}(x, y) dy dx \quad (98)$$

$X$  と  $Y$  は独立なので,  $P_{X,Y}(x, y) = P_X(x) P_Y(y)$

よって,  $P(a \leq Z \leq b) = \int_0^\infty \int_{\frac{1}{b} \frac{n}{m} x}^{\frac{1}{a} \frac{n}{m} x} P_X(x) P_Y(y) dy dx$ . さらに

$$P_X(x) = \frac{1}{2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)} x^{\frac{m}{2}-1} e^{-\frac{x}{2}} \quad (x > 0) \quad (99)$$

$$P_Y(y) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} \quad (y > 0) \quad (100)$$

すると

$$P(a \leq Z \leq b) = \frac{1}{2^{\frac{m+n}{2}} \Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \int_0^\infty \int_{\frac{1}{b} \frac{n}{m} x}^{\frac{1}{a} \frac{n}{m} x} x^{\frac{m}{2}-1} y^{\frac{n}{2}-1} e^{-\frac{x+y}{2}} dy dx \quad (101)$$

$s = \frac{1}{y} \frac{n}{m} x, y = \frac{n}{m} \frac{x}{s}, \frac{dy}{ds} = -\frac{n}{m} x s^{-2}$  とおくと

$$P(a \leq Z \leq b) = \frac{1}{2^{\frac{m+n}{2}} \Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \int_0^\infty \int_a^b x^{\frac{m}{2}-1} \left(\frac{n}{m} \frac{x}{s}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}} e^{-\frac{1}{2} \frac{n}{m} \frac{x}{s}} \frac{n}{m} x s^{-2} ds dx \quad (102)$$

$$= \frac{1}{2^{\frac{m+n}{2}} \Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \int_a^b s^{-\frac{n}{2}-1} \int_0^\infty x^{\frac{m+n}{2}-1} \exp\left(-\frac{1}{2} \left(1 + \frac{n}{ms}\right) x\right) dx ds \quad (103)$$

ここで,  $\int_0^\infty x^{\frac{m+n}{2}-1} \exp\left(-\frac{1}{2} \left(1 + \frac{n}{ms}\right) x\right) dx = \left(\frac{2}{1 + \frac{n}{ms}}\right)^{\frac{m+n}{2}} \Gamma\left(\frac{m+n}{2}\right)$  であるから

$$P(a \leq Z \leq b) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \left(\frac{n}{m}\right)^{\frac{n}{2}} \int_a^b s^{-\frac{n}{2}-1} \left(1 + \frac{n}{ms}\right)^{-\frac{m+n}{2}} ds \quad (104)$$

$$= \frac{1}{B\left(\frac{m}{2}, \frac{n}{2}\right)} \left(\frac{n}{m}\right)^{\frac{n}{2}} \int_a^b s^{-\frac{n}{2}-1} \left(1 + \frac{n}{ms}\right)^{-\frac{m+n}{2}} ds \quad (105)$$

$$= \frac{1}{B\left(\frac{m}{2}, \frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} \int_a^b s^{-\frac{n}{2}-1} \left(\frac{n}{ms}\right)^{-\frac{m+n}{2}} \left(1 + \frac{m}{ns}\right)^{-\frac{m+n}{2}} ds \quad (106)$$

$$= \int_a^b \frac{1}{B\left(\frac{m}{2}, \frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} s^{\frac{m}{2}-1} \left(1 + \frac{m}{ns}\right)^{-\frac{m+n}{2}} ds \quad (107)$$

## 6 第一種誤り・第二種誤り・検定力

### Definition 6.1. 第一種誤り・第二種誤り

1. 第一種誤り： $H_0$  が真であるのに棄却してしまう誤り
2. 第二種誤り： $H_0$  が偽であるのに棄却しない誤り
3.  $P(\text{第一種誤り}) = P(\text{棄却} | H_0 \text{が真}) = \alpha$
4.  $P(\text{第二種誤り}) = \beta$  で、 $1 - \beta$  を検定力とよぶ

標本  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$  で、 $\sigma^2$  は既知とする。帰無仮説と対立仮説をそれぞれ  $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$  とする。標本平均  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  で、 $Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$  とする。 $H_0$  のもとで  $Z \sim N(0, 1)$  である。有意水準  $\alpha$  の両側検定の棄却域を考えると、 $|Z| \geq z(\alpha)$

$$P(\text{第一種誤り}) = P(\text{棄却} | \mu = \mu_0) \quad (108)$$

$$= P(|Z| \geq z(\alpha) | \mu = \mu_0) = \alpha \quad (109)$$

真の母平均を  $\mu$  とする（対立仮説が正しいとき）： $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ 。従って

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(\delta, 1) \quad \delta = \frac{\mu - \mu_0}{\frac{\sigma}{\sqrt{n}}} \quad (110)$$

$$P(\text{第二種誤り}) = \beta(\mu) = P(\text{棄却しない} | H_1 \text{が正しい}) \quad (111)$$

$$= P(|Z| < z(\alpha) | Z \sim N(\delta, 1)) \quad (112)$$

$$= \int_{-z(\alpha)}^{z(\alpha)} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z - \delta)^2}{2}\right) dz \quad (113)$$

$$= \int_{-z(\alpha) - \delta}^{z(\alpha) - \delta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \quad (114)$$

検出力；

$$\pi(\mu) = 1 - \beta(\mu) = 1 - \int_{-z(\alpha) - \delta}^{z(\alpha) - \delta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \quad (115)$$

棄却域（両側検定）： $|Z| \geq z(\alpha)$ 、 $P(Z > z(\alpha)) = \frac{\alpha}{2}$  で、第二種誤り：

$$\beta(\mu) = \int_{-z(\alpha) - \delta}^{z(\alpha) - \delta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \quad (116)$$

で、検出力：

$$\pi(\mu) = 1 - \beta(\mu) = 1 - \int_{-z(\alpha) - \delta}^{z(\alpha) - \delta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \quad (117)$$

ここで  $\alpha$  が小さくすると  $z(\alpha)$  が大きくなる。区間  $[-z(\alpha) - \delta, z(\alpha) - \delta]$  が広がるため、 $\beta(\mu) \uparrow, \pi(\mu) \downarrow$ 。すなわち有意水準を厳しくすると検出力が下がる。 $|\delta| = \frac{|\mu - \mu_0|}{\sigma} \sqrt{n}$  なので

- 標本サイズ  $n$  を大きくすると,  $\beta(\mu) \downarrow, \pi(\mu) \uparrow$
- 母平均のずれ  $|\mu - \mu_0|$  が大きくすると,  $\beta(\mu) \downarrow, \pi(\mu) \uparrow$

母平均の検定:  $H_0: \mu = \mu_0, H_1: \mu > \mu_0$  で,  $Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ ,  $H_0$  のもとで  $Z \sim N(0, 1)$  である. 棄却域 (上側片側):  $Z \geq z(2\alpha), P(Z \geq z(2\alpha)) = \alpha$

$$P(\text{第一種誤り}) = P(Z \geq z(2\alpha) | \mu = \mu_0) = \alpha \quad (118)$$

真の平均を  $\mu > \mu_0$  とし,  $\delta = \frac{\mu - \mu_0}{\frac{\sigma}{\sqrt{n}}}, Z \sim N(\delta, 1)$  とする

$$P(\text{第二種誤り}) = \beta(\mu) = P(Z < z(2\alpha) | Z \sim N(\delta, 1)) \quad (119)$$

$$= \Phi(z(2\alpha) - \delta) \quad (120)$$

$$\pi(\mu) = 1 - \beta(\mu) = 1 - \Phi(z(2\alpha) - \delta) \quad (121)$$

ただし,  $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

## 7 回帰分析

与えられたデータ  $(x_1, y_1), \dots, (x_n, y_n)$  を  $x - y$  平面上の点とみなす

- これらの点に最もよく適合する直線  $y = ax + b$  を求め、分析することを単回帰分析という
- 各データ点  $(x_i, y_i)$  に対し、直線上の値は  $\alpha x_i + \beta$  であり、観測値  $y_i$  との差  $y_i - (\alpha x_i + \beta)$  が全ての  $i$  について小さくなるような係数  $\alpha, \beta$  を探ることが目標である
- 直線  $y = \alpha x + \beta$  がデータ  $(x_1, y_1), \dots, (x_n, y_n)$  に最もよく適合するように、 $\alpha, \beta$  の関数  $l(\alpha, \beta) = \sum_{j=1}^n (y_j - \alpha x_j - \beta)^2$  を考える。このようにして、 $l(\alpha, \beta)$  を最小とする  $\alpha, \beta$  を求める方法を最小二乗法という。最小化する  $\alpha, \beta$  を最小二乗推定量とよぶ

$l(\alpha, \beta)$  を  $\alpha, \beta$  で偏微分して 0 とおくと

$$\frac{\partial l}{\partial \beta} = -2 \sum_{j=1}^n (y_j - \alpha x_j - \beta) = 0 \quad (122)$$

$$\frac{\partial l}{\partial \alpha} = -2 \sum_{j=1}^n x_j (y_j - \alpha x_j - \beta) x_j = 0 \quad (123)$$

整理すると

$$\sum_{j=1}^n y_j - \alpha \sum_{j=1}^n x_j - n\beta = 0 \quad (124)$$

$$\sum_{j=1}^n x_j y_j - \alpha \sum_{j=1}^n x_j^2 - \beta \sum_{j=1}^n x_j = 0 \quad (125)$$

この連立方程式を解くと

$$\alpha = \frac{n \sum_{j=1}^n x_j y_j - \left( \sum_{j=1}^n x_j \right) \left( \sum_{j=1}^n y_j \right)}{n \sum_{j=1}^n x_j^2 - \left( \sum_{j=1}^n x_j \right)^2} \quad (126)$$

$$= \frac{\frac{1}{n} \sum_{j=1}^n x_j y_j - \left( \frac{1}{n} \sum_{j=1}^n x_j \right) \left( \frac{1}{n} \sum_{j=1}^n y_j \right)}{\frac{1}{n} \sum_{j=1}^n x_j^2 - \left( \frac{1}{n} \sum_{j=1}^n x_j \right)^2} \quad (127)$$

標本平均を  $\bar{X} = \frac{1}{n} \sum_{j=1}^n x_j, \bar{Y} = \frac{1}{n} \sum_{j=1}^n y_j$  とおき，分散と共分散を

$$S_{xx} = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2 \quad (128)$$

$$S_{yy} = \frac{1}{n} \sum_{j=1}^n (y_j - \bar{y})^2 \quad (129)$$

$$S_{xy} = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y}) \quad (130)$$

と定める．展開すると

$$S_{xx} = \frac{1}{n} \sum_{j=1}^n x_j^2 - \bar{x}^2 \quad (131)$$

$$S_{xy} = \frac{1}{n} \sum_{j=1}^n x_j y_j - \bar{x} \bar{y} \quad (132)$$

よって， $\alpha = \frac{S_{xy}}{S_{xx}}$  となる．一方，正規方程式の一つ目から  $\beta = \bar{y} - \alpha \bar{x} = \bar{y} - \frac{S_{xy}}{S_{xx}} \bar{x}$

#### Definition 7.1. 決定係数・残差

最小二乗法で求めた係数  $\alpha, \beta$  を用いて， $\hat{y}_j = \alpha x_j + \beta$  とおく．予測値の平均は

$$\bar{\hat{y}} = \frac{1}{n} \sum_{j=1}^n \hat{y}_j = \frac{1}{n} \sum_{j=1}^n (\alpha x_j + \beta) = \alpha \bar{x} + \beta = \bar{y} \quad (133)$$

となる．予測値の分散を次のように定める：

$$S_{pp} = \frac{1}{n} \sum_{j=1}^n (\hat{y}_j - \bar{y})^2 \quad (134)$$

決定係数  $R^2$  を  $R^2 = \frac{S_{pp}}{S_{yy}}$  で定める

残差  $e_j = y_j - \hat{y}_j$  とおく，残差の平均は

$$\bar{e} = \frac{1}{n} \sum_{j=1}^n e_j = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j) = \bar{y} - \alpha \bar{x} - \beta = 0 \quad (135)$$

となる．残差の分散を  $S_{ee} = \frac{1}{n} \sum_{j=1}^n e_j^2$  とおくと， $S_{yy} = S_{pp} + S_{ee}$ ．従って決定係数  $R^2 = \frac{S_{pp}}{S_{yy}}$

は  $R^2 = \frac{S_{pp}}{S_{pp} + S_{ee}}$  と書ける． $S_{pp} \geq 0, S_{ee} \geq 0$  なので， $0 \leq R^2 \leq 1$  である． $R^2$  が 1 に近いとき，残差の分散  $S_{ee}$  が小さく，回帰直線が  $y = \alpha x + \beta$  がデータによく適合していることを意味する

## 8 重回帰分析

与えられたデータ  $(x_{1j}, x_{2j}, \dots, x_{pj}, y_j)$  を観測したとする。これらの観測に最もよく適合する式  $y = \theta_1 x_1 + \dots + \theta_p x_p + \theta_0$  を求め、分析することを重回帰分析という。各データ点  $(x_{1j}, x_{2j}, \dots, x_{pj}, y_j)$  に対し、式の値は  $\theta_1 x_{1j} + \dots + \theta_p x_{pj} + \theta_0$  であり、観測値との差  $y_j - (\theta_1 x_{1j} + \dots + \theta_p x_{pj} + \theta_0)$  が全ての  $j$  について小さくなるような係数  $\theta_0, \theta_1, \dots, \theta_p, \theta_0$  を探すことが目標である。 $x$  と  $y$  の観測値ベクトルをそれぞれ

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \mathbf{x}_j = \begin{pmatrix} x_{j1} \\ \vdots \\ x_{jp} \end{pmatrix} \quad (136)$$

とおく。説明変数の値をまとめた行列（説明行列）を

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \quad (137)$$

とおく。各説明変数の標本平均を並べたベクトルを

$$\bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix} \quad \bar{x}_k = \frac{1}{n} \sum_{j=1}^n x_{jk} \quad (138)$$

とおく。つまり、 $\bar{\mathbf{x}} = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j$ 。説明変数の標本平均ベクトル

$$\bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{pmatrix} \quad (139)$$

を用いて、説明変数の標本分散行列（標本共分散行列）を

$$S_{xx} = \frac{1}{n} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}}) (\mathbf{x}_j - \bar{\mathbf{x}})^T \quad (140)$$

と定める。 $S_{xx}$  は  $p \times p$  対称行列である。 $y$  の標本平均を  $\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j$  とおく。説明変数と目的変数の標本共分散ベクトルを

$$S_{xy} = \frac{1}{n} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}}) (y_j - \bar{y}) \quad (141)$$

と定める。係数ベクトルは

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_p \end{pmatrix} \quad (142)$$

このとき、求めたい式は

$$y = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_p x_p + \theta_0 \quad (143)$$

各観測  $j$  について観測値との差は  $y_j - (\mathbf{x}_j^T \theta + \theta_0)$  と書ける. 説明行列を使うと  $\mathbf{y} - X\theta + \theta_0 \mathbf{1}$  とも書ける

二乗誤差を  $l(\theta, \theta_0) = \|\mathbf{y} - (X\theta + \theta_0 \mathbf{1})\|^2 = \sum_{j=1}^n (y_j - (\theta^T \mathbf{x}_j + \theta_0))^2$  と定める. ここで,  $\|\mathbf{v}\|^2 = \mathbf{v}^T \mathbf{v}$  である

### Definition 8.1.

中心化した観測値ベクトルと説明変数ベクトルをそれぞれ  $\mathbf{y}^*$  を  $\mathbf{y}^* = \mathbf{y} - \bar{y} \mathbf{1}$ ,  $\mathbf{x}_j^* = \mathbf{x}_j - \bar{\mathbf{x}}$  と定め, 説明行列は次のように定める

$$X^* = \begin{pmatrix} \mathbf{x}_1^{*T} \\ \vdots \\ \mathbf{x}_n^{*T} \end{pmatrix} = X - \mathbf{1} \bar{\mathbf{x}}^T \quad (144)$$

中心化の定義より,  $\mathbf{y} = \mathbf{y}^* + \bar{y} \mathbf{1}$ ,  $X = X^* + \mathbf{1} \bar{\mathbf{x}}^T$  が成り立つ. 従って, 二乗誤差は

$$l(\theta, \theta_0) = \|\mathbf{y}^* - X^* \theta + (\bar{y} - \bar{\mathbf{x}}^T \theta - \theta_0) \mathbf{1}\|^2 \quad (145)$$

となる. ノルムの展開より

$$l(\theta, \theta_0) = \|\mathbf{y}^* - X^* \theta\|^2 + 2(\mathbf{y}^* - X^* \theta)^T (\bar{y} - \bar{\mathbf{x}}^T \theta - \theta_0) \mathbf{1} + \|(\bar{y} - \bar{\mathbf{x}}^T \theta - \theta_0) \mathbf{1}\|^2 \quad (146)$$

$\mathbf{1}^T (\mathbf{y}^* - X^* \theta) = \mathbf{1}^T \mathbf{y}^* - \mathbf{1}^T X^* \theta$  となる.  $\mathbf{1}^T \mathbf{y}^* = \sum_{j=1}^n (y_j - \bar{y}) = \sum_{j=1}^n y_j - n\bar{y} = 0$  から,  $\forall 1 \leq k \leq p$  に対して

$$(\mathbf{1}^T X^*)_k = \sum_{j=1}^n x_{jk}^* = \sum_{j=1}^n (x_{jk} - \bar{x}_k) = 0 \quad (147)$$

従って

$$\mathbf{1}^T X^* = (0 \ \cdots \ 0) \quad \mathbf{1}^T X^* \theta = (0 \ \cdots \ 0) \theta = 0 \quad (148)$$

よって,  $\mathbf{1}^T (\mathbf{y}^* - X^* \theta) = 0$  が成り立つ. 直交性より

$$l(\theta, \theta_0) = \|\mathbf{y}^* - X^* \theta\|^2 + \|(\bar{y} - \bar{\mathbf{x}}^T \theta - \theta_0) \mathbf{1}\|^2 \quad (149)$$

$$= \|\mathbf{y}^* - X^* \theta\|^2 + n (\bar{y} - \bar{\mathbf{x}}^T \theta - \theta_0)^2 \quad (150)$$

$l_0(\theta) = \|\mathbf{y}^* - X^* \theta\|^2$  とおく.  $\theta$  を固定して  $l(\theta, \theta_0)$  を最小にする  $\theta_0$  を考えると, 右側の第2項  $\theta_0$  に関する二次関数であり, 明らかに最小値0を持つ. その最小値を与える  $\theta_0$  は  $\theta_0 = \bar{y} - \bar{\mathbf{x}}^T \theta$  である. 従って  $\forall \theta_0$  に対して  $l(\theta, \theta_0) \geq l_0(\theta)$  が成り立ち, 等号成立は  $\theta_0 = \bar{y} - \bar{\mathbf{x}}^T \theta$  のときである

•  $A$  を  $n \times n$  行列,  $x \in \mathbb{R}^n, b \in \mathbb{R}^n$  とする

$$- \nabla_x (b^T x) = \nabla_x (x^T b) = b$$

$$- \nabla_x \|Ax\|^2 = 2A^T Ax$$

$l_0$  を展開すると

$$l_0(\theta) = (X^*\theta - \mathbf{y}^*)(X^*\theta - \mathbf{y}^*)^T \quad (151)$$

$$= \theta^T (X^* X^{*T}) \theta - 2\mathbf{y}^{*T} X^* \theta + \mathbf{y}^{*T} \mathbf{y}^* \quad (152)$$

すると,  $\nabla_{\theta} l_0(\theta) = 2(X^* X^{*T})^T \theta - 2(X^*)^T \mathbf{y}^* = 2\left((X^*)^T X^* \theta - (X^*)^T \mathbf{y}^*\right)$

従って停留点は  $\nabla_{\theta} l_0(\theta) = 0 \iff (X^*)^T X^* \theta = (X^*)^T \mathbf{y}^*$  をみたす  $\theta$  である. これを  $\theta$  の正規方程式とよぶ

もう一度  $\theta$  で微分すると, ヘッセ行列は  $\nabla_{\theta}^2 l_0(\theta) = 2X^{*T} X^* = 2nS_{xx}$  となる.  $\forall v \in \mathbb{R}^p$

$$v^T \nabla_{\theta}^2 l_0(\theta) v = 2v^T X^{*T} X^* v = 2(X^* v)^T (X^* v) = 2\|X^* v\|^2 \geq 0 \quad (153)$$

よって,  $\nabla_{\theta}^2 l_0(\theta)$  は非負定値行列である