

最初から最後まで、数式、定理、証明のみ。抽象論。花文字。
具体例が一つもない。

何のためにという説明がない。歯が立たない。

「統計学 One point」は、大学で統計論を履修している学生の副読本にも活用を考えているようだ。

まえがき

通常の統計学では、あらかじめ標本の大きさを固定して考えるが、それには限界があり、そこを乗り越えるためには逐次的手法を考える必要がある。たとえば、標本の大きさをあらかじめ固定せずに、標本を抽出してそれに基づく推定量とさらに標本を追加してそれに基づく推定量について、それらの標本抽出費用（コスト）を含むリスクを比較して、追加した方のリスクが小さければ、標本抽出を継続し、そうでなければ標本抽出を停止するという規則を導入する。そのような停止（規則）と推定量からなる逐次推定方式を考えることになるが、そのためには**統計的決定論**の観点から考えるのが理解しやすい。本書はその観点から著されていて、逐次推定方式が何らかの意味で最適になることを示す際に、そのリスクが、標本が抽出される母集団分布の母数に無関係になる方が望ましい。そのためにベイズ的観点から、あるいは不変性の概念によって考察することは自然である。

本書では、まず、正規モデル、2項モデル等で逐次点推定、逐次区間推定について考え、また、逐次推定を遂行する際に必要となる停止別に関する等式、不等式について述べるとともに近似法則についても論じる（第1章）。次に、統計的決定論は、推定、検定を統一的に扱えるだけでなく、標本抽出（または実験）を行ってその（観測）結果に基づいて抽出（観測）を継続するか停止するかを決めるに伴う逐次推測方式を考察するのに適している。そこで、決定論の観点から許容性、ベイズ性、ミニマックス性、縮小推定、最小分散不偏性、十分性、不変性、推移性等について考えるとともに、標本の大きさを固定して推定すると望ましい推定量が存在しない場合があることも示す（第2章）。それから、非逐次の場合に一樣最小分散不偏（UMVU）推定量を求めるために用いた十分性、完備性のような概念が逐次の場合には、どのような状況で有用になるのかについて考える。実は、逐次推定の場合には、与えられた停止則に伴う、完備十分統計量に基づく不偏逐次推定方式は一意的とは限らない。また、逐次の場合にもクラメル・ラオ型の情報不等式について論じることができる。そしてベイズ逐次推定、不変逐次推定についても考える（第3章）。さらに、停止時刻に伴う逐次推定方式について論じ、特に、正規分布における平均や平均の差の逐次推定、多変量正規分布の平均ベクトルの逐次推定と逐次縮小推定、ガンマ分布、一樣分布の逐次推定について考え、2段階（標本抽出）法による推定についても論じる（第4章）。最後に、逐次解析において重要な逐次確率比検定について述べる（付録）。逐次推測の考え方は合理的であり、応用についても、たとえば臨床試験における実験計画等において有用になる。なお、本書では数理統計学の基盤の知識を前提とし、巻末において参考にした書籍について言及している。

記号・略号

\mathbf{R}^n	n 次元ユークリッド空間
$\mathbf{R}_+, \mathbf{R}_-$	$\mathbf{R}_+ = (0, \infty), \mathbf{R}_- = (-\infty, 0)$
\mathbb{N}	自然数全体の集合
\mathbb{N}_0	$\mathbb{N} \cup \{0\}$
a^+, a^-	$a^+ = \max\{0, a\}, a^- = \max\{-a, 0\}$
A^c	集合 A の補集合 (complementary set)
$A \times B$	集合 A, B の直積集合 $A \times B = \{(a, b) \mid a \in A, b \in B\}$
$\chi_A(x)$	集合 A の定義関数, すなわち $\chi_A(x) = 1 (x \in A);$ $= 0 (x \notin A)$
$\mathbf{a}^\top, \Sigma^\top$	ベクトル \mathbf{a} の転置ベクトル, 行列 Σ の転置行列
$\text{tr} A$	正方行列 A のトレースまたは跡 (trace) (対角成分の和)
$\ \mathbf{a}\ $	$\mathbf{a} = (a_1, \dots, a_p)$ のノルム (norm) $\ \mathbf{a}\ = \sqrt{\mathbf{a}\mathbf{a}^\top}$
$X_n \xrightarrow{L} X (n \rightarrow \infty)$	確率変数 X_n が確率変数 X に法則収束 (convergence in law) (1.5 節)
$\mathcal{L}(X_n) \rightarrow F (n \rightarrow \infty)$	確率変数 X_n が漸的に分布 F に従う (1.5 節)
$X_n \xrightarrow{P} X (n \rightarrow \infty)$	確率変数 X_n が確率変数 X に確率収束 (convergence in probability) (1.5 節)
$a.s.$	ほとんど確実に (almost surely) (1.3 節)
$a.e.$	ほとんど至るところ (almost everywhere) (1.3 節)
$\lfloor a \rfloor$	ガウスの記号 $\lfloor a \rfloor$ は a 以下の最大の整数 (1.5 節)
$\lfloor a \rfloor^*$	$\lfloor a \rfloor^*$ は a より小さい最大の整数 (1.1 節)
p.d.f.	確率密度関数 (probability density function) (1.1 節)
p.m.f.	確率量関数 (probability mass function) (1.2 節)
c.d.f.	累積分布関数 (cumulative distribution function) (1.5 節)
j.p.d.f.	同時確率密度関数 (joint probability density function) (1.2 節)
j.p.m.f.	同時確率量関数 (joint probability mass function) (1.2 節)

スク립トフォント (花文字)

A	B	C	D	E	F	G	H	I	J	K	L	M
<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	<i>K</i>	<i>L</i>	<i>M</i>
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
<i>N</i>	<i>O</i>	<i>P</i>	<i>Q</i>	<i>R</i>	<i>S</i>	<i>T</i>	<i>U</i>	<i>V</i>	<i>W</i>	<i>X</i>	<i>Y</i>	<i>Z</i>

ギリシャ文字

大文字	小文字	読み方	大文字	小文字	読み方
A	α	アルファ	N	ν	ニュー
B	β	ベータ	Ξ	ξ	クサイ (クシー, グザイ)
Γ	γ	ガンマ	O	ο	オミクロン
Δ	δ	デルタ	Π	π	パイ (ピー)
E	ε, ε	イプシロン (エプシロン)	P	ρ, ρ	ロー
Z	ζ	ゼータ (ツェータ)	Σ	σ, σ	シグマ
H	η	イータ (エータ)	T	τ	タウ
Θ	θ, θ	シータ (テータ)	Υ	υ	ウプシロン (ユープシロン)
I	ι	イオタ (イオータ)	Φ	φ, φ	ファイ (フィー)
K	κ	カッパ	X	χ	カイ
Λ	λ	ラムダ	Ψ	ψ	プサイ (プシー)
M	μ	ミュー	Ω	ω	オメガ

パラメトリックの定義

1.1 はじめに

まず、確率分布族を \mathcal{P} とし、 \mathcal{P} が母数あるいはパラメータ (parameter) によって特徴付けられ、 $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ となるとき、 \mathcal{P} をパラメトリックモデル (parametric model) という。ここで、 θ の可能な値全体の集合 Θ を母数空間 (parameter space) といい、各 $\theta \in \Theta$ について P_θ を (確率) 分布という。たとえば、平均 μ 、分散 σ^2 をもつ正規分布 $N(\mu, \sigma^2)$ 、すなわち確率密度関数 (probability density function, 略して p.d.f.)

$$p(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad (x \in \mathbf{R}^1; (\mu, \sigma^2) \in \Theta = \mathbf{R}^1 \times \mathbf{R}_+)$$

ワルド

定理 1.3.1 (ワルド (Wald) の等式)

$X_1, X_2, \dots, X_n, \dots$ をたがいに独立に、同一分布に従う確率変数列とし、 $E(|X_1|) < \infty$ とする。このとき、 $E(N) < \infty$ となる任意の N による停止則について

$$E\left[\sum_{i=1}^N X_i\right] = E(X_1)E(N) \quad (1.3.3)$$

が成り立つ。

証明¹⁴⁾ (連続型) まず、各 $i \in \mathbf{N}$ について X_i の p.d.f. を $p(x_i)$ ($x_i \in \mathbf{R}^1$) とし、 \mathbf{X}_n の同時確率密度関数 (joint probability density function, 略して j.p.d.f.) を $f_{\mathbf{X}_n}(\mathbf{x}_n) = \prod_{i=1}^n p(x_i)$ ($\mathbf{x}_n = (x_1, \dots, x_n) \in \mathbf{R}^n$) とする。このとき、停止域の列 $\{\tilde{B}_n\}$ について

¹⁴⁾ 測度論による厳密な証明については Chow, T. S. and Teicher, H. (1997). *Probability Theory, 3rd ed.*, Springer の p.143 参照。

1.7 最尤推定と非逐次推定の限界

確率ベクトル $\mathbf{X} = (X_1, \dots, X_n)$ の j.p.d.f. または j.p.m.f. を $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})$ ($\boldsymbol{\theta} \in \Theta \subset \mathbf{R}^k$) とする。ここで、 $\mathbf{X} = \mathbf{x} = (x_1, \dots, x_n)$ であるとき

$$L(\boldsymbol{\theta}; \mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})$$

において、これを $\boldsymbol{\theta}$ の関数と見なして $\boldsymbol{\theta}$ の尤度関数 (likelihood function) といひ、 $\boldsymbol{\theta}$ の尤もらしさの度合いを表す。そこで、 $\bar{\Theta}$ を Θ の閉包⁴⁶⁾とし、 $\bar{\Theta}$ において尤度関数を最大にする、すなわち

$$\max_{\boldsymbol{\theta} \in \bar{\Theta}} L(\boldsymbol{\theta}; \mathbf{x}) = L(\hat{\boldsymbol{\theta}}^*; \mathbf{x})$$

一般に、 \mathbf{R}^k の部分集合 A の閉包 (closure) とは、 A および A の集積点全体の集合で A と表される。なお、点 a が A の集積点であるとは、 A の要素からなる点列 $\{a_n\}$ で $a_n \neq a$, $\lim_{n \rightarrow \infty} a_n = a$ となるものが存在することをいう。

第4章 逐次推定

第1章で述べたように、母数の区間推定問題において固定標本に基づいて推定しても適切な精度を必ずしも達成できない場合があり、それを解決するために停止時刻を伴う逐次推定について論じる。特に、正規分布における平均や平均の差の逐次推定、多変量正規分布の平均ベクトルの逐次推定と逐次縮小推定、ガンマ分布、一様分布の逐次推定について考える。また、2段階(標本抽出)法による推定についても論じる。

4.1 正規分布の平均の逐次推定と2段階推定方式

まず、 X_1, X_2, \dots をたがいに独立に、いずれも正規分布 $N(\mu, \sigma^2)$ ($\boldsymbol{\theta} = (\mu, \sigma^2) \in \Theta = \mathbf{R}^1 \times \mathbf{R}_+$) に従う確率変数列とする。ここで、 $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$, $S_{0n}^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 / (n-1)$ ($n \geq 2$) とする。このとき、 μ を \bar{X}_n によって推定するときに生ずる損失を費用も含めて

$$L(\mu, \bar{X}_n) = (\bar{X}_n - \mu)^2 + cn \quad (4.1.1)$$

とすると、 \bar{X}_n のリスクは

$$R(\boldsymbol{\theta}, \bar{X}_n) = E_{\boldsymbol{\theta}} [L(\mu, \bar{X}_n)] = \frac{\sigma^2}{n} + cn$$

となる。ただし、 $c \in \mathbf{R}_+$ を1標本当たりの費用とする。ここで、 σ^2 が既知のとき、 $R(\boldsymbol{\theta}, \bar{X}_n)$ を最小にする n は $n_* = \sigma / \sqrt{c}$ となって、 σ^2 に