統計解析手法（初回除く）

9/16

(2) 連続型

累積分布関数

連続

F(x)が右からも左からも連続なとき $\left(\lim_{\delta \to 0} F(x+\delta) = \lim_{\delta \to 0} F(x-\delta) = F(x)\right)$

連続確率変数

・F(x)が連続　・F(x)の導関数 $F'(x) = \frac{dF(x)}{dx} = f(x)$ が存在

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

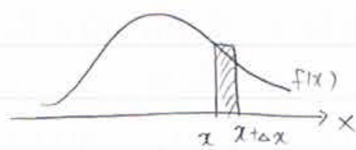
f(x)：確率密度関数　$f(x) \geqq 0$, $\int_{-\infty}^{\infty} f(x)dx = 1$

（離散型の確率分布としっかり区別）

ここで微小な $\Delta x > 0$ に対して、

$P(x \leqq X \leqq x+\Delta x) \fallingdotseq f(x)\cdot\Delta x$

と近似できる。図のように連続では



確率＝面積、離散は $f(x_i) = P(X = x_i)$

f(x)は $P(X=x)$を表さない　(ex) $P(X=a)=0$, $P(a<X<b) = F(b) - F(a) = \int_a^b f(x)dx$)

2.2 期待値と分散

(1) 定義と性質

○期待値　$E(X) = \begin{cases} \sum_i x_i f(x_i) & ：離散 \\ \int_{-\infty}^{\infty} x f(x)dx & ：連続 \end{cases}$

母集団分布の期待値：μで表現

・分散　$var(X) = E(X-\mu)^2 = \begin{cases} \sum_i (x_i - \mu)^2 f(x_i) & ：離散 \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx & ：連続 \end{cases}$

母集団分布の分散：σ^2で表現　標準偏差 $\sigma = \sqrt{var(X)}$　（E(X)と次元を合わせた）

・性質

① $E(a+bX) = a + bE(X)$

② $E(a) = a$

③ $E(X+Y) = E(X) + E(Y)$

④ $var(a+bX) = b^2 var(X)$

⑤ $var(a) = 0$

無次元量（異種のデータや，同種の多数のデータを比較するのに役立つ）

(2) 標準化

標準化変数 $z = \frac{x-\mu}{\sigma}$ （$a = -\frac{\mu}{\sigma}$, $b = \frac{1}{\sigma}$ としたときの線形変換 $a+bx$）

$E(z) = E(a+bx) = a + b\mu = 0$

$var(z) = var(a+bx) = b^2 var(x) = b^2\sigma^2 = 1$

平均は0，分散は1

2.3. 多変量分布

○ 同時確率分布と周辺確率分布

ex) 1枚のコインを3回投げ，表H，裏Tとする。

標本空間 $\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$

確率変数 X：最初の2回が表の回数

Y：最後の 〃

$(X=x) \cap (Y=y)$ の確率：

$f(x,y)$ 同時確率

HHH なら $x=y=2$，HHT なら $x=2$ $y=1$

$f(0,1) = p[(X=0)\cap(Y=1)] = p(TTH) = 1/8$

$f(1,1) = p[(X=1)\cap(Y=1)] = p(HTH) + p(THT) = 2/8$

X＼Y	0	1	2	
0	1/8	1/8	0	2/8
1	1/8	2/8	1/8	4/8
2	0	1/8	1/8	2/8
	2/8	4/8	2/8	1

$p(X=1) = g(1) = f(1,0) + f(1,1) + f(1,2) = 4/8$

$p(Y=2) = h(2) = \cdots\cdots = 2/8$

$g(x)$, $h(y)$ は周辺確率分布

離散確率変数：X, Y

$f(x_i, y_j) \geq 0$，$\sum_i \sum_j f(x_i, y_j) = 1$，$g(x) = \sum_j f(x, y_j)$，$h(y) = \sum_i (x_i, y)$

連続確率変数：X, Y $\longrightarrow$ $f(x,y)$：同時確率密度関数

$f(x,y) \geq 0$，$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$，$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$，$h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$

周辺確率密度関数

多変量での色々

2.4. 共分散と相関係数

(1) 共分散

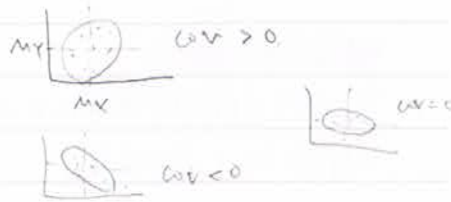
X, Y がそれぞれの期待値 μ_X, μ_Y から互いに関連しながらばらつく程度

$\mathrm{cov}(X, Y) = E[(X-\mu_X)(Y-\mu_Y)]$

$\mathrm{cov}(X, Y) > 0$ → X, Y は大小が同傾向

< 0 → 〃 反対傾向

$= 0$ → 上記の関係なし



(2) 相関係数

・共分散は変化の方向を表すので単位の異なる2つの共分散は比較できない

→ 標準化 $Z_X = \dfrac{X-\mu_X}{\sigma_X}$, $Z_Y = \dfrac{Y-\mu_Y}{\sigma_Y}$

・相関係数 ρ

$\rho = \mathrm{cov}(Z_X, Z_Y) = E[(Z_X - \overset{=0}{\mu_{Z_X}})(Z_Y - \overset{=0}{\mu_{Z_Y}})]$

$= E(Z_X, Z_Y) = E\left[\left(\dfrac{X-\mu_X}{\sigma_X}\right)\left(\dfrac{Y-\mu_Y}{\sigma_Y}\right)\right] = \dfrac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} = \dfrac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}$

・ρ は $-1 \leqq \rho \leqq 1$ の値をとる

1に近いほど強い正の相関, -1 に近いほど強い負の相関, 0: 無相関

・もし, $\rho = \pm 1$ なら X と Y は線形関係にある ($Y = a + bX$ で表される)

・相関係数は X と Y の1次結合の強さを測る尺度

2.5. 条件付確率分布

(1) 条件付確率分布

3回コインを投げる例 (cf 2.3.)

$P(X=0 \mid Y=1) = \dfrac{P(X=0, Y=1)}{P(Y=1)} = \dfrac{1/8}{4/8} = 1/4$

$g(0 \mid 1) = \dfrac{f(0, 1)}{h(1)}$

条件付確率分布 $g(x \mid y) = \dfrac{f(x, y)}{h(y)}$, $h(y \mid x) = \dfrac{f(x, y)}{g(x)}$ $(h(y) \neq 0,\ g(x) \neq 0)$

(2) 独立

・事象Aの確率が事象Bに影響されない、つまり $P(A|B)=P(A)$, 逆に $P(B|A)=P(B)$ も成立するならば、$P(A\cap B)=P(A)P(B)$ このとき、事象AとBは独立であるという

→従属

・確率分布の場合

$f(x,y)$ のすべての x,y について $f(x,y)=g(x)\cdot h(y)$ が成立 =独立

一般化すると、$f(x_1,x_2,\cdots,x_n)=f_1(x_1)\cdot f_2(x_2)\cdot\cdots\cdot f_n(x_n)=\prod_{i=1}^{n} f_i(x_i)$

添字?

。共分散と独立

独立(離散): $f(x_i,y_j)=g(x_i)h(y_j)$ このとき、

$$E(XY)=\sum_i\sum_j x_i y_j f(x_i,y_j)=\sum_i\sum_j x_i y_j g(x_i)h(y_j)$$

$$=\sum_i x_i g(x_i)\cdot\sum_j y_j h(y_j)=E(X)E(Y)$$

よって共分散 $\mathrm{cov}(X,Y)=E(XY)-E(X)E(Y)=0$.

やってみよ。

X、Yが独立 ⇒ 共分散=0, 相関係数=0　※ 逆は成り立たない

独立: 変数間に関連がない → 確率分布そのものの仮定

共分散・相関係数: 平均的な性質 → 確率分布から決まる量の仮定

/23

§3. 確率分布

・観測データ=標本 $\xleftarrow{\text{from}}$ 母集団: 確率分布によって表現

。確率分布を特徴づける定数 = パラメータ

3.1. 離散確率変数の確率分布

(1) ベルヌーイ試行 ← まだ分布じゃない

条件

① 結果が2種類 (ex) 「成功(S)」, 「失敗(F)」)

② 確率 $p=P(S)$ は一定 → $P(F)=1-p$

③ 試行は独立

いま、$S \to x=1$, $F \to x=0$ とすると、

ベルヌーイ分布　$f(x) = p^x(1-p)^{1-x}$　$(x=0,1)$
$E(x) = p$, $var(x) = p(1-p)$

これは1回の試行のみ。n回おこなうと

(2) 二項分布

ベルヌーイ試行をn回行う。

S: x回, F: $n-x$回 起こったとする。$(x=0,1,\cdots,n)$

二項分布: $f(x) = {}_nC_x \, p^x(1-p)^{n-x}$　$(x=0,1,\cdots,n)$
パラメータ: n, p $\xrightarrow{明記}$ $B(n,p)$ と表すこともある
$\Sigma f(x) = (p+1-p)^n = 1$　← 二項定理より
$E(x) = np$, $var(x) = np(1-p)$

∵) n回の独立な試行 ⟶ x回の成功確率 p^x, $(n-x)$回の失敗確率 $(1-p)^{n-x}$
最初のx回が成功なら $P(\underbrace{S,S,\cdots,S}_{x回}, \underbrace{F,\cdots,F}_{n-x回}) = p^x(1-p)^{n-x}$

x回の成功はどこで起きてもよい = 組み合わせ ${}_nC_x$ をかける。

ちょっと応用

(3) ポアソン分布

二項分布の期待値

めったに起こらない

二項分布において $np = \lambda$ (一定) とおく。$n \to$ 大, $p \to$ 小 (稀小現象) であれば、

$$ {}_nC_x \, p^x(1-p)^{n-x} = \frac{n\cdot(n-1)\cdots(n-x+1)}{x!}\cdot\left(\frac{\lambda}{n}\right)^x\cdot\left(1-\frac{\lambda}{n}\right)^{n-x} $$
$$ = \frac{1}{x!}\lambda^x\left(1-\frac{\lambda}{n}\right)^{n-x}\cdot\frac{n(n-1)\cdots(n-x+1)}{n^x} $$
$$ = \frac{1}{x!}\lambda^x\left(1-\frac{\lambda}{n}\right)^{n-x}\cdot\underbrace{1\cdot\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)}_{n\to\infty で 1} $$

$\frac{\lambda}{n} = \frac{1}{z}$ とおくと、$n = \lambda z$. $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n-x} = \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = \lim_{z\to\infty}\left\{\left(1-\frac{1}{z}\right)^{-z}\right\}^{-\lambda} = e^{-\lambda}$

∴ $n\to\infty$, $p\to 0$ の極限で、${}_nC_x \, p^x(1-p)^{n-x} = \frac{\lambda^x e^{-\lambda}}{x!}$: ポアソンの少数の法則

ポアソン分布: $f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$
パラメータ: λ → $P_o(\lambda)$ と表すこともある。$E(x) = \lambda$, $var(x) = \lambda$

下記の説明に主に使われる。
① 稀少現象の生起回数
② 到着数

ex) 交通事故が1年間で起こる回数

1年間を8760時間に分割 → 事故が起きるか否か ベルヌーイ試行

1時間内に複数回生起 → 分割を増やしベルヌーイ試行

↓ 分割を限りなく増やす

$n \to \infty$ の二項分布 = ポアソン分布

3.2 連続確率変数の確率分布

(1) 指数分布

ポアソン分布：稀少現象がある基準時間に起こる回数

↓

ある稀少現象が起きるまでの時間(待ち時間)をtとする (ポアソン分布で $\lambda \to \lambda t$)

= tまでに1回も現象が起きていない： $x=0$

累積確率密度関数 $F(t) = P(X \leqq t) = 1 - e^{-\lambda t}$

指数分布：$f(t) = \frac{dF(t)}{dt} = \lambda e^{-\lambda t}$

パラメタ：λ 　 $E(X) = \frac{1}{\lambda}$, $var(X) = \frac{1}{\lambda^2}$

(2) 正規分布

二項分布の $n \to \infty$ のときの極限の分布

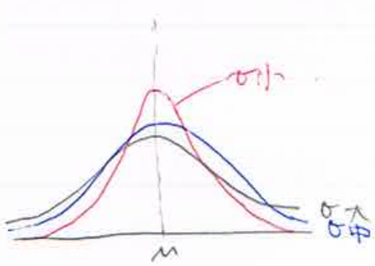
正規分布：$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$

パラメタ：$\mu, \sigma^2 \to X \sim N(\mu, \sigma^2)$ 　 $E(X) = \mu$, $var(X) = \sigma^2$

・標準化 $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$ — 標準正規分布

$E(X) = E(\mu + \sigma Z) = \mu + \sigma E(Z) = \mu$, $var(X) = var(\mu + \sigma Z) = \sigma^2 var(Z) = \sigma^2$

・μを中心に対称



対称性から

① $P(\mu + a \leqq X) = P(X \leqq \mu - a)$

② $P(X \leqq \mu + a) = P(\mu - a \leqq X)$

③ $P(\mu - a \leqq X \leqq \mu) = P(\mu \leqq X \leqq \mu + a)$

- $Y = a + bX \sim N(a + b\mu, b^2\sigma^2)$ （ただし $X \sim N(\mu, \sigma^2)$）

  ∵) $x = \frac{y-a}{b}$, $\left|\frac{dx}{dy}\right| = \frac{1}{|b|}$

  $f_Y(y) = f_X\left(\frac{y-a}{b}\right) = \frac{1}{\sqrt{2\pi}|b|\sigma} \exp\left\{-\frac{(y-a-b\mu)^2}{2b^2\sigma^2}\right\}$

- いま $P(a \le X \le b)$ を求めたい. $P(b) - P(a)$

  標準正規分布を利用 $\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}$ （$Z \sim N(0,1)$）

  $f(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}$

  逆にzが与えられたとき $x = \sigma z + \mu$

$$\left[\begin{array}{llll} z = \pm 1 \to & x = \mu \pm \sigma & \text{"}1\sigma\text{範囲"} & P(\mu - \sigma \le x \le \mu + \sigma) = 0.68 \ (\fallingdotseq \tfrac{2}{3}) \\ z = \pm 2 \to & x = \mu \pm 2\sigma & \text{"}2\sigma\text{ 〃 "} & P(\mu - 2\sigma \le x \le \mu + 2\sigma) = 0.95 \\ z = \pm 3 \to & x = \mu \pm 3\sigma & \text{"}3\sigma\text{ 〃 "} & P(\mu - 3\sigma \le x \le \mu + 3\sigma) = 0.99 \end{array}\right.$$

cf). 偏差値: $\mu = 50$, $\sigma^2 = 10^2$ に調整したもの.

(3) 多変量正規分布

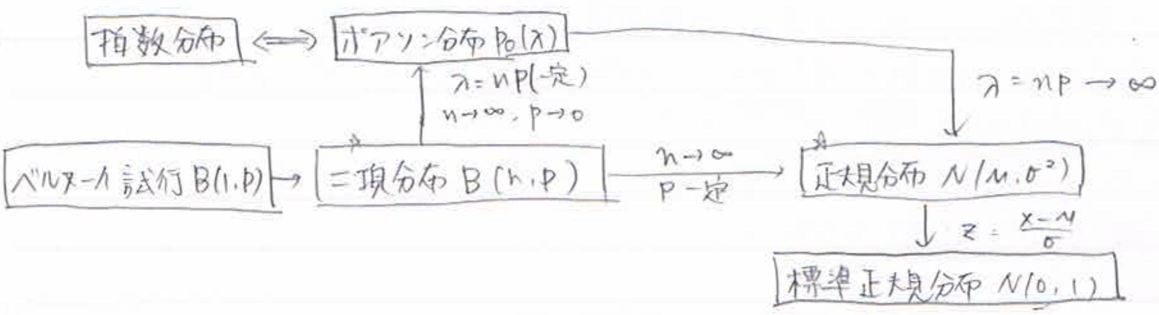
・n次元のデータ

$$f(\boldsymbol{x}) = \frac{1}{(\sqrt{2\pi})^n \sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right\} \sim N(\boldsymbol{\mu}, \Sigma)$$

$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_n)^T$: 平均ベクトル.

$\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix}$: 分散共分散行列

3.3 主要な確率分布



4/30.

§4. 大数の法則と中心極限定理

4.1 チェビシェフの不等式

このことが多い

確率分布の期待値と分散しかわかっていないとき、確率変数のとりうる値（区間）に対する確率の値の見当をつけたい。

確率分布に依存しない

↓

$$\boxed{\text{チェビシェフの不等式} \quad P(|X-\mu| \leqq k\sigma) \geqq 1-\frac{1}{k^2}}$$

下限値を見積もる

ex) $k=2$ $\quad P(\mu-2\sigma \leqq X \leqq \mu+2\sigma) \geqq 1-\frac{1}{2^2} = \frac{3}{4} = 0.75$

$k=3$ $\quad P(\mu-3\sigma \leqq X \leqq \mu+3\sigma) \geqq 1-\frac{1}{3^2} = \frac{8}{9} \fallingdotseq 0.89$

証明（X 連続のとき）

$$\sigma^2 = \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx \geqq \int_{|x-\mu|\geqq k\sigma}(x-\mu)^2 f(x)dx \geqq \int_{|x-\mu|\geqq k\sigma}(k\sigma)^2 f(x)dx = (k\sigma)^2 P(|X-\mu|\geqq k\sigma)$$

$\therefore \quad \frac{1}{k^2} \geqq P(|X-\mu| \geqq k\sigma)$ 　余事象をとって、$P(|X-\mu| \leqq k\sigma) \geqq 1-\frac{1}{k^2}$ ∎

意義 ① μ と σ によって、X の分布の状態が示される。

*② X がどのような確率分布であろうと成立し、$|X-\mu| \leqq k\sigma$ の確率の下限を与える。

③ 大数の（弱）法則を導くことができる。

4.2 大数の法則

標本から母集団を推定するときに利用

$P(|X-\mu| \geqq k\sigma) \leqq \frac{1}{k^2}$ 　　$k\sigma = c$ とおくと、$P(|X-\mu| \geqq c) \leqq \sigma^2/c^2$

X の標本 $\{x_1, \cdots, x_n\}$ があり、この平均 $\bar{x}$ を考える。（標本平均）

母平均 → $E(X) = \mu$, $var(X) = \sigma^2$ のとき、$E(\bar{x}) = \mu$, $var(\bar{x}) = \frac{\sigma^2}{n}$ となる。

$P(|\bar{x}-\mu| \geqq c) \leqq \frac{(\sigma^2/n)}{c^2} \xrightarrow{n\to\infty} 0$

$$\boxed{\text{大数の法則} \quad \lim_{n\to\infty} P(|\bar{x}-\mu| \geqq c) = 0}$$

$\bar{x}$ は μ に確率収束する

⇒ 十分な大きさの標本を調べれば、母集団の特性を知ることができる。

⇒ 統計的推測の理論

直接試験には出にくいが理解する

平均みたいなもん

和をとれば正規分布を仮定できる

Cauthy分布など

$E(x)$と$var(x)$の分からない分布には適用できない

4.3. 中心極限定理

Xの標本$\{x_1, \cdots, x_n\}$があり、$E(x) = \mu$, $var(x) = \sigma^2$

$\lim_{n\to\infty} \sum x_i \sim N(n\mu, n\sigma^2)$

厳密に表すと、$\lim_{n\to\infty} P\left(a \leq \underbrace{\frac{\sum x_i - n\mu}{\sqrt{n}\,\sigma}}_{\text{(基準化)}} \leq b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$

証明

和の期待値は期待値の和

$E(\sum x_i) = n\mu$, $var(\sum x_i) = n\sigma^2$　標準化すると、$\left(z_i = \frac{x_i - \mu}{\sigma} \to E(z_i) = 0, var(z_i) = 1\right)$

$\frac{\sum x_i - n\mu}{\sqrt{n}\,\sigma} = \frac{\sum z_i}{\sqrt{n}}$

ここで、e^xの展開式　$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots\cdots$　のxにtXを代入。

$e^{tX} = 1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots\cdots$　両辺の期待値をとる

$M_X(t) = E(e^{tX})$：モーメント母関数　← 左辺

$= 1 + tE(X) + \frac{t^2 E(X^2)}{2!} + \cdots\cdots$　← 右辺

$E(z_i) = 0$, $E(z_i^2) = var(z_i) - \{E(z_i)\}^2 = 1$　だから、

$M_{z_i}(t) = 1 + \frac{t^2}{2} + \cdots$

(標本を考えるので、x_1とx_2は独立)

また、$M_{x_1+x_2}(t) = E(e^{t(x_1+x_2)}) = E(e^{tx_1} \cdot e^{tx_2}) = E(e^{tx_1})E(e^{tx_2})$

$= M_{x_1}(t) \cdot M_{x_2}(t)$

よって、$M_{\sum z_i}(t) = \{M_{z_i}(t)\}^n = \left(1 + \frac{t^2}{2} + \cdots\right)^n$

$t \to \frac{t}{\sqrt{n}}$におきかえると、

$M_{\sum z_i}\left(\frac{t}{\sqrt{n}}\right) = \left\{M_{z_i}\left(\frac{t}{\sqrt{n}}\right)\right\}^n = \left(1 + \frac{t^2}{2n} + \cdots\right)^n = \exp\left(\frac{t^2}{2}\right)$

標準正規分布のモーメント母関数

∵ 分布とモーメント母関数は1対1に対応する

$\left(X \sim N(\mu, \sigma) \to M_X(t) = \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}\right)$

∴ $\frac{\sum z_i}{\sqrt{n}}$は標準正規分布

$\Rightarrow \sum x_i \sim N(n\mu, n\sigma^2)$となる

∎

(モーメント母関数をn階微分すると色々分かるらしい)

ex) 二項分布 → 正規分布

$B(n, p)$　$f_x = {}_nC_x\, p^x (1-p)^{n-x}$

$X \sim B(1, p)$

成功回数：$\sum x_i$，$E(\sum x_i) = np$，$var(\sum x_i) = np(1-p)$　← 前回やった。

中心極限定理より、$\lim_{n\to\infty} \frac{\sum x_i - np}{\sqrt{np(1-p)}} \sim N(0, 1)$

↑ ここまでが統計的推測の準備

§5 標本分布

5.1 母集団と標本 (cf 1.2)

(1) 母集団と標本

（悉皆調査）

調査方法　① 全数調査　② 標本調査　　無作為抽出

母集団 —標本抽出→ 標本

標本 —統計的推論→ 母集団

母集団：母集団分布，母数(パラメータ)（母平均 μ，母分散 σ^2 など）
同じもの

標本：標本分布，統計量（標本平均，標本分散など）

母数推定の重要な手がかり（cf. 大数の法則）

標本を要約し、母数の推定に使われるもの

標本 $(x_1, \cdots, x_n)$ の関数 $t(x_1, \cdots, x_n)$

未知パラメータは含まれない

観測値 $(x_1, \cdots, x_n)$ の $t(x_1, \cdots, x_n)$ ：統計値 ← 統計量は計算方法

5/13

(2) 統計量

代表的な母数：μ, σ^2 → 標本平均 $\bar{X}$，標本分散 s^2

以下，標本 $(X_1, \cdots, X_n)$ は母集団分布 (μ, σ^2) に従う独立な確率変数

◦ 標本平均

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}, \quad E(\bar{X}) = E\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{n\mu}{n} = \boxed{\mu}$$

$$var(\bar{X}) = var\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2} var(X_1 + \cdots + X_n) = \frac{n\sigma^2}{n^2} = \boxed{\frac{\sigma^2}{n}}$$

◦ 標本分散　　証明略

$$s^2 = \frac{1}{n-1}\left\{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\right\}, \quad E(s^2) = \sigma^2$$

期待値が母数のものを、不偏推定量という

∴ s^2：不偏分散

$n-1$：自由度（自由に動かせる変数の数）

$(X_1 - \bar{X}) + (X_2 - \bar{X}) + \cdots + (X_n - \bar{X}) = 0$ より、$X_n = g(X_1, \cdots, X_{n-1})$

よって自由度は $n-1$

統計量の仲間的な

(3) 標本比率

2値確率変数 $\{0, 1\}$ の場合で、1をとった回数が x 回であったとする。

→ 二項分布 (cf. 3.1.(2))

標本比率 (確率): $p = \frac{x}{n}$ を考える

$E(p) = E\left(\frac{x}{n}\right) = \frac{1}{n}E(x) = \frac{np}{n} = p$

$var(p) = var\left(\frac{x}{n}\right) = \frac{1}{n^2}var(x) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$

大事

(4) 統計量の標本分布

以下、母集団が正規分布に従うとして、次の分布を扱う。

- 標本平均 (σ^2 が既知) の分布 → 正規分布
- 標本分散の分布 → χ^2 分布 ← σ^2 は大体未知 (だから推定する) ☆
- σ^2 が未知のときの標本平均の分布 → t分布
- 標本分散比の分布 → F分布

以下これらを詳しくみていく。

5.2. 標本平均の分布 (σ^2 既知)

- 正規分布の再生性について

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

c_i を定数列とすれば $\sum_i c_i x_i \sim N\left(\Sigma c_i \mu,\ \Sigma (c_i \sigma)^2\right)$

(cf. 3.2.(2) での $Y = ax + b$ の一般化)

- $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ ← σ^2 既知なら正規分布

($\bar{x}$ は $x_1 \sim x_n$ の線形和)

(推定, 精度) 的な

- ここでの分散は精度を表す → n が増加すれば、$\bar{x}$ は μ のより正確な推定値
- 推定精度 (分散) は、標本数に対して $1/\sqrt{n}$ のオーダーでしか減少しない。
- $\bar{x}$ も正規分布 → 単独の x_i よりもすぐれた推定値

幾何平均な 高いとない

5.3. 標本分散の分布

σ^2 は不明の場合が多い → 分散の推定値の精度を知りたい。

$s^2 = \frac{1}{n-1}\Sigma(x_i - \bar{x})^2 \quad \therefore (n-1)s^2 = \Sigma(x_i - \bar{x})^2 \quad \cdots (*)$

ここで

↓

NO. 12.

DATE

・χ^2分布について

標準化 $z_i = \frac{x_i - \mu}{\sigma} \sim N(0,1)$ について考える. $\chi^2 = z_1^2 + z_2^2 + \cdots + z_k^2$ とすると.

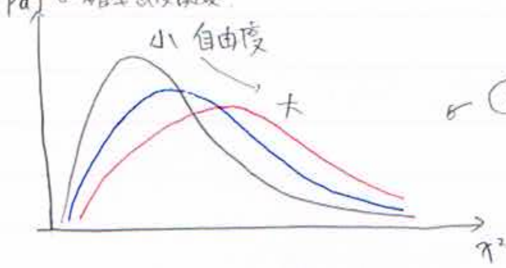
確率変数 $\chi^2 \sim$ 自由度kのχ^2分布 $\chi^2(k)$ (と定義), $E(\chi^2)=k$, $var(\chi^2)=2k$

$E(z_i^2)=1$ より. $f(x) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$ $(x>0)$ より

$(\Gamma(z) = \int_0^\infty e^{-x}x^{z-1}dx \ (z>0))$

どーでもいい...

pdf ← 確率密度関数



← このイメージは大事

$\chi^2 = \Sigma z_i^2 = \Sigma \frac{(x_i-\mu)^2}{\sigma^2}$ (*)より, $\frac{(n-1)s^2}{\sigma^2} = \Sigma\frac{(x_i-\bar{x})^2}{\sigma^2} = \chi^2$.

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

$\mu \to \bar{x}$ におきかえたので. $(x_1-\bar x)+\cdots+(x_n-\bar x)$ から自由度1減る : $n-1$

∴ $E(s^2) = \frac{\sigma^2}{n-1}E(\chi^2) = \frac{\sigma^2}{n-1}\cdot(n-1) = \sigma^2$ ← 不偏推定量 ⇒ 母分散の推定に利用

☆ 5.4 σ^2未知のときの標本平均の分布　　自然なケース

$\bar x \sim N(\mu, \frac{\sigma^2}{n})$　標準化 $z_i = \frac{\bar x - \mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$

σ^2は不明の場合が多い → 標本分散s^2で代替させた方が現実的

・t統計量

$t = \frac{\bar x-\mu}{\sqrt{s^2/n}}$　※ tは$N(0,1)$には従わない ← でも

・標本平均と標本標準偏差の比の分布

$t = \frac{\frac{\bar x - \mu}{\sqrt{\sigma^2/n}}}{\sqrt{s^2/\sigma^2}} = \frac{\frac{\bar x-\mu}{\sqrt{\sigma^2/n}}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}/n-1}}$

$\frac{\bar x-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$ …①

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

③ : $\bar x$とs^2は独立

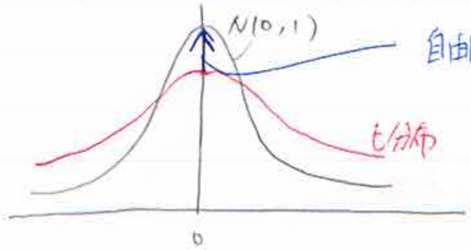
ここで.

t分布

一般に. ① $Z \sim N(0,1)$ ② $Y \sim \chi^2(k)$ ③ Z, Yは独立 のとき

$t = \frac{Z}{\sqrt{Y/k}} \sim$ 自由度kのt分布 $t(k)$　$E(t)=0$ $(k>1)$, $var(t) = \frac{k}{k-2}$ $(k>2)$

pdf : $f(t) = \frac{\Gamma[(k+1)/2]}{\sqrt{\pi k}\,\Gamma(k/2)} \cdot \frac{1}{\{(t^2/k)+1\}^{\frac{k+1}{2}}}$ $(-\infty < t < \infty)$ ← どーでもいい…



自由度→∞ で t分布は $N(0,1)$ に近づく.

∵ $n \to \infty \Rightarrow S^2 \fallingdotseq \sigma^2$

（外れた値をとる確率が大きめ）

$E(t)=0$ で対称, 分布の両すそが $N(0,1)$ より厚い.

※t分布は小標本の厳密な標本分布

以上より, $\boxed{t = \frac{\bar{x}-\mu}{\sqrt{s^2/n}} \sim t(n-1)}$

☆ 標本平均の分布のまとめ

σ^2	n \ 母集団	正規分布	非正規分布
既知	大	①	⑤
既知	小	②	⑥
未知	大	③	⑦
未知	小	④	⑧

←5.2より

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

⑤ $\bar{X} \approx N(\mu, \frac{\sigma^2}{n})$, $Z = \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \approx N(0,1)$ （≈ : 中心極限定理による）

③,⑦ $\bar{X} \approx N(\mu, \frac{S^2}{n})$, $Z = \frac{\bar{X}-\mu}{S/\sqrt{n}} \approx N(0,1)$

④ $t = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

⑥,⑧ → $\bar{X}$ の分布は分からない

5.5. 2標本問題

← F分布

第1の標本 $\{X_1, \cdots X_{n_1}\} \sim N(\mu_1, \sigma_1^2)$

〃2 〃 $\{Y_1, \cdots, Y_{n_2}\} \sim N(\mu_2, \sigma_2^2)$

$X_1 \cdots X_{n_1}, Y_1 \cdots Y_{n_2}$ は独立

DATE

◦ 標本分散比の分布

$s_1^2 = \frac{1}{n_1-1}\Sigma(x_i-\bar{x})^2$, $s_2^2 = \frac{1}{n_2-1}\Sigma(y_j-\bar{y})^2$

$\boxed{\text{F統計量} \quad F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}}$

$$F = \frac{\overbrace{\frac{(n_1-1)s_1^2}{\sigma_1^2}}^{\sim \chi^2(n_1-1)\ ①} \Big/ n_1-1}{\underbrace{\frac{(n_2-1)s_2^2}{\sigma_2^2}}_{\sim \chi^2(n_2-1)\ ②} \Big/ n_2-1}$$

③ : s_1^2, s_2^2 は独立.

ここで、

F分布

一般に ① $U \sim \chi^2(k_1)$ ② $V \sim \chi^2(k_2)$ ③ U, V は独立 のとき

$F = \frac{U/k_1}{V/k_2} \sim$ 自由度 (k_1, k_2) の F分布 : $F(k_1, k_2)$

$E(F) = \frac{k_2}{k_2-2}$ $(k_2>2)$, $var(F) = \frac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}$ $(k_2>4)$

E(F) と var(F) は参考で

以上より、 $\boxed{F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)}$

5/28

§6 パラメータの推定

母集団 →(標本抽出) 標本

推測 { パラメータ推定 { 点推定 / 区間推定 } §6
　　　仮説検定 (§7)

◦ パラメータ推定のための標本から求めた統計量 : 推定量 ― 確率変数　(計算方法)

　〃　計算値 (または) → 統計値 : 推定値 ― 実現値.

・点推定 : パラメータを1つの値で定めようとする方法

区間推定 : パラメータが存在する区間を推定する方法

↑
仮説検定と
一緒につかう.

6.1 点推定の推定法

推定量には ^ (ハット) をつける　ex) $\hat{\mu}$, $\hat{\sigma}^2$

(1) モーメント法

X の k次モーメント $= E(X^k)$ によって推定する方法

・期待値：$\mu = E(X)$，$\bar{X} = \frac{1}{n}\sum_{i} X_i$ により推定 → $\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i} X_i$

・分散：$\sigma^2 = E(X^2) - \{E(X)\}^2$，$\frac{1}{n}\sum_{i} X_i^2$ により推定 → $\hat{\sigma}^2 = \frac{1}{n}\sum_{i} X_i^2 - (\bar{X})^2 = \frac{1}{n}\sum_{i}(X_i - \bar{X})^2$

※ $s^2 = \frac{1}{n-1}\sum_{i}(X_i - \bar{X})^2$: 不偏分散 より，$\hat{\sigma}^2$ は真の分散を過小推定している

∴ $\hat{\sigma}^2$ は不偏ではない

※ この手法は確率分布によらない．

☆ (1)より精度↑

(2) 最尤法　正規分布なのか etc

母分布を必要とする．パラメータを θ (定数) と表現する

同時確率分布 $f(x_1, \overset{\text{変数}}{x_2, \cdots}, x_n : \underset{\text{定数}}{\theta})$

(パラメータ θ 固定したときに $x_1, \cdots x_n$ の値をとる確率)

ここで．

$(x_1, \cdots, x_n)$: 観測ごとに異なった値をとる確率変数

いま，1組の観測値 $(x_1, \cdots, x_n)$ が得られた (所与)

↓

$(x_1, \cdots, x_n)$ はもう確率変数ではなく，その実現値，

$(x_1, \cdots, x_n)$ をもたらした θ は様々な値が考えられる

この考え方は重要．

likelyhood ⇒ θ を変数と考えると，

$L(\underset{\text{変数}}{\theta} : \underbrace{x_1, \cdots x_n}_{\text{定数}})$ or 単に $L(\theta)$

得られた観測値 $(x_1, \cdots, x_n)$ の確率 が θ によってどう変化するか示す関数

$\theta = \theta_0$ のとき　(f と同じ値をとるが確率分布ではない．値ではなく大小が重要になる)

$L(\theta_0 : x_1, \cdots, x_n) = f(x_1, \cdots x_n : \theta_0)$

$L(\theta)$ の大小関係が重要

同じ観測結果 に対し $L(\theta_1) > L(\theta_2)$ なら，$\theta = \theta_1$ から得られた標本を考える方が

もっともらしい

⇓

$\boxed{\text{最尤法：尤度関数を最大にする } \theta \text{ を } \hat{\theta} \text{ とする．}}$　最尤推定量 (推定値)　maximum likelyhood : ML

$L(\theta)$ の最大化　$\frac{dL(\theta)}{d\theta} = 0$

パラメータが $\theta_1, \cdots \theta_k$ の場合　$\frac{\partial L(\theta)}{\partial \theta_j} = 0 \quad (j=1,\cdots,k) \rightarrow$ 連立方程式

あるいは　(計算方法をラクにする)

対数変換：単調増加変数　　$\max L(\theta) \rightarrow \max \log L(\theta)$

対数尤度関数を用いてもよい

ex) 正規分布

$\theta_1, \theta_2 : \mu, \sigma^2 \qquad X_i \sim N(\mu, \sigma^2) \quad , (x_1, \cdots, x_n)$ 所与

確率変数 X_i と x_i 実現値

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

$(x_1, \cdots x_n)$ が実現する確率

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

かけ算だし対数とるか

$$\log L(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2$$

(丸め誤差の解消にもなる)

$$\begin{cases} \frac{\partial \log L(\mu,\sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_i (x_i-\mu) = 0 \\ \frac{\partial \log L(\mu,\sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(x_i-\mu)^2 = 0 \end{cases}$$

(σ^2 で1階微分)

いつも同じとは限らない

不偏じゃない

$\therefore \hat{\mu} = \bar{x} \quad , \hat{\sigma}^2 = \frac{1}{n}\sum_i^n (x_i-\bar{x})^2$ ← モーメント法と同じ(正規分布のときは)

6.2 望ましい推定量の特性

・μ の推定量の候補：標本平均，メディアン，モード……

・$\hat{\sigma}^2$，s^2 の相違

⇒ 推定量の望ましい特性が必要

小標本特性：n の大きさに関わりなくもっているべき特性
(不偏性，有効性，最良線形不偏性)

大標本特性：n が大きくなっていくときもっているべき特性(漸近的特性)
(漸近的不偏性，一致性，漸近的有効性)

(1) 不偏性

θ が過大・過小推定でないこと　$\therefore E(\hat{\theta}) = \theta$ ← 不偏推定量

ex) $E(\bar{X}) = \mu$ (cf 5-1(2))

$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 = \sigma^2 \underbrace{- \frac{1}{n}\sigma^2}_{\text{過小推定}}$ ← 不偏推定量じゃない

$s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$　不偏標本分散

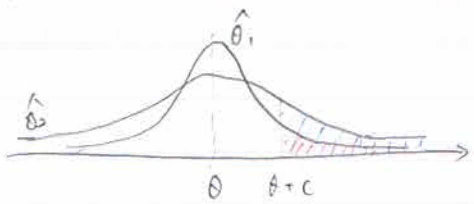
$E(s^2) = \sigma^2$　← 不偏推定量

(だけどなんか平均より劣りそう)

標本モードなども不偏 ⟶ 別の基準も必要

(2) 有効性

不偏推定量 $\hat{\theta}_1, \hat{\theta}_2$,　$var(\hat{\theta}_1) < var(\hat{\theta}_2)$



外れ値をとる確率が $\hat{\theta}_1$ の方が小さい.

$\hat{\theta}_1$ の方が望ましい.

∴ あらゆる不偏推定量の中で分散最小の推定量：有効推定量 or 最小分散不偏推定量

$X \sim N(\mu, \sigma^2)$ のとき, $\left(\begin{array}{l}\bar{X} \text{ は } \mu \text{ の有効推定量} \\ s^2 \text{ は } \sigma^2 \quad 〃\end{array}\right.$

もっとも厳しい条件

(3) 最良線形不偏性

$E(x) = \mu$,　$var(x) = \sigma^2$

$\hat{\mu} = \sum_i w_i x_i$ は μ の不偏推定量 (とすると)

$E(\hat{\mu}) = E(\sum w_i x_i) = \sum w_i E(x_i) = \mu \sum w_i$　　∴ $\sum w_i = 1$ でなければならない.

μ の任意の線形不偏推定量 $\hat{\mu}$ の中で最小分散 = 最良線形不偏推定量

($\sum w_i = 1$)　　best linear unbiased estimator (BLUE)

$\hat{\mu} - E(\hat{\mu}) = \hat{\mu} - \mu = \sum w_i x_i - \mu = \sum w_i (x_i - \mu)$　(∵ $\sum w_i = 1$)

$var(\hat{\mu}) = E[(\hat{\mu} - \mu)^2] = E[\sum w_i (x_i - \mu)]^2 = E[\sum_i \sum_j w_i w_j (x_i - \mu)(x_j - \mu)]$

分散の定義通り

$= w_i^2 E(x_i - \mu)^2 = \sigma^2 \sum w_i^2$　← こいつを最小にしたい

$\left(E(x_i - \mu)(x_j - \mu) = cov(x_i, x_j) = 0 \right)$ 独立を仮定

ラグランジュの未定乗数法

$\phi = \sum w_i^2 - \underline{2\lambda}(\sum w_i - 1)$ ← ラグランジュ関数

都合のため λ でなく 2λ とした

$$\begin{cases} \dfrac{\partial \phi}{\partial w_i} = 2w_i - 2\lambda = 0 \quad (\text{for } \forall i) \\ \dfrac{\partial \phi}{\partial \lambda} = -2(\sum w_i - 1) = 0 \end{cases} \Rightarrow w_i = \lambda, \ \lambda = \frac{1}{n}$$

$\therefore w_1 = w_2 = \cdots = \frac{1}{n}$ ⇒ よって $\bar{X}$ は BLUE.

大標本特性

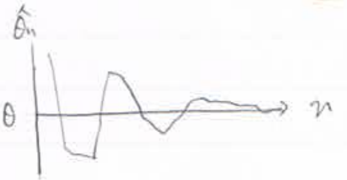
(4) 一致性

大きさ n の標本からの推定量：$\hat{\theta}_n$

$\hat{\theta}$ の方がいいと思うんだけど…

$\forall \varepsilon > 0$ に対して、$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \geq \varepsilon) = 0$ ：一致推定量

確率収束　(cf 4.2 大数の法則)



こんなイメージ

(十分条件： $\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$ (漸近的不偏性) かつ $\lim_{n\to\infty} var(\hat{\theta}_n) = 0$

ex) $E(\bar{X}) = \mu$, $var(\bar{X}) = \frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0$ ∴ $\bar{X}$ は μ の一致推定量

$\lim_{n\to\infty} (\hat{\sigma}^2) = \sigma^2$, $\lim_{n\to\infty} var(\hat{\sigma}^2) = \lim_{n\to\infty} \frac{2(n-1)\sigma^4}{n^2} = 0$

∴ $\hat{\sigma}^2$ は σ^2 の一致推定量, S^2 も σ^2 の一致推定量.

6/4

6.3. 区間推定

仮説検定とセットにすると分かりやすい

パラメータが存在すると予想される区間を確率的に推定

⇓

「θ が区間 $[\theta_1, \theta_2]$ に存在する確率が $1-\alpha$」という推定

$\therefore P(\theta_1 \leq \theta \leq \theta_2) = 1 - \alpha$

これはαの方

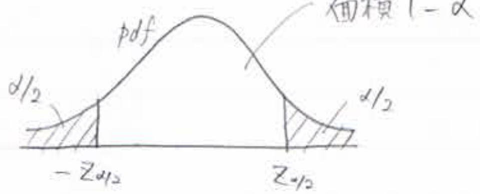
信頼係数 (有意水準)　← 通常 0.95, 0.99 が多い
(2σ, 3σ)

区間 $[\theta_1, \theta_2]$ = $100(1-\alpha)$% 信頼区間

→ θ を区間内に含むものの割合が $1-\alpha$

※ θ_1, θ_2 は実験で毎回変わる。
その区間に含まれる確率が $1-\alpha$
(定数区間に入る確率じゃない)

標準正規分布では、



($z_{\alpha/2}$：その点より上側の確率が $100 \times \alpha/2$ % となる z ＝パーセント点)

(1) 平均に関する推定

◎標本平均の分布 (cf 5.4：標本平均の分布のまとめの表)

① (σ^2 既知，母集団分布：正規分布，n：大)

②：(〃 〃 n：小) では、

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

・区間推定

$P(-z_{\alpha/2} \leq z \leq z_{\alpha/2}) = 1-\alpha$

限界誤差

$P(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$

上下限の差：区間幅　$2 z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

区間幅が小さいほど $\bar{x}$ の推定精度が上がる

- σ が小さいほど
- n が大きいほど (←限界誤差 $D = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ が所与なら $n = \left(\frac{z_{\alpha/2}\sigma}{D}\right)^2$)
- $z_{\alpha/2}$ が小さいほど (信頼度は低下)

覚えなくてよい (95% 信頼区間　$z_{\alpha/2} = 1.96$

③,② は $\sigma \to s$ のだけ.

④ t分布 $\sigma \to s$, $z_{\alpha/2} \to t_{\alpha/2}(n-1)$

①,②

ex) $X_1, \cdots, X_5 \sim N(\mu, 0.2^2)$

$\bar{x} = 19.5$ を観測　μ の 99% 信頼区間?

$\alpha = 0.001$, $z_{\alpha/2} = 2.576$, $n = 5$, $\sigma = 0.2$

$\therefore \bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 19.5 \pm 2.576 \cdot \frac{0.2}{\sqrt{5}} = 19.5 \pm 0.2304$

(2) 分散に関する推定（正規分布）

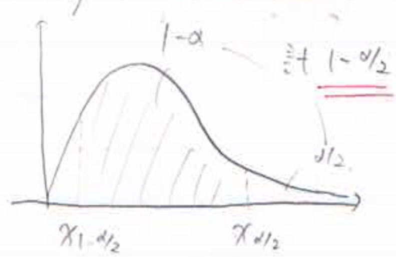
$\chi^2 = \sum_i \frac{(x_i - \bar{x})^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$ とおく

$\chi^2_\alpha(n-1)$ を χ^2_α とおく

$P(\chi^2_{1-\alpha/2} \leq \chi^2 \leq \chi^2_{\alpha/2}) = 1-\alpha$

注 右図

グラフは左右非対称



$\therefore P\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right) = 1-\alpha$

(3) 比率の区間推定（おまけ）

結果が2事象の n コの標本のうち，着目する事象の標本が k コであったとする

標本比率 $\hat{p} = \frac{k}{n}$

標本数 n が十分大きい場合（二項 → 正規）　$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$

標準化 $z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \sim N(0,1)$

分母の p を $\hat{p}$ で近似すれば，母比率 p の区間推定

$P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1-\alpha$

大事

§7 仮説検定

7.1. 検定の考え方

母集団について仮定された命題を標本に基づいて検証

母集団 → 標本

仮説H　・Hから著しくはずれる → Hを否定する（棄却する）
　　　　・Hから著しくはずれない → Hを認める（採択する）
　　　　　有意水準 α によって判断

ex) 仮説：「地球温暖化が進行」
観測：「18年間連続で異常高温が続いている」
ここでの異常：「たかだか30年に1度の頻度」
⇔ 対立する仮説：「地球平均気温は定常」
→ ある年に「異常気温」が生じる確率 1/30
各年の気温は独立と仮定

⇒ $p = 1/30$, $n = 18$ の二項分布

$f(x) = {}_nC_x\, p^x (1-p)^{n-x} = {}_{18}C_{18} (1/30)^{18} \cdot (1 - 1/30)^0 = 2.58 \times 10^{-27}$

解釈

① 観測事象は稀ではなく、誤った前提により稀な確率が計算された。

② 非常に稀だが全く起こりえないわけではない事実が起こった。

] ex) ここまで

仮定の有意性の検証 = 仮説検定

検定の流れ

(1) 仮説の設定 ← (1)これが注

(2) 検定統計量とその分布の決定

(3) 有意水準と棄却域の決定

(4) 検定の実施

ex) 母分散既知(あまりないけど)の正規分布からの小標本に対する検定 (cf 5.4 の表の②)

新製品の寿命のメーカー公称値: 9hour

旧製品: $\mu = 8$hour, $\sigma = 1$hour 〜 正規分布

(新製品の旧製品より性能upしてない?)

新製品で10回実験 → $n = 10$, $\bar{x} = 8.8$

(1) 仮説の設定 ← 知りたいことと逆のことを仮定

新製品も旧製品 $\mu = 8$ と同じではないか?? ← これを検証

H_0: $\mu = 8$ 帰無仮説

データが H_0 から著しく外れたときのみ H_0 を棄却

H_1: $\mu > 8$ 対立仮説

	H_0 が正しい	H_0 は誤り
H_0 を棄却しない	①	③
H_0 を棄却する	②	④

①, ④ は正しい検証

② H_0 は正しい → H_0 棄却 : 第I種の過誤

③ H_0 は誤り → H_0 採択 : 第II種の過誤

NO. 22
DATE

☆☆ 正しいというのは難しい(証明難)
☆ → 別のものをもってきて間違いだよ
検定の考え方 というのがラク

上記では、③より②の仮説を大きな誤りと考える。
↑ 背理法の考え(棄却することが目的)
H_0採択でも、積極的に支持したわけではなく「矛盾はしない」だけ

(2) 検定統計量とその分布の決定
μの良い推定量：$\bar{X}$ → 検定にも$\bar{X}$を用いる：検定統計量
n：小、σ^2既知　$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$　(cf 6.3(1))
(σ^2未知ならt分布)

6/11

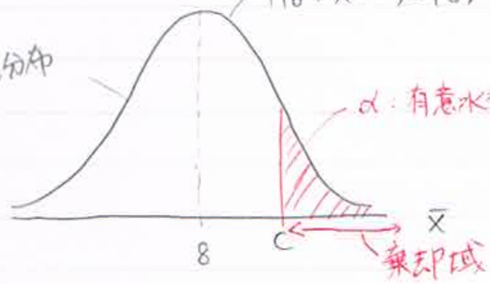
復 検定の流れ
(1) 仮説の設定
(2) 検定統計量とその分布の決定
(3) 有意水準と棄却域の決定　今日はここから
(4) 検定の実施

(1) 仮説の設定　これを棄却していく
帰無仮説 H_0：$\mu = 8$、　対立仮説：$\mu > 8$
(2) 検定統計量とその分布
検定統計量：$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$　例の8パターンで分類

(3) 有意水準と棄却域の決定
$\bar{X}$の値が8を越え大きいほど、H_0：$\mu = 8$を棄却する証拠
H_0を棄却すべき$\bar{X}$の領域：棄却域R
||
$\bar{X}$がある値cを越える領域
$R = \{\bar{x} ; \bar{x} > c\}$ ← 片側検定
(cf $R = \{\bar{x} ; |\bar{x}| > c\}$：両側検定 ($H_1$：$\mu \neq 8$))
H_0を棄却 ← データがH_0から著しくはずれる
きわめて小さな確率αでしか生じない (0.05か0.01が多い)
= 有意水準

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ だった　前回の値から

$H_0 : \bar{X} \sim N(8, \frac{1}{10})$



α：有意水準 → 今回は0.01とする

cf) 両側検定

注 有意水準 $\alpha = 0.01$ のとき、面積は0.005ずつ

前提　αに依存

$P(\bar{X} > c \mid H_0 \text{正しい}) = 0.01$ となる c を計算

$Z = \frac{\bar{X} - 8}{1/\sqrt{10}} \sim N(0,1)$ 　$\left(\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}\right)$　$\therefore P\left(Z > \frac{c-8}{1/\sqrt{10}}\right) = 0.01$　$(Z > Z_{0.01})$

所与

標準正規分布の右片側0.01の確率を考える．$Z_{0.01} = 2.326$

$Z = \frac{c-8}{1/\sqrt{10}} = 2.326$　$\therefore c = 8.74$

$\therefore$ 棄却域 $R = \{\bar{X} ; \bar{X} > 8.74\}$ ← 棄却域が決まった!!

(4) 検定の実施 ← 実際のデータから行う

$$\begin{cases} \bar{X} \in R \Rightarrow H_0 \text{棄却} \ (H_1 \text{採択}) \\ \bar{X} \notin R \Rightarrow H_0 \text{棄却しない} \end{cases}$$

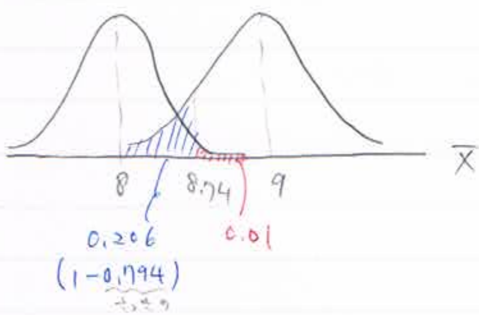
今回は、$\bar{x} = 8.8 \in R \{\bar{x} ; \bar{x} > 8.74\}$ ⇒ 有意水準1%で $H_0 : \mu = 8$ は棄却される．

$P\left(Z = \frac{8.8-8}{1/\sqrt{10}}\right) = 0.0059$：P値

おまけ 前提を変えてみる．　H_1を帰無仮説にしてみる　（一応出してみた）

$P(\bar{X} > 8.74 \mid H_1 \text{正しい（例えば} \mu = 9 \text{とする））} = P\left(Z > \frac{8.74-9}{1/\sqrt{10}}\right) = 0.794$

・過誤について



第1種の過誤の確率

$$\begin{cases} P(\mathrm{I}) = P(\bar{X} > 8.74 \mid \mu = 8) = 0.01 = \alpha \\ P(\mathrm{II}) = P(\bar{X} \leq 8.74 \mid \mu = 9) = 0.206 = \beta \end{cases}$$

αを小さくしたい ⇒ βは大きくなる．
（αとβは同時に小さくできない）
$\therefore$ 第I種の過誤の危険 → 小
⇒ 第II種の過誤の危険 → 大

α：有意水準 = 危険率

$1-\beta$：検出力

H_0が誤っているときにちゃんと棄却できる確率．

7.2 母平均の検定

7.1の通りにやってく

(1) 仮説の設定

(A) $H_0: \mu = \mu_0,\ H_1: \mu \neq \mu_0$ → 両側検定
(B) $H_0: \mu = \mu_0,\ H_1: \mu > \mu_0$ → 右片側検定
(C) $H_0: \mu = \mu_0,\ H_1: \mu < \mu_0$ → 左片側検定

(2) 検定統計量と分布

ノートP.13

5.4.「標本平均の分布のまとめ」の表で分類

(a) ①, ②, ⑤ $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ （③ ④では $\sigma \to s$）

(b) ④ $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} \sim t(n-1)$ ← t検定 平均の検定

(3) 棄却域

	a	b
A	$\lvert Z\rvert > Z_{\alpha/2}$	$\lvert t\rvert > t_{\alpha/2}(n-1)$
B	$Z > Z_\alpha$	$t > t_\alpha(n-1)$
C	$Z < -Z_\alpha$	$t < -t_\alpha(n-1)$

t分布は正規分布と同様左右対称
（すそがちょっと重い）

7.3 正規分布の母分散の検定

(1) 仮説

(A) $H_0: \sigma^2 = \sigma_0^2,\ H_1: \sigma^2 \neq \sigma_0^2$ → 両側検定
(B) $H_0: \sigma^2 = \sigma_0^2,\ H_1: \sigma^2 > \sigma_0^2$ or $\sigma^2 = \sigma_a^2$ → 右片側検定 （$\sigma_a^2 > \sigma_0^2$）
(C) $H_0: \sigma^2 = \sigma_0^2,\ H_1: \sigma^2 < \sigma_0^2$ or $\sigma^2 = \sigma_b^2$ → 左片側検定 （$\sigma_b^2 < \sigma_0^2$）

(2) 検定統計量と分布

$\chi^2 = \sum \dfrac{(x_i - \bar{x})^2}{\sigma_0^2} = \dfrac{(n-1)s^2}{\sigma_0^2} \sim \chi^2(n-1)$ ← χ^2検定

(3) 棄却域 (cf. 6.3(2))

χ^2のグラフは左右対称じゃないのでこんな書き方になる

A: $\chi^2 < \chi^2_{1-\frac{\alpha}{2}},\ \chi^2 > \chi^2_{\alpha/2}$
B: $\chi^2 > \chi^2_\alpha$
C: $\chi^2 < \chi^2_{1-\alpha}$

ex). 正規分布と仮定したものが本当に正規分布か

※ χ^2検定は適合度検定にも用いられる
※ 母分散の比の検定 → F検定
母平均の差の検定 → t検定
（おまけ）

空間情報学でも使うよ.

§8. 回帰分析

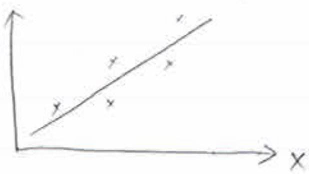
8.1. 回帰モデル

2変数：X, Y

Yの変動をXで説明：XとYの定量的な関係(モデル)を知りたい.

X：説明変数，独立変数 ← 横軸が多い
Y：被説明変数，従属変数 ← 縦軸 〃

ex). Y：ばねの長さ，X：おもりの重さ



誤差項

$Y = \beta_0 + \beta_1 X + \varepsilon$ ← 回帰式

回帰式が直線のとき，線形回帰 ↔ 非線形回帰モデル

β_0, β_1：回帰係数，ε：誤差項

(注) ここでの線形は，β_0, β_1に関してである

$Y = \beta_0 + \beta_1^2 X$ ← 非線形
$Y = \beta_0 + \beta_1 X^2$ ← 線形

説明変数 1つ：単純回帰モデル
〃 複数：重回帰モデル

◦ (線形)単純回帰モデル ← 最も簡単なもの

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ (β_0, β_1：パラメータ)　(i：サンプル)

☆☆☆ ちゃんとおさえる.

仮定

cf). コロケーション (Xも確率変数)

〈Xへの仮定〉 所与のデータ

① Xは非確率変数

〈εへの仮定〉

② $E(\varepsilon_i) = 0$ ($i = 1, 2, \cdots, n$)

③ 分散一定：$var(\varepsilon_i) \overset{②より}{=} E(\varepsilon_i^2) = \sigma^2$ ($i = 1, 2, \cdots, n$)

④ 異なる誤差項は無相関(共分散0)：$cov(\varepsilon_i, \varepsilon_j) \overset{②より}{=} E(\varepsilon_i \varepsilon_j) = 0$ ($i \neq j$, $i, j = 1, 2, \cdots, n$)

⑤ ε_iは正規分布

誤差の諸要因：$\varepsilon_{i1}, \varepsilon_{i2}, \cdots$ → $\varepsilon_i = \varepsilon_{i1} + \varepsilon_{i2} + \cdots$

中心極限定理より，ε_iは正規分布

②〜⑤より，$\varepsilon_i \sim N(0, \sigma^2)$

ここで、$E(Y_i) = \beta_0 + \beta_1 x_i$ （①，②より）

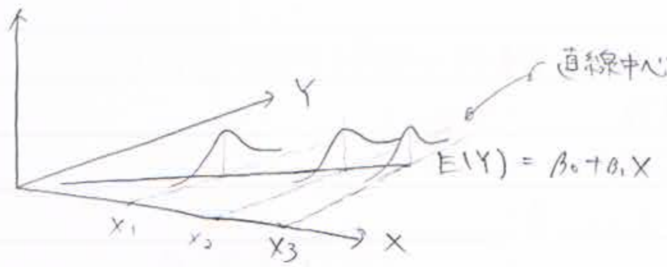
∴ $Y_i = E(Y_i) + \varepsilon_i$ ← Y_iの平均的な大きさ＋誤差

$var(Y_i) = E[\underbrace{Y_i - E(Y_i)}_{\varepsilon_i}]^2 = E(\varepsilon_i^2) = var(\varepsilon_i) = \sigma^2$ （③より）

$cov(Y_i, Y_j) = E[(Y_i - E(Y_i))(Y_j - E(Y_j))] = \varepsilon_i \varepsilon_j = 0$ （④より）

以上から、仮定のイメージ

直線中心の正規分布で分散同じ.



・目的.

① 未知パラメータ $\beta_0, \beta_1, \sigma^2$を推定　　最小二乗法

② モデルの説明力の確認

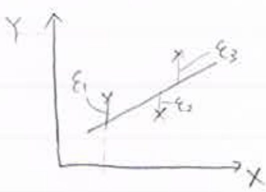
③ 推定量の統計的推測

→ 次回.

6/18

8.2. 最小二乗法

(1) 回帰係数の推定



$\varepsilon_i = Y_i - (\beta_0 + \beta_1 X_i)$　　全部正にする

誤差の二乗和を最小にしたい

$\sum \varepsilon_i^2 = \sum \{Y_i - (\beta_0 + \beta_1 X_i)\}^2 \longrightarrow \min$: 最小二乗法 (OLS)　　cf). GLS

$$\begin{cases} \dfrac{\partial \sum \varepsilon_i^2}{\partial \beta_0} = -2\sum(Y_i - \beta_0 - \beta_1 X_i) = 0 \\ \dfrac{\partial \sum \varepsilon_i^2}{\partial \beta_1} = -2\sum(Y_i - \beta_0 - \beta_1 X_i)\cdot X_i = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} n\beta_0 + \beta_1 \sum X_i = \sum Y_i \\ \beta_0 \sum X_i + \beta_1 \sum X_i^2 = \sum X_i Y_i \end{cases}$$

正規方程式

$\bar{X} = \frac{1}{n}\sum X_i$, $\bar{Y} = \frac{1}{n}\sum Y_i$ とおくと、　$\beta_0 = \bar{Y} - \beta_1 \bar{X}$　これを代入して、

$\beta_1 (\sum X_i^2 - \bar{X} \sum X_i) = \sum X_i Y_i - \bar{Y} \sum X_i$

$\Leftrightarrow \beta_1 \sum (X_i - \bar{X})^2 = \sum (X_i - \bar{X})(Y_i - \bar{Y})$

$(\bar{X}, \bar{Y})$は必ず通る

$$\boxed{\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}}$$ 最小二乗推定量

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ ← データX_iからY_iを推測する式

(残差と説明変数は直交)

$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$: 残差 $\neq$ 誤差ε_i

$\sum e_i = 0$, $\sum e_i X_i = 0$ ←(多次元 $e_i \cdot X_i = 0$)

$E(\hat{\beta}_0) = \beta_0$, $E(\hat{\beta}_1) = \beta_0$ → 不偏推定量　理由は省略

$\mathrm{var}(\hat{\beta}_0) = \frac{\sigma^2 \sum X_i^2}{n \sum (X_i - \bar{X})^2}$, $\mathrm{var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}$ ← 最小分散(有効性)

(計算過程より)

※ 最小二乗推定量は BLUE (cf 6.2)

P.25で仮定したσ^2. 誤差の分散を残差の分散から推定する

(2) 分散の推定

σ^2の推定量　$s^2 = \frac{\sum e_i^2}{n-2}$ ← パラメータ β_0, β_1 の2つより自由度は $n-2$

∵ $\sum e_i = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$ ← $\frac{\partial \sum \varepsilon_i^2}{\partial \beta_0} = 0$ から導かれる

$\sum e_i X_i = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i = 0$ ← $\frac{\partial \sum \varepsilon_i^2}{\partial \beta_1} = 0$ 〃

上記2つの制約によって、自由度は2減る

一般的には、データ数n、回帰係数(パラメータ)の数P → 自由度 $n - P$

(cf 最尤法)

(3) 最尤推定量との関係

P.25

仮定より $\varepsilon_i \sim N(0, \sigma^2)$, Y_iはε_iの線形関数, 均一分散

∴ $Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$ ←(P.26の図)

$Y_1, \dots, Y_n$ の pdf (同時確率密度関数)は、

nコかけてる

$$(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i)^2 \right\} = L(\beta_0, \beta_1, \sigma^2)$$ ← 尤度関数

対数とって考えて、

$$\begin{cases} \frac{\partial \log L}{\partial \beta_0} = \frac{1}{\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i) = 0 \\ \frac{\partial \log L}{\partial \beta_1} = \frac{1}{\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i) X_i = 0 \\ \frac{\partial \log L}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (Y_i - \beta_0 - \beta_1 X_i) = 0 \end{cases}$$

(σ²で一階微分)

最小二乗法と同じ式!!

(誤差項に正規分布を仮定していたため)

∴ $\hat{\beta}_{0u} = \hat{\beta}_0$, $\hat{\beta}_{1u} = \hat{\beta}_1$, $\hat{\sigma}_u^2 = \frac{1}{n} \sum e_i^2$

(最尤法のu)

→ 不偏でない (cf. P.16)

ε_iが正規分布という仮定のもとでは、β_0, β_1の最尤推定量と最小二乗推定量は一致(分散は異なる)

8.3 モデルの説明力

X が Y をどの程度説明できるか ⟶ 回帰モデルの当てはまりのよさ

ここで、残差 $e_i = Y_i - \hat{Y}_i$ ⟶ $Y_i = \hat{Y}_i + e_i$　両辺から $\bar{Y}$ 引いて、

$Y_i - \bar{Y} = \hat{Y}_i - \bar{Y} + e_i$　この2乗和をとると、

$$\Sigma(Y_i - \bar{Y})^2 = \Sigma(\hat{Y}_i - \bar{Y})^2 + \Sigma e_i^2 + 2\Sigma(\hat{Y}_i - \bar{Y})e_i$$

$$(\text{右辺第3項}) = \Sigma(\hat{Y}_i - \bar{Y})e_i = \Sigma \hat{Y}_i e_i \quad (\because \Sigma \bar{Y} e_i = \bar{Y}\Sigma e_i = 0)$$

$$= \Sigma(\hat{\beta}_0 + \hat{\beta}_1 X_i)e_i = 0 \quad (\because \Sigma e_i = 0,\ \Sigma X_i e_i = 0)$$

$$\therefore \underbrace{\Sigma(Y_i - \bar{Y})^2}_{\text{全変動}} = \underbrace{\Sigma(\hat{Y}_i - \bar{Y})^2}_{\text{モデルによる説明}} + \underbrace{\Sigma e_i^2}_{\text{残差平方和}}$$

全変動をモデルでどれくらい説明できてるか

$$\boxed{\therefore \text{説明力の尺度} = \text{決定係数 } r^2 = \frac{\Sigma(\hat{Y}_i - \bar{Y})^2}{\Sigma(Y_i - \bar{Y})^2} = 1 - \frac{\Sigma e_i^2}{\Sigma(Y_i - \bar{Y})^2} = \rho_{Y\hat{Y}}^2 \quad 0 \le r^2 \le 1}$$

(相関係数の2乗)

$r^2 = 1$ のとき、$\Sigma e_i^2 = 0 \Rightarrow \hat{Y}_i = Y_i$

$r^2 = 0$ のとき、$\Sigma(\hat{Y}_i - \bar{Y})^2 = 0 \Rightarrow \hat{Y}_i = \bar{Y}$

8.4 統計的推測

各パラメータ X_i の力がそれぞれどのくらいあるか

$\varepsilon_i \sim N(0, \sigma^2)$　$\hat{\beta}_0, \hat{\beta}_1$: ε_i の線形関数 ⟶ 正規分布（再生性）

$$E(\hat{\beta}_0) = \beta_0,\ E(\hat{\beta}_1) = \beta_1,\ var(\hat{\beta}_0) = \frac{\sigma^2 \Sigma X_i^2}{n\Sigma(X_i - \bar{X})^2},\ var(\hat{\beta}_1) = \frac{\sigma^2}{n\Sigma(X_i - \bar{X})^2}$$

$$\therefore \hat{\beta}_0 \sim N(\beta_0, var(\hat{\beta}_0)),\quad \hat{\beta}_1 \sim N(\beta_1, var(\hat{\beta}_1))$$

σ^2 未知 ⟶ $s^2 = \dfrac{\Sigma e_i^2}{n-2}$ でおきかえる。

$\hat{\beta}_0, \hat{\beta}_1$ の分散の不偏推定量は、

$$S_{\hat{\beta}_0}^2 = \frac{s^2 \Sigma X_i^2}{n\Sigma(X_i - \bar{X})^2},\quad S_{\hat{\beta}_1}^2 = \frac{s^2}{n\Sigma(X_i - \bar{X})^2}$$

母集団正規分布，σ^2 未知 より (cf P.13)

$\therefore \hat{\beta}_0, \hat{\beta}_1$ を標準化したものは、自由度 $n-2$ の t分布 になる。

$$t_0 = \frac{\hat{\beta}_0 - \beta_0}{S_{\hat{\beta}_0}} \sim t(n-2),\quad t_1 = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t(n-2)$$

($H_0 : \hat{\beta}_0 = 0$ とすることが多い)

⇒ パラメータに対する区間推定 (cf 6.3) や仮説検定 (cf 7) が可能

(この講義では $A^t = A'$ とする)

単回帰 (x が1つ)
(x がたくさん)

8.5. 重回帰モデル

線形回帰モデル

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_q X_{iq} + \varepsilon_i$$

↓ 行列にする

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix},\quad X = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1q} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{nq} \end{bmatrix},\quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_q \end{bmatrix},\quad \boldsymbol{\varepsilon} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$ とすると、

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

ゼロベクトル

おくべき仮定は 8.1 と同じ。ただし、$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I)$

単位行列

転置

最小二乗法 $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \to \min$ ∴ $\frac{\partial \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta}} = 0$

$\longrightarrow \hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$ ☆☆☆ 覚える , $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, $var(\hat{\boldsymbol{\beta}}) = \sigma^2(XX')^{-1}$

残差: $\mathbf{e} = \mathbf{y} - X\hat{\boldsymbol{\beta}}$

不偏分散 $s^2 = \frac{\mathbf{e}'\mathbf{e}}{n-p} = \frac{(\mathbf{y}-X\hat{\boldsymbol{\beta}})'(\mathbf{y}-X\hat{\boldsymbol{\beta}})}{n-p}$ $\quad (p = q+1,\quad p = rank(X))$

一般に、重回帰式の方が決定係数高い(説明力が上がる)が、自由度が減少するため不偏分散が大きくなり回帰直線の汎化性が低下する。

Akaike Information Criteria

説明変数の数も含めたモデルの決定 ⇒ 情報量規準 (AIC)

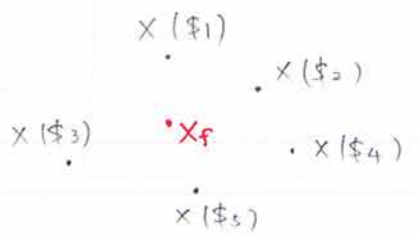
未御は 赤池先生

6/25

§9 時系列解析 (中西さん) → プリント

§10 空間データの統計解析

降雨量, 標高などをイメージ



$X(s_1) \sim X(s_5)$ は既知

未知点 X_f の値を予測したい

いいすぎ

観測値 $X(s_i)$ が互いに独立 → X_f の予測値は μ

通常は独立ではなく、近い所ほど近い値を持つ傾向 = 空間相関

空間相関を考慮し、よりよい予測値を求める。アプローチは2つ。

- 空間計量経済学 (10.2)
- 空間統計学, 地球統計学 (10.3, 10.4)

DATE

（分散，共分散をまとめて行列にしてる）

2つのアプローチで共通　試験出るかも

10.1. 一般化最小二乗法

重回帰分析（cf 8.5）：通常最小二乗法（ordinary least squares : OLS）

誤差項において空間相関を考慮：一般化最小二乗法（generalized least squares : GLS）

- 線形回帰モデル

$$y = X\beta + \varepsilon$$

自由回帰と仮定がやや異なる

- 仮定

① $E(\varepsilon) = 0$ ← これは一緒

分散 ② $E(\varepsilon\varepsilon') = V = \sigma^2\Omega$ ← 共分散も考慮

(cf) OLS : $\varepsilon \sim N(0, \sigma^2 I)$　共分散なし

Ω：正値定符号行列 $\Leftrightarrow \forall x,\ x'\Omega x > 0$

③ ε：多変量正規分布 ← 一緒

Ωは正値定符号行列だから $P\Omega P' = I$ なる P が存在（$\Omega^{-1} = P'P$）　単位行列

$Py = PX\beta + P\varepsilon$　← Pかけただけ

ここで $E(P\varepsilon) = 0$，$var(P\varepsilon) = E(P\varepsilon\varepsilon'P') = \sigma^2 I$　← OLS的だ!!

∴ Pyを被説明変数，PXを説明変数としてOLSで推定可能

∴ $$\boxed{\begin{aligned}\hat{\beta}_G &= [(PX)'(PX)]^{-1}(PX)'Py = (X'P'PX)^{-1}X'P'Py \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \ \left(= (X'V^{-1}X)^{-1}X'V^{-1}y\right)\end{aligned}}$$

p.129みて　$\Omega^{-1} = P'P$　← $V = \sigma^2\Omega$

平均と分散

代入

$\hat{\beta}_G = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}(X\beta + \varepsilon) = \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon$　だから，

$E(\hat{\beta}_G) = E(\beta) = \beta$　（∵ $E(\varepsilon) = 0$）

$var(\hat{\beta}_G) = \sigma^2(X'\Omega^{-1}X)^{-1} = (X'V^{-1}X)^{-1}$　← 覚えなくてよい

$$S_G^2 = \frac{(y - X\hat{\beta}_G)'\Omega^{-1}(y - X\hat{\beta}_G)}{n - p}$$

OLSとの違いをチェック

V（or Ω）が定義できれば，空間相関を考慮したモデルができる

この定義の仕方がアプローチの違いとなる．

10.2～10.4はあまり試験に出ないよ
存在だけ知っておこう

10.2 空間自己回帰モデル

誤差項をモデル化することにより、間接的に V をモデル化

$y = X\beta + u$

$u = \lambda W u + \varepsilon$ 　λ：係数，W：空間重み行列　ex). $W = [w_{ij}] = [d_{ij}^{-\alpha}]$ 　キョリの関数

(自己で回帰してる)　error component model

$$\left.\begin{aligned} \mathrm{var}(u) &= V = \sigma^2\Omega \\ \mathrm{var}(\varepsilon) &= \sigma^2 I \end{aligned}\right\} \text{と仮定}$$

$u = (I - \lambda W)^{-1}\varepsilon$ 　OLSとかGLSみて

$\therefore V = E(uu') = \sigma^2[(I-\lambda W)'(I-\lambda W)]^{-1}$

これを用いてGLSによりパラメータ推定

10.3 確率場と定常性

位置に依存した確率変数

位置 $s \in \mathbb{R}^2$ に対して確率変数 $Z(s)$ が対応

$\{Z(s) : s \in \mathbb{R}^2\}$：確率場　← cf). 確率過程 (§9)

統計解析のための $Z(s)$ への仮定：定常性

◦ 強定常　← おまけ程度

同時分布関数 $F_{s_1,\ldots,s_n}(Z(s_1),\ldots,Z(s_n))$ とする。

$\forall s_1,\ldots,s_n, h\,(\in\mathbb{R}^2)$ に対して

$F_{s_1,\ldots,s_n}(Z(s_1),\ldots,Z(s_n)) = F_{s_1+h,\ldots,s_n+h}(Z(s_1),\ldots,Z(s_n))$

⇓

・$Z(s)$ の分布関数は任意の位置 s で同じ

・$Z(s_1),\ldots,Z(s_n)$ の同時分布関数は，位置の差 h のみに依存

強定常：分布そのものについての仮定

通常 $F_{s_1,\ldots,s_n}$ は未知 → 仮定をゆるめる（弱定常）← 大事

ムシっていーよ

◦ 2次定常性 ← ここからまた大事

$$\begin{cases} E[Z(s)] = \mu \\ \mathrm{cov}[Z(s_i), Z(s_j)] = C(s_i - s_j) = C(h) \end{cases}$$

h：点 s_i, s_j の相対的位置（ベクトル）

$C(h)$：共分散関数 = コバリオグラム ⟷ 空間重み行列

さらに簡単にしたい

次ページ

NO. 32

DATE

☆ ・等方的2次定常性

$$\begin{cases} E[z(s)] = \mu \\ cov[z(s_i), z(s_j)] = c(h) \end{cases}$$

スカラー.

h：点 s_i, s_j の距離（スカラー）

$C(h)$ を定義 → V (or Ω) が定義できる → 次

任意の位置の推定値を調べようとした

10.4 クリギング

(1) 一般的な予測

既知点：z_i $(i=1,\ldots,n)$

未知点：z_f ← forecast

周りから未知のものを推定

キョリの

重みつき平均（内挿）

$$\hat{z}_f = \lambda_1 z_1 + \cdots \lambda_i z_i + \cdots + \lambda_n z_n = \lambda' z \qquad (\lambda：重み)$$

(2) 誤差項の予測

空間相関を誤差項で考える → 未知点の残差を既知点の残差で予測

重みつき平均

$$\hat{\varepsilon}_f = \lambda_1 \varepsilon_1 + \cdots + \lambda_n \varepsilon_n = \lambda' \varepsilon$$

既知点の残差：$\varepsilon = (y - X\hat{\beta}_G)$

ここで、既知点間の分散共分散行列：$V = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{bmatrix}$

既知点と未知点の共分散ベクトル：$c = \begin{bmatrix} \sigma_{1f} \\ \vdots \\ \sigma_{nf} \end{bmatrix} = \begin{bmatrix} c(|s_f - s_1|) \\ \vdots \\ c(|s_f - s_n|) \end{bmatrix}$ キョリ（コバリオグラム）なのでノルムをとる

誤差と予測残差の差を最小とする重み　ε_f ($\lambda'\varepsilon$ と λ)

$$E[(\varepsilon_f - \hat{\varepsilon}_f)^2] = E(\varepsilon_f^2) - 2E(\varepsilon_f \cdot \hat{\varepsilon}_f) + E(\hat{\varepsilon}_f^2)$$

$$= \sigma^2 - 2\lambda' c + \lambda' V \lambda \longrightarrow \min$$

とおく；　c に関係するのと ε_f の式から考えて.

$$\frac{\partial E[\cdots]}{\partial \lambda} = -2c + 2V\lambda = 0 \qquad \therefore \lambda' = c' V^{-1}$$

以上より，

$$\hat{\varepsilon}_f = \lambda' \varepsilon = c' V^{-1}(y - X\hat{\beta}_G)$$

テスト
- 確率分布はキホン
- 点推定(けっこう出されてる)
- 区間推定 or 仮説検定
- 回帰分析(通常の→一般最小二乗法)の主成分との違い

説明変数のもの 出るかも

ここで 線形回帰モデル

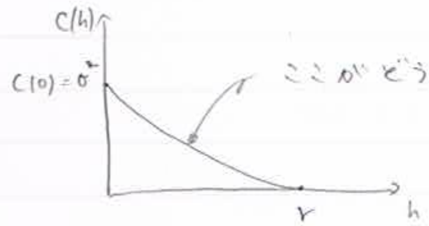
x_f : 未知点の説明変数ベクトル
y_f : 〃 被説明変数
ε_f : 〃 誤差

とすると、

$$\boxed{\hat{y}_f = x_f'\hat{\beta}_G + \hat{\varepsilon}_f = x_f'\hat{\beta}_G + c'V^{-1}(y - X\hat{\beta}_G)}$$

モデル推定の残差

(3) コバリオグラムの例



ここがどういう曲線かは色々ある.

- 指数型 $C(h) = \sigma^2 \exp\left(-\frac{h}{r}\right)$
- ガウス型 $C(h) = \sigma^2 \exp\left[\left(\frac{h}{r}\right)^2\right]$
- 球状型 $C(h) = \sigma^2\left[1 - \frac{3}{2}\left(\frac{h}{r}\right) + \frac{1}{2}\left(\frac{h}{r}\right)^3\right]$

覚えなくてよい.

7/7

§11 多変量解析

11.1 主成分分析 (PCA)

principal component analysis

統計由来

ex) ある人の各科目の得点、$x_1, \cdots, x_p$

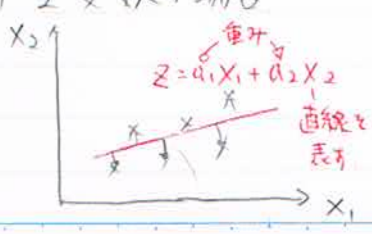
総合得点 $z = x_1 + \cdots + x_p$ ← 本当に良い総合評価か?

各科目の得点分布にもばらつきがある. これを考慮 (重みをつける).

各点に重みづけして、総合得点を求める. より一般的には、

「各変数に重みづけし、より少ない新たな変数で総合化する方法
= 主成分分析 」

(1) 2変数の場合



x_2だけで評価 → 差がない

無意味

x_1 〃 → 総合化にならない

個人差がつく総合得点 ← なるべく全ての変数を使う

2次元 → 1次元

$\lambda^2 - (S_{11}+S_{22})\lambda + S_{11}S_{22} - S_{12}S_{21} = 0$

DATE

個人差をはっきりさせる

軸を変換したとき、その軸上での値の分散が最大になる軸を求める

$z = a_1 x_1 + a_2 x_2$ を考える

x_1, x_2 の（標本）分散共分散行列 $\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$ → z を考える

$\mathrm{var}(z) = \frac{1}{n-1} \sum (z_i - \bar{z})^2$ ← 標本からとってるので自由度1おちる

$= \frac{1}{n-1} \sum \{(a_1 x_{1i} + a_2 x_{2i}) - (a_1 \bar{x}_1 + a_2 \bar{x}_2)\}^2$

$= \frac{1}{n-1} \sum \{a_1^2 (x_{1i} - \bar{x}_1)^2 + 2a_1 a_2 (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) + a_2^2 (x_{2i} - \bar{x}_2)^2\}$

$= a_1^2 S_{11} + 2a_1 a_2 S_{12} + a_2^2 S_{22}$ → 最大化したい

制約条件：$a_1^2 + a_2^2 = 1$（式自体にイミはないが、制約ないと $\mathrm{var}(z)$ 無限に大きくできる）

制約条件つき最大化 → ラグランジュの未定乗数法

$f(a_1, a_2, \lambda) = a_1^2 S_1 + 2a_1 a_2 S_{12} + a_2^2 S_{22} - \lambda \underbrace{(a_1^2 + a_2^2 - 1)}_{\text{制約条件}}$

$$\begin{cases} \frac{\partial f}{\partial a_1} = 2a_1 S_1 + 2a_2 S_{12} - 2a_1 \lambda = 0 \\ \frac{\partial f}{\partial a_2} = 2a_1 S_{12} + 2a_2 S_{22} - 2a_2 \lambda = 0 \\ \frac{\partial f}{\partial \lambda} = a_1^2 + a_2^2 - 1 = 0 \end{cases}$$

上の2式から、 $\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \lambda \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ ：固有値問題になる

∴ λ：固有値，$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$：固有ベクトル を表す

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \Leftrightarrow \begin{pmatrix} S_{11} - \lambda & S_{12} \\ S_{21} & S_{22} - \lambda \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$a_1 = a_2 = 0$ 以外の解を持つためには、

$$\begin{vmatrix} S_{11} - \lambda & S_{12} \\ S_{21} & S_{22} - \lambda \end{vmatrix} = 0 \qquad \therefore \lambda = \frac{(S_{11} + S_{22}) \pm \sqrt{(S_{11} - S_{22})^2 + 4S_{12}^2}}{2}$$

$\lambda_1 > \lambda_2$ とすると、 $a_{1(1)} = \frac{S_{12}}{\sqrt{S_{12}^2 + (\lambda_1 - S_{11})^2}}$, $a_{2(1)} = \frac{\lambda_1 - S_{11}}{\sqrt{S_{12}^2 + (\lambda_1 - S_{11})^2}}$

（a_2 消去して $\frac{\partial f}{\partial \lambda} = 0$ に代入するのが良い）

同様に $\lambda = \lambda_2$ → $a_{1(2)}, a_{2(2)}$ が求まる

新しい軸 z_1：第1主成分，z_2：第2主成分 ← 固有ベクトルは直交（幾何学的には回転）

$z_1 = a_{1(1)} x_1 + a_{2(1)} x_2$ ：主成分得点　（z_2 は略する）

固有値問題　　↙対角化

$$\begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

cf) 2つの最小二乗法

分散共分散行列の対角化 → z_1とz_2は無相関　　処理を簡単にする

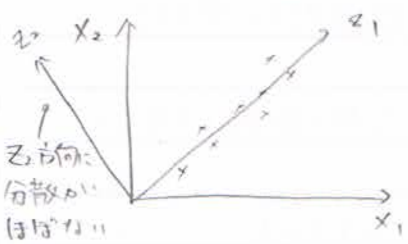
∴主成分分析：相関のあるデータを無相関のデータに変換

・z_1で全体をどれだけ表現できるか

z_1軸上の分散：λ_1、z_2軸上の分散：λ_2

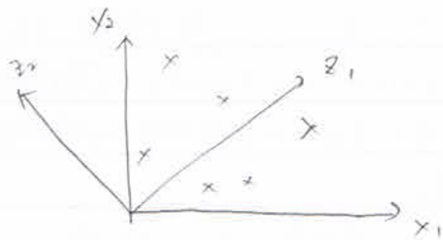
寄与率（第1主成分）：$\frac{\lambda_1}{\lambda_1+\lambda_2}$

→ 大：元のデータの違いを十分に表現
　 小：　　〃　　できていない



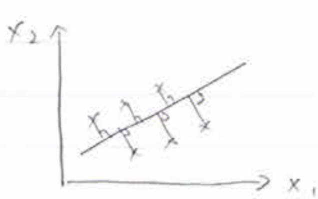
寄与率（大）

⇒ 次元圧縮



寄与率（小）

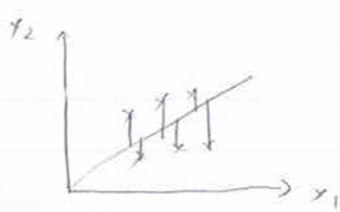
・回帰分析との相違　　同じデータ



主成分

x_1とx_2同等

→ 主成分へのキョリ



回帰

x_1でx_2を説明

→ x_2軸上での差

(2) 多変数の場合

$z = a_1 x_1 + \cdots + a_p x_p$

$var(z) = \frac{1}{n-1} \sum \{ a_1(x_{1i} - \bar{x}_1) + \cdots + a_p(x_{pi} - \bar{x}_p) \}^2 = \sum_j \sum_k S_{jk} a_j a_k$

制約条件 $a_1^2 + \cdots + a_p^2 = 1$

ラグランジュの未定乗数法から、

$$\begin{pmatrix} s_{11} & \cdots & s_{1p} \\ \vdots & & \vdots \\ s_{p1} & \cdots & s_{pp} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} = \lambda \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix}$$

: 固有値問題を解く（対角化）

11.2. 因子分析

心理学由来、さわりだけ

p 変数の測定値、データ数：n

$n > p+1$ ならば、最大限 p コの主成分が得られる

第1主成分の寄与率が著しく大きく、第2主成分以下の寄与率が低くなることが望まれる

しかしこれはきわめてまれ → 主成分をいくつで止めたらよいか？

そこで、寄与率の小さい主成分は誤差成分としてまとめ、できる限り少数の意味ある主成分を得ることを考える。

変数 $x_1, \cdots, x_p$、主成分 $z_1, \cdots, z_r$ として、次のモデルを考える。

$$\begin{cases} x_1 = a_{11} z_1 + \cdots + a_{1r} z_r + \varepsilon_1 \\ \vdots \\ x_p = a_{p1} z_1 + \cdots + a_{pr} z_r + \varepsilon_p \end{cases}$$

x を z で表している
（主成分と全く逆!!）

主成分 z → 因子分析では 共通因子 とよぶ

以下 f とする

$x_j = a_{j1} f_1 + \cdots + a_{jr} f_r + \varepsilon_j \quad (j = 1, \cdots, p)$

x_j の変動をいくつかの共通因子の加重和に分解

a_{jk} : 変数への共通因子の影響度合 = 因子負荷量

ε_j : 誤差 = 独自因子

複数の変数からなるデータを少数の共通因子を使って表現し、データが現れる仕組みのモデルを見つけ出す方法 = 因子分析（因子負荷量と共通因子の値を求める）

cf) 主成分：与えられた変数を合成して全体の変動をなるべく少数の主成分で説明
　因子分析：ある変数の共通変動のみをいくつかの因子に分解