4セメ　応用確率統計学（後半）

配付資料

担当：　林　俊介（テキスト原作：井上　亮）
Email:　s_hayashi@plan.civil.tohoku.ac.jp
講義ウェブサイト: http://www.plan.civil.tohoku.ac.jp/opt/hayashi/lectures/AppPrbStt/
（この講義資料は，上記ウェブサイトよりダウンロードして印刷持参すること.）

目次

参考図書

・東京大学教養学部統計学教室編 (1991) 「統計学入門」東京大学出版会
・松原望 (2007) 「入門　統計解析」東京図書
・蓑谷千凰彦 (1997) 「計量経済学」多賀出版
・渡辺澄夫・村田昇 (2005)「確率と統計　ー情報学への架橋ー」コロナ社
・豊田秀樹(訳) (2006)「数理統計学ハンドブック」朝倉書店

Ⅲ．統計的推論

１．標本と標本分布

(1)　母集団(population)と標本(sample)

いま，ある母集団の性質を知りたいとする．

例) 母集団：日本国内の全世帯の所得　　　　関心のある性質：平均・分散・分布

一つの方法は，全数調査 [(例) 国勢調査]に基づく方法．(←記述統計)．

もう一つの方法は，標本抽出に基づく方法．(←統計的推論)

ここでは，標本抽出を通して，母集団の性質を探る方法について説明する．

まず，標本抽出を通して母集団の性質を知る上では，以下の 2 種類の問題があることに注意．
興味の対象である母集団の確率変数 X に関して，
(1) そもそも $f(x)$（：確率変数 X が従う確率密度関数）が不明．
⇒大標本なら，中心極限定理を用いて議論できる
(2) $f(x)$の形状は分かっている（例えば，正規分布に従う）が，その分布の形状を定める母数 θ が不明（例えば，$N(\mu, \sigma^2)$の μ, σ^2 が不明)．
⇒ X は $f(x; \theta)$に従うとし，未知の母数 θ を推定する．

標本抽出方法には，以下の二通りが考えられる．

・復元抽出
（抽出した標本を母集団に戻して，標本抽出を繰り返す方法．同じ標本が複数回選ばれる可能性あり）
・非復元抽出
（抽出した標本は母集団に戻さず，標本抽出を繰り返す方法．同じ標本が複数回選ばれることはない）

望ましい標本抽出方法は，母集団に関する情報を得られる無作為標本(random sample)を作ること．

無作為標本
確率変数 $X_1, \ldots, X_n$ が iid　(独立同分布に従う: independent and identically distributed)
⇒$X_1, \ldots, X_n$ は確率変数 X のもとで，無作為標本を構成．
無作為標本の同時確率密度関数　$f_{X_1,\ldots X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} f(x_i)$　$\because X_1,\ldots,X_n$が独立

非復元抽出では，2 番目以降の標本は「母集団から既に選ばれた標本を除外した」条件の下で抽出された標本になり，得られた標本は $X_1, \ldots, X_n$ は独立にならないため適切な標本抽出方法ではない．

(2)　母数と統計量

統計量：「標本データの要約」．

n 個の標本 $X_1, \ldots, X_n$ が確率変数 X の分布から得られた標本を構成しているとき，
あらゆる標本の関数 $T = T(X_1, \ldots, X_n)$ は「統計量」
（例）標本平均，標本分散，最大値，最小値，中央値，第 1 四分位値，…

母集団の分布を表す未知の母数 θ に関する情報を得ることができる統計量が重要．

統計量 T が持つべき重要な二つの統計的性質

$E(T) = \theta$　　　T の期待値は母数 θ と一致する．⇒「T は θ の不偏推定量」

$T \xrightarrow{p} \theta$ 　$\left(\text{すべての}\varepsilon > 0\text{に対して}\ \lim_{n\to\infty} P\left(\left|T_n - \theta\right| < \varepsilon\right) = 1\right)$ 　n:標本数

標本数が大きくなるにつれて，統計量 T が θ に確率的に収束する．

⇒「T は θ の一致推定量」

それでは，代表的な統計量である「標本平均」・「標本分散」について以下に性質を示す．

a）標本平均

X は未知の平均 μ を持つ確率変数であるとし，$X_1, \ldots, X_n$ を X の分布から得られた無作為標本とする．

標本平均 $\bar{X}$ は，以下の式で表される．

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

標本平均 $\bar{X}$ の期待値 $E(\bar{X})$は

$$E\left(\bar{X}\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) = \mu$$

となるので，標本平均 $\bar{X}$ は母集団の平均 μ の「不偏推定量」であることが確認できた．

また，大数の法則より，

$$\bar{X} \xrightarrow{p} \mu$$ 　($\bar{X}$ は μ に確率的に収束する)

となるので，標本平均 $\bar{X}$ は母集団の平均 μ の「一致推定量」であるといえる．

cf.) 大数の(弱)法則

確率変数 $X_1, \ldots, X_n$ は，共通の平均 μ，分散 σ^2 に従い，独立に分布とする(iid).

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ とすると，　$\bar{X}_n \xrightarrow{p} \mu$ 　($\bar{X}$ は μ に確率収束する)

すべての $\varepsilon > 0$ に対して $\lim_{n\to\infty} P\left(\left|\bar{X}_n - \mu\right| \geq \varepsilon\right) = 0$ あるいは同様に $\lim_{n\to\infty} P\left(\left|\bar{X}_n - \mu\right| < \varepsilon\right) = 1$

(参考) $X \sim N(\mu, \sigma^2)$に従う場合，

標本平均 $\bar{X}$ は平均 μ，分散 σ^2/n の正規分布 $\bar{X} \sim N\left(\mu, \sigma^2/n\right)$ に従う．

すなわち，$\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N\left(0,1\right)$ 　(平均 0，分散 1 の標準正規分布)

b）標本分散

標本分散 s^2 は，以下の式で表される．（分母が n ではないことに注意!!）

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

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) \qquad \because \sum_{i=1}^{n} X_i = n\bar{X}$$

標本分散 s^2 の期待値 $E(s^2)$は

$$E\left(s^2\right)=\frac{1}{n-1}\left(\sum_{i=1}^{n}E\left(X_i^2\right)-nE\left(\bar{X}^2\right)\right)$$

$$=\frac{1}{n-1}\left(n\left(\sigma^2+\mu^2\right)-n\left(\frac{\sigma^2}{n}+\mu^2\right)\right)$$

$$\because\quad E\left(X_i^2\right)=E\left[\left(X_i-\mu\right)^2+2\mu X_i-\mu^2\right]=E\left[\left(X_i-\mu\right)^2\right]+\mu^2=\sigma^2+\mu^2$$

$$\because\quad E\left(\bar{X}^2\right)=E\left[\left(\bar{X}-\mu\right)^2\right]+\mu^2=E\left[\left(\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right)^2\right]+\mu^2=\frac{1}{n^2}E\left[\sum_{i=1}^{n}\left(X_i-\mu\right)^2\right]+\mu^2=\frac{n\sigma^2}{n^2}+\mu^2$$

$$=\sigma^2$$

となり，標本分散 s^2 は母集団の未知の分散 σ^2 の「不偏推定量」である.

なお，s^2 は標本不偏分散，あるいは，不偏分散と呼ばれることもある.

なお，この講義の範囲を超えるため詳細は省くが，標本分散 s^2 の分散は $Var\left(s^2\right)=\dfrac{2\sigma^4}{n-1}$ となる.

チェビシェフの不等式を用いると，

$$P\left(\left|s^2-E\left(s^2\right)\right|\geq\varepsilon\right)\leq\frac{Var\left(s^2\right)}{\varepsilon^2}$$

となるので，

$$P\left(\left|s^2-\sigma^2\right|\geq\varepsilon\right)\leq\frac{2\sigma^4}{\left(n-1\right)\varepsilon^2}$$

となり，n が大きくなると右辺は 0 に近づくため，標本分散 s^2 は母集団の未知の分散 σ^2 の「一致推定量」であることが示される.

cf.) チェビシェフの不等式

X が期待値 μ，有限の分散 σ^2 の確率変数であるとすると，任意の $\varepsilon>0$ について，

$$P\left(\left|X-\mu\right|\geq\varepsilon\right)\leq\frac{\sigma^2}{\varepsilon^2}$$ （あるいは $P\left(\left|X-\mu\right|\geq k\sigma\right)\leq\dfrac{1}{k^2}$ ）

が成立する.

なお，$S^2=\dfrac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ （こちらを標本分散と呼ぶ場合もある）で計算を行うと，$E\left(S^2\right)=\dfrac{n-1}{n}\sigma^2$ となり，母集団の分散を過小評価してしまうので注意.

$\left(X_1-\bar{X}\right)+\left(X_2-\bar{X}\right)+\cdots+\left(X_n-\bar{X}\right)=0$ の制約より, 自由に動ける変数の数は, $(n-1)$個. この値を自由度と呼ぶ.

なお，母集団の平均 μ が既知の下で，母集団の分散を推定する場合には，自由度は n となるので，n で割った値が不偏推定量となる.（確認すること）

２．点推定と信頼区間推定

前節では，標本から母集団の平均・分散を推定する方法として，標本平均・標本分散について紹介した．標本平均・標本分散は，母集団の平均・分散に関する情報をある値（点）として与える（点推定）．この標本平均・標本分散は，それぞれある確率分布に従っている確率変数であるが，点推定の結果からはそのばらつきについては示していない．

そこで，標本平均・標本分散の確率分布をもとに，信頼区間（例えば，「95%信頼区間」なら，推定したい母集団の母数が 95%の確率で入っている区間）を推定する方法について紹介する．

(1)　大標本の場合

大標本の場合には，中心極限定理より，統計量の分布は正規分布に従うこととして議論することができる．

未知の母数 θ を推定するため，統計量 $T = T(X_1, \ldots, X_n)$を用いて推定する場合を考える．

ここで，母数の真値（実際には誰も知らない値）を θ_0 と書く．統計量 T が母数の真値 θ_0 を平均，σ_T^2/n を分散とする正規分布に分布収束（標本数が多いと確率分布が収束する←確率収束より弱い）すると仮定する．

$$T \xrightarrow{D} N\left(\theta_0, \sigma_T^2/n\right)$$

通常，σ_T^2 は未知だが，ここではまず σ_T^2 が既知であると仮定する．そのとき，次式の Z は平均 0，分散 1 の標準正規分布に分布収束する．

$$Z = \frac{\sqrt{n}(T-\theta_0)}{\sigma_T} \xrightarrow{D} N(0,1)$$

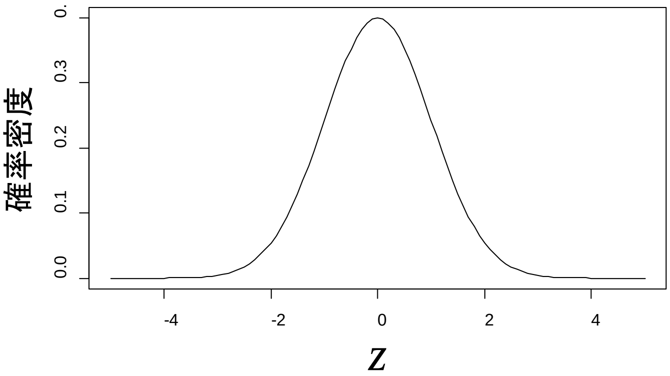


図　標準正規分布 $N(0,1)$の確率密度関数

標準正規分布では，z の値が-1.96～1.96 の範囲に入る確率が約 0.95 であるので

$$\begin{aligned}
&P(-1.96 < Z < 1.96) \\
&= P\left(-1.96 < \frac{\sqrt{n}(T-\theta_0)}{\sigma_T} < 1.96\right) \\
&= P\left(T - 1.96\frac{\sigma_T}{\sqrt{n}} < \theta_0 < T + 1.96\frac{\sigma_T}{\sqrt{n}}\right)
\end{aligned}$$

の確率は約 95%である．

標本抽出の結果，実際に得られた標本の値 $x_1, \ldots, x_n$ から $t = T(x_1, \ldots, x_n)$を算出すると，θ_0 の点推定 t の 95%信頼区間は

$$\left(t - 1.96\frac{\sigma_T}{\sqrt{n}}, t + 1.96\frac{\sigma_T}{\sqrt{n}}\right)$$

と得られる．

一方，σ_T^2 が未知の場合，s_T を σ_T の一致推定量（s_T は σ_T に確率収束する）とすると

$$\frac{\sqrt{n}(T-\theta_0)}{s_T} \xrightarrow{D} N(0,1)$$

と同じく書けるので，θ_0 の点推定 t の 95%信頼区間は

$$\left(t - 1.96\frac{s_T}{\sqrt{n}}, t + 1.96\frac{s_T}{\sqrt{n}}\right)$$

と表すことができる．

○ 母集団の平均 μ の信頼区間

$X_1, \ldots, X_n$ は未知の平均 μ，分散 σ^2 に従う X からの無作為標本とする．

標本平均を $\overline{X}$，標本分散を s^2 とすると，大標本の場合 $\sqrt{n}\left(\overline{X}-\mu\right)/s \xrightarrow{D} N(0,1)$ となるので，

母集団 X の平均 μ の 95%信頼区間は $\left(\overline{x}-1.96\cdot s/\sqrt{n}, \overline{x}+1.96\cdot s/\sqrt{n}\right)$ と表される．

ただし，$\overline{x}$：標本の実現値 $x_1, \ldots, x_n$ から計算した平均

より一般的に書けば，信頼係数 1－α の信頼区間は

$$\left(t-z_{\alpha/2}\cdot s_t/\sqrt{n}, t+z_{\alpha/2}\cdot s_t/\sqrt{n}\right)$$

ただし $z_{\alpha/2}$：標準正規分布の上側 α/2 分位数.

と表される.

信頼係数 大 (α 小) ⇒ 信頼区間　広

標本数 大 ($\sqrt{n}$ 大) ⇒ 信頼区間　狭

分散 小　(σ_T 小) ⇒ 信頼区間　狭

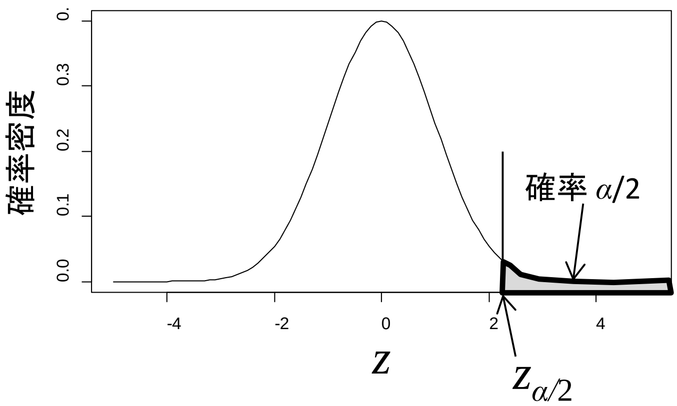


図　標準正規分布の上側 α/2 分位数

(2)　正規分布からの（小）標本の場合

$X_1, \ldots, X_n$ は，未知の平均 μ，分散 σ^2 の正規分布(N(μ, σ^2))に従う X からの無作為標本とする．この無作為標本を用いて，未知の平均 μ の信頼区間推定を行う方法について説明する．

a) 母集団の分散 σ^2 が既知の場合

通常，母集団の分散 σ^2 は未知だが，既知の場合には，標本平均 $\overline{X}$ は前節の議論から以下の通り．

標本平均：$\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$

標本平均は正規分布に従うため，

$$Z=\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$$

から，大標本の場合と同様に，信頼区間推定を行うことができる．

b) 母集団の分散 σ^2 が未知の場合

母集団の分散 σ^2 が未知の場合，標本分散は前節の議論から以下の通り．

標本分散：$s^2=\dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2 \qquad E\left(s^2\right)=\sigma^2$

さて，母集団の未知分散 σ^2 の代わりに標本分散 s^2 を利用すると

$$T=\frac{\overline{X}-\mu}{\sqrt{s^2/n}}=\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}}\Bigg/\sqrt{\frac{(n-1)s^2}{\sigma^2}\Big/(n-1)}$$

が得られる.

ここでまず，χ^2 分布 (カイ二乗分布)に注目する.

χ^2 分布

$Z_1,\ldots,Z_k$ は標準正規分布に従う確率変数であるとする.

$$\left(Z_1,\ldots,Z_k \sim N(0,1)\right)$$

このとき，$\chi^2 = Z_1^2 + \ldots + Z_k^2$ は自由度 k の χ^2 分布に従う.

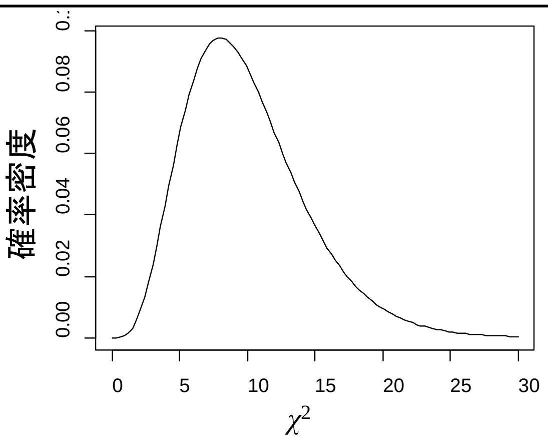


図　自由度 10 の χ^2 分布

$(n-1)\, s^2/\sigma^2$ は下式の展開により

$$\begin{aligned}
\frac{(n-1)s^2}{\sigma^2} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i - \mu\right)^2 - \frac{n}{\sigma^2}\left(\bar{X} - \mu\right)^2 \qquad \because \sum_{i=1}^{n} X_i = n\bar{X} \\
&= \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 - \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2 \\
&\sim \chi^2(n-1) \qquad \because \frac{X_i - \mu}{\sigma} \sim N(0,1),\ \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)
\end{aligned}$$

$n-1$ 個の標準正規分布に従う確率変数の二乗和として記述でき，$\chi^2 = (n-1)s^2/\sigma^2$ は自由度 $n-1$ の χ^2 分布に従うことが分かる.

次に，t 分布に着目する.

t 分布

確率変数

$$\begin{aligned}
&Z \sim N(0,1) \\
&Y \sim \chi^2(k) \quad \Rightarrow \quad t = \frac{Z}{\sqrt{Y/k}} \sim t(k) \\
&Z, Y \text{が独立}
\end{aligned}$$

自由度 k の t 分布

自由度が大きくなると，
t 分布は標準正規分布に近づく.

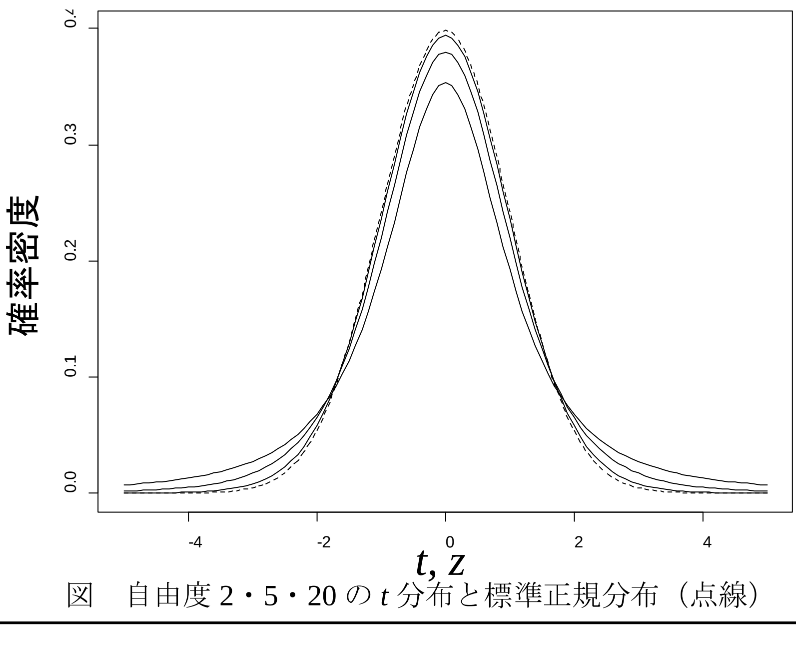


図　自由度 2・5・20 の t 分布と標準正規分布（点線）

すなわち，

$$T=\frac{\bar{X}-\mu}{\sqrt{s^2/n}}=\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}\bigg/\sqrt{\frac{(n-1)s^2}{\sigma^2}\bigg/(n-1)}$$

$$\sim t(n-1) \quad \because \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1),\ \ \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

となり，$\left(\bar{X}-\mu\right)\Big/\sqrt{s^2/n}$ は自由度 n－1 の t 分布に従うことが分かる．

t 分布の確率密度関数から，自由度 n－1 の t 分布の上側 α/2 分位数 $t_{\alpha/2,n-1}$ を求めれば，母集団の平均 μ の信頼係数 1－α の信頼区間は，標本の実現値 $x_1, \ldots, x_n$ から計算した平均を $\bar{x}$ とすると

$$\left(\bar{x}-t_{\alpha/2,n-1}\cdot s\big/\sqrt{n},\ \bar{x}+t_{\alpha/2,n-1}\cdot s\big/\sqrt{n}\right)$$

と求められる．

なお，自由度が小さい t 分布と標準正規分布を比べると，t 分布の方が中心部の確率密度が小さく，裾が厚い分布をしている．すなわち，母集団の分散 σ^2 を推定する必要が無い場合に比べて，分散を推定する場合には信頼区間が広がり，得られた標本平均 $\bar{X}$ のばらつきが大きくなることを示す．

(3)　二標本問題

二つの母集団の差に関心がある場合（例えば，薬の効き目を調べたい）の区間推定について説明する．

いま，二種類の確率変数 X, Y に関心があるとする．平均をそれぞれ μ_X, μ_Y，分散を σ_X^2, σ_Y^2 とし，それぞれの確率変数からの無作為標本を $X_1,\ldots,X_{n_X}$，$Y_1,\ldots,Y_{n_Y}$ 標本平均を $\bar{X},\bar{Y}$ とする. 関心の対象は, $\Delta=\mu_X-\mu_Y$ である. ここで, 標本平均の差を，$\hat{\Delta}=\bar{X}-\bar{Y}$ とする．

まず，標本平均の差の期待値 $E\left(\hat{\Delta}\right)$ は

$$E\left(\hat{\Delta}\right)=E\left(\bar{X}\right)-E\left(\bar{Y}\right)=\mu_X-\mu_Y=\Delta$$

となるため，$\hat{\Delta}$ は二つの母集団の平均値の差の不偏推定量であることが分かる．

次に，$\hat{\Delta}$ の漸近分布による Δ の大標本信頼区間を知りたい. 標本サイズを $n=n_X+n_Y$ とし，$n_X/n=\lambda_X$，$n_Y/n=\lambda_Y$ とおく．中心極限定理より

$$\sqrt{n_X}\left(\bar{X}-\mu_X\right)\xrightarrow{D}N\left(0,\sigma_X^2\right)$$

$$\sqrt{n_Y}\left(\bar{X}-\mu_Y\right)\xrightarrow{D}N\left(0,\sigma_Y^2\right)$$

となる．

> cf.) 中心極限定理
>
> 確率変数 $X_1, \ldots, X_n$ は，平均 μ，分散 σ^2 をもつ分布に従うとき
>
> $$Y_n=\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}=\frac{\sqrt{n}\left(\bar{X}_n-\mu\right)}{\sigma}$$ は平均 0，分散 1 の正規分布に分布収束（$Y_n\xrightarrow{D}N(0,1)$）する．

さらに

$$\sqrt{n}\left(\bar{X}-\mu_X\right)=\sqrt{\frac{n}{n_X}}\sqrt{n_X}\left(\bar{X}-\mu_X\right)=\sqrt{\frac{1}{\lambda_X}}\sqrt{n_X}\left(\bar{X}-\mu_X\right)\xrightarrow{D}N\left(0,\frac{\sigma_X^2}{\lambda_X}\right)$$
$$\sqrt{n}\left(\bar{Y}-\mu_Y\right)=\sqrt{\frac{n}{n_Y}}\sqrt{n_Y}\left(\bar{Y}-\mu_Y\right)=\sqrt{\frac{1}{\lambda_Y}}\sqrt{n_Y}\left(\bar{Y}-\mu_Y\right)\xrightarrow{D}N\left(0,\frac{\sigma_Y^2}{\lambda_Y}\right)$$

となり，標本が互いに独立なら，

$$\sqrt{n}\left[\left(\bar{X}-\bar{Y}\right)-\left(\mu_X-\mu_Y\right)\right]\xrightarrow{D}N\left(0,\frac{\sigma_X^2}{\lambda_X}+\frac{\sigma_Y^2}{\lambda_Y}\right)$$

$$\because A\sim N\left(\mu_A,\sigma_A^2\right),B\sim N\left(\mu_B,\sigma_B^2\right),A\text{と}B\text{が独立}\Rightarrow A-B\sim N\left(\mu_A-\mu_B,\sigma_A^2+\sigma_B^2\right)$$

$$E\left(A-B\right)=E\left(A\right)-E\left(B\right)=\mu_A-\mu_B$$
$$\begin{aligned}Var\left(A-B\right)&=E\left[\left\{\left(A-B\right)-E\left(A-B\right)\right\}^2\right]\\&=E\left[\left\{\left(A-E\left(A\right)\right)-\left(B-E\left(B\right)\right)\right\}^2\right]\\&=E\left[\left(A-E\left(A\right)\right)^2-2\left(A-E\left(A\right)\right)\left(B-E\left(B\right)\right)+\left(B-E\left(B\right)\right)^2\right]\\&=Var(A)-2Cov(A,B)+Var(B)\\&=\sigma_A^2+\sigma_B^2\end{aligned}$$

となる．そこで，近似的な信頼係数 1－α の信頼区間は，

$$\left(\left(\bar{x}-\bar{y}\right)-z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}},\left(\bar{x}-\bar{y}\right)+z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}\right)$$

となる．ただし，確率変数 X, Y が正規分布に従うなら，近似ではない．

もし X, Y の分散は未知だが等しいことが分かっており，正規分布に従う場合，

$$\bar{X}\sim N\left(\mu_X,\frac{\sigma^2}{n_X}\right),\ \bar{Y}\sim N\left(\mu_Y,\frac{\sigma^2}{n_Y}\right)\quad\Rightarrow\quad\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_X-\mu_Y\right)}{\sigma\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}}\sim N\left(0,1\right)$$

となる．ここで，s_p^2 を合併した分散(pooled variance)の推定量とすると，

$$s_p^2=\frac{\left(n_X-1\right)s_X^2+\left(n_Y-1\right)s_Y^2}{n_X+n_Y-2}\qquad E\left(s_p^2\right)=\sigma^2$$

より，

$$t=\frac{\left(\left(\bar{X}-\bar{Y}\right)-\left(\mu_X-\mu_Y\right)\right)\Big/\sigma\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}}{\sqrt{\frac{(n-2)s_p^2}{\sigma^2}\Big/\left(n-2\right)}}=\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_X-\mu_Y\right)}{s_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}}\sim t\left(n-2\right)$$

となるため，自由度 $n-2$ の t 分布を用いて信頼区間推定が可能である．

なお，母集団の分散が未知で，等しいとは限らない場合には，正確な分布を知ることはできず，近似的な信頼区間しか推定することができない．

３．最尤推定

(1)　最尤法による推定

最尤推定は，「最尤原理(現実に得られた標本は，確率が最大のものが実現した)」という仮定に基づく推定.

尤度：標本が得られる確率.

最尤推定量：尤度が最大となる母数の値.

(例)赤玉・白玉が入った袋から標本抽出を行い，赤玉の割合 p（母数）(ただし $0 \leqq p \leqq 1$)を推定したい.
・1 回目の標本抽出では「赤玉」⇒尤度は p ⇒尤度を最大化する p の値（最尤推定量）は 1.
・2 回目の標本抽出では「白玉」⇒尤度は $p(1-p)$ ⇒p の最尤推定量は 1/2.
・3 回目の標本抽出では「白玉」⇒尤度は $p(1-p)^2$ ⇒p の最尤推定量は 1/3.

$X_1, \ldots, X_n$ が未知の母数ベクトル $\boldsymbol{\theta}$ に基づく確率密度関数 $f(x; \boldsymbol{\theta})$に従う確率変数だとする．（**「確率密度関数の形は既知」が前提**であることに注意．例えば，「正規分布に従っている」ことが事前に分かっている．）

尤度関数(Likelihood function) $L(\boldsymbol{\theta})$は，確率変数 $X_1, \ldots, X_n$ から得られた標本の値 $\mathbf{x} = (x_1, \ldots, x_n)'$ を用いると

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}; \mathbf{x}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta})$$

と書ける．尤度関数は，確率分布の母数が $\boldsymbol{\theta}$ の時，標本 $x_1, \ldots, x_n$ が観測される確率を表す.

最尤原理では，この尤度関数 $L(\boldsymbol{\theta})$を最大化する母数 $\boldsymbol{\theta}$ が求めるべき推定値 $\hat{\boldsymbol{\theta}}$ であると考える．そこで，尤度関数を $\boldsymbol{\theta}$ に関して最大化する.

通常，尤度関数と一対一対応する対数尤度関数

$$l(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}; \mathbf{x}) = \sum_{i=1}^{n} \ln f(x_i; \boldsymbol{\theta})$$

を用いて，$\dfrac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$ となる $\hat{\boldsymbol{\theta}}$ を(点)推定する．最尤法による母数の推定値を「最尤推定量」と呼ぶ.

(例) 母集団の分布が正規分布の場合

母集団の分布が正規分布であることを仮定する場合，求めたい母集団の未知母数は，平均 μ と分散 σ^2 なので，尤度関数は $L(\mu, \sigma^2)$で表される．対数尤度関数は

$$l(\mu, \sigma^2) = \ln L(\mu, \sigma^2) = \sum_{i=1}^{n} \ln\left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x_i - \mu)^2}{2\sigma^2}\right)\right]$$

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

と書ける.

ここでまず，この対数尤度関数 $l(\mu, \sigma^2)$を最大化する平均の最尤推定量 $\hat{\mu}$ を求める.

$\hat{\mu}$ は $l(\mu, \sigma^2)$を μ で偏微分した時に 0 となる値なので

$$\frac{\partial}{\partial\mu}l\left(\mu,\sigma^2\right)=-\frac{\partial}{\partial\mu}\left\{\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right\}=\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)$$
$$=\frac{1}{\sigma^2}\left(\sum_{i=1}^{n}x_i-n\mu\right)$$

より，$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i$ が得られる．（標本平均と同じ）

次に，分散の最尤推定量 $\hat{\sigma}^2$ を求める．同様に，対数尤度関数 $l(\mu,\sigma^2)$を σ^2 で偏微分し

$$\frac{\partial}{\partial\sigma^2}l\left(\mu,\sigma^2\right)=\frac{\partial}{\partial\sigma^2}\left\{-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right\}$$
$$=-\frac{n}{2\sigma^2}+\frac{1}{2\left(\sigma^2\right)^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2$$

より $\frac{n}{2\hat{\sigma}^2}=\frac{1}{2\left(\hat{\sigma}^2\right)^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\Rightarrow\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\mu\right)^2$ となる．平均 μ が未知の場合には，平均の最尤推定量 $\hat{\mu}$ を用いて $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\hat{\mu}\right)^2$ が得られる．(標本分散 s^2 とは異なり，不偏性のない推定量であることに注意！）ただし，一致性，漸近正規性（(2) 参照）は有する．

(2)　最尤推定量の性質

最尤推定量 $\hat{\boldsymbol{\theta}}$ は，以下の二つの性質を持つ.

・一致性　　最尤推定量 $\hat{\boldsymbol{\theta}}$ は $\boldsymbol{\theta}$ に確率収束する．

$$\hat{\boldsymbol{\theta}}\xrightarrow{p}\boldsymbol{\theta}\quad\left(\text{すべての}\boldsymbol{\varepsilon}>0\text{に対して}\ \lim_{n\to\infty}P\left(\left|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}\right|<\boldsymbol{\varepsilon}\right)=1\right)$$
n:標本数

標本数が大きくなるにつれて，統計量 $\hat{\boldsymbol{\theta}}$ が $\boldsymbol{\theta}$ に確率的に収束する．

・漸近正規性　　最尤推定量 $\hat{\boldsymbol{\theta}}$ は正規分布に分布収束する．

$$\hat{\boldsymbol{\theta}}\xrightarrow{D}N\left(\boldsymbol{\theta},\mathbf{I}\left(\boldsymbol{\theta}\right)^{-1}\right)$$

ただし $\mathbf{I}\left(\boldsymbol{\theta}\right)$ ：情報行列

$$\mathbf{I}\left(\boldsymbol{\theta}\right)=E\left[\left(\frac{\partial l}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial l}{\partial\boldsymbol{\theta}}\right)'\right]=-E\left[\frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]=-E\begin{bmatrix}\frac{\partial^2 l}{\partial\theta_1^2} & \cdots & \frac{\partial^2 l}{\partial\theta_1\partial\theta_p}\\ \vdots & \ddots & \vdots\\ \frac{\partial^2 l}{\partial\theta_p\partial\theta_1} & \cdots & \frac{\partial^2 l}{\partial\theta_p^2}\end{bmatrix}$$

４．仮説検定

(1)　仮説検定とは

ある仮説を立てて，その「仮説が正しい」と仮定した時に，観測された標本が得られる確率を求め，立てた仮説の真偽を検定する方法.

仮説検定を実行するためにはまず，仮説を統計的に検証可能な形で立てる必要がある.

(例)　コインを 100 回投げたら, 表が 60 回出た(標本抽出). このとき，コインには歪みがあり（細工がされてあり），表が出る確率 p は 1/2 よりも大きいか？それとも，このコインは歪みが無く（表が出る確率 p は 1/2 であり），たまたま 60 回表が出ただけなのか？

そこで，まず仮説「p = 1/2」（帰無仮説）とそれに対立する仮説「p >1/2」（対立仮説）を立てる．そして，仮説「p = 1/2」の下で，標本抽出結果が得られる確率を計算する．もしこの確率が有意水準を下回れば，帰無仮説が棄却され，検証したい仮説（対立仮説）「p >1/2」を採択する.

実際，帰無仮説「p =1/2」の下で 100 回中 60 回表が出る確率は，二項分布の式より，$_{100}C_{60}\,\frac{1}{2}^{60}\left(1-\frac{1}{2}\right)^{40}=0.0108$ となり，確率 1%程度の希少な現象だと言える．もし有意水準 5%で仮説検定を行うと，仮説「p =1/2」は棄却され，仮説「p >1/2」が採択される.

仮説検定とは，確率変数 X が確率密度関数 $f(x;\theta)$に従っている時，X の分布から得られる標本 $X_1, \dots, X_n$ から推定された母数 θ が，母数の標本空間の部分集合 ω_1，ω_0 のいずれに含まれるかを検定すること.

H$_0$: $\theta \in \omega_0$ (帰無仮説, null hypothesis),　H$_1$: $\theta \in \omega_1$ (対立仮説, alternative hypothesis)

検定を行う際には，棄却域 C を設定し，標本が棄却域に入っているかを調べる.

$\left(X_1,\dots,X_n\right)\in C \Rightarrow$ H$_0$ を棄却，H$_1$ を採択する　　$\left(X_1,\dots,X_n\right)\in C^C \Rightarrow$ H$_0$ を保留する

しかし，下表の 2 種類の誤った検定を行う可能性が存在する.

	H$_0$ が真	H$_1$ が真
H$_0$ を棄却	第 1 種の過誤 (Type I error) 誤った対立仮説 H$_1$ を正しいと判断する	○
H$_0$ を保留	○	第 2 種の過誤 (Type II error) 正しい対立仮説 H$_1$ を誤っていると判断する

望ましいのは，誤りの確率を最小化することだが，第 1 種・第 2 種の過誤はトレードオフの関係にあるため，双方を同時に減らすことはできない.

通常は，第 1 種の過誤（誤った対立仮説 H$_1$ を正しいと判断）を抑える棄却域を選択する.（分析者が設定した誤った仮説を採択することを抑える.）

標本が棄却域に入っている$\left(\left(X_1,\dots,X_n\right)\in C\right)$が，母数は帰無仮説を満たす$\left(\theta\in\omega_0\right)$確率（すなわち，第 1 種の過誤を生じる確率）を危険率(有意水準)と呼び，α と記述する.

(2)　仮説検定の例　平均・分散の検定

a)　平均の大標本検定

X は，未知の平均 μ，分散 σ^2 のなんらかの確率分布に従う確率変数であるとする．いま X の平均が，ある値 μ_0 より大きいかを，X の確率分布に従う標本を用いて検定したい．

$\mathrm{H}_0: \mu = \mu_0$,　　$\mathrm{H}_1: \mu > \mu_0$

直観的推測では，標本平均 $\overline{X}$ が μ_0 を大きく上回っていると H_0 を棄却，H_1 を採択すると予想される．すなわち，

$P\left(\overline{X} \geq \mu_0 + h\right) = \alpha$ となるある値 h を求めればよい．

大標本の場合は，中心極限定理を用いて，標本平均 $\overline{X}$ が正規分布に従っていると仮定することができる．標本分散を s^2 とすると

$$Z = \frac{\overline{X} - \mu}{s/\sqrt{n}} \xrightarrow{D} N(0,1)$$

と表せるので，

$$\frac{\overline{X} - \mu_0}{s/\sqrt{n}} \geq z_\alpha$$

ただし，z_α は標準正規分布の上側 α 分位数　$\left(P\left(Z \geq z_\alpha\right) = \alpha\right)$

ならば H_0 を棄却，H_1 を採択することにより，近似的に危険率 α の仮説検定ができる．

b)　正規分布に従う確率変数に関する平均の検定

X は未知の平均 μ，分散 σ^2 の正規分布に従う確率変数であるとする．先ほどと同じ仮説の検定を行いたい．

$\mathrm{H}_0: \mu = \mu_0$,　　$\mathrm{H}_1: \mu > \mu_0$

信頼区間推定の議論より，統計量 T は自由度 n - 1 の t 分布に従う．

$$T = \frac{\overline{X} - \mu}{\sqrt{s^2/n}} \sim t(n-1)$$

したがって，

$$\frac{\overline{X} - \mu_0}{\sqrt{s^2/n}} \geq t_\alpha(n-1)$$

ただし，$t_\alpha(n-1)$ は自由度 $n-1$ の t 分布の上側 α 分位数　$\left(P\left(T \geq t_\alpha(n-1)\right) = \alpha\right)$

なら，H_0 を棄却，H_1 を採択する．なお，$t_\alpha(n-1) > z_\alpha$ より t 分布の方が安全側，$n \to \infty$ で収束する．

c)　母分散 σ^2 の検定

確率変数 X の分散 σ^2 が，ある値 σ^2_0 よりも大きいかを，X の確率分布に従う標本を用いて検定したい．

$\mathrm{H}_0: \sigma^2 = \sigma^2_0$,　　$\mathrm{H}_1: \sigma^2 > \sigma^2_0$

信頼区間推定の際に示したカイ二乗分布の議論より，

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

に従うことから，

$$\frac{(n-1)s^2}{\sigma_0^2} \geq c_\alpha^2(n-1)$$

$c_\alpha^2(n-1)$ は自由度 n–1 のカイ二乗分布の上側 α 分位数 $\left(P\left(C^2 \geq c_\alpha^2(n-1)\right) = \alpha\right)$

なら，H_0 を棄却，H_1 を採択する．

d) 片側検定と両側検定

平均の大標本検定を例にすると，対立仮説 H_1 が $\mu > \mu_0$（あるいは $\mu < \mu_0$）では，帰無仮説 H_0 を $\mu = \mu_0$ と設定し，危険率 α について $P(Z \geq z_\alpha) = \alpha$（あるいは $P(Z \leq -z_\alpha) = \alpha$）となる z_α に対して検定を行えば良い（片側検定）.

しかし，対立仮説 H_1 が $\mu \neq \mu_0$ の場合には，帰無仮説 H_0 は $\mu = \mu_0$ と同じであるが，検定は $\mu > \mu_0$ および $\mu < \mu_0$ の双方の場合に対して検定を行なう必要がある（両側検定）.

(例) 平均の大標本検定

片側検定　H_0: $\mu = \mu_0$　H_1: $\mu > \mu_0 \Rightarrow \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}} \geq z_\alpha$ なら H_0 を棄却　　ただし，$P(Z \geq z_\alpha) = \alpha$

両側検定　H_0: $\mu = \mu_0$　H_1: $\mu \neq \mu_0 \Rightarrow \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}} \geq z_{\alpha/2}$ あるいは $\dfrac{\bar{X} - \mu_0}{s/\sqrt{n}} \leq -z_{\alpha/2}$ なら H_0 を棄却

ただし，$P(Z \geq z_{\alpha/2}) = \dfrac{\alpha}{2}$

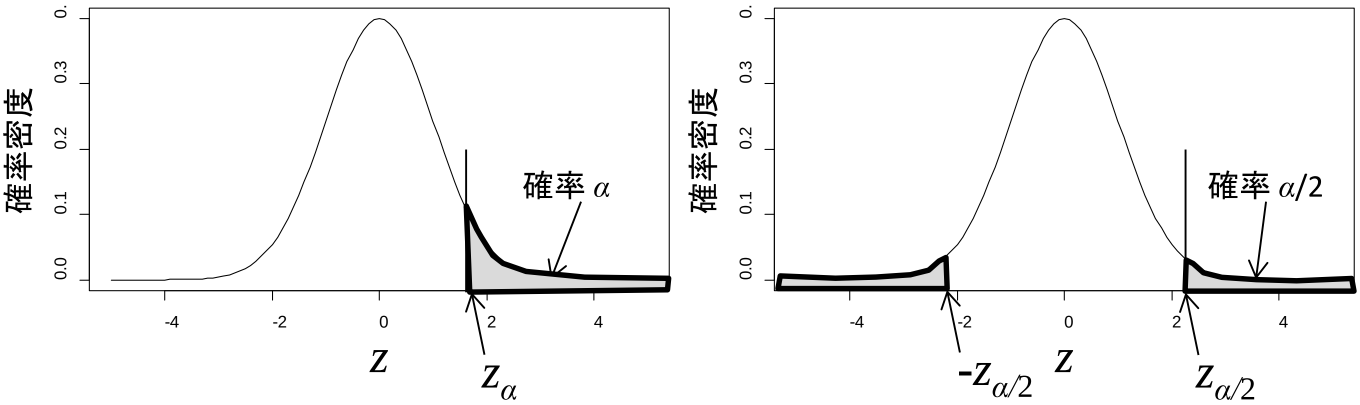


図　(a) 片側検定　　　　　(b) 両側検定

(3)　適合度検定

ある確率分布を仮定しそこから生じる理論的な値と，標本抽出で得られた値とを比較し，

H_0: 仮定した確率分布に適合している　H_1: 仮定した確率分布に適合していない

とした仮説検定を行う方法.

(例) サイコロを 60 回投げたところ，各目が下表に示す回数出た.
このサイコロは全ての目が等確率で出るサイコロと言えるか？

目	1	2	3	4	5	6
回数	13	19	11	8	5	4

この仮説検定では，

H_0: 全ての目が等確率(1/6)で出る.　H_1: 各目が出る確率は同じではない.

という帰無仮説・対立仮説を設定した上で，カイ二乗検定を行う.

なぜ，カイ二乗検定を使うか？

確率変数 X_1 が確率 p_1，試行回数 n の二項分布 $Bi(n, p_1)$に従うとき，

$n \to \infty$なら，
$$Y=\frac{X_1-np_1}{\sqrt{np_1\left(1-p_1\right)}}\sim N\left(0,1\right)\Rightarrow Y^2\sim\chi^2\left(1\right)$$

まず始めに単純な場合として，生起する背反な事象が二つ（例　コインの表裏）の場合を考える．総試行回数を n とし，表が出る回数を X_1，確率を p_1，裏が出る回数を X_2 ($=n - X_2$)，確率を p_2 ($=1 - p_1$)とする．

ここで，$Q_1=Y^2$ と記述すると

$$\begin{aligned}Q_1&=\frac{\left(X_1-np_1\right)^2}{np_1\left(1-p_1\right)}\\&=\frac{\left(X_1-np_1\right)^2}{np_1}+\frac{\left(X_1-np_1\right)^2}{n\left(1-p_1\right)}\\&=\frac{\left(X_1-np_1\right)^2}{np_1}+\frac{\left(n-X_2-n+np_2\right)^2}{np_2}\\&=\frac{\left(X_1-np_1\right)^2}{np_1}+\frac{\left(X_2-np_2\right)^2}{np_2}\end{aligned}$$

となり，$n \to \infty$なら $Q_1\sim\chi^2\left(1\right)$ となることが分かる．

さて，k 個の背反な事象が生起する（例　サイコロの目なら $k = 6$）場合について考える．各事象が起こる回数 $X_1,\ldots,X_k$ は試行回数 n と生起確率 $p_1,\ldots,p_k$ を母数とする多項分布（二項分布を 3 以上の事象が起こる場合に一般化した分布，i 番目の事象に関する確率変数 X_i は $Bi(n, p_i)$）に従うとする．

ただし，$X_k=n-\left(X_1+\cdots+X_{k-1}\right),\ p_k=1-\left(p_1+\cdots+p_{k-1}\right)$ である．

$Q_{k-1}=\sum_{i=1}^{k}\frac{\left(X_i-np_i\right)^2}{np_i}$ とすると，$n \to \infty$なら $Q_{k-1}\sim\chi^2\left(k-1\right)$ となる．

ここで，H$_0$: $p_1=p_{1,0},p_2=p_{2,0},\ldots,p_{k-1}=p_{k-1,0}$ を考え，全ての対立仮説（H$_1$: $\exists i\ \ p_i\neq p_{i,0}$）に対して検定する．

$P\left(Q_{k-1}\geq c\right)=\alpha$ (危険率・有意水準)となる c を見つければよい．

サイコロ(例)を有意水準 5％で検定する場合，自由度 5 のカイ二乗分布より $P\left(Q_5\geq 11.1\right)=0.05$ が得られる．

一方, Q_5 の観測値は
$$Q_5=\sum_{i=1}^{6}\frac{\left(X_i-np_i\right)^2}{np_i}=\sum_{i=1}^{5}\frac{X_i^{\ 2}}{np_i}-n=\frac{13^2+19^2+11^2+8^2+5^2+4^2}{10}-60=15.6>11.1$$
となるので有意水準 5％で「H$_0$: 全ての目が等確率で出る」は棄却される．

○ クロス集計表の独立性の検定

(人)		属性 A　血液型			
		O	A	B	AB
属性 B 人種	英国	467	417	86	30
	日本	305	382	219	94
	インド	310	190	412	85

http://www2.ttcn.ne.jp/honkawa/9450.html

左図のクロス集計表の結果から，人種によって血液型の分布が異なるかどうかを調べたいとする．

この場合，クロス集計表の独立性の検定を行う．

カテゴリ ij に入る確率を $p_{ij} = P\left(A_i \cap B_j\right) \quad i=1,\cdots,a,\ j=1,\cdots,b$，ij のカテゴリ内の頻度を X_{ij} とすると

H_0: $p_{ij} = p_{ij,0} \quad \forall ij$ 　　H_1: $\exists ij \quad p_{ij} \neq p_{ij,0}$ に対しては $Q_{ab-1} = \sum_{j=1}^{b}\sum_{i=1}^{a}\frac{\left(X_{ij}-np_{ij}\right)^2}{np_{ij}} \sim \chi^2\left(ab-1\right)$ で検定できる．カテゴリ数 ab に対して $\sum p_{ij} = 1$ の 1 本の制約があるため，自由度 ab-1 のカイ二乗分布で検定する．

しかしいま興味があるのは，二種類の属性の統計的独立性である．すなわち

H_0: $p_{ij} = p_{i\cdot}p_{\cdot j} \quad \forall ij$ 　　ただし $p_{i\cdot} = P\left(A_i\right), p_{\cdot j} = P\left(B_j\right) \quad \forall ij \qquad 1=\sum_i\sum_j p_{ij} = \sum_i p_{i\cdot} = \sum_j p_{\cdot j}$

に対して検定を行いたい．

もし，$p_{i\cdot}, p_{\cdot j}$ が既知なら，$Q_{ab-1} = \sum_{j=1}^{b}\sum_{i=1}^{a}\frac{\left(X_{ij}-np_{i\cdot}p_{\cdot j}\right)^2}{np_{i\cdot}p_{\cdot j}} \sim \chi^2\left(ab-1\right)$ で検定すればよい．しかし，通常 $p_{i\cdot}, p_{\cdot j}$ は未知なので，推定値 $\hat{p}_{i\cdot}, \hat{p}_{\cdot j}$ を用いて検定を行う必要がある．そこで，$\hat{p}_{i\cdot} = \frac{X_{i\cdot}}{n}, \hat{p}_{\cdot j} = \frac{X_{\cdot j}}{n}$, ただし $X_{i\cdot} = \sum_{j=1}^{b} X_{ij}, X_{\cdot j} = \sum_{i=1}^{a} X_{ij}$ を利用する．$\sum p_{i\cdot} = \sum p_{\cdot j} = 1$ という a + b 本の制約があるため，自由度は ab-1-(a+b)=(a-1)(b-1) となる．

自由度に関する別の説明：推定するパラメータは $\hat{p}_{i\cdot}, \hat{p}_{\cdot j}$ の a+b 個だが，$\sum p_{i\cdot} = \sum p_{\cdot j} = 1$ の制約から推定すべきパラメータの数は a+b-2．推定すべきパラメータの数だけもとの自由度から減少するため，(ab-1)-(a+b-2) = (a-1)(b-1) となる．

結局，
$$Q_{(a-1)(b-1)} = \sum_{j=1}^{b}\sum_{i=1}^{a}\frac{\left(X_{ij}-n\left(\frac{X_{i\cdot}}{n}\right)\left(\frac{X_{\cdot j}}{n}\right)\right)^2}{n\left(\frac{X_{i\cdot}}{n}\right)\left(\frac{X_{\cdot j}}{n}\right)} \sim \chi^2\left(\left(a-1\right)\left(b-1\right)\right)$$
に対して検定を行えばよい．

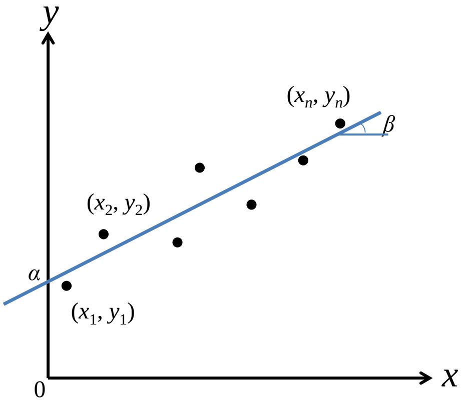
Ⅳ．回帰分析

１．線形回帰モデル

(1)　　回帰分析とは

被説明変数（従属変数・目的変数ともいう）と説明変数（独立変数ともいう）の関係を式に当てはめ，被説明変数の変動が説明変数によって説明できる割合を定量的に分析する方法.

回帰分析の基礎として，被説明変数と説明変数の関係を線形の式で表す線形回帰モデルについて説明する.



a)　単回帰

被説明変数 y_i（例：世界の平均気温）の変動を，一つの説明変数 x_i（例：大気中 CO_2 濃度）との線形関係として表すモデルを立て，説明変数が被説明変数の値に及ぼす影響を分析する方法.

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

α：切片，β：直線の傾き

ε_i：撹乱項（あるいは，誤差.　y の観測誤差やこのモデルでは表現できない y の変動など）

このモデルでは，推定すべきパラメータ（変数）は，α と β. ここで，モデルをデータに当てはめて得られる，パラメータの推定値を $\hat{\alpha}, \hat{\beta}$ と表すとする.

$$y_i = \hat{\alpha} + \hat{\beta} x_i + e_i$$

e_i：残差（パラメータ推定を行った結果，モデルでは表現できなかった y の変動）

最もデータがモデルに当てはまっている状態とは，残差 e_i が小さくなる状態．そこで，パラメータの推定値 $\hat{\alpha}, \hat{\beta}$ は，残差二乗和を最小化するように求める.

$$\min_{\alpha,\beta} \sum_i e_i^2 = \min_{\alpha,\beta} \sum_i \left(y_i - \alpha - \beta x_i \right)^2 \to \hat{\alpha}, \hat{\beta}$$

この推定法を最小二乗法といい，得られた推定値を最小二乗推定量という.

b)　重回帰

被説明変数の変動を，複数の説明変数で表すモデルを立て，各説明変数が被説明変数に与える影響を分析する方法．以降では，

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

x_{i1}: 通常は 1 （定数項）

$\beta_{1,} \cdots, \beta_k$：パラメータ（$k$個）

ε_i: 撹乱項（確率変数）

(2)　線形回帰モデルの仮定

線形回帰におけるモデルの仮定について以下に記す.

1.　線形性　線形のモデルで表される.

ただし，例えば対数変換した値を用いれば，非線形な関係を表すことも可能.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

2.　X の列（各説明変数）は線形独立.

$$rank\left(\mathbf{X}\right) = k \text{ (パラメータ数)} < n \text{ (観測数)}$$

3.　攪乱項の期待値は 0

$$E\left(\boldsymbol{\varepsilon} \mid \mathbf{X}\right) = \mathbf{0} \rightarrow E\left(\mathbf{y} \mid \mathbf{X}\right) = \mathbf{X}\boldsymbol{\beta}$$

4.　攪乱項の分散は均一，共分散は 0.

$$Var\left(\varepsilon_i \mid \mathbf{X}\right) = \sigma^2 \quad \forall i, \qquad Cov\left(\varepsilon_i, \varepsilon_j\right) = 0 \quad \forall i \neq j$$

行列表現で書くと

$$E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \mathbf{X}\right) = \begin{bmatrix} E\left(\varepsilon_1^2 \mid \mathbf{X}\right) & E\left(\varepsilon_1\varepsilon_2 \mid \mathbf{X}\right) & \cdots & E\left(\varepsilon_1\varepsilon_n \mid \mathbf{X}\right) \\ E\left(\varepsilon_2\varepsilon_1 \mid \mathbf{X}\right) & E\left(\varepsilon_2^2 \mid \mathbf{X}\right) & \cdots & E\left(\varepsilon_2\varepsilon_n \mid \mathbf{X}\right) \\ \vdots & \vdots & \ddots & \vdots \\ E\left(\varepsilon_n\varepsilon_1 \mid \mathbf{X}\right) & E\left(\varepsilon_n\varepsilon_2 \mid \mathbf{X}\right) & \cdots & E\left(\varepsilon_n^2 \mid \mathbf{X}\right) \end{bmatrix}$$

$$\because Cov\left(\varepsilon_i, \varepsilon_j\right) = E\left[\left(\varepsilon_i - E\left(\varepsilon_i\right)\right)\left(\varepsilon_j - E\left(\varepsilon_j\right)\right)\right] = E\left(\varepsilon_i\varepsilon_j\right)$$

$$= \begin{bmatrix} Var\left(\varepsilon_1 \mid \mathbf{X}\right) & Cov\left(\varepsilon_1, \varepsilon_2 \mid \mathbf{X}\right) & \cdots & Cov\left(\varepsilon_1, \varepsilon_n \mid \mathbf{X}\right) \\ Cov\left(\varepsilon_2, \varepsilon_1 \mid \mathbf{X}\right) & Var\left(\varepsilon_2 \mid \mathbf{X}\right) & \cdots & Cov\left(\varepsilon_2, \varepsilon_n \mid \mathbf{X}\right) \\ \vdots & \vdots & \ddots & \vdots \\ Cov\left(\varepsilon_n, \varepsilon_1 \mid \mathbf{X}\right) & Cov\left(\varepsilon_n, \varepsilon_2 \mid \mathbf{X}\right) & \cdots & Var\left(\varepsilon_n \mid \mathbf{X}\right) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2\mathbf{I}$$

5.　説明変数 X は非確率変数.　X は既知の $n \times k$ 行列.

（厳密には普通成立しない仮定.　**X** を確率変数とした拡張は可能）

(6.　$\boldsymbol{\varepsilon} \mid \mathbf{X} \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$　**攪乱項は期待値 0，分散 σ^2，共分散 0 の正規分布に従う）**

正規分布の仮定は，最小二乗法では不要.　仮説検定や最尤推定には必要.

２．通常最小二乗法による線形回帰モデルの推定

(1)　モデル推定

a) 通常最小二乗法(OLS: Ordinary Least Squares)**によるパラメータ推定**

前節の「**線形回帰モデルの仮定**」の下で，パラメータを推定する方法である通常最小二乗法(OLS)について説明する．

ここで，モデル $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$ のパラメータ $\boldsymbol{\beta}$ の推定値を $\mathbf{b}$ と表記することとする．このパラメータ推定値 $\mathbf{b}$ を用いて，回帰モデルによる y_i の推定値を計算すると，

$$\hat{y}_i=b_1x_{i1}+b_2x_{i2}+\cdots+b_kx_{ik}=\mathbf{x}_i'\mathbf{b}\quad \text{ただし}\mathbf{x}_i'=\begin{pmatrix}x_{i1} & x_{i2} & \cdots & x_{ik}\end{pmatrix}$$

となるが，推定値と被説明変数の差（残差 e）

$$e_i=y_i-\hat{y}_i=y_i-\mathbf{x}_i'\mathbf{b}$$

をなるべく小さくして，モデルと被説明変数を合わせたい.

そこで，残差の二乗和を最小にするという基準で，パラメータの推定値を求める．→最小二乗法

$$\min_{\mathbf{b}}\sum_i e_i^2=\min_{\mathbf{b}}\mathbf{e}'\mathbf{e}=\min_{\mathbf{b}}\left(\mathbf{y}-\mathbf{X}\mathbf{b}\right)'\left(\mathbf{y}-\mathbf{X}\mathbf{b}\right)=\min_{\mathbf{b}}\mathbf{y}'\mathbf{y}-\mathbf{b}'\mathbf{X}'\mathbf{y}-\mathbf{y}'\mathbf{X}\mathbf{b}+\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$$
$$=\min_{\mathbf{b}}\mathbf{y}'\mathbf{y}-2\mathbf{y}'\mathbf{X}\mathbf{b}+\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}\quad \because \mathbf{b}'\mathbf{X}'\mathbf{y},\mathbf{y}'\mathbf{X}\mathbf{b}\text{はスカラー}$$

ここで，残差二乗和を最小化するパラメータの推定値 $\hat{\boldsymbol{\beta}}$ は $\dfrac{\partial\mathbf{e}'\mathbf{e}}{\partial\mathbf{b}}=\mathbf{0}$ となる $\mathbf{b}$.

$$\frac{\partial\mathbf{e}'\mathbf{e}}{\partial\mathbf{b}}=\frac{\partial}{\partial\mathbf{b}}\left(\mathbf{y}'\mathbf{y}-2\mathbf{y}'\mathbf{X}\mathbf{b}+\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}\right)$$
$$=-2\mathbf{X}'\mathbf{y}+2\mathbf{X}'\mathbf{X}\mathbf{b}\quad \because \frac{\partial\left(\mathbf{a}'\mathbf{x}\right)}{\partial\mathbf{x}}=\mathbf{a},\frac{\partial\left(\mathbf{x}'\mathbf{A}\mathbf{x}\right)}{\partial\mathbf{x}}=\left(\mathbf{A}+\mathbf{A}'\right)\mathbf{x}$$

となるので，$\dfrac{\partial\mathbf{e}'\mathbf{e}}{\partial\mathbf{b}}=\mathbf{0}$ となる $\mathbf{b}$ の最適値（$\hat{\boldsymbol{\beta}}$ と記す）は

$$-2\mathbf{X}'\mathbf{y}+2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}=0\quad\Rightarrow\quad \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}=\mathbf{X}'\mathbf{y}$$

ここで，**線形回帰モデルの仮定 2.**より，$\mathbf{X}$ の列は線形独立で $rank(\mathbf{X})=k$.
$rank(\mathbf{X'X})=k$ となり，$k\times k$ の行列 $\mathbf{X'X}$ はランク落ちしないので，逆行列が存在．したがって，次式の最小二乗法によるパラメータの推定値（最小二乗推定量）が得られる．

$$\hat{\boldsymbol{\beta}}=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

b) 最小二乗推定量 $\hat{\boldsymbol{\beta}}$ と残差 e の関係

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}-\mathbf{X}'\mathbf{y}=\mathbf{0}\quad\Rightarrow\quad \mathbf{X}'\left(\mathbf{X}\hat{\boldsymbol{\beta}}-\mathbf{y}\right)=-\mathbf{X}'\mathbf{e}=\mathbf{0}$$

となることから，定数項のあるモデル（$\mathbf{X}$ の一列目がすべて 1）なら

・最小二乗残差の和は 0　$e_1+\cdots e_n=0$

c) 残差二乗和の意味解釈・決定係数

ここでは，残差二乗和が持つ意味の解釈を行う．

$\mathbf{y}=\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{e}$より

$$\begin{aligned}\mathbf{y}'\mathbf{y}&=\left(\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{e}\right)'\left(\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{e}\right)=\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{e}'\mathbf{X}\hat{\boldsymbol{\beta}}+\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{e}+\mathbf{e}'\mathbf{e}\\&=\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{e}'\mathbf{e}\quad\because\mathbf{X}'\mathbf{e}=0\end{aligned}$$

ここで

$$\begin{aligned}\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2&=\sum_{i=1}^{n}y_i^2-2\sum_{i=1}^{n}y_i\overline{y}+n\overline{y}^2\quad ただし\overline{y}=\frac{1}{n}\sum_{i=1}^{n}y_i\\&=\sum_{i=1}^{n}y_i^2-n\overline{y}^2\\&=\mathbf{y}'\mathbf{y}-n\overline{y}^2\end{aligned}$$

となり，各項は

$$\left(\mathbf{y}'\mathbf{y}-n\overline{y}^2\right)=\left(\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}-n\overline{y}^2\right)+\mathbf{e}'\mathbf{e}$$

被説明変数 y_i の平均からの変動	モデルで表現される被説明変数の変動	残差二乗和
TSS	ESS	RSS
(Total Sum of Squares)	(Estimated Sum of Squares)	(Residual Sum of Squares)

と呼ばれる．

線形回帰モデルの適合度の指標として，決定係数

$$R^2=\frac{ESS}{TSS}=1-\frac{RSS}{TSS}$$

が利用される．しかし，決定係数によるモデルの評価には問題が存在する．

説明変数を追加すると，たとえほとんど意味の無い説明変数であったとしても，$\mathbf{X}$ がランク落ちしない限りにおいては R^2 は必ず増える．そのため，説明変数を追加する意味の有無は R^2 からは正しく判断することはできない．

そこで，説明変数の追加による適合度の見かけ上の上昇を，自由度を用いて補正した

自由度修正済み決定係数 $$\overline{R^2}=1-\frac{RSS/(n-k)}{TSS/(n-1)}$$

を用いて評価することが望ましい．

(2)　最小二乗推定量

a) 最小二乗推定量$\hat{\boldsymbol{\beta}}$の統計的性質

最小二乗推定量の持つ，統計的性質について解説する．

$\hat{\boldsymbol{\beta}}$は被説明変数 **y** の線形結合で表されている線形推定量．式展開を行うと以下の通り．

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})\\ &= \boldsymbol{\beta}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\end{aligned}$$

ここで，**線形回帰モデルの仮定 5.**より，**X** は非確率変数なので

$$E(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}) = \mathbf{0} \Rightarrow E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

となり，$\hat{\boldsymbol{\beta}}$は **β** の<u>不偏推定量</u>であることが確認できる．このとき，$\hat{\boldsymbol{\beta}}$の分散共分散行列は

$$\begin{aligned}Var(\hat{\boldsymbol{\beta}}) &= E\left[(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\right]\\ &= E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right]\\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \quad \because E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\mathbf{I}\end{aligned}$$

と表される．

ここで，**β** の他の線形不偏推定量 $\mathbf{b}_0$ について考えてみよう．

$\mathbf{b}_0 = \mathbf{Cy}$

不偏性(期待値が **β**)より $E(\mathbf{b}_0) = E(\mathbf{Cy}) = E(\mathbf{CX}\boldsymbol{\beta}+\mathbf{C}\boldsymbol{\varepsilon}) = \mathbf{CX}\boldsymbol{\beta} = \boldsymbol{\beta}$ となるので $\mathbf{CX} = \mathbf{I}$.

ここで，$\mathbf{b}_0$ の分散について考える．

$$\begin{aligned}Var(\mathbf{b}_0) &= E\left[(\mathbf{b}_0-E(\mathbf{b}_0))(\mathbf{b}_0-E(\mathbf{b}_0))'\right]\\ &= E\left[(\mathbf{Cy}-\boldsymbol{\beta})(\mathbf{Cy}-\boldsymbol{\beta})'\right]\\ &= E\left[(\mathbf{CX}\boldsymbol{\beta}+\mathbf{C}\boldsymbol{\varepsilon}-\boldsymbol{\beta})(\mathbf{CX}\boldsymbol{\beta}+\mathbf{C}\boldsymbol{\varepsilon}-\boldsymbol{\beta})'\right]\\ &= \mathbf{C}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\mathbf{C}'\\ &= \sigma^2\mathbf{CC}'\end{aligned}$$

ここで$\mathbf{D} = \mathbf{C}-(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$とすると

$$Var(\mathbf{b}_0) = \sigma^2\left[\left(\mathbf{D}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\left(\mathbf{D}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)'\right]$$

また，$\mathbf{CX}=\mathbf{I}$ より $\mathbf{DX}=\mathbf{0}$ となるので

$$
\begin{aligned}
Var(\mathbf{b_0}) &= \sigma^2\left[\mathbf{DD}'+(\mathbf{X}'\mathbf{X})^{-1}\right] \\
&= Var(\hat{\boldsymbol{\beta}})+\sigma^2\mathbf{DD}'
\end{aligned}
$$

$\mathbf{DD}'$ の二次形式 $\mathbf{q}'\mathbf{DD}'\mathbf{q}$ は任意のベクトル $\mathbf{q}$ に対して $\mathbf{q}'\mathbf{DD}'\mathbf{q}\geq 0$ となるので，$\mathbf{DD}'$ は非負定符号行列となり，

$Var(\mathbf{b_0})$ の二次形式は，必ず $Var(\hat{\boldsymbol{\beta}})$ の二次形式より大きいか等しい．

$$
Var(\mathbf{b_0})\geq Var(\hat{\boldsymbol{\beta}})
$$

$\Rightarrow$ $\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の分散が最小となる線形不偏推定量.

$\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の最良線形不偏推定量(BLUE: Best Linear Unbiased Estimator) であるという.

b) σ^2 と $\hat{\boldsymbol{\beta}}$ の分散の推定

$\boldsymbol{\beta}$ の推定値 $\hat{\boldsymbol{\beta}}$ の仮説検定や信頼区間推定を行うためには，$\hat{\boldsymbol{\beta}}$ の分散に関する情報（分散共分散行列）が必要である．そこで，$\hat{\boldsymbol{\beta}}$ の分散共分散行列の推定について説明する.

$\hat{\boldsymbol{\beta}}$ の分散共分散行列は

$$
Var(\hat{\boldsymbol{\beta}})=\sigma^2(\mathbf{X}'\mathbf{X})^{-1}
$$

で表されるため，誤差項の分散 σ^2 を推定する必要がある.

そこで，σ^2 の推定を行う．残差 $\mathbf{e}$ は

$$
\begin{aligned}
\mathbf{e} &= \mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}} \\
&= \mathbf{y}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= \left(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{y}
\end{aligned}
$$

と表すことができる．ここで，$\mathbf{M}=\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ とおく．この $\mathbf{M}$ は対称行列かつ巾等行列（$\mathbf{M}^2=\mathbf{M}$）という性質を持つ．この $\mathbf{M}$ を用いると，

$$
\mathbf{e}=\mathbf{My}=\mathbf{M}(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})=\mathbf{M}\boldsymbol{\varepsilon} \quad \because \mathbf{MX}=\mathbf{0}
$$

と表せる．ここで，残差二乗和 $\mathbf{e}'\mathbf{e}$ は $\mathbf{e}'\mathbf{e}=\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$ と表すことができるので，その期待値を取ると次式の通り.

$$
E(\mathbf{e}'\mathbf{e})=E(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon})
$$

いま，$\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}$ はスカラーなので，$\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon} = tr\left(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}\right)$ となる．トレースの循環的交換則より

$$\begin{aligned}
tr\left(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}\right) &= tr\left(\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right) \\
&= tr\left(\mathbf{M}\sigma^2\mathbf{I_n}\right) \qquad \text{ただし，}\mathbf{I_n}\text{は}n\times n\text{の単位行列} \\
&= \sigma^2 tr\left(\mathbf{M}\right) \\
&= \sigma^2 tr\left(\mathbf{I_n} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right) \\
&= \sigma^2\left\{tr\left(\mathbf{I_n}\right) - tr\left(\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)\right\} \\
&= \sigma^2\left\{tr\left(\mathbf{I_n}\right) - tr\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\right)\right\} \\
&= \sigma^2\left\{tr\left(\mathbf{I_n}\right) - tr\left(\mathbf{I_k}\right)\right\} \\
&= \sigma^2\left(n-k\right)
\end{aligned}$$

となるので，$E\left(\mathbf{e}'\mathbf{e}\right) = \sigma^2\left(n-k\right)$ となり，σ^2 の不偏推定量 s^2 は

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k} \qquad \because E\left(s^2\right) = \sigma^2$$

と表すことができる．

σ^2 の不偏推定量 s^2 を用いると，$\hat{\boldsymbol{\beta}}$ の分散共分散行列は $s^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ と推定される．

(3)　最小二乗推定量の仮説検定

a) パラメータの仮説検定

ここまでの議論では，**線形回帰モデルの仮定 6. 撹乱項が正規分布に従う**は必要としていないことに注意．しかし，得られたパラメータについて仮説検定を行うには，撹乱項が正規分布に従うことを仮定しなければならない．

○ 撹乱項の分散 σ^2 が既知の場合

$$z_i = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2 S_{ii}}} \sim N\left(0,1\right)$$

（z_i は平均 0，分散 1 の正規分布に従う）　ただし $S_{ii} : \left(\mathbf{X}'\mathbf{X}\right)^{-1}$ の第 i 対角要素

○ 撹乱項の分散 σ^2 が未知の場合

σ^2 の不偏推定量 s^2 を用いて

$$t_i = \frac{\hat{\beta}_i - \beta_i}{\sqrt{s^2 S_{ii}}} \sim t\left(n-k\right)$$

（t_i は自由度 n-k の t 分布に従う）

<t 分布の定義>

$$\frac{\left(n-k\right)s^2}{\sigma^2} \sim \chi^2\left(n-k\right),\ \frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2 S_{ii}}} \sim N\left(0,1\right) \Rightarrow t = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2 S_{ii}}} \Bigg/ \sqrt{\frac{\left(n-k\right)s^2}{\sigma^2} \Big/ \left(n-k\right)} \sim t\left(n-k\right)$$

通常，Excel などを使った回帰分析で出力される t 値は，

$$t_i = \frac{\hat{\beta}_i}{\sqrt{s^2 S_{ii}}} = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}}$$
（推定されたパラメータを，推定されたパラメータの標準偏差で割った値）

パラメータの推定値 $\hat{\beta}_i$ が，0 と有意に異なるか（すなわち，意味のないパラメータではないか）を調べる．

$t_{\lambda/2}$（自由度 n-k の t 分布の 100(1-λ/2)値）と比較して，$t_i = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}} > t_{\lambda/2}$ となれば，帰無仮説は棄却され，$\hat{\beta}_i$ は有意に 0 と異なることが検定される．

信頼区間は $P\left(\hat{\beta}_i - t_{\lambda/2} s_{\hat{\beta}_i} \le \beta_i \le \hat{\beta}_i + t_{\lambda/2} s_{\hat{\beta}_i}\right) = 1-\lambda$ と表される．

b）回帰モデルの有意性検定

モデルが有意かどうかの検定⇒「定数項以外のパラメータ **β** が 0」という帰無仮説の同時検定.

攪乱項が正規分布していることを仮定すると，下式は自由度 k-1, n-k の F 分布に従う.

$$F\left[k-1, n-k\right] = \frac{R^2 / (k-1)}{(1-R^2) / (n-k)}$$

R^2: 決定係数

３．最尤法による線形回帰モデルの推定

最尤推定とは（復習）

「最尤原理(現実に得られた標本は，確率が最大のものが実現した)」という仮定に基づく推定.

尤度：標本が得られる確率.

最尤推定量：尤度が最大となる母数の値.

最尤法では，標本が従う確率密度関数を仮定した上で，実際に観測された標本の値が得られる確率（尤度）を求めて，この尤度を最大にする確率密度関数の母数を求める.

⇒　事前に，確率密度関数を設定する必要がある.

そこで，線形回帰モデルの攪乱項が正規分布に従うという仮定を設定する.

(最小二乗推定の場合には必要のない仮定を設定することに注意

ただし，最小二乗推定の場合でも仮説検定を行う際には，同じ仮定を置いている）

線形回帰モデル

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

の攪乱項$\boldsymbol{\varepsilon}$が，平均 $\mathbf{0}$，分散σ^2（未知変数），共分散 0（分散共分散行列 $\sigma^2\mathbf{I}$）の正規分布に従っていると仮定する.

$$\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$$

多次元正規分布の同時確率密度関数は

$$f\left(\boldsymbol{\varepsilon}\right) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right) \quad \because \text{一次元 } f\left(\varepsilon\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\varepsilon^2\right)$$

である．ここで，変数変換より

$$f\left(\mathbf{y} \mid \mathbf{X}\right) = f\left(\boldsymbol{\varepsilon}\right)\left|\frac{\partial\boldsymbol{\varepsilon}}{\partial\mathbf{y}}\right| = f\left(\boldsymbol{\varepsilon}\right) \quad \because \frac{\partial\boldsymbol{\varepsilon}}{\partial\mathbf{y}} = \begin{bmatrix} \frac{\partial\varepsilon_1}{\partial y_1} & \cdots & \frac{\partial\varepsilon_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial\varepsilon_n}{\partial y_1} & \cdots & \frac{\partial\varepsilon_n}{\partial y_n} \end{bmatrix} = \mathbf{I}$$

となるので，尤度関数 L は

$$L = f\left(\mathbf{y} \mid \mathbf{X}\right) = f\left(\boldsymbol{\varepsilon}\right)$$

と表せるので，対数尤度関数 l は

$$\begin{aligned} l = \ln L = \ln f\left(\boldsymbol{\varepsilon}\right) &= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \\ &= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right) \end{aligned}$$

となる．この対数尤度関数の未知変数は，$\boldsymbol{\beta}, \sigma^2$ である．これらの未知変数に関して，対数尤度関数を最大化すれば良い．すなわち，下式から最尤推定量$\hat{\boldsymbol{\beta}}, \hat{\sigma}^2$が得られる.

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = -\frac{1}{\sigma^2}\left(-\mathbf{X}'\mathbf{y} + \mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right) = \mathbf{0} \qquad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right) = 0 \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{1}{n}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right) = \frac{\mathbf{e}'\mathbf{e}}{n}$$

最小二乗推定量と比較すると，$\boldsymbol{\beta}$ の推定量は一致し不偏性を持つことが確認できるが，

σ^2 の推定量は，不偏性を持つ最小二乗推定量 $s^2 = \dfrac{\mathbf{e}'\mathbf{e}}{n-k}$ に対して，最尤推定量 $\hat{\sigma}^2 = \dfrac{\mathbf{e}'\mathbf{e}}{n}$ となり，最尤推定量は分散を過小推定しており，不偏性を持たないことが分かる．

また，下記の最尤推定量の性質を利用する．

・一致性　　最尤推定量 $\hat{\boldsymbol{\theta}}$ は $\boldsymbol{\theta}$ に確率収束する．

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta} \quad \left(\text{すべての}\boldsymbol{\varepsilon} > 0\text{に対して}\ \lim_{n\to\infty} P\left(\left|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right| < \boldsymbol{\varepsilon}\right) = 1\right)$$

n:標本数

標本数が大きくなるにつれて，統計量 $\hat{\boldsymbol{\theta}}$ が $\boldsymbol{\theta}$ に確率的に収束する．

・漸近正規性　　最尤推定量 $\hat{\boldsymbol{\theta}}$ は正規分布に分布収束する．

$$\hat{\boldsymbol{\theta}} \xrightarrow{D} N\left(\boldsymbol{\theta}, \mathbf{I}(\boldsymbol{\theta})^{-1}\right)$$

ただし $\mathbf{I}(\boldsymbol{\theta})$ ：情報行列

$$\mathbf{I}(\boldsymbol{\theta}) = E\left[\left(\frac{\partial l}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial l}{\partial \boldsymbol{\theta}}\right)'\right] = -E\left[\frac{\partial^2 l}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right] = -E\begin{bmatrix} \dfrac{\partial^2 l}{\partial \theta_1^2} & \cdots & \dfrac{\partial^2 l}{\partial \theta_1 \partial \theta_p} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 l}{\partial \theta_p \partial \theta_1} & \cdots & \dfrac{\partial^2 l}{\partial \theta_p^2} \end{bmatrix}$$

まず，最尤推定量 $\hat{\boldsymbol{\beta}}, \hat{\sigma}^2$ は一致性を持ち，

$$\begin{aligned} \frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} &= -\frac{\mathbf{X}'\mathbf{X}}{\sigma^2} & -E\left(\frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\right) &= \frac{\mathbf{X}'\mathbf{X}}{\sigma^2} \\ \frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial \sigma^2} &= -\frac{1}{\sigma^4}\left(\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right) = -\frac{1}{\sigma^4}\mathbf{X}'\boldsymbol{\varepsilon} \Rightarrow & -E\left(\frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial \sigma^2}\right) &= \mathbf{0} \\ \frac{\partial^2 l}{\partial \left(\sigma^2\right)^2} &= \frac{n}{2\sigma^4} - \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\sigma^6} & -E\left(\frac{\partial^2 l}{\partial \left(\sigma^2\right)^2}\right) &= \frac{n}{2\sigma^4} \end{aligned}$$

より，

$$\mathbf{I}(\boldsymbol{\theta}) = \mathbf{I}\begin{pmatrix} \boldsymbol{\beta} \\ \sigma^2 \end{pmatrix} = \begin{bmatrix} \dfrac{\mathbf{X}'\mathbf{X}}{\sigma^2} & \mathbf{0} \\ \mathbf{0} & \dfrac{n}{2\sigma^4} \end{bmatrix} \Rightarrow \mathbf{I}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \dfrac{2\sigma^4}{n} \end{bmatrix}$$

が導かれ，$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\sigma}^2 \end{pmatrix} \xrightarrow{D} N\left(\boldsymbol{\theta}, \mathbf{I}(\boldsymbol{\theta})^{-1}\right)$ という漸近正規性を持つことが導かれる．

４．回帰分析時の諸問題　多重共線性・系列相関・分散不均一

前述のように，線形回帰モデルの推定量$\hat{\boldsymbol{\beta}}$は最良線形不偏推定量で，統計的に優れた性質を持つ．しかし，線形回帰モデルの前提となる仮定が満たされない時には，必ずしも良い性質を持った推定量を出力しない．

回帰分析の際には，下記の現象が問題となることが多い．

- ・多重共線性 (Multi-colinearity)　　説明変数間に相関がある
- ・系列相関 (Serial-correlation)　　例えば，被説明変数間に相関がある
- ・分散不均一 (Heteroscedasticity)　　例えば，被説明変数の値の桁数が大きく異なるものがある

(・誤差が正規分布に従わない　　データ数が増えれば中心極限定理より正規分布に漸近)

ここでは，線形回帰モデルの前提が満たされない時に生じる問題について説明する．

(1)　多重共線性

多重共線性とは，説明変数間に相関がある場合にパラメータ推定結果が不安定になるという問題．

(例) ある駅周辺の地価を表現するモデルを作りたい．「交通利便性が高い土地は価格が高い」という現象を表すため，「駅からの道路距離」と「徒歩時間」を利用した下式のモデルを設定したとする．

$$(\text{地価}(\text{円}/\text{㎡}))=\beta_1+\beta_2\times(\text{駅からの道路距離}(\text{m}))+\beta_3\times(\text{駅からの徒歩時間}(\text{分}))$$

しかし，このモデルの説明変数は

$$(\text{駅からの道路距離}(\text{m}))=(\text{駅からの徒歩時間}(\text{分}))\times 80(\text{m}/\text{分})$$

の関係が成り立っており，極めて強い相関がある．

実際のデータからβ_2，β_3を推定して得られた値は，意味があるか？

$(\beta_2, \beta_3) = (0, -80)$，$(-1, 0)$という 2 組の推定値は，地価に対する二つの説明変数の当てはまりという点では差が無い．また，$(\beta_2, \beta_3) = (-6, 400)$（駅からの道路距離が 1m 延びたら 6 円/㎡安くなり，駅からの徒歩時間が 1 分延びると 400 円/㎡高くなる）という明らかに不自然な推定結果でも当てはまりは同じ．このような推定結果が得られる可能性も存在する．

上記の例のように説明変数間に完全な相関があると，$\mathbf{X}'\mathbf{X}$はランク落ちし逆行列が求まらないため，パラメータは推定できないが，完全な相関でなければパラメータは求まってしまう．しかし，その推定結果は不安定で，説明変数と被説明変数の関係に関して誤った結論を産む可能性がある．

多重共線性を回避するには，相関関係のある説明変数を用いてはならない．

(2)　系列相関・分散不均一

線形回帰モデルの仮定 4.では，撹乱項に対して分散均一・共分散 0 を仮定しているが，これらは本当に成立をしているのか？成立しないときの対処法について紹介する．

a) 系列相関

系列相関とは，撹乱項に相関があり，共分散が 0 という仮定が成立しないという問題．

時系列データの場合

(例) GDP を被説明変数として分析することを考える．「去年の景気が悪ければ，今年の景気も悪い可能性が高い」と考えるのが自然で，毎年ランダムに GDP の値が変動するとは考えにくい．すなわち，被説明変数は時系列の相関を持っている．

空間データの場合

(例) 地価を被説明変数として分析することを考える．「隣接地の地価が，他地域の地価水準よりも高ければ，自分の家の地価も高い」と考えるのが自然．被説明変数は，空間相関を持っている．

このように時系列相関や空間相関を持っている被説明変数を用いて回帰分析を行う時，攪乱項には相関がない（共分散が 0）という仮定が本当に成り立つのか？

あくまで攪乱項に対する仮定なので，モデルで相関を説明できておれば問題ないのでは？

> ここで，（神のみぞ知る）真のモデルが，下記の構造をしており，攪乱項は完全に無相関だとする．
>
> (GDP)＝β_1+β_2×(企業の生産額)+ β_3×(物価水準)＋β_4×(<u>国民のやる気</u>)＋…＋ε
>
> (地価)＝β_1+β_2×(駅からの距離)+ β_3×(容積率)＋β_4×(<u>町内活動の活発さ</u>)＋…＋ε
>
> しかし，(国民のやる気)や(町内活動の活発さ)など，説明変数には観測不可能なものが存在しているとしよう．もし，これらの観測不可能な説明変数もまた，時系列相関や空間相関を持っている場合には，観測可能な説明変数だけを使ってモデルを作ると，必然的に攪乱項に相関が残ってしまう．

時系列データや空間データを扱う分析を行う際には，系列相関への対応が不可欠である．そこで，計量経済学によるアプローチ，および，空間統計学によるアプローチについて簡単に紹介する．

○計量経済学によるアプローチ

<u>時系列相関への対処法</u>

$\mathbf{y}=\mathbf{X\beta}+\mathbf{u}$ の攪乱項 $\mathbf{u}$ の相関を自己相関モデル・移動平均モデルなどを用いて，現在の攪乱項と 1 期前の攪乱項との間の関係をモデル化する方法が提案されている．

・自己相関モデル　$u_t=\varphi u_{t-1}+\varepsilon_t$，$\varepsilon_t \sim N\left(0,\sigma_\varepsilon^2\right) \Rightarrow u_t=\varepsilon_t+\varphi\varepsilon_{t-1}+\varphi^2\varepsilon_{t-2}+\cdots$

$$Var\left(\mathbf{u}\right)=\sigma_u^2\begin{bmatrix}1 & \varphi & \cdots & \varphi^{n-1}\\ \varphi & 1 & \cdots & \varphi^{n-2}\\ \vdots & \vdots & \ddots & \vdots\\ \varphi^{n-1} & \varphi^{n-2} & \cdots & 1\end{bmatrix},\ \sigma_u^2=\frac{\sigma_\varepsilon^2}{1-\varphi^2}$$

・移動平均モデル　$u_t=\varepsilon_t+\theta\varepsilon_{t-1}$，$\varepsilon_t \sim N\left(0,\sigma_\varepsilon^2\right)$

$$Var\left(\mathbf{u}\right)=\sigma_\varepsilon^2\begin{bmatrix}1+\theta^2 & \theta & \cdots & 0\\ \theta & 1+\theta^2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1+\theta^2\end{bmatrix}$$

<u>空間相関への対処法</u>

（空間）計量経済学では，時系列相関と同様に，自己相関モデル・移動平均モデルなどが提案されている．

計量経済学的アプローチでは，行政区単位などの地域集計データを分析することを想定し，攪乱項の空間相関を地域の繋がりや近さを表現する空間重み行列 $\mathbf{W}$ を用いてモデル化する．

(例 1)　空間重み行列 $\mathbf{W}$ を地域の隣接関係で定義する場合，

地域 i, j が隣接→$\mathbf{W}$ の要素 w_{ij}=w_{ji}=1，地域 i, j が隣接していない→$\mathbf{W}$ の要素 w_{ij}=w_{ji}=0 として定義．

$$\mathbf{W}=\begin{bmatrix}0 & w_{12} & \cdots & w_{1n}\\ w_{21} & 0 & \cdots & w_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ w_{n1} & w_{n2} & \cdots & 0\end{bmatrix}$$

(例 2)　空間重み行列 $\mathbf{W}$ を地域の重心間距離で定義する場合

地域 i, j 間の関係を表す $\mathbf{W}$ の要素 w_{ij}, w_{ji} を距離の$-\alpha$ 乗として定義．

$$\mathbf{W}=\begin{bmatrix} 0 & d_{12}^{-\alpha} & \cdots & d_{1n}^{-\alpha} \\ d_{21}^{-\alpha} & 0 & \cdots & d_{2n}^{-\alpha} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1}^{-\alpha} & d_{n2}^{-\alpha} & \cdots & 0 \end{bmatrix},\ \ \alpha>0$$

これらの空間重み行列 **W** を利用して，時系列相関への対処と同様に攪乱項の相関をモデル化.

自己相関モデル　$\mathbf{u}=\varphi\mathbf{W}\mathbf{u}+\boldsymbol{\varepsilon},\quad \boldsymbol{\varepsilon}\sim N\left(\mathbf{0},\sigma_{\varepsilon}^{2}\mathbf{I}\right)\Rightarrow \mathbf{u}=\left(\mathbf{I}-\varphi\mathbf{W}\right)^{-1}\boldsymbol{\varepsilon}$

$$Var\left(\mathbf{u}\right)=E\left(\mathbf{u}\mathbf{u}'\right)=\sigma_{\varepsilon}^{2}\left(\mathbf{I}-\varphi\mathbf{W}\right)^{-1}\left[\left(\mathbf{I}-\varphi\mathbf{W}\right)^{-1}\right]'$$

移動平均モデル　$\mathbf{u}=\boldsymbol{\varepsilon}+\theta\mathbf{W}\boldsymbol{\varepsilon},\quad \boldsymbol{\varepsilon}\sim N\left(\mathbf{0},\sigma_{\varepsilon}^{2}\mathbf{I}\right)$

$$Var\left(\mathbf{u}\right)=E\left(\mathbf{u}\mathbf{u}'\right)=\sigma_{\varepsilon}^{2}\left(\mathbf{I}+\theta\mathbf{W}\right)\left(\mathbf{I}+\theta\mathbf{W}\right)'$$

限界

・地域単位の集計データへの適用が前提で，集計単位が変わらないことを暗黙の内に想定している.

空間重み行列 **W** を利用した攪乱項の相関のモデル化では，一部地域の集計単位が変わる（例えば，市町村が合併）と，全体の相関構造が変化してしまう（市町村合併と関係のない地域間の関係まで変わる）.

・観測された場所以外の値を予測する（例えば，アメダスの観測結果から自宅の気温を推定する）ことは，通常できない．統計解析の一つの大きな目的は「予測」だが，この枠組みでは対応できない.

○空間統計学によるアプローチ

攪乱項の共分散に対して，二次定常性を仮定する.

二次定常性とは，ある空間内で観測される確率変数（例えば，気温を観測）の共分散の大きさは，空間的な相対位置だけで決まるとの仮定（等方性を仮定すると，1km 離れた地点ペアで観測される気温の共分散は，空間内のどこのペアをとっても等しい）.

（なお，一次定常性は，平均（一次モーメント）は空間内のどこでも等しいことを意味する.）

二次定常性を仮定すると，共分散を距離の関数として定義することができる．$Cov(u_i, u_j) = C(d_{ij})$

$$Var\left(\mathbf{u}\right)=\begin{bmatrix} C(0) & C(d_{12}) & \cdots & C(d_{1n}) \\ C(d_{21}) & C(0) & \cdots & C(d_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ C(d_{n1}) & C(d_{n2}) & \cdots & C(0) \end{bmatrix}$$

長所

観測された場所と任意の場所の間に生じるモデルの攪乱項の共分散を定義することが可能であるため，攪乱項の相関構造を考慮して観測された場所以外のデータを「予測」することが可能.

短所

二次定常性の仮定は，攪乱項に対する仮定としてはかなり強い仮定で，実際に分析に使用するデータが仮定を満たしていることは少ない.

b）分散不均一

分散不均一とは，攪乱項の分散が，観測によって異なるという問題.

(例) 2013 年の公示地価データを用いて，東京都区部の商業地地価を回帰分析する場合を考える.
最高額　2700 万円/㎡　　東京都中央区銀座 4－5－6
最低額　30.3 万円/㎡　　東京都江戸川区中央 2－15－12

地価の違いが非常に大きいため，モデルを作って分析をすると，価格が高い地点ではモデルで説明できない残差の絶対値が大きく，安い地点では残差の絶対値が小さくなることが予想される.

対数を取った値を被説明変数として用いて攪乱項の分散不均一を回避する，あるいは，被説明変数の大きさに比例した分散を攪乱項が持つことを仮定しモデル化する，などで対応する.

c）一般化最小二乗法(GLS: Generalized Least Squares)によるパラメータ推定

以上では，攪乱項の相関のモデル化や分散不均一のモデル化などを行い，攪乱項の分散共分散行列 $Var\,(\mathbf{u})$の構造を与えるという方法について紹介した．ここでは，$Var\,(\mathbf{u})$が与えられたとき，どのようにパラメータ推定を行うかを示す.

ここで，表記の都合上，分散共分散行列 $Var(\mathbf{u}) = \sigma^2\mathbf{\Omega}$ と置く．なお，$\mathbf{\Omega}$ は正値定符号行列（任意のベクトル x に対して $\mathbf{x}'\mathbf{\Omega}\mathbf{x}>0$）という性質を持つ.

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{u},\quad \mathbf{u} \sim \mathrm{N}\left(\mathbf{0}, \sigma^2\mathbf{\Omega}\right)$$

$\mathbf{\Omega}$ が正値定符号行列であるので，$\mathbf{\Omega}^{-1} = \mathbf{P}'\mathbf{P}$ となる正則な行列 $\mathbf{P}$ が必ず存在する．この $\mathbf{P}$ を用いて

$$\mathbf{Py} = \mathbf{PX\beta} + \mathbf{Pu}$$

と表すと $\mathbf{Pu}$ の分散共分散行列 $Var(\mathbf{Pu})$は

$$Var\left(\mathbf{Pu}\right) = E\left(\mathbf{Puu}'\mathbf{P}'\right) = \mathbf{P}E\left(\mathbf{uu}'\right)\mathbf{P}' = \sigma^2\mathbf{P\Omega P}' = \sigma^2\mathbf{P}\left(\mathbf{P}'\mathbf{P}\right)^{-1}\mathbf{P}' = \sigma^2\mathbf{PP}^{-1}\left(\mathbf{P}'\right)^{-1}\mathbf{P}' = \sigma^2\mathbf{I}$$

と表すことができる．すなわち，$\mathbf{Py}$ を被説明変数，$\mathbf{PX}$ を説明変数とすると，攪乱項が分散均一，共分散 0 の仮定を満たすので，通常最小二乗法(OLS)より

$$\hat{\boldsymbol{\beta}} = \left(\left(\mathbf{PX}\right)'\mathbf{PX}\right)^{-1}\left(\mathbf{PX}\right)'\mathbf{Py} = \left(\mathbf{X}'\mathbf{P}'\mathbf{PX}\right)^{-1}\mathbf{X}'\mathbf{P}'\mathbf{Py} = \left(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y}$$

とパラメータが推定できる.

ただし，系列相関・分散不均一をモデル化する場合には，分散共分散行列は推定が必要であることには注意が必要.