数理統計の理論と応用

野口千明　著

序

筆者はもともと臨床家である. 経験則が強く, 根拠に乏しい「セオリー」がまかり通っていた世界で, 大学生の頃からいつも妙な違和感を覚えていた. そういうわけで, 10 年ほど前の大学院進学の際に, 不遜にも「医学をより客観的に語り直したい」という野心 (見栄 ?) から医学統計の勉強をしてみようと思った. 近年, EBM なる概念がしばしば強調されるが, evidence たる文献の検証手法である統計がお座なりでは, まさに「絵に書いた餅」であろう.

以前に「ケーススタディでみえる医学統計学」(中外医学社) という医学統計の「教科書」(もどき !?) を執筆させていただいたが, コンセプトが数式をあまり用いずに, 統計手法の概念とソフトの利用法を紹介するというものだったので, かなり舌足らずなものになってしまった. 自分では不完全燃焼の印象が非常に強い. 発行部数 2000 と医学書 (?) にしては大きな部数だったので, あまり他社様に迷惑をかけるわけにもいかないという事情もあった.

そこで, 今回は今までに学び, 解析してきた統計手法を, 理論的に表現することに努めた. 自社出版なので気兼ねがない.

第 I 部では, 数理統計学の基本事項を, 第 II 部では, 検定, 推定, 統計モデルなどの統計学の基礎概念を記述した. 既に多くの教科書に記述されている内容ではあるが, notation の問題もあり, 必要最小限の事項は省略できないと判断した. 第 III 部では, 実際に解析する必要性の高い手法を解説した. 紙面の都合で解析例を記載できなかったのが心苦しいが, 先のコンセプトから作成しているのでご容赦いただきたい.

前東京大学第三外科教授大原毅先生には, 大学院進学の際に, 医学統計の勉強をしたいという筆者のわがままを聞き入れていただいた. 当時, 外科の院生は薄暗くホルマリン臭い病理学教室で毎日顕微鏡を覗いて暮らすのが宿命であった. 心から感謝したい.

東京大学消化管外科教授上西紀夫先生にも感謝の意を表したい. 奔放な筆者を, 実験デザイナーもしくはデータアナリストとして, 医局員のまま残していただいている.

放送大学教授長岡亮介先生には, 前回に拙著の数学的補遺の原稿に目を通していただいた. 今回もこの部分にはほとんど手を加えていない. 先生には数学者の立場から, 医学統計の問題点や理想像について日頃からご意見をいただいており, 改めて感謝したい.

協立医療生活協同組合の三浦一泰理事長, 佐藤賢一事務長の両氏にもこの場を借りて御礼申し上げたい. 同生協のご理解とご協力がなければ, 大学院進学は断念していたかもしれない. 筆者は「煩悩の人」なので, 「清貧」に全く耐性がないのである.

畏友小林誠先生 (東京大学整形外科), 八尾厚史先生 (東京大学循環器内科) とは, 日頃からご自身の研究データの分析や論文の解釈についてフランクにディスカッションさせていただいている. 臨床の先端でご活躍の両先生との交流は, 日々の雑務に追われる筆者には数理的イメージを膨らませる貴重な体験である.

本書が科学に携わる研究者の方々のお役に立てればこれ以上の喜びはない.

2003 年 7 月
野口 千明

目次

序 iii

第 I 部　確率変数 1

第 1 章　確率 3

1.1 確率の定義 3

1.2 条件つき確率 5

第 2 章　確率変数 7

2.1 確率変数 7

2.2 積率 9

2.3 母関数 12

第 3 章　確率変数 (多次元) 17

3.1 同時分布と周辺分布 17

3.2 条件つき分布 19

3.3 積率 20

3.4 n 次元の確率変数 24

3.5 母関数 31

第 4 章　代表的な確率分布 33

4.1 二項分布 33

4.2 正規分布 37
4.3 その他の連続分布 39
4.4 多項分布 44
4.5 多変量正規分布 48

第 5 章 確率変数の関数の分布 53
5.1 確率変数の関数の分布 53
5.2 標本平均と標本分散の分布 60
5.3 大数の法則と中心極限定理 61

第 II 部 統計学的推論 65

第 6 章 検定 67
6.1 仮説と検定 67
6.2 ランダム検定 71
6.3 最強力検定 72
6.4 検出力関数 74
6.5 複合仮説の検定 75
6.6 尤度比検定 78

第 7 章 推定 79
7.1 点推定 79
7.2 有効推定量 80
7.3 最尤推定法 86
7.4 区間推定 89
7.5 十分統計量 90

第 8 章 ノンパラメトリック法 95
8.1 ノンパラメトリックな推論 95
8.2 並べ替え検定 95

8.3 順序統計量 97
8.4 順位検定 102

第 9 章 モデルに基づく推論 (1) 109
9.1 線型モデル 109
9.2 最小二乗法 110
9.3 モデルの検証 114
9.4 回帰係数の区間推定 117

第 10 章 モデルに基づく推論 (2) 119
10.1 ロジスティックモデル 119
10.2 回帰係数の推定 121
10.3 モデルの検証 122
10.4 関連モデル 125

第 III 部 実験計画とデータ解析 131

第 11 章 二標本問題 133
11.1 特性値が正規分布に従う場合 133
11.2 ノンパラメトリック法 138
11.3 特性値が名義変数の場合 153
11.4 特性値が順序カテゴリの場合 159
11.5 対応のある場合 160

第 12 章 一元配置 165
12.1 オムニバスな検定 165
12.2 多重比較 171
12.3 特性値が名義変数の場合 179
12.4 特性値が順序カテゴリの場合 183
12.5 水準に順序のついている場合 186

第 13 章　ブロック実験 191
13.1　乱塊法 191
13.2　ノンパラメトリック法 194

第 14 章　二元配置 203
14.1　オムニバスな検定 203
14.2　繰り返しのない場合 213
14.3　交互作用解析 214
14.4　繰り返しがブロックをなす場合 216
14.5　特性値が名義変数の場合 218
14.6　特性値が順序カテゴリの場合 225
14.7　水準に順序のついている場合 226

第 15 章　分割実験 231
15.1　分割実験 231
15.2　枝分れ実験 234
15.3　反復測定データ 235

付録 A　数学的補遺 237
A.1　線型代数学 237
A.2　微積分学 248
A.3　複素解析 261

第 I 部

確率変数

第 1 章

確率

1.1　確率の定義

有限事象の場合

実験, 観測, 調査などを総称して試行という. 試行の結果ひとつひとつを標本点といい, 標本点の全体から成る集合を標本空間という. たとえば, 正しいサイコロを 1 回振るという試行においては, 標本点は 1, 2, 3, 4, 5, 6 であり, 標本空間は $\{1, 2, 3, 4, 5, 6\}$ である. ある試行 T の標本空間 Ω が有限集合である場合, T を有限試行といい, Ω が無限集合である場合, T を無限試行という. 当面, 試行のみを考えることにする.

ある有限試行 T の標本空間を Ω とするとき, Ω の任意の部分集合 A を事象という. T の結果, A の要素が得られることを事象 A が起こるという. A の補集合 A^c を A の余事象という. A, B の和集合 $A \cup B$, A, B の積集合 $A \cap B$ をそれぞれ A, B の和事象, 積事象という. $A \cap B = \emptyset$ のとき, A と B とは排反事象であるという. このとき, $A \cup B$ を直和といい, $A + B$ と書く. 以下, 単に $A + B$ と書くときには, $A \cap B = \emptyset$ が仮定されているものとする.

試行 T の標本空間を Ω とし, 任意の $A \subseteq \Omega$ に対して, 事象 A の起こる確率を $P(A)$ と書く. $P(A)$ は A の上の関数である.

確率の直観的意味から,

$$
\begin{aligned}
&(\text{P.1}) \qquad P(A) \geq 0 \\
&(\text{P.2}) \qquad P(A+B) = P(A) + P(B) \\
&(\text{P.3}) \qquad P(\Omega) = 1
\end{aligned}
$$

が成り立つ.

一般に, 集合 Ω の上の関数 P で, 上の性質をもつものを Ω の上の確率測度といい, (Ω, P) を確率空間という.

無限事象の場合

試行 T の標本空間 Ω が無限集合の場合にも, 容易に議論を拡張できる. まず, $P(A)$ の定義されるような集合 A の全体を $\mathcal{D}(P)$ と表す[*1]とき, 以下の条件を仮定する.

$$
\begin{gathered}
\Omega \in \mathcal{D}(P) \\
A \in \mathcal{D}(P) \implies A^c \in \mathcal{D}(P) \\
A_n \in \mathcal{D}(P)\ (n = 1, 2, \cdots) \implies \bigcap_{n=1}^{\infty} A_n \in \mathcal{D}(P)
\end{gathered}
$$

このとき,

$$
\begin{aligned}
&(\text{P.1}) \qquad P(A) \geq 0 \\
&(\text{P.2}) \qquad P\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \\
&(\text{P.3}) \qquad P(\Omega) = 1
\end{aligned}
$$

が成立するような $P(A)$ を A の P-確率測度という. この 3 個の条件を確率の公理ということにする. 第二の条件を加法性という. 以下, 単に $P(A)$ と書いたときには, $A \in \mathcal{D}(P)$ が仮定されているものとする. $\mathcal{D}(P)$ の任意の部分集合を P-可測集合という. Ω と P を組にした (Ω, P) を確率空間という.

[*1] 有限試行のときは, 任意の $A \subset \Omega$ に対して $P(A)$ が定義された.

1.2　条件つき確率

ふたつの事象 A, B を考える．事象 A が起こることを仮定した上で，事象 B が起こる確率を，A の条件のもとでの B の起こる条件つき確率といい，$P(B \mid A)$ で表す．

$$P(B \mid A) \stackrel{\text{def}}{=} \frac{P(A \cap B)}{P(A)} \tag{1.1}$$

■定理 1.1　条件つき確率は，確率の公理を満たす．

$$P(B \mid A) \geq 0$$
$$P\left(\sum_{n=1}^{\infty} B_n \,\middle|\, A\right) = \sum_{n=1}^{\infty} P(B_n \mid A)$$
$$P(A \mid A) = 1$$

事象 A が起こることを仮定しているので，標本空間が A となることに注意しよう．

■証明　第 1 式は自明．第 3 式は (1.1) で，$B = A$ とおけばよい．つぎに，互いに排反な事象 B_n $(n = 1, 2, \cdots)$ に対して，

$$\left(\sum_{n=1}^{\infty} B_n\right) \cap A = \sum_{n=1}^{\infty} \left(B_n \cap A\right)$$

が成り立つから，この事象の起こる確率を $P(A)$ で除せば第 2 式を得る．　■

独立事象

■定義 1.2

$$P(A \cap B) = P(A)\,P(B) \tag{1.2}$$

が成り立つとき，事象 A と B は独立であるという．

式 (1.1) より直ちに

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

を得る. これと (1.2) より, 定理 1.3 および 1.4 を得る.

■定理 1.3　事象 A と B は独立 $\iff P(B) = P(B\,|\,A)$

事象 A の起こることを仮定した場合と一般の場合とで, 事象 B の起こる確率が等しいのであるから, A と B は独立であると直観的にも理解できよう.

■定理 1.4　標本空間 Ω の任意の部分集合 B に対して,

$$\sum_{n=1}^{\infty} A_n = \Omega \implies P(B) = \sum_{n=1}^{\infty} P(A_n)\,P(B\,|\,A_n)$$

が成り立つ.

■証明　確率 $P(B) = P(\Omega \cap B)$ は, (P.2) より,

$$P\biggl(\Bigl\{\sum_{n=1}^{\infty} A_n\Bigr\} \cap B\biggr) = P\biggl(\sum_{n=1}^{\infty}\{A_n \cap B\}\biggr) = \sum_{n=1}^{\infty} P(A_n \cap B)$$

と表される. これから, 直ちに定理 1.4 を得る. ■

■定理 1.5　互いに背反な事象 A_n $(n = 1, 2, \cdots)$ に対して,

$$P(A_n\,|\,B) = P(A_n)\,P(B\,|\,A_n)\Bigg/\sum_{n=1}^{\infty} P(A_n)\,P(B\,|\,A_n)$$

が成り立つ. ただし, B は標本空間 Ω の任意の部分集合である.

この定理を Bayes の定理という. $P(A_n\,|\,B)$ を事後確率, $P(A_n)$ を事前確率ということがある.

第 2 章

確率変数

2.1　確率変数

確率変数とその分布

正しいサイコロを 1 回振るとき, 出る目 Y は $1, 2, \cdots, 6$ のいずれかの値をとる. この意味で Y は変数であり, もちろんどの値をとるかは事前にはわからない. しかし, どの値をとる確率も 1/6 であることはわかっている.

一般に, ある事象に関連する変数を Y とし, Y のとりうる値 y すべてから成る集合を Ω, その部分集合を A とするとき, 任意の y および任意の A について $P(y \in A)$ が定められているような変数 Y を確率変数という.

確率 $P(y \in A)$ を Y の確率分布または単に分布という. Y の分布が D であるとき, Y は分布 D に従うといい, $Y \sim D$ と表す. 確率変数は通常アルファベットの大文字で表す. これに対して, 実際に得られた確率変数の値を観測値といい, 対応する確率変数を表すアルファベットの小文字で表す.

■例 2.1　確率変数 Y を正しいサイコロを 1 回振ったときの出る目とすると, その確率分布は次のようになる.

y	1	2	$\cdots$	6
$P(Y = y)$	1/6	1/6	$\cdots$	1/6

関数

$$F(y) = P(Y \le y) \tag{2.1}$$

を Y の分布関数という. 分布と分布関数は 1 対 1 に対応する.

確率変数 Y が離散値をとるとき, その分布を離散分布 といい,

$$p(y) = P(Y = y) \tag{2.2}$$

を確率関数という. (P.3) により

$$\sum_y p(y) = 1$$

が成立する. 和記号は y のとりうるすべての値に関する和を意味する. 分布関数との間には,

$$F(y) = \sum_{t \le y} p(t) \tag{2.3}$$

が成り立つ. したがって, 確率関数と分布は 1 対 1 に対応する.

確率変数 Y が連続的な値をとるときは, その分布を連続分布という. この場合, 任意の y に対して, $P(Y = y) = 0$ であるから, 確率関数を考えることには意味がない. そこで, 極限

$$f(y) = \lim_{h \to 0} \frac{P(y - h < Y \le y + h)}{2\,h} \tag{2.4}$$

が存在すれば, これを確率密度関数または単に密度関数という. (P.3) より

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

が成立する. 分布関数との間には,

$$F(y) = \int_{-\infty}^{y} f(t)\,dt \tag{2.5}$$

が成り立つ. したがって, 密度関数と分布は 1 対 1 に対応する.

■例 2.2　例 2.1 において, 確率関数は,

$$f(y) = \begin{cases} 1/6 & \cdots \quad y = 1, 2, \cdots 6 \\ 0 & \cdots \quad \text{otherwise} \end{cases}$$

である. 分布関数は,

$$F(y) = \begin{cases} 0 & \cdots \quad y < 1 \\ 1/6 & \cdots \quad 1 \le y < 2 \\ \vdots & \qquad \vdots \\ 5/6 & \cdots \quad 5 \le y < 6 \\ 1 & \cdots \quad 6 \le y \end{cases}$$

である.

2.2　積率

平均

　確率変数 Y が離散値 y_i $(i = 1, 2, \cdots)$ をとるとき,

$$\sum_{i=1}^{\infty} y_i \, p(y_i) \tag{2.6}$$

が存在すれば, これを Y の平均といい, $E[Y]$ や μ と表す. Y が有限個の値をとる場合は相応の修正をすればよい. 連続分布のときは,

$$\int_{-\infty}^{\infty} y \, f(y) \, dy \tag{2.7}$$

が存在するならば, この値を平均と定義する.

　一般に, 確率変数 Y の関数 $g(Y)$ の期待値を, 離散分布であれば

$$E\Big[\, g(Y) \,\Big] = \sum_{y} g(y) \, p(y) \tag{2.8}$$

で, 連続分布であれば,

$$E\Big[\, g(Y) \,\Big] = \int_{-\infty}^{\infty} g(y) \, f(y) \, dy \tag{2.9}$$

で定義する.

両者をまとめて, 次のように書くことがある.

$$E\Big[\,g(Y)\,\Big] = \int_{-\infty}^{\infty} g(y)\,dF(y) \tag{2.10}$$

とくに $g(y) = y$ とした場合が平均である. したがって, 平均を期待値と表現する場合がある. 期待値の定義より直ちに次の定理を得る.

■定理 2.1　$i = 1, \cdots, k$ に対し, $g_i(Y)$ を確率変数 Y の関数, λ_i を定数とするとき, 次が成り立つ.

$$E\Big[\sum_{i=1}^{k} \lambda_i\, g_i(Y)\Big] = \sum_{i=1}^{k} \lambda_i\, E\Big[\,g_i(Y)\,\Big]$$

分散

確率変数 Y の平均を μ とするとき, $(Y - \mu)^2$ の期待値が存在すれば, これを分散といい, $V[Y]$ や σ^2 と表す :

$$V[Y] = E\Big[\,(Y - \mu)^2\,\Big] \tag{2.11}$$

この定義より, 容易に次の性質を得る.

■定理 2.2　Y を確率変数, a, b を定数とするとき, 次が成り立つ.

$$V[Y] = E[\,Y^2\,] - E[Y]^2, \quad V\Big[\,aY + b\,\Big] = a^2\,V[Y] \tag{2.12}$$

分散の非負平方根 σ を Y の標準偏差という. とくに, 確率変数の関数の標準偏差を標準誤差ということがある.

期待値, 分散を一般化した概念が積率である. Y^k $(k = 1, 2, \cdots)$ の期待値を原点まわりの k 次の積率, $(Y - \mu)^k$ の期待値を平均まわりの k 次の積率という. これらをそれぞれ μ_k, ν_k で表す :

$$\mu_k = E\Big[\,Y^k\,\Big], \quad \nu_k = E\Big[\,(Y - \mu)^k\,\Big] \tag{2.13}$$

とくに, $\mu_1 = \mu$, $\nu_1 = 0$, $\nu_2 = \sigma^2$ である. ν_3/σ^3 を歪度, ν_4/σ^4 を尖度という. これらはそれぞれ分布の歪み, 尖りを表す指標であるが, その定義には多少の variation がある.

■定理 2.3　確率変数 Y の非負値関数 $g(Y)$ の期待値が存在するとき, 任意の正数 k に対して,

$$P\Big(g(Y) \geq k\Big) \leq \frac{1}{k} E\Big[\,g(Y)\Big] \tag{2.14}$$

が成り立つ. これを Markov の不等式という.

■証明　連続分布の場合を示す. 密度関数を $f(y)$ とすると, $g(y) \geq 0$ より

$$\int_{-\infty}^{\infty} g(y)\, f(y)\, dy \geq \int_{g(y)\geq k} g(y)\, f(y)\, dy \geq k \int_{g(y)\geq k} f(y)\, dy$$

を得る. これは結論を導く. ■

■系 2.4　確率変数 Y が平均 μ, 分散 σ^2 をもつとき, 任意の正数 ε に対して

$$P\Big(\,|Y-\mu| \geq \varepsilon\Big) \leq \frac{\sigma^2}{\varepsilon^2} \tag{2.15}$$

が成り立つ. これを Chebyshev の不等式という.

■証明　定理 2.3 で, $g(Y) = (Y-\mu)^2, k = \varepsilon^2$ とすればよい. ■

■定理 2.5　確率変数 Y が平均をもつとき, 凸関数 $g(Y)$ に対して,

$$E\Big[\,g(Y)\Big] \geq g\Big(\,E[Y]\Big) \tag{2.16}$$

が成り立つ. これを Jensen の不等式という. $g(Y)$ が狭義の凸関数であれば等号が成立するのは, Y が確率 1 で定値をとる —この場合を一点分布という— のときに限られる.

■証明　$g(Y)$ が C^2 級である場合について示す. 一般の場合も凸関数の性質から定理を得るが, やや複雑である.

Taylor の定理により,

$$g(Y) = g(\mu) + g'(\mu)(Y-\mu) + \frac{1}{2}g''(\xi)(Y-\mu)^2$$

を得る. ξ は定数 θ $(0 < \theta < 1)$ を用いて,

$$\xi = (1-\theta)Y + \theta\mu$$

と表される定数である. $g(Y)$ は凸関数であったから,

$$g(Y) \geq g(\mu) + g'(\mu)(Y-\mu) \tag{2.17}$$

が成立する.

ここで, 両辺の期待値をとると, (2.16) を得る. $g(Y)$ が狭義の凸関数であれば, (2.17) の等号が成立するのは $Y = \mu$ のときに限るので, (2.16) の等号が成立するのは, 恒等的に $Y = \mu$ となる場合のみである. ■

2.3　母関数

確率変数 Y に対して, $\exp(tY)$ の期待値

$$M(t) = E\Big[\exp(tY)\Big] \tag{2.18}$$

を積率母関数という. また, $\exp(itY)$ の期待値

$$C(t) = E\Big[\exp(itY)\Big] \tag{2.19}$$

を特性関数という. 特性関数は必ず存在する. 積率母関数は存在するとは限らないが, 存在すれば, つぎの関係が成り立つ.

$$C(-it) = M(t) \tag{2.20}$$

特性関数と分布関数は 1 対 1 に対応する. 積率母関数が存在するときは, これと分布関数も 1 対 1 に対応する.

■定理 2.6　任意の定数 a, b $(a < b)$ について

$$\frac{1}{2} P(Y = a) + P(a < Y < b) + \frac{1}{2} P(Y = b) \tag{2.21}$$
$$= \frac{1}{2\pi} \lim_{T\to\infty} \int_{-T}^{T} \frac{\exp(-i\,b\,t) - \exp(-i\,a\,t)}{-i\,t} C(t)\,dt$$

が成立する.

この証明の前に,

$$\lim_{T\to\infty} \int_0^T \frac{\sin u}{u}\,du = \frac{\pi}{2} \tag{2.22}$$

を示そう. まず, 等比数列の和

$$\sum_{k=-n}^{n} \exp(i\,k\,u) = \frac{\exp(-i\,n\,u) - \exp\{i\,(n+1)\,u\}}{1 - \exp(i\,u)}$$

に Euler の公式 $\exp(i\,u) = \cos u + i \sin u$ を用いて,

$$D_n(u) = \sum_{k=-n}^{n} \exp(i\,k\,u) = \frac{\sin\{(n+1/2)\,u\}}{\sin(u/2)}$$

を得る. $u = 0$ のときは, $D_n(u) = 2n + 1$ とする.

$k = 1, \cdots, n$ のときは

$$\int_{-\pi}^{\pi} \cos(k\,u)\,du = \int_{-\pi}^{\pi} \sin(k\,u)\,du = 0$$

であるから,

$$\int_{-\pi}^{\pi} D_n(u)\,du = 2\pi$$

を得る. ここで,

$$g(u) = \begin{cases} \cot(u/2) - 2/u & \cdots \quad u \neq 0 \\ 0 & \cdots \quad u = 0 \end{cases}$$

とおくと, $D_n(u)$ は次のように書ける.

$$D_n(u) = \cos(n\,u) + \frac{2\sin(n\,u)}{u} + g(u)\,\sin(n\,u) \tag{2.23}$$

$g(u)$ は $(-\pi, \pi)$ で連続であり, $u \to 0$ のとき $g(u) \to 0$ であるから, この範囲で有界である. したがって, (2.23) の第 3 項を $-\pi$ から π まで積分すると 0 になる.

結局, (2.23) を $-\pi$ から π まで積分すると,

$$\int_{-\pi}^{\pi} \frac{\sin(n\,u)}{u}\,du = \int_{-n\pi}^{n\pi} \frac{\sin v}{v}\,dv = 2\int_{0}^{n\pi} \frac{\sin v}{v}\,dv = \pi$$

を得る.

一般の T に対しては, $n\pi \le T < (n+1)\pi$ なる n を選べば,

$$\left|\int_0^T \frac{\sin v}{v}\,dv - \int_0^{n\pi} \frac{\sin v}{v}\,dv\right| = \left|\int_{n\pi}^T \frac{\sin v}{v}\,dv\right| \le \int_{n\pi}^T \frac{1}{v}\,dv < \frac{\pi}{n\pi}$$

となるから, $n \to \infty$ とすれば (2.22) が得られる. ■

■定理 2.6 の証明　分布が連続型である場合を考える.

$$F(T) = \int_{-T}^{T}\int_a^b \exp(-\,i\,t\,y)\,dy \int_{-\infty}^{\infty} \exp(i\,t\,y)\, f(y)\,dy\,dt$$

とおき, t について積分すると,

$$F(T) = 2\int_{-\infty}^{\infty} f(y)\,dy \int_a^b \frac{\sin(y-z)\,T}{y-z}\,dz$$

と書ける. ここで, $u = T\,(y-z)$ と置換すると, 次の式を得る.

$$F(T) = 2\int_{-\infty}^{\infty} f(y)\,dy \int_{T\,(y-b)}^{T\,(y-a)} \frac{\sin u}{u}\,du$$

いま,

$$G(y) = \int_0^y \frac{\sin u}{u}\,du$$

とおくと, $G(y)$ は連続関数で, (2.22) より有界である. したがって, 極限と積分の順番が交換ができて, 次が成り立つ.

$$\lim_{T\to\infty} F(T) = 2\int_{-\infty}^{\infty} h_{a,b}(y)\, f(y)\,dy$$

ここに,

$$h_{a,b}(y) = \lim_{T\to\infty}\Big\{G\big(T\,(y-a)\big) - G\big(T\,(y-b)\big)\Big\}$$

であるが, この値は (2.22) より, $y = a, b$ のとき $\pi/2$, $a < y < b$ のとき π , $y < a$, $y > b$ のとき 0 となる. これは結論を導く. ■

たとえば, 連続分布の場合, 密度関数を $f(y)$ とおくと,

$$M(t) = \int_{-\infty}^{\infty} \exp(t\,y)\,f(y)\,dy$$

の両辺を t について k 回微分し,

$$M^{(k)}(t) = \int_{-\infty}^{\infty} y^k\,\exp(t\,y)\,f(y)\,dy \tag{2.24}$$

を得る. ここで, $t = 0$ とおくと,

$$M^{(k)}(0) = E\big[\,Y^k\,\big] \tag{2.25}$$

となり, 原点まわりの k 次の積率が得られる. これが積率母関数の名の由来である.

最後に, 次の定理を提示して, この章を終わる.

■定理 2.7　$F_1(y)$, $\cdots$, $F_n(y)$, $\cdots$ を分布関数の無限列, $M_1(t)$, $\cdots$, $M_n(t)$, $\cdots$ をそれぞれに対応する積率母関数の無限列とする. また, $F(y)$ をある分布関数, $M(t)$ をこれに対応する積率母関数とする. このとき,

$$\lim_{n\to\infty} F_n(y) = F(y) \iff \lim_{n\to\infty} M_n(t) = M(t)$$

が成立する.

第 3 章

確率変数 (多次元)

3.1 同時分布と周辺分布

2 個の確率変数 Y_1, Y_2 について,

$$F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2) \tag{3.1}$$

を同時分布関数という. 各成分 Y_i $(i = 1, 2)$ の周辺分布関数を, 次で定義する.

$$F_i(y_i) = P(Y_i \leq y_i) = \lim_{y_j \to \infty} F(y_1, y_2) \quad (j \neq i) \tag{3.2}$$

分布が離散型の場合,

$$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2) \tag{3.3}$$

を同時確率関数という. 同時分布関数との間には,

$$F(y_1, y_2) = \sum_{t_1 \leq y_1} \sum_{t_2 \leq y_2} p(t_1, t_2) \tag{3.4}$$

の関係がある. 各成分に注目して,

$$p_i(y_i) = P(Y_i = y_i) = \sum_{y_j} p(y_1, y_2) \tag{3.5}$$

を周辺確率関数という.

周辺分布関数との間には, 次の関係が成り立つ.

$$F_i(y_i) = \sum_{t_i \le y_i} p_i(y_i) \tag{3.6}$$

同時確率関数, 周辺確率関数は確率の公理を満たす :

$$\sum_{y_1} \sum_{y_2} p(y_1, y_2) = \sum_{y_1} p_1(y_1) = \sum_{y_2} p_2(y_2) = 1$$

Y_1, Y_2 が連続的な値をとるときは, 同時密度関数を,

$$f(y_1, y_2) = \lim_{h_1 \to 0} \lim_{h_2 \to 0} \frac{P({}^\forall y_i, y_i - h_i < Y_i \le y_i + h_i)}{4\, h_1\, h_2} \tag{3.7}$$

で定義する. 同時分布関数との間には,

$$F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(t_1, t_2)\, dt_1\, dt_2 \tag{3.8}$$

が成り立つ.

各成分に注目して, Y_i $(i = 1, 2)$ の周辺密度関数を

$$f_i(y) = \lim_{h_i \to 0} \frac{P(y_i - h_i < Y_i \le y_i + h_i)}{2\, h_i} \tag{3.9}$$

で定義する. 周辺分布関数との間には, 次の関係がある.

$$F_i(y_i) = \int_{-\infty}^{y_i} f_i(t_i)\, dt_i \tag{3.10}$$

また, 同時密度関数, 周辺密度関数との間には, 次が成り立つ.

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, t_2)\, dt_2 \tag{3.11}$$

$$f_2(y_2) = \int_{-\infty}^{\infty} f(t_1, y_2)\, dt_1 \tag{3.12}$$

同時密度関数, 周辺密度関数は確率の公理を満たす :

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2)\, dy_1\, dy_2 = \int_{-\infty}^{\infty} f_i(y_i)\, dy_i = 1$$

3.2　条件つき分布

y_j を与えたときの Y_i の条件つき分布を考える. $i = 1, j = 2$ の場合を議論すれば十分であろう.

離散分布のとき, y_2 を与えたときの Y_1 の条件つき分布関数を,

$$F(y_1 \,|\, y_2) = \frac{P(Y_1 \leq y_1, Y_2 = y_2)}{P(Y_2 = y_2)} = \frac{\sum_{t_1 \leq y_1} p(t_1, y_2)}{p_2(y_2)} \tag{3.13}$$

で定義する. 連続分布のときは分母子とも 0 になるので修正を要する.

条件つき確率関数は,

$$p(y_1 \,|\, y_2) = \frac{P(Y_1 = y_1, Y_2 = y_2)}{P(Y_1 = y_2)} = \frac{p(y_1, y_2)}{p_2(y_2)} \tag{3.14}$$

で定義する.

連続分布のときは, 条件つき分布関数を,

$$F(y_1 \,|\, y_2) = \lim_{h_2 \to 0} \frac{P(Y_1 \leq y_1, y_2 - h_2 < Y_2 \leq y_2 + h_2)}{P(y_2 - h_2 < Y_2 \leq y_2 + h_2)} \tag{3.15}$$

で定義する. 実際は, 同時密度関数と周辺密度関数から,

$$F(y_1 \,|\, y_2) = \int_{-\infty}^{y_1} f(t_1, y_2)\, dt_1 \Big/ f_2(y_2) \tag{3.16}$$

で計算すればよい.

(3.16) の y_1 による偏導関数として, 条件つき密度関数

$$f(y_1 \,|\, y_2) = \frac{\partial}{\partial y_1} F(y_1 \,|\, y_2) = \frac{f(y_1, y_2)}{f_2(y_2)} \tag{3.17}$$

を定義する.

条件つき確率関数, 条件つき密度関数は確率の公理を満たす :

$$\sum_{y_1} p(y_1 \,|\, y_2) = 1\,, \quad \int_{-\infty}^{\infty} f(y_1 \,|\, y_2)\, dy_1 = 1$$

確率変数の独立性

■**定義** 3.1　2 つの確率変数 Y_1, Y_2 が独立であるとは, Y_i の標本空間をそれぞれ Ω_i $(i = 1, 2)$ として,

$$P(Y_1 \in A_1, Y_2 \in A_2) = P(Y_1 \in A_1)\, P(Y_2 \in A_2) \tag{3.18}$$

が, 任意の部分集合 $A_1 \subset \Omega_1$, $A_2 \subset \Omega_2$ について成立することである.

■**定理** 3.2

$$(3.18) \Longleftrightarrow \begin{cases} p(y_1, y_2) = p_1(y_1)\, p_2(y_2) & \cdots \quad \text{離散} \\ f(y_1, y_2) = f_1(y_1)\, f_2(y_2) & \cdots \quad \text{連続} \end{cases}$$

が成り立つ.

■**証明**　離散分布の場合は直ちにわかる. 連続分布の場合, (3.18) を仮定すると, $A_i = \{Y_i \mid Y_i \leq y_i\}$ とすれば,

$$F(y_1, y_2) = F_1(y_1)\, F_2(y_2) \tag{3.19}$$

となるので, この両辺を y_1, y_2 について偏微分すれば,

$$f(y_1, y_2) = f_1(y_1)\, f_2(y_2) \tag{3.20}$$

を得る. 逆に, (3.20) が成立すれば, 任意の A_1, A_2 について,

$$\int_{A_1}\int_{A_2} f(y_1, y_2)\, dy_1\, dy_2 = \int_{A_1} f_1(y_1)\, dy_1 \int_{A_2} f_2(y_2)\, dy_2$$

が成立する. ■

3.3　積率

平均

一般に, 確率変数 Y_1, Y_2 の関数 $g(Y_1, Y_2)$ の期待値を, 次で定義する.

$$E\Big[\, g(Y_1, Y_2) \Big] = \int_{-\infty}^{\infty} g(y_1, y_2)\, dF(y_1, y_2) \tag{3.21}$$

とくに, $\mu_i = E[Y_i]$ を Y_i の期待値という. 期待値の定義より直ちに次の定理を得る.

■定理 3.3　$i = 1, \cdots, k$ に対し, $g_i(Y_1, Y_2)$ を確率変数 Y_1, Y_2 の関数とし, λ_i を定数とするとき, 次が成り立つ.

$$E\Big[\sum_{i=1}^{k} \lambda_i\, g_i(Y_1, Y_2)\Big] = \sum_{i=1}^{k} \lambda_i\, E\Big[\, g_i(Y_1, Y_2)\,\Big]$$

2 次元確率変数の平均は同時分布に基づく期待値として定義されるが, 周辺分布に基づく期待値に等しい. 実際, 連続分布であれば, (3.20) に注意して,

$$E[Y_i] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_i\, f(y_1, y_2)\, dy_1\, dy_2 = \int_{-\infty}^{\infty} y_i\, f_i(y_i)\, dy_i$$

を得る.

分散, 共分散

つぎに, 多次元の場合の分散および共分散を定義しよう.

$$\sigma_{ij} = E\Big[\,(Y_i - \mu_i)\,(Y_j - \mu_j)\,\Big] \tag{3.22}$$

において, $i = j$ の場合を Y_i の分散といい, $V[Y_i]$, σ_i^2 などと表す. $i \neq j$ の場合を Y_i, Y_j の共分散といい, Cov $[Y_i, Y_j]$ と表す.

分散も, 平均同様, 同時分布に基づく期待値として定義されるが, 周辺分布に基づく期待値に等しい. 分散の非負平方根 σ_i を標準偏差という.

分散, 共分散の定義より,

$$\sigma_{ij} = E\Big[\,Y_i Y_j\,\Big] - E[Y_i]\,E[Y_j] \tag{3.23}$$

が成り立つ. この式は, 定理 2.2 と同様にして容易に導かれる.

また, 定理 3.3 より,

$$V[Y_1 + Y_2] = V[Y_1] + 2\,\mathrm{Cov}\,[Y_1, Y_2] + V[Y_2] \tag{3.24}$$

を得る.

とくに, Y_1, Y_2 が独立な場合は,

$$E\Big[\,Y_1 Y_2\,\Big] = E[Y_1]\,E[Y_2] \tag{3.25}$$

が成り立つ. したがって, この場合には (3.24) は

$$V\Big[\,Y_1 + Y_2\,\Big] = V[Y_1] + V[Y_2]$$

となる.

相関係数

確率変数 Y_1, Y_2 の相関係数を,

$$r_{12} = \mathrm{Corr}\,[Y_1, Y_2] = \frac{\mathrm{Cov}\,[Y_1, Y_2]}{\sqrt{V[Y_1]\,V[Y_2]}} \tag{3.26}$$

で定義する. これを Pearson の相関係数ということがある. 一般に, 2 個の確率変数が独立であれば相関係数は 0 となるが, 逆は必ずしも成り立たない.

■定理 3.4　相関係数はつねに $|r_{12}| \le 1$ を満たし, $|r_{12}| = 1$ となるのは, Y_1, Y_2 の間には完全な線型関係

$$Y_2 - \mu_2 = \pm\frac{\sigma_2}{\sigma_1}\,(Y_1 - \mu_1) \tag{3.27}$$

が成り立つときであり, かつそのときに限る.

■証明　μ_1, μ_2 を 0 としても一般性を失わない. 自明な不等式

$$0 \le E\Big[\,(Y_1 - t\,Y_2)^2\,\Big] = \sigma_2^2\,t^2 - 2\,\sigma_{12}\,t + \sigma_1^2$$

が任意の実数 t について成立するから, 次の不等式が成り立つ.

$$D/4 = \sigma_{12}^2 - \sigma_1^2\,\sigma_2^2 \le 0$$

これは $|r_{12}| \leq 1$ を導く. 等号が成立するとき, t に関する方程式

$$E\Big[\,(Y_1 - t\,Y_2)^2\,\Big] = 0 \iff Y_1 - t\,Y_2 = 0$$

は, 重解 $\sigma_{12}/\sigma_2^2 = \pm\sigma_1/\sigma_2$ をもつから, (3.27) が得られる. 逆に,

$$(3.27) \iff \frac{Y_2 - \mu_2}{\sigma_2} \mp \frac{Y_1 - \mu_1}{\sigma_1} = 0$$

の二乗の期待値をとると, 定理 3.3 と (3.24) より, $|r_{12}| = 1$ が得られる. ■

ここで, 行列表現をしておこう. 2 次元の確率変数を $\boldsymbol{Y} = (Y_1, Y_2)'$ と, その平均および分散を次のように書くことがある. $'$ は転置を表す.

$$E[\boldsymbol{Y}] = (\mu_1, \mu_2)', \quad V[\boldsymbol{Y}] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \tag{3.28}$$

$E[\boldsymbol{Y}], V[\boldsymbol{Y}]$ をそれぞれ, $\boldsymbol{\mu}, \Sigma$ と書くことがある. 以後, この表現をとくに断わらずに用いる.

条件つき平均, 条件つき分散

条件つき分布に基づく期待値も同様に定義することができる. y_j を与えたときの Y_i $(i = 1, 2\,;\; i \neq j)$ の条件つき平均は,

$$E[Y_i\,|\,y_j] = \int_{-\infty}^{\infty} y_i\, dF(y_i\,|\,y_j) \tag{3.29}$$

で定義される. 同様に, 条件つき分散は,

$$V[Y_i\,|\,y_j] = \int_{-\infty}^{\infty} \Big(y_i - E[Y_i\,|\,y_j]\Big)^2\, dF(y_i\,|\,y_j) \tag{3.30}$$

で定義される.

3.4　n 次元の確率変数

同時分布と周辺分布

n 次元の確率変数も 2 次元の場合から容易に拡張できる. n 個の確率変数の組を列ベクトルとして, $\boldsymbol{Y} = (Y_1, \cdots, Y_n)'$ で表す. このとき, $\boldsymbol{Y}$ の同時分布関数を

$$F(y_1, \cdots, y_n) = P(Y_1 \leq y_1, \cdots, Y_n \leq y_n) \tag{3.31}$$

で定義する.

離散分布の場合は, 同時確率関数を

$$p(y_1, \cdots, y_n) = P(Y_1 = y_1, \cdots, Y_n = y_n) \tag{3.32}$$

で定義する. 同時分布関数との間には, 次の関係が成り立つ.

$$F(y_1, \cdots, y_n) = \sum_{t_1 \leq y_1} \cdots \sum_{t_n \leq y_n} p(t_1, \cdots, t_n) \tag{3.33}$$

同時確率関数は確率の公理を満たす :

$$\sum_{y_1} \cdots \sum_{y_n} p(y_1, \cdots, y_n) = 1$$

連続分布のときは, 同時密度関数を

$$f(y_1, \cdots, y_n) = \lim_{{}^\forall i, h_i \to 0} \frac{P({}^\forall i, y_i - h_i < Y_i \leq y_i + h_i)}{2^n h_1 \cdots h_n} \tag{3.34}$$

で定義する. 同時分布関数との間には,

$$F(y_1, \cdots, y_n) = \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_n} f(t_1, \cdots, t_n)\, dt_1 \cdots dt_n \tag{3.35}$$

が成り立つ. 同時密度関数は確率の公理を満たす :

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_1, \cdots, y_n)\, dy_1 \cdots dy_n = 1$$

つぎに, n 個の確率変数のうち, 任意の m 個のみの確率分布を考えよう. この m 個を, $Y_1, \cdots, Y_m$ としても一般性を失わない.

分布の型を問わず, 周辺分布関数は,

$$F^*(y_1, \cdots, y_m) = P(Y_1 \leq y_1, \cdots, Y_m \leq y_m) \tag{3.36}$$

で定義される. この定義から直ちに

$$F^*(y_1, \cdots, y_m) = \lim_{y_{m+1} \to \infty} \cdots \lim_{y_n \to \infty} F(y_1, \cdots, y_n) \tag{3.37}$$

を得る.

離散分布の場合は, 次で周辺確率関数を定義する.

$$p^*(y_1, \cdots, y_m) = P(Y_1 = y_1, \cdots, Y_m = y_m) \tag{3.38}$$

同時確率関数との間には,

$$p^*(y_1, \cdots, y_m) = \frac{p(y_1, \cdots, y_n)}{\sum_{y_{m+1}} \cdots \sum_{y_n} p(y_1, \cdots, y_n)} \tag{3.39}$$

が成立する. 周辺分布関数と周辺確率関数の間には,

$$F^*(y_1, \cdots, y_m) = \sum_{t_1 \leq y_1} \cdots \sum_{t_m \leq y_m} p^*(t_1, \cdots, t_m) \tag{3.40}$$

の関係がある. 周辺確率関数は確率の公理を満たす :

$$\sum_{y_1} \cdots \sum_{y_m} p^*(y_1, \cdots, y_m) = 1$$

連続分布の場合は, 周辺密度関数を,

$$f^*(y_1, \cdots, y_m) = \lim_{{}^\forall i, h_i \to 0} \frac{P(1 \leq {}^\forall i \leq m, y_i - h_i < Y_i \leq y_i + h_i)}{2^m h_1 \cdots h_m} \tag{3.41}$$

で定義する. 同時密度関数との間には, 次が成立する.

$$f^*(y_1, \cdots, y_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_1, \cdots, y_n)\, dy_{m+1} \cdots dy_n \tag{3.42}$$

周辺分布関数と周辺密度関数の間には, 次の関係がある.

$$F^*(y_1, \cdots, y_m) = \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_m} f^*(t_1, \cdots, t_m)\, dt_1 \cdots dt_m \tag{3.43}$$

周辺密度関数は確率の公理を満たす :

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f^*(y_1, \cdots, y_m)\, dy_1 \cdots dy_m = 1$$

とくに $m = 1$ の場合, 周辺分布関数を $F_i(y_i)$ と表し, 周辺確率関数を $p_i(y_i)$ と, 周辺密度関数を $f_i(y_i)$ と表す.

条件つき分布

n 個の確率変数 Y_i $(i = 1, \cdots, n)$ のうち, 任意の l 個の値を与えたときの, m 個 $(l + m = n)$ の確率変数の条件つき分布を考える. このとき, m 個を $Y_1, \cdots, Y_m$ としても一般性を失わないので, $y_{m+1}, \cdots, y_n$ を与えたときの, $(Y_1, \cdots, Y_m)'$ の分布を考える.

離散型のとき, 条件つき分布関数 $F(y_1, \cdots, y_m \,|\, y_{m+1}, \cdots, y_n)$ を,

$$\frac{P(Y_1 \leq y_1, \cdots, Y_m \leq y_m, Y_{m+1} = y_{m+1}, \cdots, Y_n = y_n)}{P(Y_{m+1} = y_{m+1}, \cdots, Y_n = y_n)} \tag{3.44}$$

で定義する. 条件つき確率関数は

$$p(y_1, \cdots, y_m \,|\, y_{m+1}, \cdots, y_n) = \frac{P(Y_1 = y_1, \cdots, Y_n = y_n)}{P(Y_{m+1} = y_{m+1}, \cdots, Y_n = y_n)} \tag{3.45}$$

で定義される. これは確率の公理を満たす :

$$\sum_{y_1} \cdots \sum_{y_m} p(y_1, \cdots, y_m \,|\, y_{m+1}, \cdots, y_n) = 1$$

連続型のとき, 条件つき分布関数 $F(y_1, \cdots, y_m \,|\, y_{m+1}, \cdots, y_n)$ を,

$$\lim_{{}^\forall j, h_j \to 0} \frac{P({}^\forall i \leq m,\ {}^\forall j > m,\ Y_i \leq y_i, Y_j \in I_j)}{P(Y_{m+1} \in I_{m+1}, \cdots, Y_n \in I_n)} \tag{3.46}$$

で定義する. ここに, $I_j = \{y \,|\, y_j - h_j < y \leq y_j + h_j\}$ である.

条件つき密度関数は

$$f(y_1,\cdots,y_m\,|\,y_{m+1},\cdots,y_n) = \frac{\partial^m}{\partial y_1\cdots\partial y_m}F(y_1,\cdots,y_m\,|\,y_{m+1},\cdots,y_n) \quad (3.47)$$

で定義する. これは確率の公理を満たす :

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(y_1,\cdots,y_m\,|\,y_{m+1},\cdots,y_n)\,dy_1\cdots dy_m = 1$$

確率変数の独立性

■定義 3.5　確率変数の組 $Y_1,\cdots,Y_m$ および $Y_{m+1},\cdots,Y_n$ が独立であるとは, $\boldsymbol{Y}_1 = (Y_1,\cdots,Y_m)'$, $\boldsymbol{Y}_2 = (Y_{m+1},\cdots,Y_n)'$ の標本空間をそれぞれ Ω_1,Ω_2 とし, Ω_i の部分集合を A_i $(i = 1,2)$ として,

$$P(\boldsymbol{Y}_1 \in A_1, \boldsymbol{Y}_2 \in A_2) = P(\boldsymbol{Y}_1 \in A_1)\,P(\boldsymbol{Y}_2 \in A_2) \quad (3.48)$$

が任意の A_1, A_2 について成立することである.

さらに, n 個の確率変数 $Y_1,\cdots,Y_n$ が独立であるとは, $Y_1,\cdots,Y_n$ を 2 組に分けたときに, 任意の 2 組が独立になることをいう.

2 次元の場合と同様にして, 次の定理が得られる.

■定理 3.6

$$Y_1,\cdots,Y_n \text{ が互いに独立} \iff F(y_1,\cdots,y_n) = \prod_{i=1}^{n} F_i(y_i)$$

この条件は, 分布が離散型のときは,

$$p(y_1,\cdots,y_n) = \prod_{i=1}^{n} p_i(y_i)$$

と表される. 連続型のときは,

$$f(y_1,\cdots,y_n) = \prod_{i=1}^{n} f_i(y_i)$$

となる.

積率

一般に, n 次元の確率変数の関数 $g(Y_1, \cdots, Y_n)$ の期待値は,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(y_1, \cdots, y_n)\, dF(y_1, \cdots, y_n) \tag{3.49}$$

で定義される.

■定理 3.7　$i = 1, \cdots, k$ に対し, n 個の確率変数 $(Y_1, \cdots, Y_n)'$ の関数 $g_i(Y_1, \cdots, Y_n)$ を考え, λ_i を定数とするとき,

$$E\Big[\sum_{i=1}^{k} \lambda_i\, g_i(Y_1, \cdots, Y_n)\Big] = \sum_{i=1}^{k} \lambda_i\, E\Big[\, g_i(Y_1, \cdots, Y_n)\,\Big]$$

が成り立つ.

とくに,

$$\mu_i = E[Y_i] \tag{3.50}$$

を Y_i の平均という. 平均は同時分布に基づく期待値として定義されるが, 2 次元の場合と同様に, 周辺分布による平均と一致する.

また,

$$\sigma_{ij} = E\Big[\,(Y_i - \mu_i)\,(Y_j - \mu_j)\,\Big] \tag{3.51}$$

において, $i = j$ の場合を Y_i の分散といい, σ_i^2 や σ_{ii}, $V[Y_i]$ などと表す. $i \neq j$ の場合を Y_i, Y_j の共分散といい, σ_{ij} や Cov $[Y_i, Y_j]$ で表す. 分散, 共分散も, 定義は同時確率による期待値であるが, これらはそれぞれの周辺分布による期待値と一致する. σ_i^2 の非負平方根 σ_i を Y_i の標準偏差という.

n 次元確率変数 $\boldsymbol{Y} = (Y_1, \cdots, Y_n)'$ の平均を

$$E[\boldsymbol{Y}] = \boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)' \tag{3.52}$$

と書く. これを平均ベクトルということがある.

また, $\boldsymbol{Y}$ の分散共分散行列 $V[\boldsymbol{Y}]$ を, σ_{ij} を i, j 成分にもつ行列として定義する.

$$V[\boldsymbol{Y}] = \Sigma = \left(\sigma_{ij}\right) = E\left[(\boldsymbol{Y} - \boldsymbol{\mu})(\boldsymbol{Y} - \boldsymbol{\mu})'\right] \tag{3.53}$$

rank $\Sigma < n$ のとき, この確率変数は退化しているという.

条件つき期待値と条件つき分散についても, 2 次元の場合と同様に定義することができる. たとえば, $y_{m+1}, \cdots, y_n$ を与えたときの, $Y_1, \cdots, Y_m$ の条件つき平均ベクトルは, 次で定義される.

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (y_1, \cdots, y_m)' \, dF(y_1, \cdots, y_m \,|\, y_{m+1}, \cdots, y_n) \tag{3.54}$$

これを $\boldsymbol{e}_m = E[Y_1, \cdots, Y_m \,|\, y_{m+1}, \cdots, y_n]$ と表し, つぎに, 条件つき分散共分散行列 $V[Y_1, \cdots, Y_i \,|\, y_{m+1}, \cdots, y_n]$ を

$$\int_{-\infty}^{\infty} (\boldsymbol{Y}_m - \boldsymbol{e}_m)(\boldsymbol{Y}_m - \boldsymbol{e}_m)' dF(y_1, \cdots, y_m \,|\, y_{m+1}, \cdots, y_n) \tag{3.55}$$

で定義する. ここに, $\boldsymbol{Y}_m = (Y_1, \cdots, Y_m)'$ である.

次の公式は定理 3.7 より容易に導かれる. ここに, A は任意の定行列である.

$$E[A\boldsymbol{Y}] = A\,E[\boldsymbol{Y}], \quad V[A\boldsymbol{Y}] = A\,V[\boldsymbol{Y}]\,A' \tag{3.56}$$

■定理 3.8　n 次確率変数 $\boldsymbol{Y}$ の二次形式について,

$$E[\boldsymbol{Y}' C \boldsymbol{Y}] = E[\boldsymbol{Y}]' C\, E[\boldsymbol{Y}] + \mathrm{trace}\, C\, V[\boldsymbol{Y}] \tag{3.57}$$

が成り立つ. ただし, $C = (c_{ij})$ は定対称行列である.

■証明　次の等式から直ちに得られる.

$$E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}\, Y_i\, Y_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}\, E[Y_i]\, E[Y_j] + \sum_{i=1}^{n}\left\{\sum_{j=1}^{n} c_{ij}\,\sigma_{ij}\right\}$$

確率変数 $\boldsymbol{Y}, \boldsymbol{Z}$ の (相互) 共分散行列を,

$$\mathrm{Cov}\,[\boldsymbol{Y}, \boldsymbol{Z}] = \Big(\mathrm{Cov}\,[Y_i, Z_j]\Big) \tag{3.58}$$

で定義する. 再度, 定理 3.7 を用いて

$$\mathrm{Cov}\,[A\,\boldsymbol{Y}, B\,\boldsymbol{Z}] = A\,\mathrm{Cov}\,[\boldsymbol{Y}, \boldsymbol{Z}]\,B' \tag{3.59}$$

を得る. ただし, A, B は任意の定行列である.

■定理 3.9 Σ は非負定値行列である.

■証明 Σ は対称行列であるから, 適当な直交行列 Γ を用いて,

$$\Gamma'\Sigma\Gamma = \Lambda = \mathrm{diag}\,(\lambda_i)$$

と対角化できる. ここで, $\boldsymbol{v} = \Gamma'\boldsymbol{u} = (v_1, \cdots, v_n)'$ とおくと, (3.56) より,

$$^\forall\boldsymbol{u},\ \boldsymbol{u}'\Sigma\,\boldsymbol{u} \ge 0 \iff {}^\forall\boldsymbol{v},\ \boldsymbol{v}'\Lambda\,\boldsymbol{v} = \sum_{i=1}^{n} \lambda_i\, v_i^2 \ge 0$$

を得る. ■

つぎに, 確率変数 $\boldsymbol{Y}$ の相関行列 $R[\boldsymbol{Y}]$ を,

$$r_{ij} = \mathrm{Corr}\,[Y_i, Y_j] = \frac{\mathrm{Cov}\,[Y_i, Y_j]}{\sqrt{V[Y_i]\,V[Y_j]}} \tag{3.60}$$

を i, j 成分にもつ行列として定義する. r_{ij} は, 同時分布に基づいて定義されるが, 分散, 共分散のみで表されることから, Y_i, Y_j の周辺分布に基づいて求めてもよい.

相関行列は, 各変数の分散の違いを補正したときの分散共分散行列にほかならない. 実際,

$$\boldsymbol{Z} = \mathrm{diag}\,(\sigma_i^{-1})\,\boldsymbol{Y}$$

とおくと, (3.56) より

$$V[\boldsymbol{Z}] = \mathrm{diag}\,(\sigma_i^{-1})\,V[\boldsymbol{Y}]\,\mathrm{diag}\,(\sigma_i^{-1}) = R[\boldsymbol{Y}]$$

を得る. これから, 相関行列 R も非負定値行列であることがわかる.

多次元の場合も, 1 次元の場合とまったく同じように, 原点まわりおよび平均まわりの積率を, それぞれ

$$\mu_{k_1,\cdots,k_n} = E\Big[\prod_{i=1}^{n} Y_i^{k_i}\Big], \quad \nu_{k_1,\cdots,k_n} = E\Big[\prod_{i=1}^{n}(Y_i - \mu_i)^{k_i}\Big] \tag{3.61}$$

で定義する.

3.5　母関数

$\boldsymbol{Y}$ の積率母関数を

$$M(\boldsymbol{t}) = E\Big[\exp(\boldsymbol{t}'\boldsymbol{Y})\Big] = E\Big[\exp\Big(\sum_{i=1}^{n} t_i Y_i\Big)\Big] \tag{3.62}$$

で定義する. ここに, $\boldsymbol{t} = (t_1, \cdots, t_n)'$ である.

積率母関数から原点まわりの積率 $\mu_{k_1,\cdots,k_n}$ が得られる :

$$\mu_{k_1,\cdots,k_n} = \frac{\partial^{k_1+\cdots k_n}}{\partial t_1^{k_1} \cdots \partial t_n^{k_n}} M(\boldsymbol{t})\Big|_{\boldsymbol{t}=\boldsymbol{o}} \tag{3.63}$$

この式は 1 次元の場合とまったく同様にして得られる.

分布関数と積率母関数が 1 対 1 に対応するのも 1 次元の場合と同様である. 分布関数列の極限と, 積率母関数列の極限も対応する.

■定理 3.10　Y_i $(i = 1, \cdots, n)$ が互いに独立であることは

$$M(\boldsymbol{t}) = \prod_{i=1}^{n} M(t_i) \tag{3.64}$$

と同値である. ここに, $M(t_i)$ は Y_i の積率母関数である.

■証明　Y_i $(i = 1, \cdots, n)$ が互いに独立であれば,

$$M(\boldsymbol{t}) = E\Big[\exp\Big(\sum_{i=1}^{n} t_i\, Y_i\Big)\Big] = \prod_{i=1}^{n} E[\exp(t_i\, Y_i)] = \prod_{i=1}^{n} M(t_i)$$

が成り立つ. 逆に, (3.64) を仮定すると, 積率母関数と分布関数の 1 対 1 対応性から, 各 Y_i が独立であることがわかる. ■

第 4 章

代表的な確率分布

4.1　二項分布

一般に, 成功の確率が p , 失敗の確率が $1 - p$ であるような試行を Bernoulli 試行という. 成功なら 1, 失敗なら 0 の 2 値をとる確率変数 $Y_i\,(i = 1, \cdots, n)$ を考え, これらの独立性を仮定すると, 成功の回数

$$Y = \sum_{i=1}^{n} Y_i \tag{4.1}$$

は確率変数で, その確率関数は,

$$p(y) = {}_n\mathrm{C}_y\, p^y\,(1-p)^{n-y} \tag{4.2}$$

である. この分布を二項分布といい, $B(n, p)$ で表す.

(4.2) が確率の公理を満たすことは容易に確認できる :

$$\sum_{y=0}^{n} {}_n\mathrm{C}_y\, p^y\,(1-p)^{n-y} = \left\{p + (1-p)\right\}^n = 1$$

つぎに,

$$E[Y_i] = E\left[Y_i^2\right] = 1 \cdot p + 0 \cdot (1-p) = p \tag{4.3}$$

と, (3.56) より, Y の平均, 分散を得る :

$$E[Y] = n\,p, \quad V[Y] = n\,p\,(1-p) \tag{4.4}$$

積率母関数は次の式で与えられる．

$$M(t) = \sum_{y=0}^{n} \exp(t\,y)\ \cdot {}_n\mathrm{C}_y\, p^y\,(1-p)^{n-y} = \{p\ \exp t + (1-p)\}^n \tag{4.5}$$

Poisson 分布

二項分布 $B(n, p)$ において, (4.2) を

$$p(y) = \frac{1}{y\,!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{y-1}{n}\right)(\,np\,)^y\,(\,1-p\,)^{n-y}$$

と変形して, $n\,p = \lambda$ を一定としたまま, $n \to \infty$ とすると,

$$p(y) = \exp(-\,\lambda)\,\frac{\lambda^y}{y\,!} \tag{4.6}$$

を得る. $\exp\lambda$ の Taylor 展開

$$\exp\lambda = \sum_{y=0}^{\infty}\frac{\lambda^y}{y\,!}$$

を考えれば, (4.6) が確率の公理を満たすことはすぐにわかる.

確率関数が (4.6) で与えられる分布を Poisson 分布といい, $\mathrm{Po}(\lambda)$ で表す. パラメタ λ は Poisson 分布の強度といわれるが, 実はこれは分布の平均を表す. 実際,

$$E[Y] = \sum_{y=0}^{\infty} y\ \exp(-\lambda)\,\frac{\lambda^y}{y\,!} = \sum_{y=1}^{\infty}\exp(-\lambda)\,\frac{\lambda^{y-1}\lambda}{(y-1)\,!} = \lambda \tag{4.7}$$

同様に, $E\Big[Y(Y-1)\Big] = \lambda^2$ より, 分散も強度に一致する：

$$V[Y] = E[Y] + E\Big[\,Y(Y-1)\,\Big] - E[Y]^2 = \lambda \tag{4.8}$$

Poisson 分布は, 「稀な事象の大量観測」の場合にしばしば仮定される. 式 (4.6) の導出の過程を考えれば, このことは当然である. 1 時間にオフィスにかかってくる電話の数, 古い例では 1 年間に馬に蹴られて死んだ兵隊の数などに Poisson 分布を仮定するとあてはまりがよいといわれている.

超幾何分布

全部で N 個のもののうち, ある性質 C をもつものが m 個あるとする. N 個の中から無作為に n 個を取り出すとき, この n 個の中に C をもつものの個数 Y は確率変数で, その確率関数は

$$p(y) = \frac{{}_m\mathrm{C}_y \cdot {}_{N-m}\mathrm{C}_{n-y}}{{}_N\mathrm{C}_n} \tag{4.9}$$

である. この分布を超幾何分布といい, $HG(N, n, p)$ で表す. ただし, $p = m/N$ である.

ところで, 同時に n 個を取り出すのではなく, 1 個すつ取り出しては元に戻すようにして, これを n 回繰り返すと, C をもつものを取り出した回数 Y は $B(n, p)$ に従う. 一般に, サンプルを集団から取り出すときに, ひとつずつ戻すときを復元抽出, 戻さないときを非復元抽出という. N が十分大きい場合は, ひとつを戻す効果が相対的に薄れるので, 復元抽出も非復元抽出も実質的に変わらなくなることは直観的に理解できるだろう. 実際, (4.9) を

$$\begin{aligned} p(y) = &\frac{n\,!}{y\,!\,(n-y)\,!} \cdot \frac{m}{N} \cdot \frac{m-1}{N-1} \cdot \cdots \cdot \frac{m-y+1}{N-y+1} \\ &\cdot \frac{N-m}{N-y} \cdot \frac{N-m-1}{N-y-1} \cdot \cdots \cdot \frac{N-m-n+y+1}{N-n+1} \end{aligned}$$

と変形し, $m/N = p$ 一定のもとで $N \to \infty$ とすると, (4.2) を得る.

超幾何分布は 2×2 表の解析の際にしばしば仮定される. 2 値の 2 因子の値別に頻度を集計し, 下のような表を考える.

	反応あり	反応なし	計
要因あり	y	$n-y$	n
要因なし	$m-y$	$N-n-m+y$	$N-n$
計	m	$N-m$	N

「要因と結果は関連がない」という仮定のもとでは, m を定数視できて, 要因のある n 個のサンプルの中で y 個のサンプルに反応が得られる確率は, (4.9) で与えられる.

ところで,

$$(a+b)^m\,(a+b)^{N-m} = (a+b)^N$$

の両辺を展開し, $a^n\,b^{N-n}$ の係数を比較することによって,

$$\sum_{y=0}^{n} {}_m\mathrm{C}_y \cdot {}_{N-m}\mathrm{C}_{n-y} = {}_N\mathrm{C}_n \tag{4.10}$$

を得る. これから直ちに (4.9) が確率の公理を満たすことがわかる.

つぎに, 平均と分散を導く. まず, (4.10) と同様に

$$\sum_{y=1}^{n} {}_{m-1}\mathrm{C}_{y-1} \cdot {}_{N-m}\mathrm{C}_{n-y} = {}_{N-1}\mathrm{C}_{n-1}$$

を得る. これから, 平均 $E[Y]$ を導くことができる.

$$\begin{aligned}\sum_{y=0}^{n} y\,\frac{{}_m\mathrm{C}_y \cdot {}_{N-m}\mathrm{C}_{n-y}}{{}_N\mathrm{C}_n} &= \frac{n\,m}{N}\sum_{y=1}^{n}\frac{{}_{m-1}\mathrm{C}_{y-1} \cdot {}_{N-m}\mathrm{C}_{n-y}}{{}_{N-1}\mathrm{C}_{n-1}} \\ &= \frac{n\,m}{N} \end{aligned}\tag{4.11}$$

同様に,

$$E\Big[\,Y(Y-1)\,\Big] = \frac{n\,(n-1)\,m\,(m-1)}{N\,(N-1)}$$

が得られるから, 分散は

$$\begin{aligned} V[Y] &= E[Y] + E\Big[\,Y(Y-1)\,\Big] - E[Y]^2 \\ &= \frac{n\,m\,(N-n)\,(N-m)}{N^2\,(N-1)} \end{aligned}\tag{4.12}$$

で与えられる.

一般に, 標本が十分大きい状態を漸近的という. 統計学的では, あまり (∞ における) 極限という表現を用いない. この節で議論したことから, Poisson 分布 $\mathrm{Po}(\lambda)$ も超幾何分布 $HG(N,n,p)$ も漸近的には二項分布 $B(n,p)$ に等しいことがわかる. このことに対応して, 三者の平均は等しい. 実際, この節での notation を用いると, $n\,p = \lambda = n\,m/N$ である. 分散は多少異なるが, 漸近的には一致する.

4.2　正規分布

密度関数が

$$f(y) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} \tag{4.13}$$

で与えられる分布を正規分布といい, $N(\mu, \sigma^2)$ で表す. μ, σ は正値のパラメタである. 正規分布は, いわゆる釣り鐘型の分布の代表であり, 発見者の名にちなんでガウス分布ともいわれる.

多くの分布が漸近的に正規分布になることや既に多くの数表が正規分布に基づいて作成されていることなどにより, この分布は統計学において最も重要な分布とされている.

とくに, $\mu = 0, \sigma = 1$ のときを標準正規分布という.

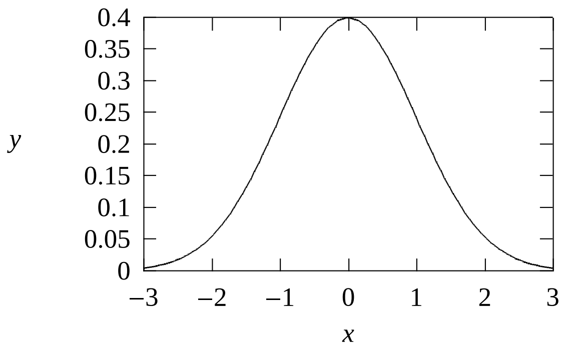


図 4.1: 標準正規分布のグラフ

このとき, (4.13) が確率の公理 (P.3) を満たすことは,

$$\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) dy = \int_0^{\infty} \int_0^{2\pi} r \exp\left(-\frac{r^2}{2}\right) dr\, d\theta = 2\pi$$

からわかる. 第 1 の等号は極座標による変数変換である.

一般の場合は, 次の置換により, (P.3) の成立が直ぐにわかる.

$$\frac{y-\mu}{\sigma} = z \tag{4.14}$$

実は, (4.13) において, パラメタ μ, σ^2 はそれぞれ平均, 分散を表す. まず, 置換 (4.14) により,

$$E[Y-\mu] = \int_{-\infty}^{\infty} \frac{y-\mu}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} dy = 0 \tag{4.15}$$

を得る. 再び, 置換 (4.14) を行うと,

$$V[Y] = \sigma^2 \int_{-\infty}^{\infty} \frac{y^2}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{y^2}{2\,\sigma^2}\right\} dy \tag{4.16}$$

を得る. ここで,

$$y^2 \exp\left(-\frac{y^2}{2}\right) = y \cdot y \exp\left(-\frac{y^2}{2}\right)$$

に注意して, 部分積分

$$\left[y\left\{-\exp\left(-\frac{y^2}{2}\right)\right\}\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\left\{-\exp\left(-\frac{y^2}{2}\right)\right\} dy$$

を行う. 第 1 項の値は 0 であり, 既にみたように, 第 2 項の値は $\sqrt{2\pi}$ である. これを (4.16) に用いて, 結論を得る. ■

なお, 式 (4.14) は, 平均を減じ, 標準偏差で除すという置換である. この操作を標準化という.

積率母関数は定義より, 次のようになる.

$$M(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \tag{4.17}$$

■**定理** 4.1　確率変数 Y が二項分布 $B(n, p)$ に従うとすると, n が十分大きいとき, 次が成り立つ.

$$Z = \frac{Y - n p}{\sqrt{n p (1-p)}} \sim N(0, 1) \tag{4.18}$$

■証明　(4.5) より,

$$E\Big[\exp(tZ)\Big] = \exp\Big(-\frac{\sqrt{np}}{\sqrt{1-p}}\,t\Big)\Big\{p\,\exp\frac{t}{\sqrt{np(1-p)}} + 1 - p\Big\}^n$$

を得る. これを $M_n(t)$ とおくと,

$$\log M_n(t) + \sqrt{\frac{np}{1-p}}\,t = n\log\Big\{1 + p\Big(\exp\frac{t}{\sqrt{np(1-p)}} - 1\Big)\Big\}$$

が成り立つ. $\exp t$ に Taylor の定理を用いて,

$$n\log\Big[\,1 + p\,\Big\{\frac{t}{\sqrt{np(1-p)}} + \frac{t^2}{2\,np(1-p)} + o(n^{-1})\Big\}\,\Big]$$

を得る. 今度は, $\log(1+t)$ に Taylor の定理を用いて,

$$\begin{aligned} &np\,\Big\{\frac{t}{\sqrt{np(1-p)}} + \frac{t^2}{2\,np(1-p)} + o(n^{-1})\Big\} \\ &\quad -\frac{np^2}{2}\Big\{\frac{t}{\sqrt{np(1-p)}} + \frac{t^2}{2\,np(1-p)} + o(n^{-1})\Big\}^2 + o(1) \end{aligned}$$

と変形できる. ここで, $n\to\infty$ とすると,

$$\log M_n(t) \to \frac{t^2}{2}$$

を得る. これは, $N(0,1)$ の積率母関数の対数に一致する. ■

この定理は, 二項分布 $B(n,p)$ が正規分布 $N(\,np, np(1-p)\,)$ に漸近的に等しいことを示している. (4.18) が標準化を表すことに注意しよう.

4.3　その他の連続分布

Γ 分布

Γ 分布は, 次の密度関数で定義される.

$$f(y) = \begin{cases} y^{\alpha-1}\exp(-y/\beta)\Big/\Gamma(\alpha)\beta^{\alpha} & (y>0) \\ 0 & (y\le 0) \end{cases} \tag{4.19}$$

α, β は正値のパラメタである. $\Gamma(\alpha)$ は Γ 関数で,

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} \exp(-y)\, dy \tag{4.20}$$

で定義される. とくに, $\alpha = n/2, \beta = 2$ の場合を自由度 n の χ^2 分布という.

$$f(y) = \frac{1}{\Gamma(n/2)\, 2^{n/2}}\, y^{n/2-1} \exp\left(-\frac{y}{2}\right) \tag{4.21}$$

■定理 4.2　α を実数, n を自然数とすると, 次が成り立つ.

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha), \quad \Gamma(n+1) = n\,! \tag{4.22}$$

■証明　第 1 式は, 部分積分

$$\Gamma(\alpha+1) = \Big[y^{\alpha}\{-\exp(-y)\}\Big]_0^{\infty} - \int_0^{\infty} \alpha\, y^{\alpha-1}\{-\exp(-y)\}\, dy$$

により得られる. これを繰り返し用いると, $\Gamma(1) = 1$ より, 第 2 式

$$\Gamma(n+1) = n\,\Gamma(n) = \cdots = n\,!\,\Gamma(1) = n\,!$$

を得る. ■

(4.19) が確率の公理を満たすことは, 次の式からわかる.

$$\int_0^{\infty} y^{\alpha-1} \exp\left(-\frac{y}{\beta}\right) dy = \int_0^{\infty} (\beta z)^{\alpha-1} \exp(-z)\beta\, dz = \Gamma(\alpha)\beta^{\alpha}$$

つぎに, 期待値と分散を求める.

$$\Gamma(\alpha)\beta^{\alpha} E[Y] = \int_0^{\infty} (\beta z)^{\alpha} \exp(-z)\beta\, dz = \beta^{\alpha+1}\,\Gamma(\alpha+1)$$

に定理 4.2 を用いて,

$$E[Y] = \alpha\beta \tag{4.23}$$

を得る.

同様に

$$E\left[\,Y^2\,\right] = (\alpha + 1)\,\alpha\,\beta^2$$

が示されるから,

$$V[Y] = E\left[\,Y^2\,\right] - E[Y]^2 = \alpha\,\beta^2 \tag{4.24}$$

を得る.

最後に, 積率母関数を導く.

$$M(t) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty y^{\alpha-1}\,\exp\left\{-\frac{1-\beta t}{\beta}\,y\right\}dy$$

において, $(1-\beta\,t)\,y = \beta\,z$ と置換して

$$M(t) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\left\{\frac{\beta}{1-\beta\,t}\right\}^\alpha\Gamma(\alpha) = (1-\beta\,t)^{-\alpha} \tag{4.25}$$

を得る.

とくに, (4.19) において, $\alpha = 1$ とおくと, $y > 0$ のときは,

$$f(y) = \frac{1}{\beta}\exp\left(-\frac{y}{\beta}\right) \tag{4.26}$$

となる. これを指数分布という. この分布は製品などの寿命にしばしば仮定される. また, 密度関数

$$f(y) = \frac{1}{2\beta}\,\exp\left(-\frac{|\,y\,|}{\beta}\right) \tag{4.27}$$

で与えられる分布を両側指数分布という.

B 分布

密度関数が

$$f(y) = \begin{cases} y^{\alpha-1}(1-y)^{\beta-1}\big/B(\alpha,\beta) & (0 < y < 1) \\ 0 & (\text{otherwise}) \end{cases} \tag{4.28}$$

で与えられる分布を B 分布という. $B(\alpha,\beta)$ は B 関数である :

$$B(\alpha,\beta) = \int_0^1 y^{\alpha-1}\,(1-y)^{\beta-1}\,dy \tag{4.29}$$

B 関数の性質

$$B(\alpha,\beta)=\frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{4.30}$$

を示しておこう. それには,

$$\Gamma(\alpha)\,\Gamma(\beta)=\int_0^\infty\int_0^\infty x^{\alpha-1}\ \exp(-x)\,y^{\beta-1}\ \exp(-y)\,dx\,dy$$

において, 次の変数変換を行う.

$$\begin{cases} x+y=u \\ x/(x+y)=v \end{cases} \Longleftrightarrow \begin{cases} x=u\,v \\ y=u\,(1-v) \end{cases}$$

Jacobian は

$$\det\begin{bmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{bmatrix}=\det\begin{bmatrix} v & u \\ 1-v & -u \end{bmatrix}=-u$$

となるから,

$$\begin{aligned}
&\int_0^\infty\int_0^1 (u\,v)^{\alpha-1}\{u\,(1-v)\}^{\beta-1}\exp(-u)\,u\,du\,dv \\
&\quad=\int_0^\infty u^{\alpha+\beta-1}\ \exp(-u)\,du\int_0^1 v^{\alpha-1}\,(1-v)^{\beta-1}\,dv
\end{aligned}$$

より (4.30) を得る. ■

密度関数が確率の公理を満たすことは, B 分布の定義から直ちにわかる. 平均および分散も (4.30) より簡単に得られる.

まず,

$$E[Y]=\frac{1}{B(\alpha,\beta)}\int_0^1 y\cdot y^{\alpha-1}\,(1-y)^{\beta-1}\,dy=\frac{B(\alpha+1,\beta)}{B(\alpha,\beta)}$$

に, 定理 4.2 と (4.30) を用いて,

$$E[Y]=\frac{\alpha}{\alpha+\beta} \tag{4.31}$$

を得る.

同様に,

$$E\Big[\,Y^2\,\Big] = \frac{\alpha\,(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

より分散が導かれる.

$$V[Y] = E[Y^2] - E[Y]^2 = \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)} \tag{4.32}$$

B 分布において, とくに $\alpha = \beta = 1$ とした場合を一様分布という.

一様分布 $U(a,b)$ は, より一般に, 密度関数

$$f(y) = \begin{cases} (b-a)^{-1} & (a < y < b) \\ 0 & (\text{otherwise}) \end{cases} \tag{4.33}$$

で与えられる分布である.

ロジスティック分布

ロジスティック分布の密度関数は,

$$f(y) = \frac{\exp y}{(1+\exp y)^2} \tag{4.34}$$

で与えられる. しかし, この分布は, 分布関数

$$F(y) = \frac{\exp y}{1+\exp y} \tag{4.35}$$

で定義した方が見通しがよい.

毒性試験のように, ある値 y を与えたときに, その値が閾値以上であれば反応が得られ, 閾値未満なら反応が起こらない 2 値変数

$$u(y) = \begin{cases} 1 & \cdots & Y \geq y \\ 0 & \cdots & Y < y \end{cases}$$

を考えよう. Y を確率変数として, y において反応の得られる確率を (4.35) とすれば,

$$P\Big(u(y) = 1\Big) = P(Y \geq y) = \frac{\exp y}{1+\exp y}$$

である. このことを用いて, 連続変数 y に, 反応の得られる確率を対応させることを考えるのが ロジスティック解析である.

4.4 多項分布

Y_{ij} ($j = 1, \cdots, b$) を, 確率 p_j で 1 を, $1 - p_j$ で 0 をとる 2 値の確率変数としよう. ただし, 各 Y_{ij} は独立ではなく,

$$\sum_{j=1}^{b} p_j = 1 \tag{4.36}$$

を満たす. 各確率 p_j は i に無関係である. いま, この試行を a 回繰り返す. それぞれの試行は独立であるとして, 新たな確率変数

$$Y_j = \sum_{i=1}^{a} Y_{ij} \quad (j = 1, \cdots, b) \tag{4.37}$$

を考える. Y_j は, a 回の施行の中で $Y_{ij} = 1$ となる回数を表す. このとき, b 次元確率変数 $\boldsymbol{Y} = (Y_1, \cdots, Y_b)'$ の同時確率関数は

$$p(y_1, \cdots, y_n) = a\,! \prod_{j=1}^{b} \frac{p_j^{y_j}}{y_j\,!} \quad \Bigl(\sum_{j=1}^{b} y_j = a\Bigr) \tag{4.38}$$

で与えられる. この分布を多項分布という. 項数を明示して, b 項分布ということがある. (4.38) は, $b = 2$ のときは, notation の違いを除いて (4.2) と一致する.

まず, (4.38) が確率の公理を満たすことを示そう.

$$\sum a\,! \prod_{j=1}^{b} \frac{p_j^{y_j}}{y_j\,!} = (p_1 + \cdots + p_b)^a = 1$$

つぎに, Y_j の平均と分散は notation の違いを除いて, 二項分布の場合と同一である :

$$E[Y_j] = a\,p_j\,, \quad V[Y_j] = a\,p_j\,(1 - p_j) \tag{4.39}$$

ここでは, 共分散を求めよう. $j \neq j'$ のとき,

$$\sum y_j\, y_{j'}\, a\,! \prod_{j=1}^{b} \frac{p_j^{y_j}}{y_j\,!} = a\,(a-1)\, p_j\, p_{j'}\, (p_1 + \cdots + p_b)^{a-2}$$

に注意して,

$$E\Big[\, Y_j Y_{j'} \,\Big] = a\,(a-1)\, p_j\, p_{j'}$$

を得るから,

$$\mathrm{Cov}\,[Y_j, Y_{j'}] = a\,(a-1)\, p_j\, p_{j'} - a\, p_j \cdot a\, p_{j'} = -a\, p_j\, p_{j'} \tag{4.40}$$

となる. Y_j の総和が一定 (a) なので, 負の共分散をもつことは直観的に理解できる.

以上の結果を行列を用いて表現しておく.

$$E[\boldsymbol{Y}] = a\, \boldsymbol{p}, \quad V[\boldsymbol{Y}] = a \Big\{ \mathrm{diag}\,(p_i) - \boldsymbol{p}\, \boldsymbol{p}' \Big\} \tag{4.41}$$

ここに, $\boldsymbol{p} = (p_1, \cdots, p_b)'$ である.

最後に, 積率母関数を導く. 独立な変数が $b-1$ 個であることに注意して, 積率母関数 $M(t_1, \cdots, t_{n-1})$ は

$$\sum \exp(t_1\, y_1 + \cdots + t_{b-1}\, y_{b-1}) \cdot \frac{a\,!}{y_1\,! \cdots y_b\,!}\, p_1^{y_1} \cdots p_b^{y_b}$$

で定義される. これから, 多項定理を用いて,

$$M(t_1, \cdots, t_{b-1}) = (p_1\ \exp t_1 + \cdots + p_{b-1}\ \exp t_{b-1} + p_b)^a \tag{4.42}$$

を得る.

多項超幾何分布

2×2 表において, 周辺分布を固定したときに仮定されたのが超幾何分布であった. 多項超幾何分布は一般の分割表 (度数分布表) で仮定される多次元確率変数の分布である.

A_i $(i = 1, \cdots, a)$ を要因側のカテゴリ, B_j $(j = 1, \cdots, b)$ を結果側のカテゴリとして次のような分割表を考えよう.

		結　果					
		B_1	$\cdots$	B_j	$\cdots$	B_b	計
	A_1	y_{11}	$\cdots$	y_{1j}	$\cdots$	y_{1b}	$n_{1.}$
要	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	A_i	y_{i1}	$\cdots$	y_{ij}	$\cdots$	y_{ib}	$n_{i.}$
因	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	A_a	y_{a1}	$\cdots$	y_{aj}	$\cdots$	y_{ab}	$n_{a.}$
	計	$n_{.1}$	$\cdots$	$n_{.j}$	$\cdots$	$n_{.b}$	n

周辺度数

$$n_{i.} = \sum_{j=1}^{b} y_{ij}\,, \quad n_{.j} = \sum_{i=1}^{a} y_{ij} \tag{4.43}$$

を固定した場合, Y_{ij} の同時確率関数 $p(y_{11}, \cdots, y_{ab})$ は

$$\frac{\prod_j (n_{.j}\,!\,/\,\prod_i y_{ij}\,!)}{n\,!\,/\,\prod_i n_{i.}\,!} = \frac{\prod_i n_{i.}\,!\prod_j n_{.j}\,!}{n\,!\prod_i\prod_j y_{ij}\,!} \tag{4.44}$$

で与えられる. この分布を多項超幾何分布 $PHG(n, \boldsymbol{r}, \boldsymbol{p})$ という.

ここに,

$$\boldsymbol{r} = (n_{1.}, \cdots, n_{a.})'\,, \quad \boldsymbol{c} = n\,\boldsymbol{p} = (n_{.1}, \cdots, n_{.b})' \tag{4.45}$$

である.

周辺度数 $n_{i.}, n_{.j}$ は,

$$\sum_{i=1}^{a} n_{i.} = \sum_{j=1}^{b} n_{.j} = n \tag{4.46}$$

を満たす.

(4.44) が確率の公理を満たすことを示そう.

$$(t_1 + \cdots + t_b)^{n_{1.}} \cdots (t_1 + \cdots + t_b)^{n_{a.}} = (t_1 + \cdots + t_b)^{n}$$

の両辺を展開して, $t_1^{n_{.1}} \cdots t_b^{n_{.b}}$ の係数を比較することにより, 次を得る.

$$\sum \frac{\prod_i n_{i.}\,!}{\prod_i\prod_j y_{ij}\,!} = \frac{n\,!}{\prod_j n_{.j}\,!}$$

和記号は, $y_{11}, \cdots, y_{ab}$ のすべてのパタンに関する和を意味する．これから直ちに次の式を得る．

$$\sum p(y_{11}, \cdots, y_{ab}) = 1$$

Y_{ij} の平均, 分散は, notation の違いを除いて超幾何分布の場合と同一である：

$$E[Y_{ij}] = \frac{n_{i.}n_{.j}}{n}, \quad V[Y_{ij}] = \frac{n_{i.}n_{.j}(n-n_{i.})(n-n_{.j})}{n^2(n-1)} \tag{4.47}$$

共分散は (4.4) と同様にして得られる．超幾何分布の分散の導出も参考にしてほしい．まず, 次の式が成り立つ．

$$\sum \frac{(n_{.j}-1)\,!}{(y_{ij}-1)\,!}\frac{(n_{.j'}-1)\,!}{(y_{ij'}-1)\,!}\prod_{k\neq j,j'}\frac{n_{.k}\,!}{y_{ik}\,!}\cdot\frac{(n_{i.}-2)\,!\prod_{k\neq i}n_{k.}\,!}{(n-2)\,!} = 1$$

したがって, $i = i'$, $j \neq j'$ のときは,

$$E[Y_{ij}Y_{i'j'}] = \sum y_{ij}\,y_{ij'}\;p(y_{11}, \cdots, y_{ab}) = \frac{n_{i.}\,(n_{i.}-1)\,n_{.j}\,n_{.j'}}{n\,(n-1)}$$

となる．$i \neq i'$, $j = j'$ の場合は, 添え字を交換することにより容易に得られる. $i \neq i'$, $j \neq j'$ の場合も同様である．

結局, 3 個の場合をまとめて, 次のように表現できる．

$$E[Y_{ij}Y_{i'j'}] = \frac{n_{i.}(n_{i'.}-\delta_{ii'})n_{.j}(n_{.j'}-\delta_{jj'})}{n\,(n-1)} \tag{4.48}$$

ただし, $\delta_{ii'}$ は, Kronecker のデルタ

$$\delta_{ii'} = \begin{cases} 1 & \cdots \quad i = i' \\ 0 & \cdots \quad i \neq i' \end{cases} \tag{4.49}$$

である．以上から, 共分散は次のように表現できる．

$$\mathrm{Cov}\,[Y_{ij}, Y_{i'j'}] = \frac{n_{i.}\,n_{.j}\,(\delta_{ii'}\,n-n_{i'.})\,(\delta_{jj'}\,n-n_{.j'})}{n^2\,(n-1)} \tag{4.50}$$

この結果を行列表現する. $\boldsymbol{Y} = (Y_{11}, \cdots, Y_{ab})'$ とおくと, その平均, 分散はそれぞれ次のようになる. $\otimes$ は行列の直積を表す．

$$E[\boldsymbol{Y}] = \frac{\boldsymbol{r}\otimes\boldsymbol{c}}{n} \tag{4.51}$$

$$V[\boldsymbol{Y}] = \frac{\{n\,\mathrm{diag}\,(n_{i.})-\boldsymbol{r}\,\boldsymbol{r}'\}\otimes\{n\,\mathrm{diag}\,(n_{.j})-\boldsymbol{c}\,\boldsymbol{c}'\}}{n^2\,(n-1)} \tag{4.52}$$

多項超幾何分布は, $p_j = n_{\cdot j}/n$ $(j = 1, \cdots, b)$ を一定としたまま, $n \to \infty$ とすると, 多項分布に近づく. (4.44) を,

$$\prod_{i}^{a-1} \frac{n_{i\cdot}\,!}{\prod_j y_{ij}\,!} \cdot \frac{n_{\cdot 1}\,(n_{\cdot 1}-1)\cdots(n_{\cdot 1}-y_{a1}+1)}{n\,(n-1)\cdots(n-y_{a1}+1)} \cdots \cdot$$
$$\cdot \frac{n_{\cdot b}(n_{\cdot b}-1)\cdots(n_{\cdot b}-y_{ab}+1)}{(n-y^*_{a(b-1)})(n-y^*_{a(b-1)}-1)\cdots(n-y^*_{ab}+1)}$$

と書き換える. ここに, $y_{a1}+\cdots+y_{aj} = y^*_{aj}$ $(j = 1, \cdots, b-1)$ である. ここで, $n \to \infty$ とすると,

$$\prod_{i=1}^{a-1} \frac{n_{i\cdot}\,!}{\prod_j y_{ij}\,!}\, p_1^{y_{i1}} \cdots p_b^{y_{ib}}$$

となり, $a-1$ 個の b 項分布の積となる.

4.5　多変量正規分布

確率変数 $\boldsymbol{Y} = (Y_1, \cdots, Y_n)'$ の同時密度関数が

$$f(\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}\,|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\,(\boldsymbol{y}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right\} \tag{4.53}$$

で与えられる場合, この分布を多変量正規分布といい, $N_n(\boldsymbol{\mu}, \Sigma)$ で表す. $\boldsymbol{\mu}, \Sigma$ はそれぞれ n 次元の定ベクトル, 定正定値対称行列である.

まず, (4.53) が (P.3) を満たすことを示す. Σ は正定値対称行列であるから,

$$\Gamma'\,\Sigma\,\Gamma = \Lambda = \mathrm{diag}\,(\lambda_i) \quad (\lambda_i > 0)$$

となる直交行列 Γ が存在して, $\boldsymbol{Z} = \Gamma'(\boldsymbol{Y}-\boldsymbol{\mu})$ とおくと,

$$\int_{-\infty}^{\infty} f(\boldsymbol{y})\,d\boldsymbol{y} = \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{n/2}|\Lambda|^{1/2}} \exp\left\{-\frac{1}{2}\,\boldsymbol{z}'\,\Lambda^{-1}\,\boldsymbol{z}\right\} d\boldsymbol{z}$$

が成り立つ. Λ は対角行列だから, 上の積分は

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\lambda_i^{1/2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{z_i^2}{2\,\lambda_i}\right\} dz_i = 1^n = 1$$

となる. ■

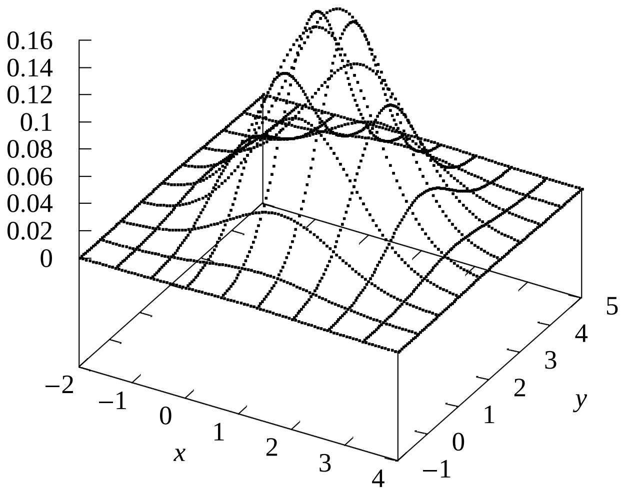


図 4.2: 2 変量標準正規分布のグラフ

1 変量の場合と同様に, $\boldsymbol{\mu}, \Sigma$ はそれぞれ平均と分散を表す. 実際, 上の証明から, Z_i $(i = 1, \cdots, n)$ は互いに独立で, それぞれ $N(0, \lambda_i)$ に従うことがわかる. したがって,

$$E[\boldsymbol{Z}] = \boldsymbol{o}, \quad V[\boldsymbol{Z}] = \Lambda$$

であり, (3.56) を用いて, 次の式を得る.

$$E[\boldsymbol{Y}] = \Gamma\,\boldsymbol{o} + \boldsymbol{\mu} = \boldsymbol{\mu}, \quad V[\boldsymbol{Y}] = \Gamma\,\Lambda\,\Gamma' = \Sigma \tag{4.54}$$

$\boldsymbol{Z}$ の積率母関数は, 各 Z_i が独立なことから,

$$\prod_{i=1}^{n} E\Big[\exp(t_i Z_i)\Big] = \prod_{i=1}^{n} \exp\left(\frac{\lambda_i}{2} t_i^2\right) = \exp\left(\frac{1}{2}\, \boldsymbol{t}' \Lambda\, \boldsymbol{t}\right) \tag{4.55}$$

とわかる. ここに, $\boldsymbol{t} = (t_1, \cdots, t_n)'$ である. したがって, $\boldsymbol{Y}$ の積率母関数は次のようになる：

$$\exp\left\{ \boldsymbol{t}' \boldsymbol{\mu} + \frac{1}{2} (\Gamma'\, \boldsymbol{t})' \Lambda\, (\Gamma'\, \boldsymbol{t}) \right\} = \exp\left(\boldsymbol{t}' \boldsymbol{\mu} + \frac{1}{2}\, \boldsymbol{t}' \Sigma\, \boldsymbol{t} \right) \tag{4.56}$$

(4.53) で多変量正規分布を定義するためには Σ が正則でなければならないが, rank $\Sigma < n$ の場合は積率母関数が (4.56) になることを多変量正規分布の定義とすればよい.

■定理 4.3　$\boldsymbol{l} = (l_1, \cdots, l_n)'$ を定ベクトルでとするとき, 次の 2 命題は同値である.
1) n 次元確率変数 $\boldsymbol{Y} = (Y_1, \cdots, Y_n)'$ が $N_n(\boldsymbol{\mu}, \Sigma)$ に従う.
2) $Y = \boldsymbol{l}'\, \boldsymbol{Y}$ が $N(\boldsymbol{l}' \boldsymbol{\mu}, \boldsymbol{l}'\, \Sigma\, \boldsymbol{l})$ に従う.

■証明 1) ⇒2)　Y の積率母関数は, (4.17) より,

$$M(t) = \exp\left(\boldsymbol{l}'\, \boldsymbol{\mu}\, t + \frac{1}{2} \boldsymbol{l}'\, \Sigma\, \boldsymbol{l}\, t^2 \right) \tag{4.57}$$

となる. これは, $N(\boldsymbol{l}'\, \boldsymbol{\mu}, \boldsymbol{l}'\, \Sigma\, \boldsymbol{l})$ の積率母関数にほかならない.

■ 2) ⇒1)　逆に, (4.57) を

$$M(t) = \exp\left\{ (t\, \boldsymbol{l})' \boldsymbol{\mu} + \frac{1}{2} (t\, \boldsymbol{l})' \Sigma (t\, \boldsymbol{l}) \right\}$$

とかき直し, $\boldsymbol{t} = t\, \boldsymbol{l}$ とおくと, (4.56) に一致する. ■

以上から, 正規分布する確率変数の線型和はやはり正規分布に従う. これを正規分布の再生性という.

■定理 4.4　$\boldsymbol{Y} = (Y_1, \cdots, Y_n)'$ が $N_n(\boldsymbol{\mu}, \Sigma)$ に従うとき, 各 Y_i が無相関であることと独立であることは同値である.

■証明　独立 ⇒ 無相関は一般に成立するので, 無相関 ⇒ 独立を示す.

各 Y_i が無相関であるとき, Σ の非対角成分が 0 となるから, (4.53) は各 Y_i の密度関数の積として表される. したがって, 定理 3.6 より Y_i は互いに独立になる. ■

最後に, 多項分布が漸近的には多変量正規分布になることを示そう. $Y_1, \cdots, Y_b$ を (4.38) に従う確率変数とする. まず,

$$Z_i = \frac{Y_i - a\,p_i}{\sqrt{a\,p_i\,(1-p_i)}}$$

と標準化する. (4.42) と (4.36) より, $(Z_1, \cdots, Z_b)'$ の積率母関数 $M(t_1, \cdots, t_{b-1})$ は,

$$\prod_{i=1}^{b-1} \exp\left(-\frac{\sqrt{a\,p_i}}{\sqrt{1-p_i}}\,t_i\right)\left\{1 + \sum_{i=1}^{b-1} p_i\left(\exp\frac{t_i}{\sqrt{a\,p_i\,(1-p_i)}} - 1\right)\right\}^a$$

となる. 両辺の対数をとると, 次の式を得る.

$$\begin{aligned}&\log M(t_1, \cdots, t_{b-1}) + \sum_{i=1}^{b-1}\frac{\sqrt{a\,p_i}}{\sqrt{1-p_i}}\,t_i\\&= a\log\left\{1 + \sum_{i=1}^{b-1} p_i\left(\exp\frac{t_i}{\sqrt{a\,p_i\,(1-p_i)}} - 1\right)\right\}\end{aligned}$$

ここで, $\exp t$ に Taylor の定理を用いると, 上の式の右辺は,

$$a\ \log\left\{1 + \sum_{i=1}^{b-1} p_i\left(\frac{t_i}{\sqrt{a\,p_i\,(1-p_i)}} + \frac{t_i^2}{2\,a\,p_i(1-p_i)} + o(a^{-1})\right)\right\}$$

と変形できる. さらに, $\log(1+t)$ の Taylor 展開を用いて,

$$\begin{aligned}&a\sum_{i=1}^{b-1} p_i\left\{\frac{t_i}{\sqrt{a\,p_i\,(1-p_i)}} + \frac{t_i^2}{2\,a\,p_i\,(1-p_i)}\right\}\\&-\frac{a}{2}\left\{\sum_{i=1}^{b-1} p_i\left(\frac{t_i}{\sqrt{a\,p_i\,(1-p_i)}} + \frac{t_i^2}{2\,a\,p_i\,(1-p_i)}\right)\right\}^2 + o(1)\end{aligned}$$

を得る.

結局, 積率母関数の対数は, 次のようになる.

$$\log M(t_1, \cdots, t_{b-1}) = \sum_{i=1}^{b-1} \frac{t_i^2}{2} + \sum_{i \neq j} r_{ij}\, t_i\, t_j + o(1)$$

ただし,

$$r_{ij} = \mathrm{Corr}\,[Y_i, Y_j] = -\frac{\sqrt{p_i\, p_j}}{\sqrt{(1-p_i)\,(1-p_j)}}.$$

ここで, $a \to \infty$ とすると,

$$\log M(t_1, \cdots, t_m) \to \frac{1}{2}\sum_{i=1}^{b-1} t_i^2 + \sum_{i \neq j} r_{ij}\, t_i\, t_j$$

を得る. これは平均 $\boldsymbol{o}$, 分散 $R = (r_{ij})$ の $b-1$ 次元多変量正規分布の積率母関数の対数にほかならない. ■

第 5 章

確率変数の関数の分布

一般に, 確率変数の関数もまた確率変数である. とくに, 統計学的な考察をする目的で定義された確率変数の関数を統計量という. この章では, t 統計量や F 統計量などの応用上重要な統計量の密度関数を導く. 後半では, 同一分布に従う確率変数の標本平均は母平均に収束し, しかもその分布は漸近的に正規分布であることを示そう.

5.1 確率変数の関数の分布

確率変数 Y の関数 $Z = g(Y)$ の分布を求めることを考える. Y の分布が離散型であれば, 明らかに

$$q(z) = \sum_{g(y)=z} p(y) \tag{5.1}$$

で, Z の確率関数が与えられる. 和記号は $g(y) = z$ なるすべての y に関する和を意味する. $p(y)$ は Y の確率関数である.

Y の分布が連続型であるときは, 若干の修正を要する. Y の密度関数を $f(y)$ で表す. $z = g(y)$ が微分可能な単調増加関数であるとすると, Z の分布関数は

$$H(z) = P(Z \leq z) = P\Big(Y \leq g^{-1}(z)\Big) = \int_{-\infty}^{g^{-1}(z)} f(y)\, dy$$

となる.

この式の両辺を z で微分すると, 密度関数

$$h(z) = f\left(g^{-1}(z)\right) \frac{d}{dz} g^{-1}(z) \tag{5.2}$$

を得る. この結果は $z = g(y)$ が微分可能な単調減少関数である場合も同じである.

つぎに, 2 次元の場合を考える. 簡単のため, (X, Y) から (U, V) への対応は 1 対 1 であり,

$$\begin{cases} U = g_1(X, Y) \\ V = g_2(X, Y) \end{cases} \Longleftrightarrow \begin{cases} X = \phi_1(U, V) \\ Y = \phi_2(U, V) \end{cases}$$

と表されるとする. $(X, Y)'$ の分布が離散型なら

$$q(u, v) = \sum_{u=g_1(x,y), v=g_2(x,y)} p(x, y) \tag{5.3}$$

で $(U, V)'$ の同時確率関数が与えられる. $p(x, y)$ は $(X, Y)'$ の同時確率関数である.

分布が連続型で, ϕ_1 , ϕ_2 が連続微分可能であれば, $(U, V)'$ の同時分布関数は

$$H(u, v) = P(U \leq u, V \leq v) = \int_{D(u,v)} f(x, y)\, dx\, dy \tag{5.4}$$

となる. ここに,

$$D(u, v) = \{(x, y) \,|\, g_1(x, y) \leq u \,,\; g_2(x, y) \leq v\} \tag{5.5}$$

である. $f(x, y)$ は (X, Y) の同時密度関数である. 重積分の変数変換公式により, (5.4) は,

$$\int_{-\infty}^{u} \int_{-\infty}^{v} f\left(\phi_1(t, s), \phi_2(t, s)\right) |\det J(t, s)|\, dt\, ds \tag{5.6}$$

と書ける. ただし,

$$J(t, s) = \begin{bmatrix} \partial x / \partial t & \partial x / \partial s \\ \partial y / \partial t & \partial y / \partial s \end{bmatrix} \tag{5.7}$$

である.

両辺を u, v で偏微分すると, $(U, V)'$ の同時密度関数

$$h(u, v) = f\Big(\phi_1(u, v), \phi_2(u, v)\Big)\Big|\det J(u, v)\Big| \tag{5.8}$$

を得る.

n 次元の場合も同様の議論ができる.

$$(Y_1, \cdots, Y_n) \xrightarrow{g} (Z_1, \cdots, Z_n)$$

なる写像が連続微分可能であるとすると, 逆写像が存在する. そこで,

$$Z_i = g_i(Y_1, \cdots, Y_n) \Longleftrightarrow Y_i = \phi_i(Z_1, \cdots, Z_n) \quad (i = 1, \cdots n)$$

とおく. 分布が離散型のときは, $(Z_1, \cdots, Z_n)'$ の同時確率関数は

$$q(z_1, \cdots, z_n) = \sum p(y_1, \cdots, y_n) \tag{5.9}$$

である. ただし, $p(y_1, \cdots, y_n)$ は $(Y_1, \cdots, Y_n)'$ の同時確率関数であり, 和記号は,

$$g_i(y_1, \cdots, y_n) = z_i \quad (i = 1, \cdots, n)$$

となるすべての $(y_1, \cdots, y_n)$ の組み合わせに関する和を表すとする.

連続型のときは, i, j 成分が $\partial y_i/\partial z_j$ である行列を $J(z_1, \cdots, z_n)$ と書くと, 同時密度関数は,

$$\begin{aligned} h(z_1, \cdots, z_n) = f\Big(&\phi_1(z_1, \cdots, z_n), \cdots, \phi_n(z_1, \cdots, z_n)\Big) \\ &\times\Big|\det J(z_1, \cdots, z_n)\Big| \end{aligned} \tag{5.10}$$

である. $f(y_1, \cdots, y_n)$ は $(Y_1, \cdots, Y_n)'$ の同時密度関数である.

χ^2 分布

確率変数 Y が標準正規分布 $N(0, 1)$ に従うとき, 確率変数 $Z = Y^2$ の従う分布を考える. Z の分布関数は, 次のようになる.

$$F(z) = P(-\sqrt{z} \leq Y \leq \sqrt{z}) = 2\int_0^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{y^2}{2}\Big) dy$$

この両辺を z で微分して, 密度関数

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z}{2}\right) z^{-1/2} \tag{5.11}$$

を得る. この分布を自由度 1 の χ^2 分布といい, χ_1^2 で表す.

つぎに, χ_1^2 に従う独立な n 個の確率変数 Y_i $(i = 1, \cdots, n)$ に対し, 新たな確率変数

$$Y = \sum_{i=1}^{n} Y_i \tag{5.12}$$

を考えると, Y は自由度 n の χ^2 分布 χ_n^2 (4.21) に従う. 実際, χ_1^2 は Γ 分布において, $\alpha = 1/2$, $\beta = 2$ とおいた場合であるから, その積率母関数は (4.25) により

$$M(t) = E\Big[\exp(t\,Y_i)\Big] = (1 - 2\,t)^{-1/2}$$

である. したがって, Y (5.12) の積率母関数は, 次のようになる.

$$E\Big[\exp(t\sum_i Y_i)\Big] = \prod_i E\Big[\exp(t\,Y_i)\Big] = (1 - 2\,t)^{-n/2}$$

これは (4.25) において $\alpha = n/2$, $\beta = 2$ とおいたものにほかならない. ■

これから χ^2 分布の再生性とよばれる次の性質を得る.

■定理 5.1　確率変数 X, Y がそれぞれ独立に自由度 χ_m, χ_n に従うとき, $X + Y$ は自由度 χ_{m+n} に従う.

$\boldsymbol{Y}$ が退縮した多変量正規分布の場合は, 次のように χ^2 統計量を構成すればよい.

■定理 5.2　n 次元確率変数 $\boldsymbol{Y}$ が, n 変量正規分布 $N_n(\boldsymbol{\mu}, \Sigma)$ に従うとき, $(\boldsymbol{Y} - \boldsymbol{\mu})'\Sigma^-(\boldsymbol{Y} - \boldsymbol{\mu}) \sim \chi_r$ が成り立つ. ここに, r は Σ のランクであり, Σ^- は Σ の一般化逆行列である.

■証明　Σ は, 非負定値対称行列であるから,

$$\Gamma'\Sigma\Gamma = \Lambda = \mathrm{diag}\,(\lambda_i)$$
$$(\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0)$$

を満たす直交行列 Γ が存在する. ここで,

$$\boldsymbol{Z} = \Gamma'(\boldsymbol{Y} - \boldsymbol{\mu}) = (Z_1, \cdots, Z_n)'$$

とおくと, (3.56) により, 各 Z_i $(i = 1, \cdots, r)$ は独立で, それぞれ $N(0, \lambda_i)$ に従うことがわかる. そこで,

$$\Lambda^- = \mathrm{diag}\,(\lambda_i^-), \quad \lambda_i^- = \begin{cases} 1/\lambda_i & \cdots \quad \lambda_i > 0 \\ 0 & \cdots \quad \lambda_i = 0 \end{cases}$$

とおくと, $\Sigma^- = \Gamma\Lambda^-\Gamma'$ であるから,

$$(\boldsymbol{Y} - \boldsymbol{\mu})'\Sigma^-(\boldsymbol{Y} - \boldsymbol{\mu}) = \boldsymbol{Y}'\Lambda^-\boldsymbol{Y} = \sum_{i=1}^{r} \frac{Z_i^2}{\lambda_i} \sim \chi_r^2$$

を得る. ■

(4.23) (4.24) より χ_n^2 の平均および分散

$$E[Y] = \frac{n}{2} \cdot 2 = n, \quad V[Y] = \frac{n}{2} \cdot 2^2 = 2n \tag{5.13}$$

を得る.

また, Y_i $(i = 1, \cdots, n)$ が互いに独立に, それぞれ正規分布 $N(\mu_i, 1)$ に従う場合,

$$\sum_{i=1}^{n} Y_i^2$$

の従う分布を非心度

$$\left(\mu_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right)^2$$

の非心 χ^2 分布という.

t 分布

X, Y が独立な確率変数で, それぞれ $N(0,1), \chi_n^2$ に従うとき,

$$T = \frac{X}{\sqrt{Y/n}} \tag{5.14}$$

の従う分布を自由度 n の t 分布といい, t_n で表す. また, 確率変数 X が正規分布 $N(\gamma, 1)$ に従い, これと独立な確率変数 Y が χ_n に従うとき, T (5.14) の従う分布を, 自由度 n , 非心度 γ の非心 t 分布という.

t_n の密度関数を導こう. X, Y がそれぞれ独立に, $N(0,1), \chi_n^2$ に従うとき, 同時密度関数は次のようになる.

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \cdot \frac{1}{\Gamma(n/2)\, 2^{n/2}}\, y^{n/2-1} \exp\left(-\frac{y}{2}\right)$$

ここで

$$\begin{cases} t = \sqrt{n}\, x / \sqrt{y} \\ s = y \end{cases} \Longleftrightarrow \begin{cases} x = t\sqrt{s/n} \\ y = s \end{cases}$$

なる変数変換をすると,

$$\det\begin{bmatrix} \partial x/\partial t & \partial x/\partial s \\ \partial y/\partial t & \partial y/\partial s \end{bmatrix} = \det\begin{bmatrix} \sqrt{s/n} & t/2\sqrt{n\,s} \\ 0 & 1 \end{bmatrix} = \sqrt{s/n}$$

を得る. したがって, T, S の同時密度関数は (5.8) より,

$$\frac{1}{\sqrt{2\,n\,\pi}} \cdot \frac{1}{\Gamma(n/2)\, 2^{n/2}}\, s^{(n+1)/2-1} \exp\left\{-\frac{1}{2}\left(1 + \frac{t^2}{n}\right)s\right\}$$

となる. T の周辺密度関数 $f(t)$ は,

$$\frac{1}{\sqrt{n\pi}\,\Gamma(n/2)\, 2^{(n+1)/2}} \int_0^\infty s^{(n+1)/2-1} \exp\left\{-\frac{1}{2}\left(1 + \frac{t^2}{n}\right)s\right\} ds$$

で与えられる.

$$u = \frac{1}{2}\left(1 + \frac{t^2}{n}\right)s$$

と置換して,

$$f(t) = \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma(n/2 + 1/2)}{\Gamma(n/2)}\left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \tag{5.15}$$

を得る.

F 分布

X, Y が独立な確率変数で, それぞれ χ^2_m, χ^2_n に従うとき,

$$W = \frac{X/m}{Y/n} \tag{5.16}$$

の従う分布を自由度 m, n の F 分布といい, F^m_n で表す.

この分布の密度関数を導こう. X, Y の同時密度関数は

$$\frac{1}{\Gamma(m/2)\,\Gamma(n/2)\,2^{(m+n)/2}}\,x^{m/2-1}\,y^{n/2-1}\exp\!\left(-\frac{x+y}{2}\right)$$

である. ここで

$$\begin{cases} w = n\,x/m\,y \\ z = y \end{cases} \Longleftrightarrow \begin{cases} x = m\,z/n \\ y = z \end{cases}$$

なる変数変換をすると, (5.8) より Z, W の同時密度関数

$$\frac{(m\,/\,n)^{m/2}\,w^{m/2-1}}{\Gamma(m/2)\,\Gamma(n/2)\,2^{(m+n)/2}} z^{(m+n)/2-1}\exp\!\left\{-\frac{1}{2}\left(1+\frac{m}{n}\,w\right)z\right\}$$

を得る. したがって, W の周辺密度関数は,

$$\begin{aligned} f(w) = &\frac{(m/n)^{m/2}\,w^{m/2-1}}{\Gamma(m/2)\,\Gamma(n/2)\,2^{(m+n)/2}} \\ &\cdot \int_0^\infty z^{(m+n)/2-1}\exp\!\left\{-\frac{1}{2}\left(1+\frac{m}{n}\,w\right)z\right\}dz \end{aligned}$$

となる.

$$v = \frac{1}{2}\left(1+\frac{m}{n}\,w\right)z$$

と置換して,

$$\frac{\Gamma(m/2+n/2)}{\Gamma(m/2)\,\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2} w^{m/2-1}\left(\frac{m}{n}\,w+1\right)^{-(m+n)/2} \tag{5.17}$$

を得る.

t 分布は 2 標本問題において, 平均の差を考察するときに構成する最も基本的な統計量の分布である. また, F 分布は分散分析における効果判定のための統計量が従う分布としてやはり非常に重要である.

5.2　標本平均と標本分散の分布

母集団とは, 仮説の対象となる集合である. 実際に実験を行う対象は, 母集団から抽出された要素であり, 標本という. 実験およびその解析は, 標本に関して得た情報を一般化して, 母集団の性質を考察することにほかならない. このことを統計学的推論という.

この際, 標本が母集団の性質をそのまま反映するように, 標本が母集団から偏りなく抽出されなければならない. これが無作為化の原則である. また, 抽出した標本の個数 —標本の大きさという— が小さすぎると, 実験データからの推論に信憑性がないことは明らかであろう.

標本平均と標本分散の分布

n 個の確率変数 Y_i $(i = 1, \cdots, n)$ に対し,

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{5.18}$$

を標本平均という. これは, n 個の観測値 $y_1, \cdots, y_n$ の算術平均に対応する確率変数である. また,

$$S_n = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2 \tag{5.19}$$

を標本分散という. 期待値のかわりに算術平均としているのは, 標本平均と同様である.

■定理 5.3　独立な n 個の確率変数 Y_i $(i = 1, \cdots, n)$ が同一分布 $N(\mu, \sigma^2)$ に従うとき, 標本平均 $\overline{Y}_n$ (5.18) と標本分散 S_n (5.19) は独立で, それぞれ次の分布に従う.

$$\overline{Y}_n \sim N(\mu, \sigma^2/n), \quad \frac{nS_n}{\sigma^2} \sim \chi^2_{n-1}$$

■証明　正規分布の再生性から, $\overline{Y}_n$ は正規分布し, (3.56) より

$$E[\overline{Y}_n] = \mu\,, \quad V[\overline{Y}_n] = \frac{\sigma^2}{n} \tag{5.20}$$

となることがわかる.

nS_n/σ^2 の分布については, まず,

$$\begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} = \begin{bmatrix} n^{-1/2}\boldsymbol{j}_n' \\ Q' \end{bmatrix} \begin{bmatrix} (Y_1-\mu)/\sigma \\ \vdots \\ (Y_n-\mu)/\sigma \end{bmatrix}$$

なる直交変換を考える. ただし, Q' は

$$Q'\boldsymbol{j}_n = \boldsymbol{o}\,, \quad Q'Q = I_{n-1} \tag{5.21}$$

を満たす $n-1$ 行 n 列行列である. このとき, 各 Z_i は正規分布の再生性からそれぞれ独立に標準正規分布に従う.

他方, 恒等式

$$\sum_{i=1}^{n}(Y_i-\mu)^2 = \sum_{i=1}^{n}(Y_i-\overline{Y}_n)^2 + \sum_{i=1}^{n}(\overline{Y}_n-\mu)^2$$

より,

$$\frac{nS_n}{\sigma^2} = \sum_{i=1}^{n}\frac{(Y_i-\mu)^2}{\sigma^2} - \left(\frac{\overline{Y}_n-\mu}{\sigma/\sqrt{n}}\right)^2 = \sum_{i=2}^{n} Z_i^2 \sim \chi_{n-1}$$

を得る. (5.21) より $\overline{Y}_n$ と S_n の独立性はすぐにわかる.

5.3　大数の法則と中心極限定理

大数の法則

標本平均は, 標本の大きさ n が大きくなると, 母平均 μ に近づく.

■定理 5.4　任意の正数 ε に対して, 次が成立する. (大数の弱法則)

$$\lim_{n\to\infty} P\Big(\,|\,\overline{Y}_n-\mu\,|\geq\varepsilon\Big) = 0$$

■証明 定理 5.3 より $V(\overline{Y}_n) = \sigma^2/n$ であるから, 系 2.4 より,

$$P\Big(|\overline{Y}_n - \mu| \geq \varepsilon\Big) \leq \frac{\sigma^2}{n\varepsilon^2}$$

を得る. ここで $n \to \infty$ とすると, 定理 5.4 を得る. ■

定理 5.4 の収束を確率収束というが, 次の定理は, もう少し強い意味での収束 — 概収束という — を保証する.

■定理 5.5 任意の正数 ε に対して, 次が成立する. (大数の強法則)

$$P\Big(\lim_{n\to\infty} |\overline{Y}_n - \mu| = 0\Big) = 1$$

中心極限定理

定理 5.3 では正規分布を仮定したが, n が十分大きいときは正規性は必要なく, 同一分布にさえ従えば同様の結論が得られる.

■定理 5.6 $Y_1, \cdots, Y_n$ が独立で, 平均が μ, 分散が σ^2 の同一分布に従うとき, 漸近的に次が成立する. (中心極限定理)

$$\frac{\overline{Y}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

■証明 $Z_i = (Y_i - \mu)/\sigma$ とおくと, 定理 2.1 より各 Z_i の平均は 0 , 分散は 1 となる. Z_i の積率母関数の存在を仮定すると, これらは i によらず一定で, これを $M(t)$ とおく. $M(t)$ が C^3 級であるときは, $M(t)$ の定義と (2.25) より,

$$M(0) = 1, \quad M'(0) = E[Z_i] = 0$$
$$M''(0) = E\Big[Z_i^2\Big] = V[Z_i] = 1$$

を得る. ただし, $'$ は t による微分を表す.

ここで,

$$W = \frac{\overline{Y}_n - \mu}{\sigma / \sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i$$

とおくと, W の積率母関数は,

$$M\left(\frac{t}{\sqrt{n}}\right)^n = \exp\left\{n \log M\left(\frac{t}{\sqrt{n}}\right)\right\} \tag{5.22}$$

である. そこで, $h(t) = \log M(t)$ とおくと, 微分公式

$$h'(t) = \frac{M'(t)}{M(t)}, \quad h''(t) = \frac{M''(t)\, M(t) - M'(t)^2}{M(t)^2}$$

より, $h(0) = h'(0) = 0\,, \quad h''(0) = 1$ より, 次のように書ける.

$$h(t) = \log M(t) = \frac{1}{2}\, t^2 + \frac{h'''(0)}{6}\, t^3 + o(t^3)$$

これを (5.22) に代入すると,

$$\exp\left\{n\left(\frac{t^2}{2n} + \frac{h'''(0)\, t^3}{6\, n^{3/2}} + O(n^{-2})\right)\right\} \to \exp\left(\frac{t^2}{2}\right) \quad (n \to \infty)$$

となる. この極限は標準正規分布 $N(0, 1)$ の積率母関数の対数にほかならないから, 定理が得られる. ■

第 II 部

統計学的推論

第 6 章

検定

第 I 部では, 分布のパラメタを既知として, その分布の性質を考察した. しかし, 多くの実験においてはパラメタの値は未知である. この部では, 得られた標本から, 母集団の性質を推論する方法を議論する.

6.1 仮説と検定

少々不謹慎な例だが, 博打の話から始める. ひとりがテラ場で丁に賭け続けるとしよう. ところが, 出目は半ばかり続く. 最初のうちは「ついてないな」と思うかもしれないが, 10 回も続けて半が出ると, 彼は「ああ, イカサマにはまった.」と思うことだろう.

このときの彼の思考回路を整理してみよう. 最初, 彼は「サイコロは正しく作られている」と暗黙のうちに認めている. この認識のもとに, 10 回続けて半が出る確率は $(1/2)^{10} \cong 10^{-3}$ であるから, この現実は, 非常に珍しいことが起こったか, もともとサイコロがイカサマだったかのどちらである. 通常はこの現実の前に, ついていないと思うよりイカサマにはまったと判断するだろう.

これほど極端でないにしても, ある程度半の回数の割合が大き (小さ) ければ, イカサマの可能性は大きくなる. どこでイカサマと判断するかは個人差があるだろうが, 客観的な基準を定めておけば大きな被害を被らずに済むであろう. たとえば, 10 回の勝負で, 丁の出た回数が 8 回以上または 2 回以下のときはイカサマと判断するルールを作る.

このルールは,「サイコロは正しい」という仮定のもとに, 観測を得る確率

$$\sum_{y=0}^{2} {}_{10}\mathrm{C}_y \left(\frac{1}{2}\right)^{10} + \sum_{y=8}^{10} {}_{10}\mathrm{C}_y \left(\frac{1}{2}\right)^{10} = 0.109\cdots \tag{6.1}$$

が十分小さいと判断できるときに正当なものとなるであろう.

一般に, 科学研究においては, ある仮説を検証するために実験や調査を行う. 統計学的検定は, 背理法的な論理を用いて, 自分が証明したい仮説 H_1 と反対の仮説 H_0 — 先の例では「サイコロは正しい」という最初の認識は否定される運命にあった — を定めておき, これを誤りであると判断できるデータが得られたときに, H_1 を支持するという手順を踏む. 否定されることを期待されている H_0 を帰無仮説といい, 興味のある仮説 H_1 を対立仮説という.

棄却域

検定のルールを決めることは, 得られた標本 $\boldsymbol{y}$ に対して, 標本空間 Ω の部分集合 W を,

$$\boldsymbol{y} \in W \text{ なら } H_1 \text{ を支持}, \quad \boldsymbol{y} \notin W \text{ なら } H_0 \text{ を支持} \tag{6.2}$$

するように定めることにほかならない. W を棄却域という. H_1 を支持する場合, 帰無仮説を棄却するといい, H_0 を支持する場合は, 帰無仮説を採択するという. 明らかに, 棄却域 W と, 検定関数

$$\phi(\boldsymbol{y}) = \begin{cases} 1 & \cdots \quad \boldsymbol{y} \in W \\ 0 & \cdots \quad \boldsymbol{y} \notin W \end{cases} \tag{6.3}$$

は 1 対 1 に対応する.

二種類の誤り

検定において, 帰無仮説を棄却するか採択するか, どちらの判断をするにしても誤りを犯す可能性は避けられない. すなわち

1. H_0 が正しいのに H_1 を支持する誤り
2. H_1 が正しいのに H_0 を支持する誤り

の二種類の誤りを犯す危険が伴う.

前者を第一種の誤りといい, 後者を第二種の誤りという. 研究者は帰無仮説を否定する目的で実験をするのだから, 第一種の誤りは, 自分の仮説が誤っているのに正しいと信じてしまうあわてもののミス, 第二種の誤りは自分の仮説が正しいのにそれに気づかないうっかりもののミスである.

上の定義から, 二種類の誤りをともに小さくすることは不可能である. 第一種の誤りを小さくしようとして, 棄却域 W を小さくとりすぎると, 第二種の誤りが大きくなってしまう. このことを検定が保守的になりすぎるという.

有意水準と危険率

このようなことを避けるために, 第一種の誤りに上限 α を定め, このもとに第二種の誤りを最小にする検定法式が推奨されている：

$$P\left(\boldsymbol{y} \notin W \middle| H_1\right) \to \mathrm{Min} \quad \text{under} \quad P\left(\boldsymbol{y} \in W \middle| H_0\right) \le \alpha \tag{6.4}$$

$P(\boldsymbol{y} \in W \mid H_0)$ を危険率といい, その上限 α を有意水準という. また, $1-P(\boldsymbol{y} \notin W \mid H_1)$ は棄却したい帰無仮説をきちんと棄却できる確率を表すので, 検出力という. これは検定の感度にほかならない.

両側検定, 片側検定

棄却域 W は通常, 標本 $\boldsymbol{y}$ のある関数 $T(\boldsymbol{y})$ —先の例では丁の回数とした—を用いて,

$$W_1 = \left\{ \boldsymbol{y} \,\middle|\, |T(\boldsymbol{y})| > |t| \right\} \tag{6.5}$$

$$W_2 = \left\{ \boldsymbol{y} \,\middle|\, T(\boldsymbol{y}) > t \right\} \tag{6.6}$$

$$W_3 = \left\{ \boldsymbol{y} \,\middle|\, T(\boldsymbol{y}) < t \right\} \tag{6.7}$$

などのように定める. $T(\boldsymbol{y})$ に対応する確率変数 $T(\boldsymbol{Y})$ を統計量, t を臨界値という. (6.5) の場合を両側検定, (6.6) (6.7) の場合を片側検定という. とくに, (6.6) を右片側検定, (6.7) を左片側検定という.

両側検定をするか片側検定をするかは対立仮説の性質による. 丁半賭博の例では, 丁ばかり出るのも半ばかり出るのもイカサマの証拠になるので両側検定が妥当であろう. これに対して, 品質検査のような場合は, 不良品の割合が低すぎる場合は問題にしなくてもよいであろう.

なお,

$$P\Big(T(\boldsymbol{Y}) > T(\boldsymbol{y}) \Big) \tag{6.8}$$

$$P\Big(|T(\boldsymbol{Y})| > |T(\boldsymbol{y})| \Big) \tag{6.9}$$

などを有意確率または p 値という.

有意水準 α の検定は有意確率が α 以下なら帰無仮説を棄却することになる. このことを単に有意水準 α で有意ということがある. 通常, 有意水準として, 0.05 や 0.01 といった値が設定 — 医学統計などでは 0.05 という数値がひとり歩きしている — されるが, 有意確率を記載しておいた方が一般に情報が多い.

6.2　ランダム検定

検出力を上げるために危険率は有意水準いっぱいにとるのが望ましい. このため, 危険率と有意水準はときに混同されるが, $\boldsymbol{y}$ が離散値をとるときは, 両者を等しくするのは一般に不可能である.

このような場合は, 統計量がちょうど臨界値をとるときに, つねに帰無仮説を棄却するのではなく, 一定の確率で棄却するようにすればよい.

たとえば, 両側検定のときは検定関数を

$$\phi(\boldsymbol{y}) = \begin{cases} 1 & \cdots \quad |T(\boldsymbol{y})| > |t| \\ p & \cdots \quad |T(\boldsymbol{y})| = |t| \\ 0 & \cdots \quad |T(\boldsymbol{y})| < |t| \end{cases} \tag{6.10}$$

とすればよい. ただし,

$$p = \frac{\alpha - P\Big(|T(\boldsymbol{y})| > |t| \,\Big|\, H_0\Big)}{P\Big(|T(\boldsymbol{y})| = |t| \,\Big|\, H_0\Big)} \tag{6.11}$$

である. このような検定をランダム検定という.

実際には, 一様分布 $U(0,1)$ に従う確率変数 U を乱数で求めて, $U < p$ なら棄却, $U \geq p$ なら採択とすればよい. このとき, (6.11) より, たしかに危険率は有意水準と等しくなる.

右片側検定の場合は

$$\phi(\boldsymbol{y}) = \begin{cases} 1 & \cdots \quad T(\boldsymbol{y}) > t \\ q & \cdots \quad T(\boldsymbol{y}) = t \\ 0 & \cdots \quad T(\boldsymbol{y}) < t \end{cases} \tag{6.12}$$

とすればよい. ただし, q は次により定める.

$$q = \frac{\alpha - P\Big(T(\boldsymbol{y}) > t \,\Big|\, H_0\Big)}{P\Big(T(\boldsymbol{y}) = t \,\Big|\, H_0\Big)} \tag{6.13}$$

■例 6.1　丁半賭博の例で, 10 回中丁の出る回数を Y とする. 有意水準を 0.05 とすると,

$$P(Y \le 1) + P(Y \le 9) = \frac{2(1+10)}{1024} = 0.021\cdots$$
$$P(Y = 2) + P(Y = 8) = \frac{2 \cdot 45}{1024} = 0.088\cdots$$

より, $y \le 1, y \ge 9$ のときはつねにサイコロはイカサマ, $y = 2, 8$ のときは確率 $(0.05 - 0.021)/0.088 = 0.33\cdots$ でイカサマと判断すると, 危険率がちょうど 0.05 となる.

ランダム検定は, 検出力の点では利点があるが, 有意か否かの判定が実験結果とまったく無関係に決まることから敬遠されることもしばしばである.

6.3　最強力検定

検出力を最大にするように棄却域を定める検定を最強力検定という. 簡単のため, 分布の型を既知とし, 唯一の未知パラメタ θ を含むとする. このとき, 検定方式を

$$\begin{array}{ll} \text{帰無仮説} & H_0 : \theta = \theta_0 \\ \text{対立仮説} & H_1 : \theta = \theta_1 \end{array} \tag{6.14}$$

とする. たとえば, 新薬と従来薬において, 有効率を比較する場合などがこれに当たる. この H_0 や H_1 のように, そのもとで, $\boldsymbol{Y}$ の分布が一意に定まるような仮説を単純仮説という.

この場合は次の定理が最強力検定の方式を与える.

■定理 6.1　標本 $\boldsymbol{y}$ に対応する確率変数 $\boldsymbol{Y}$ の密度関数を $f(\boldsymbol{y}, \theta)$ とする. このとき, 検定法式を (6.14) とすると, 最強力検定の棄却域は次で与えられる.

$$W = \left\{ \boldsymbol{y} \,\middle|\, \frac{f(\boldsymbol{y}, \theta_1)}{f(\boldsymbol{y}, \theta_0)} > c \right\} \tag{6.15}$$

ただし c は定数で, 有意水準 α から定まる. (Neyman-Pearson)

■証明 まず, 危険率を有意水準以下に設定することから,

$$P\Big(\boldsymbol{y} \in W \,\Big|\, H_0\Big) = \int \phi(\boldsymbol{y})\, f(\boldsymbol{y}, \theta_0)\, d\boldsymbol{y} \le \alpha.$$

が成り立つ. ここで, 任意の正数 c に対して,

$$\phi^*(\boldsymbol{y}) = \begin{cases} 1 & \cdots \quad f(\boldsymbol{y}, \theta_1) - c\, f(\boldsymbol{y}, \theta_0) > 0 \\ 0 & \cdots \quad f(\boldsymbol{y}, \theta_1) - c\, f(\boldsymbol{y}, \theta_0) \le 0 \end{cases}$$

とおくと, 任意の検定関数 $\phi(\boldsymbol{y})$ に対して,

$$\int \Big\{\phi^*(\boldsymbol{y}) - \phi(\boldsymbol{y})\Big\} \Big\{f(\boldsymbol{y}, \theta_1) - c\, f(\boldsymbol{y}, \theta_0)\Big\}\, d\boldsymbol{y} \ge 0$$

が成り立つ. そこで, 定数 c を

$$c \int \phi^*(\boldsymbol{y})\, f(\boldsymbol{y}, \theta_0)\, d\boldsymbol{y} = \alpha$$

となるように定めることができれば,

$$\begin{aligned} &\int \phi^*(\boldsymbol{y})\, f(\boldsymbol{y}, \theta_1)\, d\boldsymbol{y} - \int \phi(\boldsymbol{y})\, f(\boldsymbol{y}, \theta_1)\, d\boldsymbol{y} \\ &\ge c \int \{\phi^*(\boldsymbol{y}) - \phi(\boldsymbol{y})\}\, f(\boldsymbol{y}, \theta_0)\, d\boldsymbol{y} \ge 0 \end{aligned}$$

を得る. これは, $\phi^*(\boldsymbol{y})$ が最強力検定を与えることを示す. ■

■例 6.2 Y が正規分布 $N(\mu, \sigma^2)$ に従うとする. σ^2 を既知として,

$$\text{帰無仮説 } H_0 : \mu = \mu_0 \tag{6.16}$$

$$\text{対立仮説 } H_1 : \mu = \mu_1 > \mu_0 \tag{6.17}$$

なる検定を考える. このとき, 最強力検定の棄却域は

$$\begin{aligned} \exp\Big(\frac{\mu_1 - \mu_0}{\sigma^2}\, y - \frac{\mu_1^2 - \mu_0^2}{2\sigma^2}\Big) > c &\Longleftrightarrow (\mu_1 - \mu_0)\, y > c' \\ &\Longleftrightarrow y > c'' \end{aligned}$$

である. ここに, c, c', c'' は定数である. 結局, 標本 y がある程度大きな値をとるときに H_1 を支持するという方式となる.

6.4　検出力関数

検定法式において, 帰無仮説 H_0 と対立仮説 H_1 は対称に扱われていない. このことは, 有意水準を大きく設定してもそれが検出力の大きさを保証しないことを意味する.

例 6.1 で, 丁の出る確率を p として, 対立仮説を $H_1 : p = 4/5$ としてみよう. (6.1) より, 棄却域を $y \leq 2$ または $y \geq 8$ としたときは, 危険率は 0.109 であるが, 検出力は

$$\sum_{y \leq 2,\, y \geq 8} {}_{10}\mathrm{C}_y \left(\frac{4}{5}\right)^y \left(\frac{1}{5}\right)^{10-y} = 0.678\cdots$$

にすぎない.

一般に, 検出力は帰無仮説と対立仮説におけるパラメタの値の差の関数となる. この関数を検出力関数という.

実際は少し修正して, 検出力を例 6.1 の場合

$$p - 1/2$$

例 6.2 の場合

$$\delta = \frac{\mu_1 - \mu_0}{\sigma}$$

の関数とみる場合が多い. このプロットを検出力曲線という.

■例 6.3　例 6.2 で, 棄却域を

$$W = \left\{ y \,\middle|\, y \geq \mu_0 + \sigma K_\alpha \right\} \tag{6.18}$$

とすると, 危険率は α である. K_α は Y を標準正規分布 $N(0,1)$ に従う確率変数として, $P(Y \geq K_\alpha) = \alpha$ を満たす定数である.

このとき, 検出力

$$P\left(\frac{y - \mu_1}{\sigma} \geq K_\alpha - \delta \,\middle|\, \mu_1 \right) = 1 - \Phi^{-1}(K_\alpha - \delta)$$

は δ の増加関数である. $\Phi(y)$ は標準正規分布の分布関数である.

一般に, Y を分布 D に従う確率変数とするとき,

$$P\left(Y \geq d_\alpha \middle| D \right) = \alpha \tag{6.19}$$

を満たす d_α を, 分布 D の上側 α 点という. 例 6.3 の K_α は標準正規分布の上側 α 点である.

6.5　複合仮説の検定

より複雑な対立仮説を考える場合, たとえば,

$$\begin{array}{ll} \text{帰無仮説} & H_0 : \theta = \theta_0 \\ \text{対立仮説} & H_1 : \theta > \theta_1 \end{array}$$

とした場合には, 対立仮説のもとでの $\boldsymbol{Y}$ の分布が一意に定まらない. このような場合の仮説を複合仮説という.

一様最強力検定

もし, 対立仮説のもとでの任意の θ に対して, 同一の最強力検定を構成できるならば, それを一様最強力検定という.

たとえば, 例 6.3 において, 棄却域 W は μ_1 に依らないから, これは対立仮説を $\mu > \mu_0$ としたときの一様最強力検定の棄却域を与える. 同様に, 左側片側仮説 $\mu < \mu_0$ に対しては, 一様最強力検定の棄却域は $W' = \left\{ y \middle| y \leq \mu_0 - \sigma K_\alpha \right\}$ で与えられる.

ところが, 対立仮説を $\mu \neq \mu_0$ とすると, 一様最強力検定は存在しない. 実際, $\mu > \mu_0$ なる μ に対しては, 棄却域を (6.18) とすればよいが, このとき, $\mu < \mu_0$ なる μ に対しては, 検出力は

$$P\left(\frac{y - \mu_0}{\sigma} > K_\alpha \middle| \mu < \mu_0 \right) < \alpha$$

であり, 著しく小さい.

不偏検定

そこで, H_2 を検定するのに, 対立仮説のもとで, 帰無仮説を棄却する確率がつねに有意水準より小さくならないという基準を設ける. この性質を検定の不偏性という. 不偏な検定の中で, 対立仮説のもとでつねに最強力な検定があれば, それを不偏一様最強力検定という.

例 6.2 において, μ のときの検出力

$$P\Bigl(y \in W \Bigm| \mu\Bigr) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma}\,\phi(y)\ \exp\Bigl\{-\frac{(y-\mu)^2}{2\,\sigma^2}\Bigr\}dy$$

を μ の関数とみて $g(\mu)$ とおき, これを最大化することを考えよう. $\phi(y)$ は検定関数である.

不偏性は, $g(\mu) \geq \alpha$ と表現できるが, 危険率を有意水準とすると, $g(\mu_0) = \alpha$ となるから, $g(\mu)$ は $\mu = \mu_0$ のとき最小値をとり,

$$g'(\mu_0) = \frac{1}{\sqrt{2\pi}\,\sigma^3}\int_{-\infty}^{\infty}(y-\mu_0)\,\phi(y)\ \exp\Bigl\{-\frac{(y-\mu_0)^2}{2\sigma^2}\Bigr\}\,dy = 0$$

が成り立つ. この条件と, 有意水準に関する次の条件を課す.

$$g(\mu_0) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\phi(y)\exp\Bigl\{-\frac{(y-\mu_0)^2}{2\sigma^2}\Bigr\}\,dy = \alpha$$

いま, $\phi(y)$ と $\phi^*(y)$ が上の 2 条件を満たすとして, 任意の $\phi(y)$, μ に対して,

$$\int_{-\infty}^{\infty}\{\phi^*(y) - \phi(y)\}\ \exp\Bigl\{-\frac{(y-\mu)^2}{2\sigma^2}\Bigr\}\,dy \geq 0 \tag{6.20}$$

となる $\phi^*(y)$ があれば, それが求める検定関数である. c_1, c_2 を定数とすると, 2 個の制約条件から, (6.20) は

$$\int_{-\infty}^{\infty}\{\phi^*(y) - \phi(y)\}\,h(y)\,dy \geq 0$$

と書ける. ただし, $h(y)$ は次で与えられる.

$$h(y) = \exp\Bigl\{-\frac{(y-\mu)^2}{2\sigma^2}\Bigr\} - \{c_1(y-\mu_0) + c_2\}\exp\Bigl\{-\frac{(y-\mu_0)^2}{2\sigma^2}\Bigr\}$$

結局, $\phi^*(y) - \phi(y)$ と $h(y)$ の符号を一致するように $\phi^*(y)$ を定めればよい. すなわち, $h(y) \geq 0$ を棄却域にすればよい. これは, 指数関数が凸関数であることに注意すると,

$$\exp(\mu - \mu_0) y \geq c_1' y + c_2' \Longleftrightarrow y \leq c_1'', y \geq c_1'' \tag{6.21}$$

と同値である. (c_1', c_2', c_1'', c_2'' は定数)

非常に天下り的ではあるが, 棄却域を

$$W = \left\{ y \middle| y \leq \mu_0 - \sigma K_{\alpha/2},\ y \geq \mu_0 + \sigma K_{\alpha/2} \right\} \tag{6.22}$$

とすると, これは (6.21) の形であり, 2 個の制約条件を満たすので, 不偏一様最強力検定を与えることがわかる.

不偏性は, 検出力が有意水準以下にならないという条件であるが, これは検出力のレベルとしてはかなり甘い.

相似検定

今度は, 帰無仮説のもとでも $\boldsymbol{Y}$ の分布が一意に決まらない場合を考える. 帰無仮説のもとでの $\boldsymbol{Y}$ の分布を, 密度関数 $f(\boldsymbol{y}|\theta, \eta)$ で表そう. このとき, 任意の η に対して, 帰無仮説 : $\theta = \theta_0$ を検定したいとする. この場合の η のように, 推論の直接の対象とはならないが, 推論に影響を与えうる未知パラメタを撹乱母数という.

このような場合は, 危険率が撹乱母数の値によらず, つねに有意水準に等しくなるように $^\forall\eta, P\left(\boldsymbol{Y} \in W \middle| \theta_0, \eta\right) = \alpha$ とすればよい. このような検定を相似検定という. 一般に, 相似検定は撹乱母数に無関係な分布に従う統計量を用いて構成される.

■例 6.4　$\boldsymbol{Y} = (Y_1, \cdots, Y_n)'$ を多変量正規分布 $N_n(\mu_0\, \boldsymbol{j}, \sigma^2 I)$ に従う確率変数とすると, 定理 5.3 と t 分布の定義より,

$$T = \frac{\sqrt{n-1}\,(\overline{Y}_n - \mu_0)}{\sqrt{S_n}} \sim t_{n-1} \tag{6.23}$$

を利用して相似検定を構成できる. $\overline{Y}_n, S_n$ は定理 5.3 と同一である. σ^2 がこの場合の撹乱母数である. この検定を t 検定という.

6.6　尤度比検定

定理 6.1 を拡張して,

$$\text{帰無仮説 } H_0 : \theta \in \Theta_0, \quad \text{対立仮説 } H_1 : \theta \in \Theta_1$$

なる検定を考えよう. Θ_0, Θ_1 はそれぞれの仮説のもとでの θ のとりうる値の集合を表す. 一般に, 分布のパラメタのとりうる値の集合を母数空間という.

分布が連続型であり, 同時密度関数が $f(\boldsymbol{y}\,|\,\theta)$ と表される場合は, 上の検定の棄却域は, 定数 c を用いて,

$$\max_{\theta\in\Theta_1} f\Big(\boldsymbol{y}\,\Big|\,\theta\Big)\Big/\max_{\theta\in\Theta_0} f\Big(\boldsymbol{y}\,\Big|\,\theta\Big) > c \tag{6.24}$$

とすればよい. これが, Neyman-Pearson の最強力検定の自然な拡張になっていることは明らかであろう.

(6.24) の最大は, θ に関する最大を意味している. 明らかに, 同時密度関数 $f(\boldsymbol{y}\,|\,\theta)$ は, θ に依存する. そこで, これを θ の関数とみて, $L(\theta)$ とおく. $L(\theta)$ は標本 $\boldsymbol{y}$ が得られる可能性を表すが, 一般に確率の公理は満たさないので, 尤度関数という. 尤度関数の対数 $l(\theta) = \log L(\theta)$ を対数尤度関数という. 尤度関数, 対数尤度関数はしばしば単にそれぞれ尤度, 対数尤度といわれる. 未知母数が複数個存在する場合は, それらの組をベクトル $\boldsymbol{\theta}$ として, θ のかわりに $\boldsymbol{\theta}$ とすればよい. 本書を通じて, 尤度関数は $L(\theta)$ または $L(\boldsymbol{\theta})$, 対数尤度は $l(\theta)$ または $l(\boldsymbol{\theta})$ と表す. (6.24) は尤度の比を統計量としているので, 尤度比検定という.

第 7 章

推定

7.1 点推定

検定の際に棄却域を構成したように, 標本 $\boldsymbol{y}$ の関数 $T(\boldsymbol{y})$ を用いて未知パラメタ θ を推定することを考える. $\boldsymbol{y}$ を対応する確率変数 $\boldsymbol{Y}$ で置き換えた統計量 $T(\boldsymbol{Y})$ を推定量といい, その実現値を推定値という. θ の推定量であることを強調して, $T(\boldsymbol{Y}) = \hat{\theta}$ と表すことがある. このように, パラメタの値を数値として推定することを点推定という.

不偏推定量

推定量 $T(\boldsymbol{Y})$ の性質の良さを, しばしば $T(\boldsymbol{Y})$ とパラメタの真の値 θ の差の二乗の期待値

$$E\left[\{T(\boldsymbol{Y}) - \theta\}^2\right] = \int_{-\infty}^{\infty} \{T(\boldsymbol{y}) - \theta\}^2 dF\left(\boldsymbol{y}, \theta\right) \tag{7.1}$$

で評価する. (7.1) を平均二乗誤差という. Mean Square Error を略して, 以下平均二乗誤差を MSE と表す.

もし, 任意の θ に対して, MSE をつねに最小にする推定量が存在するならば, それが一様に最良な推定量となるが, そのような推定量は存在しない.

実際, $\theta = \theta_1$ に対しては, 恒等的に $T(\boldsymbol{Y}) = \theta_1$ となる意味のない推定量が最小の MSE を与えるが, それは, 別の値 $\theta = \theta_2$ のときは, 明らかに $T(\boldsymbol{Y}) = \theta_2$ より大きな MSE を与える.

そこで, 推定量 $T(\boldsymbol{Y})$ に, 条件

$$E\Big[\,T(\boldsymbol{Y})\,\Big] = \theta \tag{7.2}$$

を課すことが多い.

統計量 $T(\boldsymbol{Y})$ が (7.2) を満たすとき, 不偏推定量という. このとき, MSE は推定量の分散 $V[\boldsymbol{Y}]$ に一致する. 不偏推定量の中で, 分散が最小となる推定量を最良不偏推定量という.

■例 7.1　確率変数 $Y_1, \cdots, Y_n$ がそれぞれ独立に, 正規分布 $N(\mu, \sigma^2)$ に従うとき, 標本平均 (5.18) は母平均 μ の不偏推定量である. しかし, 標本分散 (5.19) は母分散 σ^2 の不偏推定量ではない. 実際, (5.13) と定理 5.3 より

$$E\Big[\,\overline{Y}_n\,\Big] = \mu, \quad V[S_n] = \frac{n-1}{n}\,\sigma^2 \tag{7.3}$$

が成り立つ.

例 7.1 より,

$$\hat{\sigma}_*^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2 \tag{7.4}$$

は母分散の不偏推定量である. これを標本不偏分散という. 実は, (7.4) は最良不偏推定量であるが, このことは後に示そう.

7.2　有効推定量

一般の場合に最良不偏推定量を求めるのは容易ではないが, 適当な条件のもとで, 分散の下限は与えられる.

■定理 7.1　$\boldsymbol{Y} = (Y_1, \cdots, Y_n)'$ の同時密度関数を $f(\boldsymbol{y}, \theta)$ とするとき, 不偏推定量 $T(\boldsymbol{Y})$ の分散は次の不等式を満たす. この不等式を Cramer-Rao の不等式という.

$$V[\,T(\boldsymbol{Y})\,] \geq I(\theta)^{-1} \tag{7.5}$$

ただし, $I(\theta)$ は, Fisher の情報量

$$I(\theta) = E\left[-\frac{\partial^2 \log f(\boldsymbol{Y}, \theta)}{\partial \theta^2}\right] = E\left[\left(\frac{\partial \log f(\boldsymbol{Y}, \theta)}{\partial \theta}\right)^2\right] \tag{7.6}$$

である. 離散分布の場合は密度関数を確率関数に置き換えればよい. なお, (7.6) の第 2 の等号は下の証明の中で示す.

■証明　簡単のため, $T(\boldsymbol{Y}), T(\boldsymbol{y})$ を T , $f(\boldsymbol{Y}, \theta)$, $f(\boldsymbol{y}, \theta)$ を f と表す. また,

$$S = \frac{\partial}{\partial \theta} \log f(\boldsymbol{Y}, \theta)$$

とおく. まず, T は θ の不偏推定量であることと, 確率の公理から

$$\int_{-\infty}^{\infty} T\, f\, d\boldsymbol{y} = \theta, \quad \int_{-\infty}^{\infty} f\, d\boldsymbol{y} = 1$$

が成り立つ. 微積分の順序交換を保証する適当な条件のもとで, θ で微分すると次の 2 式を得る.

$$\int_{-\infty}^{\infty} T\, \frac{\partial f}{\partial \theta}\, d\boldsymbol{y} = \int_{-\infty}^{\infty} T\, \frac{\partial \log f}{\partial \theta} \cdot f\, d\boldsymbol{y} = 1 \tag{7.7}$$

$$\int_{-\infty}^{\infty} \frac{\partial f}{\partial \theta}\, d\boldsymbol{y} = \int_{-\infty}^{\infty} \frac{\partial \log f}{\partial \theta} \cdot f\, d\boldsymbol{y} = 0 \tag{7.8}$$

(7.8) が $E[S] = 0$ を意味することに注意しよう.

(7.7) (7.8) より,

$$\int_{-\infty}^{\infty} (T - \theta) \cdot \frac{\partial \log f}{\partial \theta} \cdot f\, d\boldsymbol{y} = \mathrm{Cov}\,[\,T, S\,] = 1 \tag{7.9}$$

を得る.

ところで, 一般に, 確率変数 Z, W の間には,

$$\mathrm{Corr}\,[Z, W] \le 1 \Longleftrightarrow \mathrm{Cov}\,[Z, W] \le V[Z]\,V[W]$$

が成り立つので, (7.9) より

$$1 \le V[T]\,V[S] \tag{7.10}$$

を得る.

ここで, 再度 (7.8) の両辺を θ で微分すると

$$\int_{-\infty}^{\infty} \frac{\partial^2 f}{\partial \theta^2}\, d\boldsymbol{y} = 0 \tag{7.11}$$

となる. さらに, 微分公式

$$\frac{\partial^2 \log f}{\partial \theta^2} = \frac{1}{f}\frac{\partial^2 f}{\partial \theta^2} - \left(\frac{1}{f}\frac{\partial f}{\partial \theta}\right)^2$$

の両辺の期待値をとると, (7.11) に注意して, (7.6) の第 2 の等号が成り立つことがわかる. (7.8) より $E[S] = 0$ であったから, これは

$$V[S] = E\left[S^2\right] = I(\theta)$$

を意味する. これと, (7.10) より定理を得る. ■

Cramer-Rao の不等式の下限の分散をもつ統計量を有効推定量という. 一般の推定量 $\hat{\theta}$ に対し,

$$100 \cdot \left(I(\hat{\theta})\,V\left[\hat{\theta}\right]\right)^{-1} (\%) \tag{7.12}$$

を推定量の効率という.

■例 7.2　$Y_1, \cdots, Y_n$ を正規分布 $N(\mu, \sigma^2)$ に従う独立な確率変数であるとする. このとき, 標本平均 $\overline{Y}_n$ (5.18) は母平均 μ の不偏推定量であった. (⇒ 例 7.1)

実は, 標本平均 $\overline{Y}_n$ は母平均 μ の有効推定量である. 実際, 対数尤度

$$l(\mu) = \text{const} - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2 \tag{7.13}$$

の両辺を μ で 2 回偏微分すると,

$$\frac{\partial^2 l(\mu)}{\partial\theta^2} = -\frac{n}{\sigma^2}$$

となるので,

$$V\Big[\,\overline{Y}_n\,\Big] = \sigma^2/n \tag{7.14}$$

は Cramer-Rao の下限である.

Efficient Score

定理 7.1 の証明の中で, 統計量 S は対数尤度の微分として定義された :

$$S = \frac{\partial}{\partial\theta}\log f(\boldsymbol{Y},\theta) \tag{7.15}$$

この統計量を Efficient Score という.

なお, 上の証明の中で, 次が示された.

$$E[S] = 0\,,\quad V[S] = I(\theta) \tag{7.16}$$

未知パラメタが複数の場合

未知パラメタが 2 個以上の場合,

$$E\left[-\frac{\partial^2\log f(\boldsymbol{Y},\boldsymbol{\theta})}{\partial\theta_i\,\partial\theta_j}\right] = E\left[\left(\frac{\partial\log f(\boldsymbol{Y},\boldsymbol{\theta})}{\partial\theta_i}\right)\left(\frac{\partial\log f(\boldsymbol{Y},\boldsymbol{\theta})}{\partial\theta_j}\right)\right]$$

を i, j 成分にもつ行列として Fisher の情報行列を定義する. 等号成立は, 1 次元の場合と同様に示される. 情報行列は

$$I(\boldsymbol{\theta}) = E\left[-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] = E\left[\left(\frac{\partial l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)'\right] \tag{7.17}$$

と表現できる. このとき, 定理 7.1 は次のように拡張される.

■定理 7.2　確率変数 $\boldsymbol{Y}$ の同時密度関数を $f(\boldsymbol{Y}, \boldsymbol{\theta})$ とするとき, 未知パラメタ $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)'$ の推定量 $\boldsymbol{T}(\boldsymbol{Y})$ の分散 $V[\,\boldsymbol{T}(\boldsymbol{Y})\,]$ に関して, 次が成立する.

$$^{\forall}\boldsymbol{u} \in \boldsymbol{R}^k, \quad \boldsymbol{u}'\{V[\,\boldsymbol{T}(\boldsymbol{Y})\,] - I(\boldsymbol{\theta})^{-1}\}\boldsymbol{u} \geq 0 \tag{7.18}$$

すなわち, $V[\,\boldsymbol{T}(\boldsymbol{Y})\,] - I(\boldsymbol{\theta})^{-1}$ は非負定値行列である.

■証明　Efficient Score を

$$\boldsymbol{S} = \frac{\partial \log f(\boldsymbol{Y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{7.19}$$

で定義すると, 1 次元の場合と同様に,

$$E[\boldsymbol{S}] = \boldsymbol{o}, \quad V[\boldsymbol{S}] = I(\boldsymbol{\theta}) \tag{7.20}$$

を得る. また, (7.9) と同様にして, Cov $[\,\boldsymbol{T}, \boldsymbol{S}\,] = I_n$ が成り立つ.

したがって,

$$V[\,\boldsymbol{T} - I(\boldsymbol{\theta})^{-1}\boldsymbol{S}\,] = V[\boldsymbol{T}] - 2\,I(\boldsymbol{\theta})^{-1} + I(\boldsymbol{\theta})^{-1} = V[\boldsymbol{T}] - I(\boldsymbol{\theta})^{-1}$$

は非負定値である. ■

パラメタが 2 次元以上の場合, $\Sigma = I(\theta)^{-1}$ を Cramer-Rao の下限という. このとき, $\boldsymbol{T}$ を同時有効推定量という. これは一般の場合, 存在するとは限らない.

■例 7.3　$Y_1, \cdots, Y_n$ が互いに独立に, 正規分布 $N(\mu, \sigma^2)$ に従う場合において, μ, σ^2 を推定する. このとき, 対数尤度は (7.13) であり, Fisher の情報行列は, 簡単な計算の後, 次のようになる.

$$I(\mu, \sigma^2) = \begin{bmatrix} n/\sigma^2 & \sum_i (y_i - \mu)/\sigma^4 \\ \sum_i (y_i - \mu)/\sigma^4 & \{2\sum_i (y_i - \mu)^2 - n\sigma^2\}/\sigma^6 \end{bmatrix}$$

他方, 例 7.2 で μ の有効不偏推定量は (5.18) であった. また, 定理 5.3 と (5.13) より, $\hat{\sigma}^2_*$ (7.4) の分散は, 次のようになる.

$$V[\hat{\sigma}^2_*] = \frac{\sigma^4}{(n-1)^2} \cdot 2\,(n-1) = \frac{2\sigma^4}{n-1} \tag{7.21}$$

これは, 後に示すように分散を最小にするので, 結局,

$$V\left[\,(\overline{Y}_n, \hat{\sigma}_*^2)'\,\right] = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/(n-1) \end{bmatrix} \tag{7.22}$$

が最良不偏推定量の分散共分散行列である. しかし, これは

$$I(\mu, \sigma^2)^{-1}$$

と一致しないので, 同時有効推定量は存在しない.

ここで, $\hat{\sigma}_*^2$ (7.4) が最小分散を与える不偏推定量であることを示そう. まず, その準備として次の定理を示す.

■定理 7.3　確率変数 $\boldsymbol{Y}$ の従う分布の未知パラメタを $\boldsymbol{\theta}$ とする. このとき, $\boldsymbol{\theta}$ の関数 $g(\boldsymbol{\theta})$ の不偏推定量 $\hat{g}$ が最良不偏推定量であるための必要十分条件は, $\hat{g}$ が 0 の任意の不偏推定量 h と無相関であることである. (Rao)

■証明　必要性) k を任意の定数とすると, $\hat{g} + kh$ もやはり g の不偏推定量となるから,

$$V[\hat{g} + kh] = V[\hat{g}] + 2k\,\mathrm{Cov}\,[\hat{g}, h] + V[h]$$

より, $\mathrm{Cov}\,[\hat{g}, h] \neq 0$ の場合は,

$$V[\,\hat{g} + kh] \leq V[\hat{g}]$$

となるように k を定めることができる.

十分性) $\hat{g}$ が 0 の任意の不偏推定量と無相関であるとする. $\hat{g}^*$ を $\hat{g}$ 以外の任意の g の不偏推定量であるとすると, $\hat{g}^* - \hat{g}$ は 0 の不偏推定量であるから,

$$\mathrm{Cov}\,[\,\hat{g}, \hat{g}^*\,] = \mathrm{Cov}\,[\,\hat{g}, \hat{g}^* - \hat{g}\,] + V[\hat{g}]$$

において, 右辺の第 1 項は 0 になる. したがって,

$$\begin{aligned} 0 \leq V[\,\hat{g} - \hat{g}^*\,] &= V[\hat{g}] - 2\,\mathrm{Cov}\,[\,\hat{g}, \hat{g}^*\,] + V[\,\hat{g}^*\,] \\ &= V[\,\hat{g}^*\,] - V[\hat{g}] \end{aligned}$$

が成り立つ. ■

確率変数 $Y_1, \cdots, Y_n$ が互いに独立に正規分布 $N(\mu, \sigma^2)$ に従うとする．このとき, $\boldsymbol{Y} = (Y_1, \cdots, Y_n)'$ の同時密度関数は

$$f(\boldsymbol{y}) = \frac{1}{(2\pi\sigma)^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right\} \tag{7.23}$$

である. いま, 0 の任意の不偏推定量を $h(\boldsymbol{Y})$ とすると,

$$\int_{-\infty}^{\infty} h(\boldsymbol{y})\, f(\boldsymbol{y})\, d\boldsymbol{y} = 0$$

が成り立つ. この両辺を σ^2 で偏微分すると,

$$\int_{-\infty}^{\infty} \frac{n}{2\,\sigma^4}\left\{-\sigma^2 + \frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2\right\} h(\boldsymbol{y})\, f(\boldsymbol{y})\, d\boldsymbol{y} = 0$$

となる. 上の 2 式より

$$\int_{-\infty}^{\infty} \sum_{i=1}^{n}(y_i-\mu)^2 h(\boldsymbol{y})\, f(\boldsymbol{y})\, d\boldsymbol{y} = 0$$

を得る. これは,

$$\sum_{i=1}^{n}(y_i-\mu)^2 \Big/ (n-1)$$

と $h(\boldsymbol{y})$ が無相関であることを示す. したがって, 定理 7.3 を用いて不偏標本分散 (7.4) が最良不偏推定量であることがわかる. ■

7.3 最尤推定法

尤度と最尤推定法

尤度を最大にする $\boldsymbol{\theta}$ を最尤推定量という．以下, Maximum Likelihood Estimator を略して, 最尤推定量を MLE と書く.

MLE は, 対数尤度をも最大にすることから, 通常

$$\frac{\partial}{\partial\boldsymbol{\theta}} l(\boldsymbol{\theta}) = \boldsymbol{o} \tag{7.24}$$

の解として求める．(7.24) を尤度方程式という．このような推定方法を最尤推定法という.

■例 7.4　A, B 両者があるゲームをする. 引き分けはなく, A が勝つ確率を $p\,(\neq 0, 1)$, B が勝つ確率を $1-p$ とする. いま, 10 ゲームを戦って, A の 7 勝 3 敗という結果が得られたとしよう. このとき, 対数尤度は,

$$l(p) = \mathrm{const} + 7\ \log p + 3\ \log(1-p)$$

であり, 尤度方程式

$$\frac{7}{p} - \frac{3}{1-p} = 0$$

より, $p = 7/10$ のときに対数尤度は最大となる.

■例 7.5　$Y_1, \cdots, Y_n$ が互いに独立に正規分布 $N(\mu, \sigma^2)$ に従うとする. このとき, 対数尤度は (7.13) である. μ, σ^2 で偏微分すると,

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu)\,, \quad \frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\,\sigma^2} + \frac{1}{2}\sum_{i=1}^{n}\frac{(y_i-\mu)^2}{\sigma^4}$$

となる. 対応する確率変数で置き換え, 0 とおくと, 次の最尤推定量を得る.

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \overline{Y}_n\,, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2 \tag{7.25}$$

$\hat{\sigma}^2$ は不偏推定量になっていないことに注意しよう.

最尤推定量の性質

最尤推定量の「良さ」は直観的に認めていた. しかし, 実際に, k 次元未知パラメタ $\boldsymbol{\theta}$ の最尤推定量 $\hat{\boldsymbol{\theta}}$ は,

$$\lim_{n\to\infty} \hat{\boldsymbol{\theta}} = \boldsymbol{\theta} \tag{7.26}$$

を満たし, Cramer-Rao の下限を達成し, さらに漸近的に正規分布となることが示される. (7.26) は一致性といわれる. これは, 小標本における不偏性に対応する概念である.

ここで, 最尤推定量の一致性を示そう. $Y_1, \cdots, Y_n$ が互いに独立に同一分布に従うとし, その密度関数を $f(y, \boldsymbol{\theta})$ とする. ここでは, f は同時密度でない. まず, 尤度方程式の左辺の $1/n$ 倍

$$\frac{1}{n} \cdot \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log f(y_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

は, n が十分大きくなると, 大数の法則により,

$$E_0\left[\frac{\partial \log f(y, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] \overset{\text{def}}{=} \varepsilon_0(\boldsymbol{\theta})$$

に確率収束する. ただし, 添え字の 0 は $\boldsymbol{\theta}$ が真値 $\boldsymbol{\theta}_0$ をとる場合, すなわち, $f(y, \boldsymbol{\theta}_0)$ に基づく期待値をとることを表す. ところが,

$$\int_{-\infty}^{\infty} f(y, \boldsymbol{\theta}_0)\, dy = 1$$

の両辺を $\boldsymbol{\theta}$ で偏微分すると,

$$\int_{-\infty}^{\infty} \frac{1}{f(y, \boldsymbol{\theta}_0)} \cdot \frac{\partial f(y, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\, f(y, \boldsymbol{\theta}_0)\, dy = \boldsymbol{o}$$

を得る. この値は ε_0 にほかならない.

これは, $\boldsymbol{\theta}_0$ が方程式 $\varepsilon_0(\boldsymbol{\theta}) = \boldsymbol{o}$ の解であることを示すから, 最尤推定値 $\hat{\boldsymbol{\theta}}$ は真値 $\boldsymbol{\theta}_0$ に確率収束する. これで一致性が示された.

つぎに, 尤度方程式 (7.24) に, Taylor の定理を用いて,

$$\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{\partial^2 l(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \boldsymbol{o} \tag{7.27}$$

を得る. ただし, $\boldsymbol{\theta}^* = (\theta_1^*, \cdots, \theta_k^*)'$ は

$$^{\forall} i, (\theta_i^* - \theta_i)(\theta_i^* - \hat{\theta}_i) < 0$$

を満たすある k 次元ベクトルであり, $\theta_i, \hat{\theta}_i$ はそれぞれ $\theta, \hat{\theta}$ の第 i 成分である. (7.27) の左辺の第 1 項

$$\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \frac{\partial \log f(Y_i, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$$

は, 真値 $\boldsymbol{\theta}_0$ で評価した Efficient Score である.

この統計量は, 中心極限定理により漸近的に k 変量正規分布に従う. また, (7.20) より, その期待値は $\boldsymbol{o}$ であり, 分散は $I(\boldsymbol{\theta}_0)$ である. 情報行列は同時密度関数から定義されるので, 分散は $n\,I(\boldsymbol{\theta}_0)$ にはならないことに注意しよう.

他方, (7.27) の左辺の第 2 項の $\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}$ の係数は, 大数の法則により,

$$E\left[\frac{\partial^2 l(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]=-I(\boldsymbol{\theta}_0)$$

に確率収束する.

以上と式 (3.56) より, $\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0$ は, 漸近的に k 変量正規分布に従い, その分散は, $I(\boldsymbol{\theta}_0)^{-1}$ である :

$$\hat{\boldsymbol{\theta}}\sim N_k(\boldsymbol{\theta}_0, I(\boldsymbol{\theta}_0)^{-1}) \tag{7.28}$$

ここでは, 各 Y_i $(i=1,\cdots,n)$ が独立で, 同一分布に従うことを仮定したが, より一般に, 観測 $\boldsymbol{y}=(y_1,\cdots,y_n)'$ が得られたときの尤度を $L(\boldsymbol{\theta})$ とするとき, (7.28) が適当な正則条件のもとで成立する.

7.4 区間推定

区間推定と信頼域

通常用いる推定量は, 今までみてきたように, 望ましい性質をもってはいるものの, その値が真の値に等しいかどうかは依然として明らかでない. このような推定方法を点推定といった.

これに対して, 確率変数 $\boldsymbol{Y}$ に, 未知パラメタ $\boldsymbol{\theta}$ のある範囲 $W(\boldsymbol{Y})$ を対応させ,

$$^\forall\boldsymbol{\theta},\ P\Big(\boldsymbol{\theta}\in W(\boldsymbol{Y})\Big)\geq 1-\alpha \tag{7.29}$$

となるようにする方法がある.

このような推定方法を区間推定という. $W(\boldsymbol{Y})$ を信頼係数 $1-\alpha$ の信頼域という. とくに, 信頼域が区間の形で与えられる場合は, 信頼区間といい, その端点を信頼限界という. 信頼係数 $1-\alpha$ の信頼区間を単に $1-\alpha$ 信頼区間ということがある. ある信頼域 $W_0(\boldsymbol{y})$ を得たとき, 真値 θ_0 が $W_0(\boldsymbol{y})$ に含まれる確率は 1 か 0 のいずれかであることに注意しよう.

区間推定と検定の関係

区間推定は, 検定と密接な関係がある. 実際, 未知パラメタ θ の $1-\alpha$ 信頼区間は, 帰無仮説 $\theta = \theta_0$ が有意水準 α で棄却されないような θ_0 の値の集合として得られる.

信頼区間は真でない値を含む確率をできるだけ小さくするように, 狭いことが望まれる. 信頼区間が狭いほど, 精度が高いといわれる. θ の真の値を θ_0 として, $\theta_1 \neq \theta_0$ に対して,

$$1 - P\Big(\theta_1 \in W(\boldsymbol{Y})\Big) \tag{7.30}$$

を検出力ということがある.

任意の (θ_0, θ_1) に対して, 検出力が常に最大となる信頼域を一様最強力信頼域というが, それは一般に存在しない. 検出力がつねに α 以上となる信頼域を不偏信頼域という. これは不偏検定から得られる.

■例 7.6　確率変数 $Y_1, \cdots, Y_n$ が互いに独立に $N(\mu, \sigma^2)$ に従うとき, σ^2 が未知であるとすると, 母平均 μ の $1-\alpha$ 信頼区間は, 定理 5.3 より,

$$\overline{Y}_n - \frac{t_{n-1}(\alpha/2)\,\hat{\sigma}_*}{\sqrt{n}} \leq \mu \leq \overline{Y}_n + \frac{t_{n-1}(\alpha/2)\,\hat{\sigma}_*}{\sqrt{n}} \tag{7.31}$$

で与えられる. ただし, $t_{n-1}(\alpha)$ は t_{n-1} の上側 α 点であり, $\hat{\sigma}_*$ は $\hat{\sigma}_*^2$ (7.4) の正の平方根である.

7.5　十分統計量

十分統計量の定義

確率変数 $Y_1, \cdots, Y_n$ が互いに独立に正規分布 $N(\mu, \sigma^2)$ に従うとする. このとき, 分布はパラメタ μ, σ^2 に依存するが, 標本平均 $\overline{Y}_n$ (5.18) および標本分散 S_n (5.19) の値がわかれば, 条件つき分布は一意に定まる.

実際, 同時密度関数は

$$\frac{1}{(\sqrt{2\pi}\,\sigma)^n}\ \exp\!\left[-\frac{1}{2\sigma^2}\{n(\overline{y}_n-\mu)^2+nS_n\}\right]$$

と書けるので, $\overline{Y}_n, S_n$ を与えたときの条件つき密度関数は 1 である.

話を一般化しよう.

■定義 7.4　同時密度関数 $f(\boldsymbol{y},\boldsymbol{\theta})$ をもつ確率変数 $\boldsymbol{Y}$ について, 統計量の組 $\boldsymbol{t}=(t_1,\cdots,t_m)'$ を与えたときの $\boldsymbol{Y}$ の条件つき分布がパラメタ $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_k)'$ に依らないとき, $\boldsymbol{t}$ を十分統計量という.

■定理 7.5　分布が連続型の場合, $\boldsymbol{t}(\boldsymbol{y})=\left(t_1(\boldsymbol{y}),\cdots,t_m(\boldsymbol{y})\right)'$ が十分統計量であるための必要十分条件は, 同時密度関数が

$$f(\boldsymbol{y},\boldsymbol{\theta})=g(\boldsymbol{y},\boldsymbol{t})\,h(\boldsymbol{t},\boldsymbol{\theta}) \tag{7.32}$$

と分解されることである. (分解定理)

■証明　必要性) $\boldsymbol{t}(\boldsymbol{y})$ が十分統計量であれば, 条件つき密度関数は $\boldsymbol{\theta}$ に依らないから,

$$f(\boldsymbol{y},\boldsymbol{\theta})\bigg/\int_{\boldsymbol{T}(\boldsymbol{y})=\boldsymbol{t}} f(\boldsymbol{y},\boldsymbol{\theta})\,d\boldsymbol{y}=g(\boldsymbol{y},\boldsymbol{t})$$

と表現できる. 左辺の分母は $\boldsymbol{T}$ の密度関数であり, $\boldsymbol{y}$ に依らないから, $h(\boldsymbol{t},\boldsymbol{\theta})$ と表すことができる. これから, (7.32) を得る.
十分性) 密度関数が (7.32) と表される場合は,

$$\int_{\boldsymbol{T}(\boldsymbol{y})=\boldsymbol{t}} f(\boldsymbol{y},\boldsymbol{\theta})\,d\boldsymbol{y}=G(\boldsymbol{y})\,h(\boldsymbol{t},\boldsymbol{\theta})$$

と表される. このとき, 条件つき密度関数は,

$$\frac{g(\boldsymbol{y},\boldsymbol{t})\,h(\boldsymbol{t},\boldsymbol{\theta})}{G(\boldsymbol{y})\,h(\boldsymbol{t},\boldsymbol{\theta})}=\frac{g(\boldsymbol{y},\boldsymbol{t})}{G(\boldsymbol{y})}$$

となり, $\boldsymbol{\theta}$ に依らない. ■

離散型への修正は容易である.

■定義 7.6　同時密度関数が

$$f(\boldsymbol{y},\boldsymbol{\theta}) = g(\boldsymbol{y})\,\phi(\boldsymbol{\theta})\,\exp\left\{-\sum_{i=1}^{m}\theta_i t_i(\boldsymbol{y})\right\} \tag{7.33}$$

と表されるとき, この分布を指数分布族 という.

確率変数 $Y_1,\cdots,Y_n$ が互いに独立に指数分布族に属し, その密度関数が (7.33) と表される場合は, 定理 7.5 より

$$T_i = \sum_{j=1}^{n} t_j(Y_j) \quad (i = 1,\cdots,m)$$

が十分統計量を与える.

十分統計量の性質

実は, 最良推定量は, 十分統計量のみから構成することができる. 次の定理はそれを保証する. 簡単のため, 未知パラメタが 1 次元の場合について述べておこう.

■定理 7.7　T が十分統計量であるとき, θ の任意の不偏推定量 $\hat{\theta}$ に対して, $T = t$ を与えたときの条件つき期待値 $\hat{\theta}^*$ も θ の不偏推定量であり, 次の不等式が成り立つ. (Rao-Blackwell)

$$V\left[\hat{\theta}^*\right] \le V\left[\hat{\theta}\right] \tag{7.34}$$

■証明　$\boldsymbol{\theta}$ と θ の違いを除き, 定理 7.5 と同一の notation を用いる. $\hat{\theta}^*$ は θ に依らないから,

$$\hat{\theta}^*(t) = \int_{\boldsymbol{T}(\boldsymbol{y})=\boldsymbol{t}} \hat{\theta}(\boldsymbol{y})\,g(\boldsymbol{y},\boldsymbol{t})\,d\boldsymbol{y}$$

と表すことができる.

T の分布による $\hat{\theta}^*$ の平均は,

$$\int \hat{\theta}^*(\boldsymbol{t})\, h(\boldsymbol{t}, \theta)\, d\boldsymbol{t} = \iint_{\boldsymbol{T}(\boldsymbol{y})=\boldsymbol{t}} \hat{\theta}(\boldsymbol{y})\, g(\boldsymbol{y}, \boldsymbol{t})\, h(\boldsymbol{t}, \theta)\, d\boldsymbol{y}\, d\boldsymbol{t}$$

であるが, これに定理 7.5 を用いると,

$$\int \hat{\theta}(\boldsymbol{y})\, f(\boldsymbol{y}, \theta)\, d\boldsymbol{y} = \theta$$

となる. これで, $\hat{\theta}^*$ の不偏性が証明された.

つぎに,

$$(\hat{\theta} - \theta)^2 = \{(\hat{\theta} - \hat{\theta}^*) + (\hat{\theta}^* - \theta)\}^2$$

の期待値を同様に求める.

$(\hat{\theta} - \hat{\theta}^*)\,(\hat{\theta}^* - \theta)$ の期待値が,

$$\int (\hat{\theta}^* - \theta)\, h(\boldsymbol{t} \,|\, \theta)\, d\boldsymbol{t} \cdot \int_{\boldsymbol{T}=\boldsymbol{t}} (\hat{\theta} - \hat{\theta}^*)\, g(\boldsymbol{y}, \boldsymbol{t})\, d\boldsymbol{y} = 0$$

となることに注意すると,

$$V\Big[\,\hat{\theta}\,\Big] = E\Big[\,(\hat{\theta} - \hat{\theta}^*)^2\,\Big] + E\Big[\,(\hat{\theta}^* - \theta)^2\,\Big]$$

を得る. 第 1 項は非負であるから, (7.34) が成り立つ. ■

第 8 章

ノンパラメトリック法

8.1 ノンパラメトリックな推論

実際の解析において, 分布型がわかっている場合はむしろ稀である. 分布型が分かっていない場合は, 標本を並び替えたり, 大きさの意味での順位に関する情報のみを使ったりして解析すればよい. このように, 特定の確率分布を仮定せずに行う統計学的推論をノンパラメトリックな推論という. とくに, 解析手法に注目した場合は, ノンパラメトリック法ということがある. ノンパラメトリックとは文字通りパラメタに依存しないという意味である.

本章では, 正規分布が仮定できない場合を考えよう. たとえば, n 個の標本のうち, かなりの割合で μ_0 より大きい観測値が得られた場合は, 帰無仮説より対立仮説を支持するのが直観的に考えても妥当であろう. 極端な例だが, 分布が中心 μ_0 に関して対称であれば, n 個すべての y_i が μ_0 より大きい確率は $1/2^n$ である.

8.2 並べ替え検定

上の直観的議論を定式化しよう. $z_i = y_i - \mu_0$ を与えたときに, 帰無仮説のもとで, 対応する確率変数 Z_i が等確率 $1/2$ で z_i または $-z_i$ をとると仮定する.

対立仮説のもとでは, 正の値をとる観測の個数が大きくなる傾向があるから, 棄却域を条件つき確率に基づいて,

$$\sum_{i=1}^{n} Z_i > \sum_{i=1}^{n} z_i \tag{8.1}$$

とすれば検定を構成できる. この不等式を満たす $(Z_1, \cdots, Z_n)$ の組が $N(n)$ 個であるとすると, 有意確率は $N(n)/2^n$ である.

上の検定は, これまでのように, データを得る前に検定方式を定めておき, それに基づいて仮説を棄却するか採択するかを決めるのではないことに注意しよう. このように, 得られた観測値を, ある条件のもとに仮想的に並べ替え, 有意確率を計算する検定方式を並べ替え検定という.

各 Y_i が正規分布に従う場合には, この検定の検出力は, 当然最強力検定である t 統計量に基づく検定に劣る. しかし, 並べ替え検定の検出力は相当によいことが知られている. 実際, t 検定と並べ替え検定は漸近的に等しい. これについては後に示そう.

正規近似

実際に $N(n)$ を数え上げることは, n がある程度大きくなってくると相当に面倒である. このような場合は正規近似が便利である. 統計量

$$Z = \sum_{i=1}^{n} Z_i \tag{8.2}$$

は, 標本平均の n 倍であり, 漸近的に正規分布に従う. この証明はここでは省略するが, この事実は, いわば中心極限定理の延長といえよう.

各 Z_i の分布は, z_i の値は異なるものの, 等確率 1/2 で正または負の符号をとる類似性をもつ. Z_i の平均と分散が

$$E[Z_i] = 0\,, \quad V[Z_i] = E\left[\,Z_i^2\,\right] = z_i^2 \tag{8.3}$$

となることはすぐにわかる.

したがって, 漸近的に

$$Z\Big/\Big(\sum_i z_i^2\Big)^{1/2} \sim N(0,1) \tag{8.4}$$

となる. そこで, 棄却域を

$$Z - \frac{1}{2} > K_\alpha \Big(\sum_{i=1}^n z_i^2\Big)^{1/2} \tag{8.5}$$

とすればよい. ここに, K_α は標準正規分布 $N(0,1)$ の上側 α 点である.

なお, 左辺から 1/2 を減じているのは, 近似の精度を上げるためで, 連続補正といわれる.

t 検定との漸近同等性

実は並べ替え検定と t 検定は漸近的に同等である. 以下にこのことを示そう. 確率変数 $Y_1, \cdots, Y_n$ の分布については, μ_0 に関して対称であることだけを仮定する. このとき, T 統計量 (6.23) は

$$\sqrt{n-1}\,Z\Big/\Big(\sum Z_i^2 - Z^2/n\Big)^{1/2} \tag{8.6}$$

と表現できる.

$$\sum Z_i^2 = \sum z_i^2 \tag{8.7}$$

が定数であることに注意すると, (8.6) は Z の増加関数であり, 結局 t 検定と並べ替え検定は漸近的に同等であることがわかる.

このことは, t 検定は漸近的にはノンパラメトリック検定であり, 正規分布からのずれには頑健であることを意味する. しかし, これは検定の妥当性の問題であり, 必ずしも検出力を保証するものではない.

8.3　順序統計量

$Y_1, \cdots, Y_n$ を互いに独立に同一分布に従う確率変数とする. これらを大きさの順に小さい方から並べたときに, $Y_{(1)}, \cdots, Y_{(n)}$ となるとしよう.

すなわち, i 番目に小さい確率変数を $Y_{(i)}$ で表す. これを i 番目の順序統計量という. この notation は本書を通じて共通とする.

$Y_1, \cdots, Y_n$ が連続分布に従う場合は, 分布関数の連続性から, 同順位 (タイ) の存在はあり得ない. しかし, 実際には, 分布の離散性, 測定の精度や観測値の丸めなどによりさらに同順位が生じる. このような場合は, たとえば一様分布 $U(0,1)$ に従う $U_1, \cdots, U_n$ を乱数を用いて抽出し, (Y_i, U_i) の組を考える方法が提案されている. 同順位 Y_i, Y_j に対しては, U_i, U_j の大きさに応じて確率的に大小を決定しようというわけである. しかし, これでは研究対象とはまったく無関係なランダム性に解析結果が依存することになる. これよりは, 平均順位を与える方法の方が一般的である.

中央値, 四分位点, パーセント点

観測値を大きさの順に並べたとき, 中央に位置する値を中央値または median という. より厳密には, $Y_{(i)}$ の実測値を $y_{(i)}$ として,

$$\begin{cases} y_{(n/2+1/2)} & \cdots \quad n: \text{奇数} \\ (y_{(n/2)} + y_{(n/2+1)})/2 & \cdots \quad n: \text{偶数} \end{cases} \tag{8.8}$$

で定義する. 中央値は, 分布の中心の指標であるが, 平均と比較して外れ値の影響を受けにくいという長所がある. すなわち, 外れ値に対して頑健である.

より一般に, 100α パーセント点または α 点を,

$$\{1 - (n\alpha - [n\alpha])\}\, y_{([n\alpha])} + (n\alpha - [n\alpha])\, y_{([n\alpha]+1)} \tag{8.9}$$

で定義する. ここに, $[y]$ は y を超えない最大の整数である. α パーセント点は, データを大きさの順に小さい方から並べたときに, 順位がサンプルサイズの α 倍となる値 (に線型補完を加えた値) にほかならない.

とくに, $\alpha = 1/4, 1/2, 3/4$ の場合を, それぞれ第一四分位数, 第二四分位数, 第三四分位数という. 第二四分位数は中央値と一致することに注意しよう. 第三四分位数と第一四分位数の差を四分偏差または四分位差という. これは分布の散らばりの指標であり, 分散と比べて外れ値に対して頑健である.

順位相関係数

確率変数 X, Y の観測値を各々 x_i, y_j $(i, j = 1, \cdots, n)$ とするとき,

$$\rho(x, y) = \frac{\sum_i (r_i - \overline{r})(s_i - \overline{s})}{\left\{\sum_i (r_i - \overline{r})^2 \sum_j (s_j - \overline{s})^2\right\}^{1/2}} \tag{8.10}$$

を Spearman の標本順位相関係数という. ただし, r_i, s_j はそれぞれ大きさの順に並べたときの x_i, y_j の順位である. 同順位があるときは, その平均順位を用いる. また, $\overline{r}, \overline{q}$ はそれぞれ, r_i, s_j の総平均順位

$$\overline{r} = \frac{1}{n}\sum_{i=1}^{n} r_i\,, \quad \overline{s} = \frac{1}{n}\sum_{j=1}^{n} s_j \tag{8.11}$$

である.

(8.10) は, 順位データの相関係数であるから, $-1 \leq \rho(x, y) \leq 1$ を満たす. 値が 1 となるのは, (x_i, y_i) の順位が完全に一致するときであり, そのときに限る. また, -1 になるのは, 順位が完全に逆転するとき, すなわち $r_i = s_{n+1-i}$ となるときであり, そのときに限る.

なお, (8.10) において, r_i, s_j を, それぞれに対応する確率変数で置き換えれば, 母集団における Spearman の順位相関係数が得られる.

つぎに,

$$\tau(x, y) = \frac{2}{n\,(n-1)}\sum_{i<j} \mathrm{sgn}(x_i - x_j)(y_i - y_j) \tag{8.12}$$

を Kendall の (標本) 順位相関係数はという. ここに,

$$\mathrm{sgn}\ x = \begin{cases} 1 & \cdots \quad x > 0 \\ 0 & \cdots \quad x = 0 \\ -1 & \cdots \quad x < 0 \end{cases} \tag{8.13}$$

である. 係数 $2/n(n-1)$ は同順位のある場合多少の修正を要する.

(8.12) は, 他の相関係数同様, 次を満たす.

$$-1 \leq \tau(x, y) \leq 1 \tag{8.14}$$

このことは,

$$\sum_{i<j} \mathrm{sgn}(x_i - x_j)(y_i - y_j) \tag{8.15}$$

が, (x_i, y_i) の順位が完全に一致するとき最大値 $n(n-1)/2$ を, 順位が完全に逆転する $(r_i = s_{n+1-i})$ とき, 最小値 $-n(n-1)/2$ をとることからわかる.

なお, 母集団における Kendall の順位相関係数は, x_i, y_j に対応する確率変数をそれぞれ X_i, Y_j として,

$$P\Big((X_i - X_j)(Y_i - Y_j) > 0\Big) - P\Big((X_i - X_j)(Y_i - Y_j) < 0\Big) \tag{8.16}$$

で定義する. この値は (i, j) に依らないことに注意しよう.

順位相関係数は, Pearson の相関係数 (3.26) と異なり, 外れ値の影響を受けにくい.

順序統計量の分布

確率変数 Y_i $(i = 1, \cdots, n)$ の従う同一分布の分布関数を $F(y)$, 密度関数を $f(y)$ として, $Y_{(i)}$ の密度関数を求めておこう. ここでは, 同順位の存在は考えない. $Y_{(i)}$ の分布関数は, $Y_1, \cdots, Y_n$ のうち, 少なくとも i 個が y 以下の値をとる確率にほかならないから,

$$F_{(i)}(y) = \sum_{j=i}^{n} {}_n\mathrm{C}_j \{F(y)\}^j \{1 - F(y)\}^{n-j} \tag{8.17}$$

である. この両辺を y で微分すると, $Y_{(i)}$ の密度関数はつぎのようになる.

$$\begin{aligned}
f_{(i)}(y) = \sum_{j=i}^{n} {}_n\mathrm{C}_j\, f(y) \Big[\, & j\,\{F(y)\}^{j-1}\,\{1 - F(y)\}^{n-j} \\
& -(n-j)\,\{F(y)\}^{j}\,\{1 - F(y)\}^{n-j-1} \Big] \\
= \sum_{j=i}^{n} & n\, f(y)\, {}_{n-1}\mathrm{C}_{j-1}\, \{F(y)\}^{j-1}\,\{1 - F(y)\}^{n-j} \\
& - \sum_{j=i}^{n-1} n\, f(y)\, {}_{n-1}\mathrm{C}_{n-j-1}\, \{F(y)\}^{j}\,\{1 - F(y)\}^{n-j-1}.
\end{aligned}$$

第一の和において, $j-1$ を改めて j とおくと, 次を得る.

$$f_{(i)}(y) = n\, f(y)\, {}_{n-1}\mathrm{C}_{i-1}\, \{F(y)\}^{i-1}\, \{1-F(y)\}^{n-i} \tag{8.18}$$

分布が離散型の場合は, 密度関数 $f(y)$ を確率関数を $p(y)$ で置き換えれば, $Y_{(i)}$ の確率関数が得られる.

ところで, $Y_{(i-1)} \le y-h < Y_{(i)} \le y+h < Y_{(i+1)}$ である確率は,

$$\frac{n\,!}{(i-1)\,!\,(n-i)\,!}\, \{F(y-h)\}^{i-1} \cdot 2\,h\, f(y) \cdot \{1-F(y)\}^{n-i} \tag{8.19}$$

であるから, $2\,h$ で除し $h \to 0$ とすると (8.18) を得る. これは直感的にも納得がいくだろう.

つぎに, $Y_{(i)}, Y_{(i')}\ (i < i')$ の同時分布を考えよう. まず, 条件 $Y_{(i')} \le y_2$ のもとでの $Y_{(i)}$ の条件つき密度関数は, (8.18) において, n のかわりに $i'-1$ とおき, さらに $f(y), F(y)$ をそれぞれ条件つき分布の密度関数, 分布関数で置き換えて,

$$(i'-1)\, {}_{i'-2}\mathrm{C}_{i-1}\, \frac{f(y_1)}{F(y_2)} \left\{\frac{F(y_1)}{F(y_2)}\right\}^{i-1} \left\{1 - \frac{F(y_1)}{F(y_2)}\right\}^{i'-i-1} \tag{8.20}$$

となる. これに $Y_{(i')}$ の密度関数

$$n\, f(y_2)\, {}_{n-1}\mathrm{C}_{i'-1}\, \{F(y_2)\}^{i'-1}\, \{1-F(y_2)\}^{n-i'} \tag{8.21}$$

を乗じて, $Y_{(i)}, Y_{(i')}$ の同時密度関数

$$\begin{aligned} f_{(i,i')}(y_1, y_2) = & \frac{n\,!\, f(y_1)\, f(y_2)}{(i-1)\,!\,(i'-i-1)\,!\,(n-i')\,!}\, \{F(y_1)\}^{i-1} \\ & \cdot \{F(y_2) - F(y_1)\}^{i'-i-1}\, \{1-F(y_2)\}^{n-i'} \end{aligned} \tag{8.22}$$

を得る. これも直観的には, 5 項分布から理解できよう.

同様に, $Y_{(i_1)}, \cdots, Y_{(i_m)}\ (\, i_1 < \cdots < i_m\,)$ の同時密度関数は,

$$\prod_{j=1}^{m} \frac{n\,!\, f(y_j)\, \{F(y_j) - F(y_{j-1})\}^{i_j - i_{j-1} - 1}\, \{1-F(y_m)\}^{n-i_m}}{(i_j - i_{j-1} - 1)\,!\,(n-i_m)\,!} \tag{8.23}$$

となる. 以下, これを

$$f_{(i_1,\cdots,i_m)}(y_1, \cdots, y_m) \tag{8.24}$$

と書くことにする. ここに, $i_0 = 0,\ y_0 = -\infty$ である.

8.4　順位検定

順位検定

順位検定は, 標本のもつ情報のうち, それらのもつ順位に関するもののみを用いて構成する検定である. 標本 $y_1, \cdots, y_n$ を大きさの順に小さい方から並べたときの y_i の順位を r_i とするとき, y_i, r_i に対応する確率変数をそれぞれ Y_i, R_i とすると, R_i の分布は Y_i の分布に関係なく一定であり, このことから検定を構成できる.

一般に, n 個の自然数 $1, \cdots, n$ を並べ換え, $r_1, \cdots, r_n$ となるとき, この写像を置換といい, τ_n で表す. τ_n により得られる $\boldsymbol{r} = (r_1, \cdots, r_n)'$ のすべてから成る集合を $\mathcal{T}_n$ とする.

いま, n 次元確率変数 $\boldsymbol{R} = (R_1, \cdots, R_n)'$ を $\mathcal{T}_n$ の $n\,!$ 個の要素のいずれかを等確率 $1/n\,!$ でとると仮定する :

$$P\Big(\boldsymbol{R} = \boldsymbol{r}\Big) = \frac{1}{n\,!} \quad \text{for } {}^{\forall}\boldsymbol{r} \in \mathcal{T}_n \tag{8.25}$$

この分布を $\mathcal{T}_n$ 上の一様分布という. (8.25) は,

$$P\Big(R_i = j\Big) = \frac{1}{n}, \quad P\Big(R_i = j, R_{i'} = j'\Big) = \frac{1 - \delta_{jj'}}{n\,(n-1)} \tag{8.26}$$

が任意の i, i', j, j' $(i \neq i')$ について成立することと同値である. $\delta_{jj'}$ は Kronecker のデルタ (4.49) である.

線型順位統計量

順位 R_i に対して, 任意の単調非減少関数 $s(R_i)$ をスコアという. また, c_i を定数として, スコアの線型和

$$S = \sum_{i=1}^{n} c_i\, s(R_i) = \sum_{i=1}^{n} s(i)\, c_{(i)} \tag{8.27}$$

を線型順位統計量という.

■定理 8.1 $\boldsymbol{R}$ が $\mathcal{T}_n$ 上で一様分布するとき, S (8.27) に関して,

$$E[S] = n\,\overline{c}\,\overline{s} \tag{8.28}$$

$$V[S] = \frac{1}{n-1}\sum_{i=1}^{n}\{c_i - \overline{c}\}^2 \sum_{i=1}^{n}\{s(i) - \overline{s}\}^2 \tag{8.29}$$

が成り立つ. ただし, $\overline{c}, \overline{s}$ は次で定義する.

$$\overline{c} = \frac{1}{n}\sum_{i=1}^{n} c_i\,, \quad \overline{s} = \frac{1}{n}\sum_{i=1}^{n} s(i) \tag{8.30}$$

■証明 まず, (8.26) より, S の平均は

$$E\Big[\,s(R_i)\Big] = \sum_{j=1}^{n} s(j)\,P\Big(R_i = j\Big) = \frac{1}{n}\sum_{j=1}^{n} s(j) = \overline{s}$$

であるから,

$$E[S] = \sum_{i=1}^{n} c_i\,E\Big[\,s(R_i)\Big] = \overline{s}\sum_{i=1}^{n} c_i = n\,\overline{c}\,\overline{s}$$

を得る. 同様に, 分散は, 次で与えられる.

$$V\Big[\,s(R_i)\Big] = \frac{1}{n}\sum_{i=1}^{n}\{s(i) - \overline{s}\}^2 \tag{8.31}$$

$i \neq i'$ のときは,

$$\sum_{j\neq j'}\{s(j) - \overline{s}\}\{s(j') - \overline{s}\} = \Big[\sum_{j=1}^{n}\{s(j) - \overline{s}\}\Big]^2 - \sum_{j=1}^{n}\{s(j) - \overline{s}\}^2$$

の右辺の第 1 項は 0 であり, 再び (8.26) を用いて,

$$\mathrm{Cov}\,\Big[\,s(R_i), s(R_{i'})\Big] = -\frac{1}{n\,(n-1)}\sum_{i=1}^{n}\{s(i) - \overline{s}\}^2 \tag{8.32}$$

を得る. 他方, (3.56) より,

$$V[S] = \sum_{i=1}^{n} c_i^2\,V\Big[\,s(R_i)\Big] + \sum_{i\neq i'} c_i\,c_{i'}\,\mathrm{Cov}\,\Big[\,s(R_i), s(R_i')\Big]$$

を得る.

(8.31) (8.32) を代入して,

$$V[S] = \frac{1}{n(n-1)} \sum_{i=1}^{n} \{s(i) - \overline{s}\}^2 \left\{ (n-1) \sum_{i=1}^{n} c_i^2 - \sum_{i \neq i'} c_i c_{i'} \right\}$$

を得る. これに,

$$\sum_{i \neq i'} c_i c_{i'} = \sum_{i=1}^{n} c_i \sum_{i'=1}^{n} c_{i'} - \sum_{i=1}^{n} c_i^2$$

を用いると, 結論が得られる. ■

■系 8.2　定理 8.1 において, とくに,

$$S = \sum_{i=1}^{m} s(R_i) \quad (m < n) \tag{8.33}$$

とおくと,

$$E[S] = m\overline{s}, \quad V[S] = \frac{m(n-m)}{n(n-1)} \sum_{i=1}^{n} \{s(i) - \overline{s}\}^2 \tag{8.34}$$

が成り立つ.

■証明　定理 8.1 において,

$$c_1 = \cdots = c_m = 1, \ c_{m+1} = \cdots = c_n = 0$$

とおけばよい. ■

統計量 S は適当な条件のもとで, 漸近的には正規分布に従う. 適当な条件とは, しばしば

$$\max_i \{c_i - \overline{c}\}^2 \Big/ \sum_i \{c_i - \overline{c}\}^2 \tag{8.35}$$

が十分小さいと表現される. 証明は複雑なので省略する.

対称な分布の中心に関する仮説の検定

いま, 対称な連続分布の中心 μ に関する検定 (6.17) を考えよう. 帰無仮説のもとでは, $Y_1, \cdots, Y_n$ が互いに独立に μ_0 中心に対称に分布すると仮定する.

このとき, 順位を

$$Z_i = Y_i - \mu_0 \tag{8.36}$$

の絶対値に基づいて定め, (8.27) において,

$$c_{(i)} = \begin{cases} 1 & \cdots \quad Z_{(i)} > 0 \\ 0 & \cdots \quad Z_{(i)} \leq 0 \end{cases} \tag{8.37}$$

とおく.

この確率変数は i に依らず, 1/2 の確率で 1 または 0 の 2 値をとるから,

$$E\left[\, c_{(i)} \,\right] = E\left[\, c_{(i)}^2 \,\right] = \frac{1}{2} \tag{8.38}$$

であり, これから

$$E[S\,] = \frac{1}{2}\sum_{j=1}^{n} s(j)\,, \quad V[S\,] = \frac{1}{4}\sum_{j=1}^{n} s(j)^2 \tag{8.39}$$

を得る.

線型順序統計量の漸近正規性については既に論じたが, いまの場合, 統計量 S は同一分布に従う確率変数の線型和であるから, 中心極限定理によって漸近的に正規分布することが直ちにわかる.

対立仮説のもとでは Z_i は正の値をとる傾向があると予想されるから, 棄却域は

$$\frac{S - E[S\,] - 1/2}{\sqrt{V[S\,]}} > K_\alpha \tag{8.40}$$

とすればよい. K_α は標準正規分布 $N(0,1)$ の上側 α 点であり, 分子から 1/2 を減じているのは連続補正である.

左片側検定, 両側検定を構成するのは容易である.

符号検定

とくに, スコアとして, つねに $s(i) = 1$ とした場合を符号検定という. このとき, S は正値をとる観測数にほかならない. (8.39) より直ちに次を得る.

$$E[S\,] = n/2\,, \quad V[S\,] = n/4 \tag{8.41}$$

Wilcoxon の符号順位検定

また, $s(i) = i$ とした場合を Wilcoxon の符号順位検定という. (8.39) より平均, 分散は次のようになる.

$$E[S] = \frac{1}{4}n(n+1), \quad V[S] = \frac{1}{24}n(n+1)(2n+1) \tag{8.42}$$

ところで, $Z_{(i)}$ の定義から, $(i) \leq (j)$ なる任意の i に対して,

$$Z_{(j)} \leq 0 \Longrightarrow Z_{(i)} + Z_{(j)} \leq 0, \quad Z_{(j)} > 0 \Longrightarrow Z_{(i)} + Z_{(j)} > 0$$

が成り立つ. したがって, Wilcoxon の符号順位検定の統計量は,

$$S = \sum_{Z_{(j)}>0} j = \sum_{Z_{(i)}+Z_{(j)}>0} \sum_{i=1}^{j} 1 \tag{8.43}$$

と表現することができる. 指示関数

$$u(t) = \begin{cases} 1 & \cdots \quad t > 0 \\ 0 & \cdots \quad t \leq 0 \end{cases} \tag{8.44}$$

を用いると, 次の表現を得る.

$$S = \sum_{j=1}^{n}\sum_{i=1}^{j} u\Big(Z_{(i)}+Z_{(j)}\Big) = \sum_{j=1}^{n}\sum_{i=1}^{j} u\Big(Z_i+Z_j\Big) \tag{8.45}$$

(8.45) を Mann-Whitney 型の表現という.

同順位が存在する場合

同順位が存在する場合, i ($i = 1, \cdots, k$) 番目の同順位の個数を m_i として, 次を仮定しても一般性を失わない.

$$\begin{aligned} Z_{(1)} = \cdots = Z_{(m_1)} < Z_{(m_1+1)} = \cdots = Z_{(m_1+m_2)} \\ < \cdots < Z_{(m_1+\cdots+m_{k-1}+1)} = \cdots = Z_{(n)} \end{aligned}$$

いま, i 番目のタイにおける m_i 個の観測値に, その中間順位

$$\begin{cases} (m_1+1)/2 & \cdots \quad i=1 \\ m_1+\cdots+m_{i-1}+(m_i+1)/2 & \cdots \quad i \geq 2 \end{cases} \tag{8.46}$$

に基づくスコアを与える. このとき, 統計量の平均は変わらないが, 分散は当然小さくなる. 実際, i 番目のタイにおいて, 順位の二乗和を (8.46) の二乗で置き換えると, その減少分は, Wilcoxon の符号順位検定の場合,

$$(1^2+\cdots+m_i^2)-\frac{1}{4}\,m_i\,(m_i+1)^2=\frac{1}{12}\,m_i\,(m_i^2-1)$$

である. したがって, 同順位補正後の分散は, (8.42) より,

$$V[S]=\frac{1}{24}\left\{n\,(n+1)\,(2n+1)-2\sum_{i=1}^{k}m_i\,(m_i^2-1)\right\} \tag{8.47}$$

となる. 一般の場合の分散の補正も同様に導かれる.

第 9 章

モデルに基づく推論 (1)

9.1 線型モデル

たとえば, 農作物の収穫を大きくすることを目的に, 肥料の種類をいろいろと変えて各条件における収穫量の平均を比較することを考える. 肥料の種類を A_i $(i = 1, \cdots, k)$ として, 条件 A_i のもとで n_i 区画の収穫量を比較する実験計画を立てたとしよう. この計画に対して, しばしば仮定されるモデルは,

$$Y_{ij} = \theta_i + \varepsilon_{ij} \tag{9.1}$$

である. Y_{ij} は A_i の第 j 区画における収穫量を表す確率変数, θ_i は A_i における収穫量の平均, ε_{ij} は誤差を表す確率変数である. θ_i を A_i の主効果という. ε_{ij} は互いに独立に, 正規分布 $N(0, \sigma^2)$ に従うと仮定する.

$$\hat{\theta}_{ij} = \sum_{j=1}^{n_i} \frac{Y_{ij}}{n_i}$$

が θ_i の最良推定量を与えることは既に述べた.

Y_{ij}, ε_{ij} を辞書順に並べた $n_1 + \cdots + n_k = n$ 次元ベクトルをそれぞれ $\boldsymbol{Y}, \boldsymbol{\varepsilon}$ とし, θ_i を辞書順に並べた k 次元ベクトルを $\boldsymbol{\theta}$ とおくと, (9.1) は

$$\boldsymbol{Y} = X\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{9.2}$$

と行列表現できる.

ここに,

$$X = \begin{bmatrix} 1 & \cdots & 1 & & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \ddots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & & 1 & \cdots & 1 \end{bmatrix}'$$

である. これをデザイン行列という. 一般に, (9.2) で表されるモデルを線型モデルという.

9.2　最小二乗法

正規方程式

$\boldsymbol{\theta}$ を推定するには, 標本 $\boldsymbol{y}$ を得たとして, 線型モデル (9.2) からの乖離

$$S_e = \sum_i \sum_j \varepsilon_{ij}^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\boldsymbol{y} - X\boldsymbol{\theta})'(\boldsymbol{y} - X\boldsymbol{\theta}) \tag{9.3}$$

を最小になるようにする. (9.3) を残差平方和という. 残差とは誤差の実現値である. 具体的には,

$$\frac{\partial S_e}{\partial \boldsymbol{\theta}} = -2X'\boldsymbol{y} + 2X'X\boldsymbol{\theta} = \boldsymbol{o} \Longleftrightarrow X'X\boldsymbol{\theta} = X'\boldsymbol{y} \tag{9.4}$$

とすればよい. (9.4) を正規方程式といい, (9.4) を満たす任意の推定量を最小二乗推定量という. $\boldsymbol{\theta}$ の不偏推定量を, $\boldsymbol{y}$ の各成分の線型和 $\boldsymbol{l}'\boldsymbol{y}$ に限れば, 後に示すように最小二乗推定量は最良不偏推定量である. これを最良線型不偏推定量という. 以下 Best Linear Unbiased Estimator を略して BLUE と書く.

推定可能関数

$X'X$ が正則であれば, (9.4) は一意に解けて,

$$\hat{\boldsymbol{\theta}} = (X'X)^{-1}X'\boldsymbol{y} \tag{9.5}$$

となる. しかし, 一般には正規方程式は一意に解けるとは限らず,

$${}^{\forall}\boldsymbol{\theta},\ \| \boldsymbol{y} - X\hat{\boldsymbol{\theta}} \|^2 \le \| \boldsymbol{y} - X\boldsymbol{\theta} \|^2 \tag{9.6}$$

を満たす $\boldsymbol{l}'\hat{\boldsymbol{\theta}}$ を考える.

このうち, $\boldsymbol{Y}$ に関して線型な,

$$\boldsymbol{l}'\hat{\boldsymbol{\theta}} = \boldsymbol{\lambda}'\boldsymbol{Y} \tag{9.7}$$

と表現できるものを推定可能関数という. もちろん, 任意の $\boldsymbol{l}'$ に対して (9.7) の形で表せるとは限らない.

■定理 9.1　$\boldsymbol{l}'\boldsymbol{\theta}$ が推定可能である必要十分条件は, $\boldsymbol{l}'$ がデザイン行列 X の行ベクトルの線型結合として表されることである.

■証明　(9.7) の両辺の期待値を考えると, $\boldsymbol{l}'\boldsymbol{\theta} = \boldsymbol{\lambda}'X\boldsymbol{\theta}$ となる. これが任意の $\boldsymbol{\theta}$ に関して成り立つことから, $\boldsymbol{l}' = \boldsymbol{\lambda}'X$ を得る.

■定理 9.2　モデル (9.2) において, 正規方程式 (9.4) を満たす任意の $\hat{\boldsymbol{\theta}}$ が $\boldsymbol{\theta}$ の最小二乗推定量である. また, $\boldsymbol{l}'\boldsymbol{\theta}$ を任意の推定可能関数とすると, $\boldsymbol{l}'\hat{\boldsymbol{\theta}}$ は一意に定まり, $\boldsymbol{l}'\boldsymbol{\theta}$ の BLUE を与える. (Gauss-Markov)

■証明　まず,

$$\boldsymbol{y} - X\boldsymbol{\theta} = (\boldsymbol{y} - X\hat{\boldsymbol{\theta}}) + (X\hat{\boldsymbol{\theta}} - X\boldsymbol{\theta}) \tag{9.8}$$

に注意すると, 残差平方和 (9.3) は,

$$\begin{aligned} &\| \boldsymbol{y} - X\hat{\boldsymbol{\theta}} \|^2 + \| X\hat{\boldsymbol{\theta}} - X\boldsymbol{\theta} \|^2 \\ &\qquad + 2\,(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'(X'\boldsymbol{y} - X'X\hat{\boldsymbol{\theta}}) \end{aligned} \tag{9.9}$$

と表すことができる. 第 2 項は非負, 第 3 項は (9.4) により 0 であるから, これは定理の前半を導く.

つぎに, 定理の後半を示すが, まず $X'X$ が正則の場合を考える. はじめに,

$$\boldsymbol{l}'\hat{\boldsymbol{\theta}} = \boldsymbol{l}'(X'X)^{-1}X'\boldsymbol{Y} \tag{9.10}$$

と, $\boldsymbol{l}'\boldsymbol{\theta}$ の任意の不偏な推定可能関数 $\boldsymbol{\lambda}'\boldsymbol{Y}$ の差を次のようにおく.

$$\boldsymbol{c}'\boldsymbol{Y} = \boldsymbol{\lambda}'\boldsymbol{Y} - \boldsymbol{l}'(X'X)^{-1}X'\boldsymbol{Y} \tag{9.11}$$

両辺の期待値を考えると,

$$\boldsymbol{c}'X\boldsymbol{\theta} = \boldsymbol{l}'\boldsymbol{\theta} - \boldsymbol{l}'(X'X)^{-1}X'X\boldsymbol{\theta} = \boldsymbol{o} \tag{9.12}$$

が任意の $\boldsymbol{\theta}$ について成立するから,

$$\boldsymbol{c}'X = \boldsymbol{o}' \tag{9.13}$$

を得る.

したがって,

$$\mathrm{Cov}\left[\boldsymbol{l}'(X'X)^{-1}X'\boldsymbol{Y},\ \boldsymbol{c}'\boldsymbol{Y}\right] = \sigma^2\,\boldsymbol{l}'(X'X)^{-1}X'\boldsymbol{c} = 0$$

であり,

$$V\left[\boldsymbol{\lambda}'\boldsymbol{Y}\right] = V\left[\boldsymbol{l}'\hat{\boldsymbol{\theta}} - \boldsymbol{c}'\boldsymbol{Y}\right] = V\left[\boldsymbol{l}'\hat{\boldsymbol{\theta}}\right] + V\left[\boldsymbol{c}'\boldsymbol{Y}\right] \tag{9.14}$$

と書くことができる. 右辺の第 2 項は非負であるから, これは結論を導く. ■

$X'X$ が正則とは限らない一般の場合は, 0 でない固有値を,

$$\lambda_1 \geq \cdots \geq \lambda_{k^*} > 0\,, \quad k^* = \mathrm{rank}\,(X'X)$$

とおき, λ_i に属する固有ベクトルを第 i 列に配した k 行 k^* 列行列を P とする. さらに , $\Lambda = \mathrm{diag}\,(\lambda_i)$ とおくと,

$$X'XP = P\Lambda \tag{9.15}$$

が成り立つ.

ここで,

$$\boldsymbol{z} = \begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \end{bmatrix} = \Gamma'\boldsymbol{y}\,, \quad \Gamma' = \begin{bmatrix} \Lambda^{-1/2}P'X' \\ Q' \end{bmatrix} \tag{9.16}$$

により, 標本 $\boldsymbol{y}$ を直交変換する. Q' は

$$X'Q = O\,, \quad Q'Q = I_{n-k^*} \tag{9.17}$$

なる $n - k^*$ 行 n 列行列である. Γ' のブロック分割に応じて $\boldsymbol{z}$ も分割した.

$\boldsymbol{Z}_2$ の平均は,

$$E[\boldsymbol{Z}_2] = Q'E[\boldsymbol{Y}] = Q'X\boldsymbol{\theta} = \boldsymbol{o} \tag{9.18}$$

であるから,

$$E[\boldsymbol{Z}_1] = \Lambda^{-1/2}P'X'X\boldsymbol{\theta} = \Lambda^{1/2}P'\boldsymbol{\theta} = \boldsymbol{\eta} \tag{9.19}$$

とおくと, (9.2) は

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 \\ \boldsymbol{Z}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{o} \end{bmatrix} + \boldsymbol{\xi} \quad (\boldsymbol{\xi} = \Gamma\boldsymbol{\varepsilon}) \tag{9.20}$$

となる. Γ が直交行列であることから, 誤差 $\boldsymbol{\xi}$ は,

$$E[\boldsymbol{\xi}] = \boldsymbol{o}, \quad V[\boldsymbol{\xi}] = \sigma^2 I_n \tag{9.21}$$

を満たす.

$\boldsymbol{l}'_1, \boldsymbol{l}'_2$ をそれぞれ任意の k^* 次元, $k - k^*$ 次元ベクトルとして, 推定可能関数 $\boldsymbol{l}'_1\boldsymbol{\eta}$ は一般に, $\boldsymbol{l}'_1\boldsymbol{Z}_1 + \boldsymbol{l}'_2\boldsymbol{Z}_2$ と表される. Γ は直交行列であるから, $\boldsymbol{Z}_1$ と $\boldsymbol{Z}_2$ は独立で,

$$V[\boldsymbol{l}'_1\boldsymbol{Z}_1 + \boldsymbol{l}'_2\boldsymbol{Z}_2] = \sigma^2(\boldsymbol{l}'_1\boldsymbol{l}_1 + \boldsymbol{l}'_2\boldsymbol{l}_2) \tag{9.22}$$

である. $\boldsymbol{l}'_1$ を固定した際, この分散は $\boldsymbol{l}'_2 = \boldsymbol{o}$ のとき最小値 $\sigma^2\boldsymbol{l}'_1\boldsymbol{l}_1$ をとる.

したがって,

$$\boldsymbol{l}'_1\boldsymbol{\eta} = \boldsymbol{l}'_1\Lambda^{1/2}P'\boldsymbol{\theta} \tag{9.23}$$

の BLUE は

$$\boldsymbol{l}'_1\boldsymbol{Z}_1 = \boldsymbol{l}'_1\Lambda^{-1/2}P'X'\boldsymbol{Y} \tag{9.24}$$

である. ところが, 正規方程式 (9.4) と (9.15) より

$$P'X'\boldsymbol{Y} = P'X'X\hat{\boldsymbol{\theta}} = \Lambda P'\hat{\boldsymbol{\theta}} \tag{9.25}$$

である. これを (9.24) に代入すると次を得る.

$$\boldsymbol{l}'_1\boldsymbol{Z}_1 = \boldsymbol{l}'_1\Lambda^{1/2}P'\hat{\boldsymbol{\theta}} \tag{9.26}$$

これは (9.23) に最小二乗推定量を代入したものである. ■

最小二乗法の幾何学的意味

正規方程式 (9.4) を $X'(\boldsymbol{y} - X\boldsymbol{\theta}) = \boldsymbol{o}$ と変形すると, $\boldsymbol{\theta}$ は残差 $\boldsymbol{y} - X\boldsymbol{\theta}$ が X の各列ベクトルすべてと直交するように定められることがわかる. 実際, $X'X$ が正則なときは,

$$\Pi = X(X'X)^{-1}X' \tag{9.27}$$

とおくと,

$$\Pi^2 = \Pi, \quad \Pi' = \Pi \tag{9.28}$$

が成り立つから, 最小二乗法により得られた

$$X\hat{\boldsymbol{\theta}} = \Pi\boldsymbol{y} \tag{9.29}$$

は, $\boldsymbol{y}$ を X の列ベクトルの張る線型空間に射影したベクトルである. また, (9.28) より, $\Pi\boldsymbol{y}$ と $(I - \Pi)\boldsymbol{y}$ は独立であるから, 最小二乗推定は, 観測値 $\boldsymbol{y}$ をデザイン行列 X の列ベクトルの張る線型空間の要素 $\Pi\boldsymbol{y}$ と, その直交補空間の要素 $(I - \Pi)\boldsymbol{y}$ に分解することにほかならない.

9.3 モデルの検証

以上で説明してきたことは, すべて線型モデル (9.2) が正しいという仮定のもとに成り立っている. モデルの正しさは別に検証しなければならない.

決定係数

観測値 $\boldsymbol{y}$ と予測値 $\Pi\boldsymbol{y}$ の各成分の相関係数の二乗

$$R^2 = \frac{(\boldsymbol{y}'\Pi\boldsymbol{y})^2}{\|\boldsymbol{y}\|^2\|\Pi\boldsymbol{y}\|^2} = \frac{\|\Pi\boldsymbol{y}\|^2}{\|\boldsymbol{y}\|^2} \tag{9.30}$$

はモデルの正当性の指標と考えられる. これを決定係数という.

線型モデルの正当性の検定

つぎに, モデルの正当性の検定を構成しよう. $\boldsymbol{Y}$ は多変量正規分布に従うのであったから, 誤差 $(I-\Pi)\boldsymbol{Y}$ も多変量正規分布に従い,

$$E\Big[(I-\Pi)\boldsymbol{Y}\Big]=\boldsymbol{o},\quad V\Big[(I-\Pi)\boldsymbol{Y}\Big]=\sigma^2(I-\Pi) \tag{9.31}$$

が成り立つ. したがって, 定理 5.2 により,

$$S_e=\|(I-\Pi)\boldsymbol{Y}\|^2\sim\sigma^2\chi^2_{n-k} \tag{9.32}$$

を得る. これは残差平方和 (9.3) に対応し, その自由度は

$$n-k=\mathrm{rank}\,(I-\Pi)=\mathrm{trace}\,(I-\Pi X) \tag{9.33}$$

となる.

これを以下に示そう. まず, Π の定義から $\mathrm{rank}\,\Pi\le\mathrm{rank}\,X$ であるが, $\Pi X=X$ より $\mathrm{rank}\,\Pi\ge\mathrm{rank}\,X$ でもあるので,

$$\mathrm{rank}\,\Pi=\mathrm{rank}\,X=\mathrm{rank}\,(X'X)=k$$

が成り立つ. また, (9.28) より Π の固有値は 0 か 1 であるので,

$$\mathrm{trace}\,\Pi=\mathrm{rank}\,X=k$$

である. まったく同様にして, (9.33) を得る. ■

(9.32) と (5.13) より,

$$\hat{\sigma}^2=S_e\Big/(n-k) \tag{9.34}$$

は σ^2 の不偏推定量である. これが最良不偏推定量であることは,

$$\int\frac{h(\boldsymbol{y})}{(\sqrt{2\pi}\,\sigma)^n}\,\exp\Big(-\frac{S_e}{2\sigma^2}\Big)d\boldsymbol{y}=0$$

の両辺を σ^2 で偏微分することにより, Rao の定理を用いて得られる. ここに, $h(\boldsymbol{y})$ は 0 の任意の推定量である.

また, (9.32) と同様にして,

$$\| X(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \|^2 \sim \sigma^2 \chi_k \tag{9.35}$$

を得る. $\Pi \boldsymbol{y} = X\hat{\boldsymbol{\theta}}$ と $(I - \Pi)\,\boldsymbol{y}$ は独立であったから, (9.32) と (9.35) の統計量も独立で,

$$\frac{\| X(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \|^2}{k\,\hat{\sigma}^2} \sim F^k_{n-k} \tag{9.36}$$

を用いてモデルの正当性の検定を構成できる.

デザイン行列 X が正則でない場合には $(X'X)^{-1}$ を, 一般化逆行列 $(X'X)^-$ で置き換えればよい. このとき, k もデザイン行列のランク k^* で置き換える.

モデルの選択

モデルがより低次元のパラメタにより説明できる場合もありうる. もしそうなら, パラメタの数はできるだけ小さくした方が解析結果が安定したものになる. そこで, (9.2) を,

$$\boldsymbol{Y} = [\,X_1 \quad X_2\,] \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix} + \boldsymbol{\varepsilon} \tag{9.37}$$

と表現する. $\boldsymbol{\theta}_1$ は κ 次元列ベクトル, $\boldsymbol{\theta}_2$ は $k - \kappa$ 次元列ベクトルであり, これに応じて X もブロック分割した. このとき, 次の検定方式を考える.

帰無仮説 $H_0 : \boldsymbol{\theta}_2 = \boldsymbol{o}$
対立仮説 $H_1 : \boldsymbol{\theta}_2 \neq \boldsymbol{o}$

この検定は通常の検定方式と異なり, 帰無仮説を支持する方向に興味があるので, 有意水準は高めに設定 (0.2 程度) しておく必要がある.

H_0 に対する誤差の二乗和と H_1 に対する誤差の二乗和の差は

$$G = \boldsymbol{Y}'(I - \Pi_1)\boldsymbol{Y} - \boldsymbol{Y}'(I - \Pi)\boldsymbol{Y} = \boldsymbol{Y}'(\Pi - \Pi_1)\,\boldsymbol{Y} \tag{9.38}$$

である. ここに,

$$\Pi_1 = X_1(X_1X_1)^- X_1' \tag{9.39}$$

である.

明らかに, X の張る空間 $T(X)$ と, X_1 の張る空間 $T(X_1)$ の間には,

$$T(X_1) \subset T(X), \quad T(X_1)^{\perp} \supset T(X)^{\perp}$$

の関係があるので,

$$\mathit{\Pi}\,\mathit{\Pi}_1 = \mathit{\Pi}_1\,\mathit{\Pi} = \mathit{\Pi}_1 \tag{9.40}$$

が成り立つ.

(9.32) や (9.35) とまったく同様にして,

$$G \sim \sigma^2 \chi^2_{k^*-\kappa^*} \tag{9.41}$$

が得られる. κ^* は X_1 のランクである. また, (9.40) より,

$$(\mathit{\Pi} - \mathit{\Pi}_1)(I - \mathit{\Pi}) = O \tag{9.42}$$

が成り立つので, G と $\hat{\sigma}^2$ は独立であり,

$$\frac{G}{(k^* - \kappa^*)\,\hat{\sigma}^2} \sim F^{k^*-\kappa^*}_{n-k^*} \tag{9.43}$$

が得られる. とくに, $\kappa^* = k^* - 1$ とおけば,

$$\frac{G}{\hat{\sigma}^2} \sim F^{1}_{n-k^*} \tag{9.44}$$

を得る. これより, パラメタの次元を小さくしていくことができる. 有意確率が予め設定しておいた水準より大きければ, このパラメタの要素をモデルから除外し, 除外するパラメタがなくなった時点で検定を終了すればよい.

9.4　回帰係数の区間推定

今度は, $\hat{\boldsymbol{\theta}}$ の区間推定を考える. まずその準備として, A が非負定値, B が正定値行列のとき,

$$\max_{\boldsymbol{x} \neq \boldsymbol{o}} \left(\frac{\boldsymbol{x}' A\, \boldsymbol{x}}{\boldsymbol{x}' B\, \boldsymbol{x}} \right) = \lambda_1 \tag{9.45}$$

が成り立つことを示す. ただし, λ_1 は AB^{-1} の最大固有値である.

対称行列

$$C = B^{-1/2} A B^{-1/2} \tag{9.46}$$

は適当な直交行列 Γ によって

$$\Gamma' C \Gamma = \mathrm{diag}\,(\lambda_i) \tag{9.47}$$

と対角化できる. ここで,

$$B^{1/2}\,\boldsymbol{x} = \boldsymbol{y}\,, \quad \Gamma' \boldsymbol{y} = \boldsymbol{z} \tag{9.48}$$

とおくと,

$$\frac{\boldsymbol{x}' A\,\boldsymbol{x}}{\boldsymbol{x}' B\,\boldsymbol{x}} = \frac{\boldsymbol{y}' B^{-1/2} A B^{-1/2}\,\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}} = \frac{\boldsymbol{z}' \Gamma' C \Gamma\,\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{z}} \leq \lambda_1 \tag{9.49}$$

となり, (9.45) を得た. とくに, A がベクトル $\boldsymbol{a}$ を用いて $\boldsymbol{a}\,\boldsymbol{a}'$ と表されるときは,

$$\lambda_1 = \boldsymbol{a}' B^{-1} \boldsymbol{a} \tag{9.50}$$

であるから, 任意の k 次元定ベクトル $\boldsymbol{l}'$ に対して,

$$\max_{\boldsymbol{l} \neq \boldsymbol{o}} \left\{ \frac{\boldsymbol{l}'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\,(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\boldsymbol{l}}{\boldsymbol{l}'(X'X)^{-1}\,\boldsymbol{l}} \right\} = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' X' X\,(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \tag{9.51}$$

を得る. これが (9.35) の統計量に等しいことに注意しよう.

これと (9.36) から, 任意の $\boldsymbol{l}' \neq \boldsymbol{o}$ に対し,

$$\left| \boldsymbol{l}'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \right| \leq \left\{ k\,\hat{\sigma}^2\,\boldsymbol{l}'(X'X)^{-1}\,\boldsymbol{l}\;f^k_{n-k}(\alpha) \right\}^{1/2} \tag{9.52}$$

が $\boldsymbol{l}'\boldsymbol{\theta}$ の $1-\alpha$ 信頼区間を与える. ここに, $f^k_{n-k}(\alpha)$ は F^k_{n-k} の上側 α 点である.

第 10 章

モデルに基づく推論 (2)

10.1 ロジスティックモデル

たとえば, 二項分布モデルで, 応答確率 p を要因 y の関数として,

$$p = \theta_0 + \theta_1 y \tag{10.1}$$

という線型モデルを考えると, 右辺の値が負になったり 1 より大きくなったりして都合が悪い. このときは p のかわりに,

$$\text{logit } p = \log \frac{p}{1-p} = \theta_0 + \theta_1 y \tag{10.2}$$

とモデルを修正すればそのような矛盾は生じない. logit p は全実数値をとるからである.

(10.2) を logit モデルという. logit 関数の逆関数 (4.35) を logistic 関数という. (10.2) は

$$p = \frac{\exp(\theta_0 + \theta_1 y)}{1 + \exp(\theta_0 + \theta_1 y)} \tag{10.3}$$

と同値であり, logistic モデルとも呼ばれる.

一般に, 事象 A の起こる確率 $P(A)$ に対して, $P(A) / \{1 - P(A)\}$ を, 事象 A のオッズ比 または単にオッズというので, logit はしばしば対数オッズ (比) といわれる.

logit モデルを拡張して, 応答確率の logit を要因を表す変数 $y_1, \cdots, y_k$ の線型和として,

$$\text{logit } p = \theta_0 + \theta_1 y_1 + \cdots + \theta_k y_k \tag{10.4}$$

と表現するモデルを一般化 logit モデルという. $y_1, \cdots, y_k$ を共変量といい, 未知パラメタ $\theta_0, \theta_1, \cdots, \theta_k$ を回帰係数という.

一般化線型モデル

logit モデルは, 確率 p に対して, 全実数値をとりうる線型予測子

$$\theta_0 + \theta_1 y_1 + \cdots + \theta_k y_k \tag{10.5}$$

を対応させる logit 関数の性質に基づくものであった.

このような関数はひとつではない. たとえば, 標準正規分布の分布関数

$$\Phi(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \tag{10.6}$$

を用いて,

$$\Phi^{-1}(p) = \theta_0 + \theta_1 y_1 + \cdots + \theta_k y_k \tag{10.7}$$

とするモデルを probit モデルといい,

$$\log\left\{-\log(1-p)\right\} = \theta_0 + \theta_1 y_1 + \cdots + \theta_k y_k \tag{10.8}$$

とするモデルを二重対数モデルという. (10.4) (10.7) (10.8) をまとめて一般化線型モデルという.

反応が多値の場合

logit モデルは反応が R_i $(i = 0, \cdots, a)$ と多値の場合, R_i の得られる確率を p_i とすると,

$$\log \frac{p_i}{p_0} = \theta_{i0} + \theta_{i1} y_1 + \cdots + \theta_{ik} y_k \quad (i = 1, \cdots, a) \tag{10.9}$$

とすればよい. R_0 は反応なしを表す.

$R_1 < \cdots < R_a$ のように反応に順序のついている場合は, 一種の累積確率を考え, モデルをつぎのようにすればよい.

$$\text{logit } P_i = \theta_{i0} + \theta_1 y_1 + \cdots + \theta_k y_k\,, \quad P_i = \sum_{l=i}^{a} p_l \tag{10.10}$$

各反応において, 切片以外の回帰係数は共通である. このとき,

$$\frac{P_i}{1-P_i} = \frac{\exp\theta_{i0}}{\exp\theta_{j0}} \cdot \frac{P_j}{1-P_j} \quad \text{for } {}^{\forall}i,\ {}^{\forall}j \tag{10.11}$$

が成立するので, (10.10) を比例オッズモデルということがある.

10.2　回帰係数の推定

(10.9) において, 反応確率はパラメタの一次式ではなく, 非線型モデルといわれる. 非線型モデルにおいては推定には最尤推定法を用いる. モデルを (10.9) とするとき, 尤度は

$$\prod_i \prod_j \frac{\exp(\boldsymbol{\theta}'_{i_j}\boldsymbol{y}_j)}{1+\exp(\boldsymbol{\theta}'_{i_j}\boldsymbol{y}_j)} \tag{10.12}$$

となる. i は j に依存するので, i_j と記述している. ここに,

$$\boldsymbol{\theta}_i = (\theta_{i0}, \theta_{i1}, \cdots, \theta_{ik})' \quad (i = 1, \cdots, a) \tag{10.13}$$

であり, $\boldsymbol{y}_j$ は, 第 j 標本における

$$(1, y_1, \cdots, y_k)' \tag{10.14}$$

の値である. しかし, 尤度方程式を陽的に解くことは通常困難であり, Newton-Raphson 法という反復計算により近似的に解く.

区間推定

すべての未知パラメタを, 辞書順に縦に並べて得られるベクトルを $\boldsymbol{\theta}$ とおく :

$$\boldsymbol{\theta} = (\theta_{10}, \cdots, \theta_{1k}, \theta_{20}, \cdots, \theta_{ak})' \tag{10.15}$$

$\boldsymbol{\theta}$ の近似的な分散は, (7.28) よりつぎで与えられる.

$$V\left[\hat{\boldsymbol{\theta}}\right] \simeq I\left(\hat{\boldsymbol{\theta}}\right)^{-1} \tag{10.16}$$

$\boldsymbol{\theta}$ の第 l 成分 θ_l の分散 $V[\theta_l]$ は (10.16) の第 l 対角成分として求めることができる.

結局, θ_l の $1-\alpha$ 信頼区間は

$$\hat{\theta}_l - K_{\alpha/2}\, V[\hat{\theta}_l]^{1/2} \leq \theta_l \leq \hat{\theta}_l + K_{\alpha/2}\, V[\hat{\theta}_l]^{1/2} \tag{10.17}$$

で与えられる. ここに , $K_{\alpha/2}$ は標準正規分布 $N(0,1)$ の上側 $\alpha/2$ 点である.

10.3　モデルの検証

モデルに基づく解析は, 結果が大きくモデルに依存するので, その正当性を別に検証しなければならないことは線型モデルの章で既に述べた.

一般に, モデルの検証は,

- モデル全体の正当性
- より良いモデルの検索

の両者を含む. 後者は, パラメタの次元を小さくすることにより解析結果を安定させるために重要である. このことを以下, 変数選択と表現する.

上のふたつの検証には, $\boldsymbol{\theta}$ の次元を k として, 次の検定方式を考えればよい.

$$\text{帰無仮説}: \theta_1 = \cdots = \theta_\kappa = 0 \quad (\kappa \leq k)$$
$$\text{対立仮説}: 1 \leq {}^{\exists}i \leq \kappa,\ \ \theta_i \neq 0$$

最尤推定法に基づく推定であれば, 次の 3 検定が構成できる.

Wald 検定

Wald 検定は,

$$W = \hat{\boldsymbol{\theta}}' V\left[\hat{\boldsymbol{\theta}}\right]^{-1} \hat{\boldsymbol{\theta}} \sim \chi^2_k \tag{10.18}$$

を用いる検定である. これは (7.28) から導かれる.

$\hat{\boldsymbol{\theta}}$ の特定の成分に注目した場合は,

$$\frac{\hat{\theta}_l^2}{V\left[\hat{\theta}_l\right]} \sim \chi_1^2 \quad \text{or} \quad \frac{\hat{\theta}_l}{V\left[\hat{\theta}_l\right]^{1/2}} \sim N(0,1) \tag{10.19}$$

を利用して, 該当する共変量の影響を検定することができる. これから θ_l の信頼区間が得られたことに注意しよう.

$\kappa\,(<k)$ 次元のパラメタを $\boldsymbol{\theta}^*$ とし, これに対応する Wald 統計量を W^* とすると, (10.18) より

$$W - W^* \sim \chi^2_{k-\kappa} \tag{10.20}$$

が成立する. これを用いてモデルの比較ができる. とくに, $\kappa = k-1$ とおくと, パラメタの次元をひとつずつ小さくしていくことができる.

スコア検定

スコア検定は, Efficient Score (7.15) に基づく検定である. この統計量の漸近正規性は既に述べた. これと (7.20) から, 漸近的に

$$S = \boldsymbol{S}(\boldsymbol{\theta})' I\left(\hat{\boldsymbol{\theta}}\right)^{-1} \boldsymbol{S}(\boldsymbol{\theta}) \sim \chi_k^2 \tag{10.21}$$

が成り立つ. これがスコア検定である.

$\boldsymbol{\theta}^*$ に基づくスコア検定量を S^* とおくと,

$$S - S^* \sim \chi_{k-\kappa} \tag{10.22}$$

が成り立つので, これからモデルの変数選択のための検定を構成できる.

尤度比検定

尤度比検定は,

$$-2\left\{\log L\left(\hat{\boldsymbol{\theta}}^*\right) - L\left(\hat{\boldsymbol{\theta}}\right)\right\} \sim \chi^2_{k-\kappa} \tag{10.23}$$

を利用する検定である.

まず, (10.23) を導こう. 対数尤度 $l(\boldsymbol{\theta})$ は Taylor の定理により,

$$l(\boldsymbol{\theta}) = l(\hat{\boldsymbol{\theta}}) + \frac{1}{2}\,(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \frac{\partial^2 l(\hat{\boldsymbol{\theta}}^{\sharp})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}\,(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \tag{10.24}$$

と表すことができる. この式は最尤推定値 $\hat{\boldsymbol{\theta}}$ まわりの展開なので $\boldsymbol{\theta}$ の一次の項は 0 になることに注意しよう. ここに,

$$\boldsymbol{\theta}^{\sharp} = (\theta_1^{\sharp}, \cdots, \theta_k^{\sharp})'$$

は, 任意の $\theta_l^{\sharp}$ に対して,

$$(\theta_l^{\sharp} - \theta_l)\,(\theta_l^{\sharp} - \hat{\theta}_l) < 0$$

となる k 次元ベクトルである. また, 最尤推定値量の一致性から, $\boldsymbol{\theta}$ の真値を $\boldsymbol{\theta}_0$ とすると,

$$\frac{\partial^2 l(\hat{\boldsymbol{\theta}}^{\sharp})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'} \simeq -I(\boldsymbol{\theta}_0) \tag{10.25}$$

となる.

したがって, (10.24) より

$$-2\Big\{l(\boldsymbol{\theta}_0) - l(\hat{\boldsymbol{\theta}})\Big\} \simeq (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\, I(\boldsymbol{\theta}_0)\,(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \tag{10.26}$$

が成立する. これと (7.28) より,

$$-2\Big\{l(\boldsymbol{\theta}_0) - l(\hat{\boldsymbol{\theta}})\Big\} \sim \chi_k^2 \tag{10.27}$$

を得る. これから直ちに (10.23) を得る.

変数選択

(10.23) を利用すると変数選択を行うことができる. ここでは, 尤度比検定に基づく方法を述べるが, Wald 統計量やスコア統計量に基づく方法もまったく同様に構成することができる.

変数減少法は, 最初に共変量すべてを含むモデルを想定する. パラメタの次元が i のときの最大対数尤度を l_i とおくと,

$$2\left\{l_i - l_{i-1}\right\} \sim \chi_1^2 \tag{10.28}$$

が成り立つ. 有意確率の大きい順に変数を除き, 取り除く変数がなくなったら操作を打ち切ればよい.

逆に, 定数項のみのモデルから出発して, 対数尤度を有意に増加させる変数を順次取り込んでいく方法もある. これを変数増加法という. また, このふたつの方法を交互に行い, 取り込む変数も除く変数もなくなったら打ち切るという方法もある. これを変数増減法という. 一度取り込んだ変数が, モデルで組み合わせる共変量によっては, 次のステップで取り除かれたり, その逆が起こったりすることがありうる.

どの方法を用いるかは好みがあるだろうが, 特別の理由がない限り, 筆者は変数減少法を用いることにしている. 解析モデルはデータ自身により決定されるものではないと考えるからである.

しかし, 選択方法により, 最終モデルが異なる可能性は小さくない. すべての方法で変数選択を行い, 赤池の情報基準

$$\mathrm{AIC} = -2\,l_i + 2\,i \tag{10.29}$$

を最小にするモデルを選ぶということも考えられる. これは, 対数尤度に, 変数を多くした場合のペナルティを考慮した基準である.

10.4　関連モデル

臨床試験や疫学調査のように, 反応確率を等しく評価できない場合は, 多少の修正を要する.

対数線型モデル

第 l 層の総観測数を n_l として, 反応の頻度 y_l に Poisson 分布を仮定する. 強度 λ_l を, モデル

$$\log \lambda_l = \log n_l + \theta_0 + \theta_1 y_1 + \cdots + \theta_k y_k \tag{10.30}$$

を用いて推定することを考える. これを対数線型モデルという.

反応が発病や死亡といった稀な事象の場合は, Poisson 分布 $\mathrm{Po}(\lambda_l)$ は二項分布 $B(n_l, p_l)$ で近似できる. ただし, $n_l\, p_l = \lambda_l$ である. これを (10.30) に代入すると,

$$\mathrm{logit}\ p_l \simeq \log p_l = \theta_0 + \theta_1 y_1 + \cdots + \theta_k y_k \tag{10.31}$$

となり, 近似的にはロジスティックモデルと同等になる.

比例ハザードモデル

生存時間解析のようにフォロー期間が一致しなかったり, 打ち切りがあったりする場合は Cox の比例ハザードモデルを用いる. このモデルは, 時間 t までに反応が得られていないという条件のもとで, t から $t + dt$ までに反応の得られる条件つき確率を $h(t, \boldsymbol{y})\, dt$ で表し, 比例ハザード性

$$h(t, \boldsymbol{y}) = h(t, \boldsymbol{o})\ \exp(\boldsymbol{\theta}' \boldsymbol{y}) \tag{10.32}$$

を仮定する. $\boldsymbol{y}$ は共変量, $\boldsymbol{\theta}$ は回帰係数である. $h(t, \boldsymbol{y})$ をハザード関数という. とくに, $h(t, \boldsymbol{o})$ を基準ハザード関数という.

時間を表す非負の確率変数を T とするとき,

$$S(t, \boldsymbol{y}) = P(T \geq t, \boldsymbol{y}) \tag{10.33}$$

を生存関数という. また,

$$H(t, \boldsymbol{y}) = \int_0^t h(\tau, \boldsymbol{y})\, d\tau = -\log S(t, \boldsymbol{y}) \tag{10.34}$$

を累積ハザード関数という.

T が離散値 $t_1 < \cdots < t_m$ をとる場合は, ハザード関数を

$$h(t, \boldsymbol{y}) = \sum_l h_l(\boldsymbol{y})\,\delta(t - t_l, \boldsymbol{y}) \tag{10.35}$$

で定義すればよい. ここに,

$$h_l = P\Big(T = t_l \,\Big|\, T \geq t_l\Big) = \frac{P(T = t_l)}{S(t_l)} \tag{10.36}$$

である. $\delta(t)$ は Dirac の δ 関数, すなわち次の関数である.

1) $\delta(t) = 0$ for ${}^{\forall}t \neq 0$
2) $\delta(0) = \infty$
3) $\displaystyle\int_{-\infty}^{\infty} \delta(t)\,dt = 1$

以下, $\boldsymbol{y}$ を省略する.

Kaplan-Meier **推定量**

h_l が小さいときには, $\exp(-h_l) \simeq 1 - h_l$ より,

$$S(t) = \exp\{-H(t)\} = \prod_{t_l < t}(1 - h_l) \tag{10.37}$$

と書ける. Y_l を時点 t_l におけるイベント数, n_l を同じ時点におけるリスク集合 R_l の要素の数とする. リスク集合とは, その直前までイベントの危険に曝されている集合である.

(10.37) において, h_l をその最尤推定量 Y_l/n_l で置き換えると,

$$\hat{S}(t) = \prod_{t_l \leq t}\left(1 - \frac{Y_l}{n_l}\right) \tag{10.38}$$

となる. これを Kaplan-Meier 推定量という. $\hat{S}(t_l)$ は, $S(t_l)$ の最尤不偏推定量である. 他方, 条件つき確率の定義から,

$$S(t_l) = \prod_{i=1}^{l}(1 - h_i) \tag{10.39}$$

が成り立つ.

Y_i が二項分布 $B(n_i, h_i)$ に従うと仮定する[*1]と,

$$E\left[\hat{S}(t_l)^2\right] = \prod_{i=1}^{l}\left\{\frac{h_i(1-h_i)}{n_i} + (1-h_i)^2\right\} \tag{10.40}$$

を得る. これを,

$$\prod_{i=1}^{l}(1-h_i)^2\left\{1+\sum_{i=1}^{l}\frac{h_i}{n_i(1-h_i)}\right\} = S(t_l)^2\left\{1+\sum_{i=1}^{l}\frac{h_i}{n_i(1-h_i)}\right\}$$

で近似すると, 不偏性も考えて, 次を得る. (Greenwood)

$$V\left[\hat{S}(t_l)^2\right] \simeq S(t_l)^2\sum_{i=1}^{l}\frac{h_i}{n_i(1-h_i)} \tag{10.41}$$

部分尤度とパラメタの推定

比例ハザードモデルにおけるパラメタ $\boldsymbol{\theta}$ は最尤推定法を用いて推定できる. まず, 部分尤度関数を定義しておこう. 時点 t_l , j 番目の観測において, $\boldsymbol{y}$ のとる値を $\boldsymbol{y}_j$ とし,

$$\boldsymbol{s}_l = \sum_{j\in D_l}\boldsymbol{y}_j \tag{10.42}$$

とおく. D_l は時点 t_l におけるリスク集合である. モデルのもとでの尤度関数は, t_l がいつであるかという情報を別にすれば,

$$L(\boldsymbol{\theta}) = \prod_{l=1}^{m}\frac{\exp(\boldsymbol{\theta}'\boldsymbol{s}_l)}{\left\{\sum_{j\in R_l}\exp(\boldsymbol{\theta}'\boldsymbol{y}_j)\right\}^{y_l}} \tag{10.43}$$

となる. 結局, 尤度方程式 (7.24) の適切な解 $\hat{\boldsymbol{\theta}}$ を Newton-Raphson 法で近似的に求めればよい. $\hat{\boldsymbol{\theta}}$ の分散共分散行列も Fisher の情報行列から得られる.

[*1] 実際には n_i は確率変数である.

ロジスティックモデルとの漸近同等性

時間 $0 \le t \le t_0$ までの基準ハザードがほぼ一定で小さい場合, これを h_0 とおくと, 時間 t_0 までに反応の得られない確率は,

$$\begin{aligned} 1 - S(t_0, \boldsymbol{y}) &= 1 - \exp\{-h_0 t_0 \exp(\boldsymbol{\theta}' \boldsymbol{y})\} \\ &\simeq h_0 t_0 \exp(\boldsymbol{\theta}' \boldsymbol{y}) \end{aligned}$$

となる. 両辺の対数をとると,

$$\log\{1 - S(t_0, \boldsymbol{y})\} \simeq \log h_0 t_0 + \boldsymbol{\theta}' \boldsymbol{y}$$

となり, 比例ハザードモデルも漸近的にはロジスティックモデルと同等であることがわかる.

後ろ向き研究の場合は, 反応確率を近似的にも求めることができない. たとえば, 疾患のある群 (case) とない群 (control) で, それぞれ過去にさかのぼって曝露の有無を調べ, 曝露と疾患発生の因果関係を調べるとしよう. この場合, 母集団における case の割合を, 曝露を表す共変量 $\boldsymbol{y}$ の関数として $p(\boldsymbol{y})$ と表すと, 観測される case の割合は次のようになる.

$$p^*(\boldsymbol{y}) = \frac{q_1 p(\boldsymbol{y})}{q_1 p(\boldsymbol{y}) + q_2\{1 - p(\boldsymbol{y})\}} \tag{10.44}$$

q_1, q_2 はそれぞれ, 母集団の case, control から抽出した標本の割合であり, 一般には未知である. しかし, logit に関しては,

$$\log \frac{p^*(\boldsymbol{y})}{1 - p^*(\boldsymbol{y})} = \log \frac{q_1}{q_2} + \log \frac{p(\boldsymbol{y})}{1 - p(\boldsymbol{y})} \tag{10.45}$$

が成立する. このことから, 後ろ向き研究に logit モデルを適用しても, 定数項以外は通常の場合とまったく同じ解釈ができる.

第 III 部

実験計画とデータ解析

第 11 章

二標本問題

一般に, k 個の異なる母集団からの無作為標本から, 母集団の性質の違いを考察する問題を k 標本問題という. 2 標本問題は, 3 個以上の母集団の違いを考察する際の雛型であるというだけでなく, 差を考えることにより 1 標本問題の基本ともいえる概念である.

11.1　特性値が正規分布に従う場合

t 検定

確率変数 Y_{ij} $(i = 1, 2 : \ j = 1, 2, \cdots, n_i)$ が互いに独立に正規分布 $N(\mu_i, \sigma^2)$ に従うとする. 2 個の母集団の分散が等しいことに注意しよう. このとき, 帰無仮説

$$H_0 : \mu_1 = \mu_2$$

を, 対立仮説

$$\begin{aligned} &H_1 : \mu_1 \neq \mu_2 \quad (\text{ 両側 }) \text{ または} \\ &H_2 : \mu_1 > \mu_2 \quad (\text{右片側}) \text{ または} \\ &H_2' : \mu_1 < \mu_2 \quad (\text{左片側}) \end{aligned}$$

に対して, 検定することを考える.

いま,

$$\boldsymbol{Y}_i = (Y_{i1}, \cdots, Y_{in_i})' \quad (i = 1, 2)$$

とおくと, このモデルは, 線型モデル

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{j}_{n_1} & \boldsymbol{o} \\ \boldsymbol{o} & \boldsymbol{j}_{n_2} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \boldsymbol{\varepsilon} \tag{11.1}$$

にほかならない. したがって, 標本平均

$$\overline{Y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \tag{11.2}$$

は μ_i の最小二乗推定量であり,

$$\hat{\sigma}^2 = \frac{1}{n_1 + n_2 - 2} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i.})^2 \tag{11.3}$$

は母分散 σ^2 の最良不偏推定量である.

定理 5.3 と t 分布の定義により, 次がわかる.

$$T = \frac{\overline{Y}_{1.} - \overline{Y}_{2.} - (\mu_1 - \mu_2)}{\sqrt{(1/n_1 + 1/n_2)\,\hat{\sigma}^2}} \sim t_{n_1+n_2-2} \tag{11.4}$$

以上から, 次のような棄却域を得る.

$$\begin{array}{lll} H_1 \text{ に対し } W_1: & |T| > & t_{n_1+n_2-2}(\alpha/2) \\ H_2 \text{ に対し } W_2: & T > & t_{n_1+n_2-2}(\alpha) \\ H_2' \text{ に対し } W_2': & T < & -t_{n_1+n_2-2}(\alpha) \end{array}$$

ここに, $t_{n_1+n_2-2}(\alpha)$ は $t_{n_1+n_2-2}$ の上側 α 点である.

平均の差に関する $1-\alpha$ 信頼限界は, 両側検定の場合,

$$\overline{Y}_{1.} - \overline{Y}_{2.} \pm t_{n_1+n_2-2}(\alpha/2) \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{\sigma}^2} \tag{11.5}$$

で与えられる.

$\mu_1 \neq \mu_2$ のときには, 統計量 T は非心度 $(\mu_1 - \mu_2)/\sigma$ の非心 t 分布に従う. したがって, 上の検定の検出力は非心 t 分布を用いて計算できる. これから, 片側検定は一様最強力相似, 両側検定は一様最強力不偏検定であることがわかる.

サンプルサイズの決定

2 標本の平均の差が $\mu_1 - \mu_2 = \delta\sigma > 0$ のときに, 有意水準を α に保ったまま, 2 つの母集団の平均の差を確率 $1 - \beta$ 以上で検出したいとする. このときに必要な標本の大きさを考えよう. 非常に荒い議論であるが,

$$n_1 = n_2 = n/2 \tag{11.6}$$

として, Y_{ij} が互いに独立に正規分布 $N(\mu_i, \hat{\sigma}^2)$ に従うとみなす.

右片側検定であれば, 有意水準が α であるから,

$$\frac{\overline{Y}_{1.} - \overline{Y}_{2.}}{\sqrt{4\,\hat{\sigma}^2/n}} \geq K_\alpha$$

が成り立つ. 他方, 検出力の条件は,

$$\frac{\overline{Y}_{1.} - \overline{Y}_{2.} - \delta\,\hat{\sigma}}{\sqrt{4\,\hat{\sigma}^2/n}} \leq K_{1-\beta} = -K_\beta$$

である. この 2 式から,

$$K_\alpha - \frac{\delta}{\sqrt{4/n}} \leq -K_\beta \Longleftrightarrow n \geq \frac{4(K_\alpha + K_\beta)^2}{\delta^2} \tag{11.7}$$

を得る. この n が要求される標本の大きさである.

両側検定の場合は, α のかわりに $\alpha/2$ として, 対立仮説のもとで,

$$\frac{\overline{Y}_{1.} - \overline{Y}_{2.}}{\sqrt{4\hat{\sigma}^2/n}} < -K_{\alpha/2}$$

となる確率を無視すれば, n を次のようにとればよい.

$$n \geq \frac{4(K_{\alpha/2} + K_\beta)^2}{\delta^2} \tag{11.8}$$

等分散性の検定

t 検定は 2 個の正規母集団の分散が等しいことを仮定していたので, 実際にこの仮定が満たされているかどうかを検証しなければならない. そこで今度は, Y_{ij} が互いに独立に正規分布 $N(\mu_i, \sigma_i^2)$ に従うと仮定し, 帰無仮説 $\sigma_1^2 = \sigma_2^2$ の検定を考える.

定理 5.3 より,

$$\hat{\sigma}_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i.})^2 \sim \frac{\sigma_i^2}{n_i - 1} \chi^2_{n_i - 1}$$

であり, $\hat{\sigma}_1^2, \hat{\sigma}_2^2$ は互いに独立である. したがって, F 分布の定義より,

$$W = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \sim \frac{\sigma_1^2}{\sigma_2^2} F^{n_1-1}_{n_2-1}$$

を得る. したがって, 対立仮説

$$\begin{aligned} H_1 &: \sigma_1^2 \neq \sigma_2^2 \quad (\ \text{両側}\ )\ \text{または} \\ H_2 &: \sigma_1^2 > \sigma_2^2 \quad (\text{右片側})\ \text{または} \\ H_2' &: \sigma_1^2 < \sigma_2^2 \quad (\text{左片側}) \end{aligned}$$

に対して, それぞれ次のように棄却域を定めればよい.

$$W_1 : W > f^{n_1-1}_{n_2-1}(\alpha/2),\ \ W < f^{n_1-1}_{n_2-1}(1 - \alpha/2) \tag{11.9}$$

$$W_2 : W > f^{n_1-1}_{n_2-1}(\alpha) \tag{11.10}$$

$$W_2' : W < f^{n_1-1}_{n_2-1}(1 - \alpha) \tag{11.11}$$

ここに, $f^{n_1-1}_{n_2-1}(\alpha)$ は, $F^{n_1-1}_{n_2-1}$ の上側 α 点である. 数表を用いる場合は,

$$f^{n_1-1}_{n_2-1}(1 - \alpha) = 1 \Big/ f^{n_2-1}_{n_1-1}(\alpha) \tag{11.12}$$

の関係を用いる.

Welch の検定

等分散性が保証されない場合は, 2 個の正規母集団の平均の比較には Welch の検定を用いる. 直観的には, 統計量

$$T_* = \frac{\overline{Y}_{1.} - \overline{Y}_{2.}}{\sqrt{\hat{\sigma}_1^2/n_1 + \hat{\sigma}_2^2/n_2}}$$

を考えればよいように思われるが, この分布は分散比に依存し, 形式的に $t_{n_1+n_2-2}$ を仮定した解析を行うと, 第一種の誤りが増大することが知られている.

帰無仮説のもとでの $P(|T^*| < t)$ を近似的に求めてみよう. Z を正規分布 $N(0,1)$ に従う確率変数として, 関数

$$g(t) = P(Z^2 \le t)$$

を用いると, 求める確率は, 次のように表現できる.

$$P(|T^*| < t) = E\left[g\left(t^2 \frac{\hat{\sigma}_1^2/n_1 + \hat{\sigma}_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}\right)\right] \tag{11.13}$$

期待値の線型性に注意して, (11.13) を t^2 のまわりに Taylor 展開すると, 次の式を得る.

$$g(t^2) + \frac{1}{2}g''(t^2)\,E\left[\left(t^2 \frac{\hat{\sigma}_1^2/n_1 + \hat{\sigma}_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} - t^2\right)^2\right] + \cdots$$

$\hat{\sigma}_i^2$ の不偏性により $g'(t^2)$ の項は 0 になる. 上の式に

$$V\left[\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}\right] = \frac{2\sigma_1^2}{n_1^2(n_1-1)} + \frac{2\sigma_2^2}{n_2^2(n_2-1)}$$

を用いると, 次の式を得る.

$$g(t^2) + g''(t^2)\,\frac{t^4}{2} \cdot \frac{2\sigma_1^4/\{n_1^2(n_1-1)\} + 2\sigma_2^4/\{n_2^2(n_2-1)\}}{(\sigma_1^2/n_1 + \sigma_2^2/n_2)} + \cdots$$

他方, 自由度 ν の t 分布に従う確率変数 T_ν については, χ_ν に従う確率変数を Y として,

$$P\Big(\,|\,T_\nu\,|<t\,\Big) = E\left[g\left(\frac{t^2\,Y}{\nu}\right)\right] = g(t^2) + g''(t^2)\,\frac{t^4}{2}\cdot\frac{2\nu}{\nu^2} + \cdots$$

と表される.

これと (11.13) を t^4 のオーダーまで等しいとおくと,

$$\frac{\sigma_1^4/\{n_1^2(n_1-1)\} + \sigma_2^4/\{n_2^2(n_2-1)\}}{(\sigma_1^2/n_1 + \sigma_2^2/n_2)} = \frac{1}{\nu} \tag{11.14}$$

を得る. 結局, 統計量 T^* が, 近似的に (11.14) で与えられる自由度 ν の t 分布に従うことを利用するのが Welch の検定である. ν は一般には整数値をとらないので, 四捨五入する. 実際には, σ_1^2, σ_2^2 は未知なので, それぞれ $\hat{\sigma}_1^2, \hat{\sigma}_2^2$ で代用する. 近似は, n_1, n_2 が 10 以上なら精度がよいといわれている.

11.2　ノンパラメトリック法

前節では, 特性値に正規分布を仮定した. 本節では正規性の仮定できない一般の場合についての 2 標本問題を議論する.

2 つの母集団の分布が等しいという帰無仮説 H_0 を, 母集団 1 からの標本が母集団 2 からの標本より統計的に大きな値を取り易いという対立仮説 H_1 に対して検定するとしよう. 前節と同様に, データを Y_{ij} $(i = 1, 2\ ,\ j = 1, \cdots, n_i)$ とおき, i が母集団, j がオブザベーションを表すとする.

一般に, 分布を仮定できない場合の統計学的推論には, ノンパラメトリック法を用いるが, 後で述べるスコアの取り方により, 検出力は異なってくる. したがって, 大まかな分布の傾向がわかっている場合には, どの手法を用いるか事前に決めておくことが応用上重要である. 分布が既知の場合は, 帰無仮説の近傍で, 最も検出力の大きい局所最強力順位検定を導くことができる. いくつかの代表的な分布の場合について, 本節で議論しよう.

並べ替え検定

2 標本のときの並べ替え検定は, 母集団 1 からの標本値の総和

$$Y = \sum_{j=1}^{n_i} Y_{1j} \tag{11.15}$$

を統計量とする. $n = n_1 + n_2$ 個の全データから n_1 個を選ぶすべての組み合わせ ${}_n\mathrm{C}_{n_1}$ のうち, その和が (11.15) 以上になるものの個数を m_1 として, 有意確率を次の式で求めればよい.

$$m_1 \big/ {}_n\mathrm{C}_{n_1} \tag{11.16}$$

実際はこの数え上げは容易ではなく, 正規近似を用いる. それにはまず, すべての y_{ij} を所与としたときの Y_{1j} の帰無仮説のもとでの条件つき平均を,

$$E[Y_{1j}] = \frac{1}{n} \sum_i \sum_j y_{ij} = \mu \tag{11.17}$$

とおく. これから直ちに,

$$E[Y] = n_1\, E[Y_{1j}] = n_1\, \mu \tag{11.18}$$

を得る. 同様に, y_{ij} を与えたときの Y_{1j} の帰無仮説のもとでの条件つき分散は

$$V[Y_{1j}] = \frac{1}{n} \sum_i \sum_j (y_{ij} - \mu)^2 \stackrel{\mathrm{def}}{=} \sigma^2 \tag{11.19}$$

である. 共分散 Cov $[Y_{1j}, Y_{1j'}]$ は次で与えられる.

$$\frac{1}{{}_n\mathrm{C}_2} \sum_{i=1}^{2} \sum_{j<j'} (y_{ij} - \mu)\,(y_{ij'} - \mu) = -\frac{\sigma^2}{n-1} \tag{11.20}$$

以上を,

$$V[Y] = n_1\, V[Y_{1j}] + n_1\,(n_1 - 1)\,\mathrm{Cov}\,[Y_{1j}, Y_{1j'}] \tag{11.21}$$

に代入, 整理して,

$$V[Y] = \frac{n_1\, n_2\, \sigma^2}{n - 1} \tag{11.22}$$

を得る.

結局, 近似的に

$$\frac{Y - E[Y]}{\sqrt{V[Y]}} = \frac{Y - n_1\, \mu}{\sqrt{n_1\, n_2\, \sigma^2/(n-1)}} \sim N(0, 1) \tag{11.23}$$

が成り立つことを利用して, 検定を構成できる.

1 標本の場合と同様に, 並べ替え検定は通常の t 検定量 (11.4) と漸近的に同等である. 実際,

$$\overline{Y}_{1.} - \overline{Y}_{2.} = \frac{Y}{n_1} - \frac{n\mu - Y}{n_2} = \frac{n}{n_1\, n_2}(Y - n_1\, \mu)$$

および

$$\sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2 = n\,(\sigma^2 + \mu^2) - \left\{\frac{Y^2}{n_1} + \frac{(n\mu - Y)^2}{n_2}\right\}$$

より, 簡単な計算の後,

$$T = \frac{\sqrt{n-2}\,(Y - n_1\, \mu)}{\sqrt{n_1\, n_2\, \sigma^2 - (Y - n_1\, \mu)^2}} \tag{11.24}$$

が成立する. これは Y の単調増加関数であるから, 結局, 並べ替え検定と t 検定は同等であることがわかる.

順位検定

並べ替え検定は, 計算が面倒な上に, 標本を得る前には有意確率を計算できないので, 数表を用意することもできない. 実際には, 順位検定の方が一般的である.

最強力順位検定

2 標本を合併したときに, Y_{ij} を大きさの順に小さい方から数えた順位を並べて得られるベクトルを $\boldsymbol{R}_i = (R_{i1}, \cdots, R_{in_i})'$ $(i = 1, 2)$ とすると, $\boldsymbol{R}' = (\boldsymbol{R}_1', \boldsymbol{R}_2')'$ は 2 つの母集団の分布が等しいという帰無仮説のもとで, 一様分布 (8.25) に従う.

他方, 対立仮説のもとでの Y_{1j}, Y_{2j} の密度関数をそれぞれ $f(y), g(y)$ とすると, 標本 $\boldsymbol{r}$ の得られる確率は

$$\int_{y_{(1)}<\cdots<y_{(n)}} \prod_{j=1}^{n_1} \frac{f(y_{(r_{1j})})}{g(y_{(r_{1j})})} \prod_{i=1}^{2} \prod_{j=1}^{n_i} g(y_{(r_{ij})})\, dy_{(r_{ij})} \tag{11.25}$$

と表される. (8.25) より, (11.25) は

$$P_1(\boldsymbol{R}_1 = \boldsymbol{r}_1, \boldsymbol{R}_2 = \boldsymbol{r}_2) = \frac{1}{n\,!} E\Bigl[\prod_{j=1}^{n_1} \frac{f(y_{(r_{1j})})}{g(y_{(r_{1j})})}\Bigr] \tag{11.26}$$

と表される. 結局, 最強力順位検定は次の定理により与えられる.

■定理 11.1　$1, \cdots, n$ から n_1 個を選ぶ ${}_n\mathrm{C}_{n_1}$ 個の組み合わせすべてからなる集合を $C(n, n_1)$ とする. このとき, $\boldsymbol{r}_1 \in C(n, n_1)$ に対し, 最強力順位検定は,

$$\phi^*(\boldsymbol{r}_1) = \begin{cases} 1 & \cdots \quad T(\boldsymbol{r}_1) > c \\ \lambda & \cdots \quad T(\boldsymbol{r}_1) = c \qquad (0 < \lambda < 1) \\ 0 & \cdots \quad T(\boldsymbol{r}_1) < c \end{cases} \tag{11.27}$$

なる検定関数 $\phi^*(\boldsymbol{r}_1)$ によって与えられる. ここに, $T(\boldsymbol{r}_1)$ は密度関数 $g(y)$ に従う n 個の独立な確率変数の順序統計量 $Y_{(l)}$ $(l = 1, \cdots, n)$ を用いて

$$T(\boldsymbol{r}_1) = E\Bigl[\prod_{j=1}^{n_1} f(Y_{(r_{1j})}) \Big/ g(Y_{(r_{1j})})\Bigr] \tag{11.28}$$

と表される.

また, 定数 c は次の式により定まる.

$$\sum_{\boldsymbol{r}_1 \in C(n, n_1)} \phi^*(\boldsymbol{r}_1) = \alpha\, {}_n\mathrm{C}_{n_1} = m \tag{11.29}$$

■証明　$T(\boldsymbol{r}_1)$ は, $y_{(r_{1j})}, y_{(r_{2j})}$ の順位を任意に置換しても不変であるから, $\mathcal{T}_n$ を, その部分集合 $C(n, n_1)$ に限定しても一般性を失わない. まず, 検定関数 $\phi(\boldsymbol{r}_1), \phi^*(\boldsymbol{r}_1)$ が有意水準の条件

$$\frac{1}{n!}\sum_{\boldsymbol{r}_1 \in C(n,n_1)} \phi(\boldsymbol{r}_1) = \alpha \tag{11.30}$$

を満たすと仮定する. 検定関数 $\phi^*(\boldsymbol{r}_1)$ と $\phi(\boldsymbol{r}_1)$ の検出力の差は

$$\sum_{\boldsymbol{r}_1 \in C(n,n_1)} \{\phi^*(\boldsymbol{r}_1) - \phi(\boldsymbol{r}_1)\} P_1(\boldsymbol{R}_1 = \boldsymbol{r}_1) \tag{11.31}$$

である. (11.31) は, 有意水準の条件から, 任意の k に対して,

$$\sum_{\boldsymbol{r}_1 \in C(n,n_1)} \{\phi^*(\boldsymbol{r}_1) - \phi(\boldsymbol{r}_1)\}\left\{\frac{T(\boldsymbol{r}_1)}{n!} - \frac{k}{n!}\right\} \tag{11.32}$$

と表される. ここで,

$$\phi^*(\boldsymbol{r}_1) = \begin{cases} 1 & \cdots \quad T(\boldsymbol{r}_1) > k \\ \lambda & \cdots \quad T(\boldsymbol{r}_1) = k \\ 0 & \cdots \quad T(\boldsymbol{r}_1) < k \end{cases} \tag{11.33}$$

とおけば, $0 \le \phi \le 1$ より (11.32) は 0 以上の値をとる. そこで, $k = c$ とおけばよいが, c は有意水準の条件から, (11.29) を満たすように定めなければならない. 実際的な手順は, ${}_n\mathrm{C}_{n_1}$ 個の $T(\boldsymbol{r}_1)$ を大きさの順に並べ, 大きい方から m 番目までは必ず帰無仮説を棄却し, $m+1$ 番目は確率 $m - [m]$ で帰無仮説を棄却すればよい.

局所最強力順位検定

最強力順位検定において, 平均のシフトモデル

$$f(y) = g(y - \varDelta) \tag{11.34}$$

を仮定して, 帰無仮説の近傍における局所最強力検定を導こう. 一般に, $\varDelta$ が大きければ, 平均の差を検出するのは容易であり, $\varDelta = 0$ の近傍では最もシビアに検出力が要求されるからである.

上の仮定はかなり強い仮定であるが, このもとに (11.28) は

$$T(\boldsymbol{r}_1) = E\left[\prod_{j=1}^{n_1} \frac{g(Y_{(r_{1j})} - \varDelta)}{g(Y_{(r_{1j})})}\right] \tag{11.35}$$

となる.

$\varDelta$ に依存することを強調して, 改めて $T(\boldsymbol{r}_1, \varDelta)$ と表すと,

$$T(\boldsymbol{r}_1, \varDelta) - T(\boldsymbol{r}_1, 0) = E\left[\left\{\prod_{j=1}^{n_1} \frac{g(Y_{(r_{1j})} - \varDelta)}{g(Y_{(r_{1j})})}\right\} - 1\right]$$

を得る.

ところで, 一般に, 2 つの数列 $\{a_n\}, \{b_n\}$ に対して,

$$\prod_{k=1}^{m} a_k - \prod_{k=1}^{m} b_k = \sum_{k=1}^{m} a_1 \cdots a_{k-1}\,(a_k - b_k)\,b_{k+1} \cdots b_m$$

が成り立つ. これを用いると,

$$\int_{-\infty}^{\infty} \left|g'(y)\right| dy < \infty$$

の仮定のもとで,

$$T'(\boldsymbol{r}_1, \varDelta) = \lim_{\varDelta \to 0} \frac{T(\boldsymbol{r}_1, \varDelta) - T(\boldsymbol{r}_1, 0)}{\varDelta} = E\left[-\sum_{j=1}^{n_1} \frac{g'(Y_{(r_{1j})})}{g(Y_{(r_{1j})})}\right]$$

を得る. したがって, $T(\boldsymbol{r}_1, 0) = 1$ より,

$$T(\boldsymbol{r}_1, \varDelta) = 1 + \varDelta \sum_{j=1}^{n_1} E\left[-\frac{g'(Y_{(r_{1j})})}{g(Y_{(r_{1j})})}\right] + o(\varDelta^2)$$

を得る. 有意水準を, ランダム検定にならないように適当に選べば, 検定方式は

$$\phi(\boldsymbol{r}_1) = \begin{cases} 1 & \cdots \quad T(\boldsymbol{r}_1) \geq c \\ 0 & \cdots \quad T(\boldsymbol{r}_1) < c \end{cases}$$

となる. ここに, c は有意水準から定まる定数である.

以上から, 改めて $T(\boldsymbol{r}_1)$ を

$$\sum_{j=1}^{n_1} E\left[-\frac{g'(Y_{(r_{1j})})}{g(Y_{(r_{1j})})}\right] = \sum_{j=1}^{n_1} E\left[-\frac{d\log g(Y_{(r_{1j})})}{dY_{(r_{1j})}}\right] \tag{11.36}$$

とおくと, 検定方式は

$$\phi(\boldsymbol{r}_1) = \begin{cases} 1 & \cdots \quad T(\boldsymbol{r}_1) \geq c' \\ 0 & \cdots \quad T(\boldsymbol{r}_1) < c' \end{cases} \tag{11.37}$$

となる. ここに, c' は有意水準から定まる定数である.

Fisher-Yates **検定**

f, g が標準正規分布のときは

$$-\frac{d\log f(y)}{dy} = -\frac{d}{dy}\left\{\text{const} - \frac{1}{2}y^2\right\} = y$$

より,

$$\sum_{j=1}^{n_1} E\left[\, Y_{(r_{ij})}\right] \tag{11.38}$$

により検定を構成すればよい. これを Fisher-Yates 検定という. なお, ここでは標準正規分布を仮定したが, 一般の正規分布でも局所最有力検定は Fisher-Yates 検定である. なぜなら, 順位検定の方式は標本を連続な単調増加関数 $h(y)$ により変換しても不変だからである. 実際, Y_{ij} を $h(y)$ により変換すると, $h^{-1}(y)$ が存在して, これも連続な単調増加関数であるから,

$$F(y) \geq G(y) \Longleftrightarrow F(h^{-1}(y)) \geq G(h^{-1}(y))$$

であり, しかも F, G の任意性から分布関数 $F(h^{-1}(y)), G(h^{-1}(y))$ を新たに $F(y), G(y)$ とおけばよい. ここに, $F(y), G(y)$ はそれぞれ f, g に対応する分布関数である.

van der Waerden 検定

一般に, 確率変数 Y が連続分布に従うとき, その分布関数を $F(y)$ とすると, $U = F(Y)$ は (5.2) より, 一様分布 $U(0,1)$ に従う. したがって, $\Phi(y)$ を標準正規分布の分布関数とすると, (11.38) は

$$\sum_{j=1}^{n_1} E\Big[\, \Phi^{-1}(U_{(r_{1j})})\,\Big] \tag{11.39}$$

と表現できる. ここに, $U_{(l)}$ は一様分布 $U(0,1)$ に従う独立な n 個の確率変数の第 l 順序統計量である. n が大きいときは,

$$E\Big[\, \Phi^{-1}(U_{(l)})\,\Big] \simeq \Phi^{-1}\Big(\, E[U_{(l)}]\Big) \tag{11.40}$$

で近似できる. 他方, (8.18) より,

$$E\Big[\, U_{(l)}\,\Big] = \frac{n\,!}{(l-1)\,!\,(n-l)\,!} \int_0^1 y \cdot y^{l-1}\,(1-y)^{n-l}\,dy$$

を得る. 積分の部分に (4.30) を用いると,

$$B(l+1, n-l+1) = \frac{\Gamma(l+1)\,\Gamma(n-l+1)}{\Gamma(n+2)}$$

となる. さらに, 定理 4.2 を用いると,

$$E\Big[\, U_{(l)}\,\Big] = \frac{l}{n+1} \tag{11.41}$$

を得る. 結局, 統計量を

$$\sum_{l=1}^{n_1} \Phi^{-1}\Big(\frac{r_{1j}}{n+1}\Big) \tag{11.42}$$

とすればよい. この検定を van der Waerden 検定という.

順位検定は, 標本のもつ情報のうち, 順位に関する情報のみを利用するので, 一般に検出力が低いことが懸念されるが, 正規分布を仮定した場合, van der Waerden 検定の検出力は相当に良い.

Wilcoxon 検定

$f(y)$ を (4.34) とすると, 簡単な計算により,

$$-\frac{f'(y)}{f(y)} = 2\,F(y) - 1$$

を得る. したがって, 局所最強力検定は, 定理 11.1 により,

$$\sum_{l=1}^{n_1} E\Big[\,F(Y_{(r_{1j})})\Big] \tag{11.43}$$

を統計量とした場合と同等である. van der Waerden 検定の項でもみたように,

$$U = F(Y) = \frac{\exp Y}{1 + \exp Y} \sim U(0,1)$$

であるから, 検定統計量は,

$$\sum_{j=1}^{n_1} E\Big[\,U_{(r_{1j})}\Big] = \sum_{j=1}^{n_1} \frac{r_{1j}}{n+1} \tag{11.44}$$

となる. ここに, $U_{(l)}$ は一様分布 $U(0,1)$ に従う独立な n 個の確率変数の第 l 順序統計量である. 明らかに, この検定は, $n+1$ 倍して順位そのものをスコアとした

$$\sum_{j=1}^{n_1} r_{1j} \tag{11.45}$$

により構成される検定と同等である. これを Wilcoxon 検定 という.

実は, ロジスティック分布は正規分布にかなり近い. したがって, 標本が正規分布に従う場合にも Wilcoxon 検定の検出力は相当に高く, 正規性を仮定した場合に, t 検定と同じ検出力を保つために必要な標本の大きさは, $\pi/3$ 倍であることが知られている.

中央値検定

両側指数分布

$$f(y) = \frac{1}{2}\exp -|y| \tag{11.46}$$

は正規分布より裾の重い, すなわち外れ値の出やすい分布である.

(11.46) は, 0 において微分不能であるが, 0 以外では微分可能で,

$$h(y) = -\frac{f'(y)}{f(y)} = -\frac{d}{dy}\log f(y) = \begin{cases} 1 & \cdots \quad y > 0 \\ -1 & \cdots \quad y < 0 \end{cases}$$

である.

したがって,

$$m(y) = \frac{h(y)+1}{2} = \begin{cases} 1 & \cdots \quad y > 0 \\ 0 & \cdots \quad y < 0 \end{cases}$$

とおくと, 局所最強力検定は,

$$\sum_{j=1}^{n_1} E\Big[\, m\Big(F^{-1}(U_{(r_{1j})})\Big)\Big] \simeq \sum_{j=1}^{n_1} m\Big(\, F^{-1}(E[U_{(r_{1j})}])\Big) \tag{11.47}$$

を統計量とした場合と漸近的に同等である. 実はこの統計量は, 母集団 1 において, 二標本を合併したときの中央値より大きい値をとるサンプル数を表す. これが, 中央値検定という名前の由来である.

実際,

$$l^* = E[U_{(l)}] = l/(n+1)$$

とおくと,

$$F^{-1}(y) = \begin{cases} \log 2y & \cdots \quad y \le 1/2 \\ -\log 2(1-y) & \cdots \quad y > 1/2 \end{cases}$$

より,

$$\begin{aligned} l < \frac{n+1}{2} &\iff \quad \log 2\, l^* < 0 \quad \iff m\Big(F^{-1}(l^*)\Big) = 0\ , \\ l > \frac{n+1}{2} &\iff -\log 2(1-l^*) > 0 \iff m\Big(F^{-1}(l^*)\Big) = 1 \end{aligned}$$

を得る. n が奇数のときは中央値に等しいサンプルは無視する.

Savage 検定

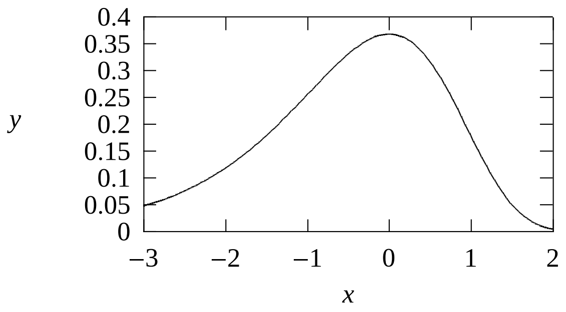


図 11.1: $y = \exp(x - \exp x)$ のグラフ

左裾が重く, 右裾が軽い非対称な次の分布を仮定する.

$$f(y) = \exp(y - \exp y) = \exp y \cdot \exp(-\exp y) \tag{11.48}$$

(11.48) より

$$-\frac{f'(y)}{f(y)} = -\frac{d \log f(y)}{dy} = \exp y - 1$$

を得るから, 局所最強力順位検定は

$$\sum_{j=1}^{n_1} E\Big[\exp Y_{(r_{1j})}\Big] \tag{11.49}$$

に基づく検定と同等である.

ここで, $Z = \exp Y$ とおくと, Z の密度関数は (5.2) より

$$z\ \exp(-z)\,\frac{dy}{dz} = \exp(-z)$$

となり, Z は指数分布に従うことがわかる.

第 l 順序統計量 $Z_{(l)}$ の平均を求めよう. まず,

$$a_l = \frac{(l-1)!\,(n-l)!}{n!}\,E\left[\,Z_{(l)}\,\right]$$

とおくと, (8.18) より,

$$a_l = \int_0^\infty z\,\{1-\exp(-z)\}^{l-1}\,\exp(-z)\,\{\exp(-z)\}^{n-l}\,dz$$

を得る. 部分積分により, 若干の計算の後, 次を得る.

$$a_l = \frac{1}{n-l+1}\,B(n-l+1, l) + \frac{l-1}{n-l+1}\,a_{l-1}$$

これに, (4.30) と定理 4.2 を用いて

$$E\left[\,Z_{(l)}\,\right] = \frac{1}{n-l+1} + E\left[\,Z_{(l-1)}\,\right]$$

を得る. この漸化式と,

$$E\left[\,Z_{(1)}\,\right] = n\int_0^\infty y\,\{\exp(-y)\}^n\,dy = \frac{1}{n}$$

より,

$$E\left[\,Z_{(l)}\,\right] = \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n-l+1} = \sum_{k=n-l+1}^{n}\frac{1}{k}$$

が得られる. 結局, 分布 (11.48) に対しては,

$$\sum_{j=1}^{n_1}\sum_{k=n-r_{1j}+1}^{n}\frac{1}{k} \tag{11.50}$$

に基づく検定が局所最強力である. これを Savage 検定という.

ノンパラメトリック検定の手順

局所最強力順位検定の統計量は, c_{ij} を $i=1$ のときは 1 を, $i=2$ のときは 0 をとる変数として, 次のように表すことができる.

$$S = \sum_{j=1}^{n_i} s(R_{1j}) = \sum_{i=1}^{2}\sum_{j=1}^{n_i} c_{ij}\,s(R_{ij}) \tag{11.51}$$

ここに, R_{ij} は n 個の Y_{ij} の順位であり, $s(R_{ij})$ は順位のスコアである.

ここで, Y_{ij} を辞書順に並べて, 改めて $Y_1, \cdots, Y_n$ と書き直すと,

$$S = \sum_{l=1}^{n} c_l\, s(R_l)$$

となって, 検定統計量は 1 標本問題の場合とまったく同じになる. すなわち, 局所最強力検定統計量は線型順位統計量である. ただし, 係数 c_l は l 番目のオブザベーションが母集団 1 からの標本であれば 1 を, 2 からの標本であれば 0 をとる.

局所最強力検定の場合, 統計量 S の平均, 分散については, 系 8.2 と,

$$\begin{aligned}\sum_{(l)=1}^{n} E\Big[\frac{f'(Y_{(l)})}{f(Y_{(l)})}\Big] &= \sum_{l=1}^{n} E\Big[\frac{f'(Y_l)}{f(Y_l)}\Big] \\ &= n\int_{-\infty}^{\infty}\frac{f'(y)}{f(y)}\cdot f(y)\,dy = 0 \qquad (11.52)\end{aligned}$$

より,

$$E[S] = 0\,, \quad V[S] = \frac{n_1\, n_2}{n\,(n-1)}\sum_{l=1}^{n} E\Big[\Big\{\frac{f'(Y_{(l)})}{f(Y_{(l)})}\Big\}^2\Big] \qquad (11.53)$$

となる.

以上から,

$$\frac{S - E[S]}{\sqrt{V[S]}} \sim N(0,1) \qquad (11.54)$$

により検定を構成すればよい.

Fisher-Yates 検定, van der Waerden 検定については, すでに (11.53) に平均, 分散を与えた. Savage 検定では, 局所最強力順位検定のスコアに 1 を加えてスコアを求めたので, 平均は 1 である.

Wilcoxon 統計量では, 局所最強力順位検定と同等の検定を導いたので, 新たに平均, 分散を導く. 系 8.2 において, $s(l) = l$ とおくと, 直ちに次を得る.

$$E[S] = \frac{1}{2}\,n_1\,(n+1)\,, \quad V[S] = \frac{1}{12}n_1\,n_2\,(n+1) \qquad (11.55)$$

タイのない場合, 一般性を失うことなく $Y_1 < \cdots < Y_{n_1}$ を仮定できて, Y_j の順位は, 指示関数 (8.44) を用いて,

$$R_j = \sum_{i=n_1+1}^{n} u(Y_j - Y_i) + j \tag{11.56}$$

と表現できる. $j = 1, \cdots, n_1$ とおき, 総和をとると,

$$U = \sum_{j=1}^{n_1} R_j = T + \frac{1}{2} n_1 (n_1 + 1), \tag{11.57}$$

$$T = \sum_{j=1}^{n_1} \sum_{i=n_1+1}^{n} u(Y_j - Y_i) \tag{11.58}$$

となる. この T が Mann-Whitney 型の表現である. これは定数 $n_1 (n_1 + 1)/2$ の違いを除いて, Wilcoxon の統計量 S と一致する. (11.55) より直ちに次を得る.

$$E[T] = \frac{1}{2} n_1 n_2 , \quad V[T] = \frac{1}{12} n_1 n_2 (n + 1) \tag{11.59}$$

中央値検定においても, 系 8.2 より, n が偶数のときは

$$E[S] = \frac{n_1}{2} , \quad V[S] = \frac{n_1 n_2}{4 (n - 1)} \tag{11.60}$$

であり, n が奇数のときは,

$$E[S] = \frac{n_1 (n - 1)}{2 n} , \quad V[S] = \frac{n_1 n_2 (n + 1)}{4 n^2} \tag{11.61}$$

となる.

ノンパラメトリックな区間推定

今度は, 母集団 $i = 1, 2$ の密度関数をそれぞれ $f(y), g(y)$ として, シフトモデル (11.34) を考える. このモデルに基づいて, ノンパラメトリックに Δ の信頼区間を構成しよう.

まず, Wilcoxon 検定から信頼区間を導く. $Z_{(l)}$ を $n_1 n_2$ 個の

$$Y_{1j} - Y_{2j'}, \quad (j = 1, \cdots, n_1\ ;\ j' = 1, \cdots, n_2)$$

を大きさの順に並べたときの第 l 順序統計量とする. このとき, n_1 個の $Y_{1j} - \mathit{\Delta}$, n_2 個の $Y_{2j'}$ から得られる Mann-Whitney 型の Wilcoxon 統計量 T は $Z_{(l)} > \mathit{\Delta}$ を満たす $Z_{(l)}$ $(l = 1, \cdots, n_1 n_2)$ の個数にほかならない.

したがって, $\mathit{\Delta}$ の $1 - \alpha$ 信頼区間は,

$$Z_{(n_1 n_2 + 1 - s_\alpha)} < \mathit{\Delta} < Z_{(s_\alpha)} \tag{11.62}$$

である. ここに,

$$s_\alpha = \frac{n_1 n_2 + 1}{2} + K_{\alpha/2} \sqrt{\frac{1}{12} n_1 n_2 (n + 1)} \tag{11.63}$$

である.

実際, 帰無仮説 $\mathit{\Delta} = 0$ のもとで, (11.59) と対称性から,

$$P(\mathit{\Delta} \geq Z_{s_\alpha}) = P(T \geq s_\alpha) = P(T \leq n_1 n_2 - s_\alpha) \leq \alpha/2$$

が成り立つ.

つぎに, 中央値検定から $\mathit{\Delta}$ の信頼区間を導く. 再び, (11.34) を仮定し, Y_{ij} の第 l 順序統計量を $Y_{i(l)}$ とおく. $(i = 1, 2\ ;\ j = 1, \cdots, n_i)$

n が偶数のときを考えよう. $Y_{1j} - \mathit{\Delta}, Y_{2j}$ を大きさの順に並べ, 中央値を m とおくと, 中央値検定の統計量 S は, 母集団 1 において m を超える標本の数であったから,

$$\begin{aligned} &\text{母集団 1}\ : Y_{1(n_1 - s + 1)} - \mathit{\Delta}, \cdots, Y_{1(n_1)} - \mathit{\Delta} \\ &\text{母集団 2}\ : Y_{2(n_2 - m + s + 1)}, \cdots, Y_{2(n_2)} \end{aligned}$$

が成り立つ. このことから次の不等式を得る.

$$\begin{aligned} S > s &\Longleftrightarrow S \geq s + 1 \Longleftrightarrow Y_{1(n_1 - s)} - \mathit{\Delta} > Y_{2(n_2 - m + s + 1)}\,, \\ S < s &\Longleftrightarrow S \leq s - 1 \Longleftrightarrow Y_{2(n_2 - m + s)} > Y_{1(n_1 - s + 1)} - \mathit{\Delta} \end{aligned}$$

以上から,

$$s_\alpha = \frac{n_1\,m}{n} + \frac{1}{2} + K_{\alpha/2}\sqrt{\frac{n_1\,n_2\,m\,(n-m)}{n^2\,(n-1)}} \tag{11.64}$$

とおくと, $\varDelta$ の $1-\alpha$ 信頼区間は,

$$\begin{aligned} &Y_{1(n_1-s_\alpha+1)} - Y_{2(n_2-m+s_\alpha)} < \varDelta \\ &\qquad < Y_{1(n_1-s_\alpha)} - Y_{2(n_2-m+s_\alpha+1)} \end{aligned} \tag{11.65}$$

で与えられる. 実際, この不等式の成立する確率は, 超幾何分布の項で述べたことから, $1-\alpha$ に等しい.

なお, 以上の結果は n が奇数のときも正しい.

11.3　特性値が名義変数の場合

今度は, 特性値が名義変数である場合に, 2 標本問題を考える. 名義変数とは, 文字通り定性的な属性しかもたない変数である.

このうち, 最も重要なのは, 良, 不良や, 改善, 悪化などといった 2 通りの属性しか持たない場合である. このような変数を 2 値変数という. 当面, 議論の対象を 2 値変数に限る. 3 値以上の場合, 名義変数を特性値とする比較は, 母集団における確率分布が等しいという帰無仮説に対する総括的な検定しか構成できず, 実際のアクションに結び付けにくいからである. 連続変数以外で, 3 値以上をとる変数は, むしろ順序カテゴリの場合が多い.

そこで, Y_{ij} を二項分布 $B(1, p_i)$ に従う確率変数とし,

$$\text{帰無仮説 } H_0 : p_1 = p_2$$

を,

$$\text{対立仮説 } H_1 : p_1 \neq p_2 \quad (\text{両側}) \text{ または } H_2 : p_1 > p_2 \quad (\text{右片側})$$

に対して, 検定することを考える. 3 値以上の場合は, 一元配置の特別な場合として議論しよう.

大標本の場合

標本の大きさ n_1, n_2 が十分大きければ, 二項分布の正規近似を用いて, 平均の差に関する 2 標本問題に帰着する. 実際,

$$\sum_{j=1}^{n_i} \frac{Y_{ij}}{n_i} \sim N\left(p_i, \frac{p_i(1-p_i)}{n_i}\right) \tag{11.66}$$

が漸近的に成り立つので, H_1, H_2, それぞれに対し, 棄却域を次のようにとればよい.

$$W_1 : \left| p_1 - p_2 \right| > K_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \tag{11.67}$$

$$W_2 : \quad p_1 - p_2 \quad > \quad K_{\alpha} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \tag{11.68}$$

ここに,

$$\hat{p}_i = \sum_{j=1}^{n_i} \frac{Y_{ij}}{n_i} \tag{11.69}$$

である. また, 比率の差に関する $1-\alpha$ 信頼区間は

$$\begin{aligned} \hat{p}_1 - \hat{p}_2 - K_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} &< p_1 - p_2 \\ &< \hat{p}_1 - \hat{p}_2 + K_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \end{aligned} \tag{11.70}$$

で与えられる.

適合度の検定

標本があまり大きくない場合は適合度の検定がしばしば用いられる. 次の分割表を想定しよう. i は処理を, j は特性のあり (1) またはなし (2) を表すとする. Y_{ij} は i, j セルの頻度に対応する確率変数である.

超幾何分布を仮定すると, 周辺度数が固定されることから, 独立な確率変数は唯一で, Y_{11} のみ考えれば十分である.

	$j=1$	$j=2$	計
$i=1$	Y_{11}	Y_{12}	$n_{1.}$
$i=2$	Y_{21}	Y_{22}	$n_{2.}$
計	$n_{.1}$	$n_{.2}$	n

このとき, 漸近的に次が成立する.

$$\frac{(Y_{11}-E[Y_{11}])^2}{V[Y_{11}]} = \frac{(Y_{11}-n_{1.}\,n_{.1}/n)^2}{n_{1.}\,n_{2.}\,n_{.1}\,n_{.2}/n^2\,(n-1)} \sim \chi^2_1 \tag{11.71}$$

分散の評価を少し変え, 分母の $n^2\,(n-1)$ のかわりに n^3 として構成する検定が Pearson の適合度の検定である.

この統計量は, 次のように表現することもできる.

$$\frac{n\,(Y_{11}\,Y_{22}-Y_{12}\,Y_{21})^2}{n_{1.}\,n_{2.}\,n_{.1}\,n_{.2}} = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(Y_{ij}-E[Y_{ij}])^2}{E[Y_{ij}]} \tag{11.72}$$

なお, 適合度の検定は両側検定のみで, 片側検定はできない.

Fisher の直接確率法

適合度の検定は各セルの平均 $n_{i.}\,n_{.j}/n$ が 5 以上のときによい近似を与えるといわれる. この条件がみたされないときは, Fisher の直接確率法という方法を用いる. これは, 超幾何分布

$$p(y) = \frac{{}_{n_{1.}}\mathrm{C}_y \cdot {}_{n_{2.}}\mathrm{C}_{n_{.1}-y}}{{}_n\mathrm{C}_{n_{.1}}} \tag{11.73}$$

から, 有意確率を以下のように計算すればよい. この検定を Fisher の正確な検定ということがある.

$$\sum_{p(y)\le p(y_{11})} p(y)\ (\text{両側})\ \text{or} \sum_{y=y_{11}}^{\min\{n_{1.},n_{.1}\}} p(y)\ (\text{右片側}) \tag{11.74}$$

なお, (11.72) に対して,

$$\chi_c^2 = \frac{n\left(\left|Y_{11}Y_{22} - Y_{12}Y_{21}\right| - n/2\right)^2}{n_{1.}n_{2.}n_{.1}n_{.2}} \tag{11.75}$$

とする方法を Yates の連続性の補正といい, χ_c^2 を補正 χ^2 統計量という. ただし, $|Y_{11}Y_{22} - Y_{12}Y_{21}| \leq n/2$ のときは補正しない. 補正を行うと, χ^2 値が小さくなり, 検出力が当然下がるが, 有意確率が Fisher の直接確率に近づくといわれている.

尤度比検定

今度は, 分割表に 2 個の二項分布を仮定しよう. すなわち, 列に関する周辺度数を固定せず, Y_{i1} $(i = 1, 2)$ を二項分布 $B(n_{i.}, p_i)$ に従う確率変数とする. このとき, 対数尤度 $l(p_1, p_2)$ は

$$\sum_{i=1}^{2}\left\{\log{}_{n_i}\mathrm{C}_{y_{i1}} + y_{i1}\log p_i + y_{i2}\log(1 - p_i)\right\} \tag{11.76}$$

となる. 対立仮説 $p_1 \neq p_2$ のもとで,

$$\frac{\partial}{\partial p_i} l(p_1, p_2) = \frac{y_{i1}}{p_i} - \frac{y_{i2}}{1 - p_i} = 0$$

とおくと, 最尤推定量

$$\hat{p}_i = \frac{y_{i1}}{n_{i.}} \tag{11.77}$$

を得る.

結局, 対数尤度は,

$$l_1 = \sum_{i=1}^{2}\log{}_{n_i}\mathrm{C}_{y_{i1}} + \sum_{i=1}^{2}\sum_{j=1}^{2} y_{ij}\log y_{ij} - \sum_{i=1}^{2} n_{i.}\log n_{i.} \tag{11.78}$$

となる.

同様に, 帰無仮説 $p_1 = p_2 = p$ のもとでは, 最尤推定量は

$$\hat{p} = \frac{y_{.1}}{n}, \quad y_{.j} = \sum_{i=1}^{2} y_{ij} \tag{11.79}$$

であり, このときの対数尤度は次で表される.

$$l_0 = \sum_{i=1}^{2} \log {}_{n_i}\mathrm{C}_{y_{i1}} + \sum_{j=1}^{2} y_{.j} \log y_{.j} - n \log n \tag{11.80}$$

以上から, (10.27) により, 漸近的に

$$2\Big(\sum_{i=1}^{2}\sum_{j=1}^{2} Y_{ij} \log Y_{ij} - \sum_{i=1}^{2} n_{i.} \log n_{i.} - \sum_{j=1}^{2} Y_{.j} \log Y_{.j} + n \log n\Big) \tag{11.81}$$

が漸近的に χ_1^2 に従う. これが尤度比検定である.

ところで, 確率変数

$$e_{ij} = \frac{n_{i.}\, Y_{.j}}{n} \tag{11.82}$$

は, $Y_{.j}$ を固定した場合, 超幾何分布を仮定したときの Y_{ij} の平均に等しい. 尤度比検定の統計量は,

$$2 \sum_{i=1}^{2}\sum_{j=1}^{2} Y_{ij} \log \frac{Y_{ij}}{e_{ij}} \tag{11.83}$$

と表すことができる. 他方,

$$2\, Y_{ij} \log\Big\{1 + \frac{Y_{ij} - e_{ij}}{e_{ij}}\Big\} \simeq 2\, Y_{ij} \frac{Y_{ij} - e_{ij}}{e_{ij}} \simeq \frac{(Y_{ij} - e_{ij})^2}{e_{ij}}$$

を考えると, 尤度比検定と適合度の検定が漸近的に同等であることがわかる.

区間推定

■前向き研究　特性値が名義変数で, 二項分布に従う場合,

$$\psi = p_1 / p_2 \tag{11.84}$$

をリスク比という.

p_i をその最尤推定量 $\hat{p}_i$ で置き換え, 対数をとると,

$$\log \hat{\psi} = \log Y_{11} - \log Y_{21} - \log n_{1.} + \log n_{2.} \tag{11.85}$$

を得るが, この統計量は漸近的に正規分布に従う. 最尤推定量の一致性から, この統計量の平均は 0 であることがわかる.

(11.85) を, Y_{11}, Y_{21} の平均まわりで Taylor 展開すると,

$$\log \hat{\psi} \simeq \frac{1}{n_{1.} p_1}(Y_{11} - n_{1.} p_1) - \frac{1}{n_{2.} p_2}(Y_{21} - n_{2.} p_2) + \text{const}$$

となるので,

$$V[\log \hat{\psi}] = \frac{n_{1.} p_1(1 - p_1)}{n_{1.}^2 p_1^2} + \frac{n_{2.} p_2(1 - p_2)}{n_{2.}^2 p_2^2} \tag{11.86}$$

を得る. これに p_i の最尤推定値 $\hat{p}_i$ の実現値を代入, 整理すると

$$V\Big[\log \hat{\psi}\Big] = \frac{1}{y_{11}} + \frac{1}{y_{21}} - \frac{1}{n_{1.}} - \frac{1}{n_{2.}} \tag{11.87}$$

を得る. したがって, $\log \psi$ の $1 - \alpha$ 信頼限界は次で与えられる.

$$\log \frac{y_{11}\, n_{2.}}{y_{21}\, n_{1.}} \pm K_{\alpha/2} \sqrt{\frac{1}{y_{11}} + \frac{1}{y_{21}} - \frac{1}{n_{1.}} - \frac{1}{n_{2.}}} \tag{11.88}$$

■後ろ向き研究　後ろ向き研究では, 逆に, 結果 $j = 1$ と 2 で, 要因 $i = 1$ と 2 に差がみられるどうかを調べる. この場合も, 形式的には同じ分割表が得られるが, 母集団における $i = 1$ の場合と 2 の場合の比はわからず, 周辺和 $n_{i.}$ は意味を成さない. したがって, リスク比のかわりに, オッズ比を次で定義する.

$$\phi = \frac{p_1/(1 - p_1)}{p_2/(1 - p_2)} = \frac{p_1\,(1 - p_2)}{p_2\,(1 - p_1)} \tag{11.89}$$

この ϕ は

$$\hat{\phi} = \frac{Y_{11}\, Y_{22}}{Y_{12}\, Y_{21}} \tag{11.90}$$

で推定できる.

ところで, 要因 i と結果 j の組み合わせが得られる確率を p_{ij} とすると,

$$\phi = \frac{p_{11}/(p_{11}+p_{12})}{p_{12}/(p_{11}+p_{12})} \cdot \frac{p_{22}/(p_{21}+p_{22})}{p_{21}/(p_{21}+p_{22})} = \frac{p_{11}\,p_{22}}{p_{12}\,p_{21}} \tag{11.91}$$

が成り立つ. したがって, 結果 1 が得られるという条件のもとで要因が i である確率を q_i で表すと,

$$\phi = \frac{q_1}{1-q_1} \cdot \frac{1-q_2}{q_2} \tag{11.92}$$

であり, q_i をその最尤推定量 $Y_{i1}/n_{.1}$ で置き換えれば (11.90) を得る. 前向き研究の場合と, 行と列の役割が逆転しているので注意されたい. 対数オッズの信頼区間も (11.88) とまったく同様に得られるが, これは次章で述べる.

11.4　特性値が順序カテゴリの場合

特性値が連続変数でなく, 順序カテゴリの場合にも, 順位検定を適用することができる. 順序カテゴリとは, 量的な順序関係のみを表す変数である. 下表のような集計データを解析する場合は少なくない.

	著効	有効	不変	悪化	計
新薬	15	23	8	2	46
対照	8	20	12	5	45

次の分割表を仮定しよう. A_i は処理, B_j は実験で得られた順序カテゴリであり, Y_{ij} $(i=1,2\,;\, j=1,\cdots,b)$ は対応する頻度を表す確率変数である.

	B_1	B_2	$\cdots$	B_b	計
A_1	Y_{11}	Y_{12}	$\cdots$	Y_{1b}	$n_{1.}$
A_2	Y_{11}	Y_{22}	$\cdots$	Y_{2b}	$n_{2.}$
計	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.b}$	n

第 j カテゴリにスコア η_j $(\eta_1 \le \cdots \le \eta_b)$ を与え, 第 i 群におけるスコア η_j の荷重つき平均

$$U_i = \sum_{j=1}^{b} \eta_j \frac{Y_{ij}}{n_{i.}} \tag{11.93}$$

を考える.

Y_{ij} は多項超幾何分布に従うと仮定する. 周辺度数が固定されているので, $i = 1$ の場合を考えれば十分で, (4.47) (4.50) により,

$$\frac{(U_1 - \eta_j\, n_{.j}/n)^2}{\sum_j \sum_{j'} \eta_j \eta_{j'}\, n_{2.}\, n_{.j}\, (\delta_{jj'} n - n_{.j'})/n_{1.}\, n^2\, (n-1)} \tag{11.94}$$

が漸近的に χ^2_1 に従う. (11.94) は,

$$\frac{n^2\,(n-1)\left\{\sum_j \eta_j \left(Y_{1j} - n_{1.}\, n_{.j}/n\right)\right\}^2}{n_{1.}\, n_{2.}\left\{n\, \sum_j \eta_j^2\, n_{.j} - \left(\sum_j \eta_j\, n_{.j}\right)^2\right\}} \tag{11.95}$$

と変形できる.

この検定を Mantel 検定という. なお, (11.95) において分子の $n^2(n-1)$ を n^3 で置き換える場合を Cochran-Armitage の検定といい, 用量反応関係の検証に利用される.

Mantel 検定においては, もちろん, スコア η_j のつけかたによって検定の結果が変わってくる. 通常はカテゴリの平均順位を割り付けるのが無難であろう. この場合, Mantel 検定は Wilcoxon 検定と同等になる. また, 連続な特性値をカテゴライズした場合には, 各カテゴリの平均値や中央値をわりつけるのが妥当であろう.

11.5　対応のある場合

対応のある t 検定

データが対になっている場合は, 形式的に通常の t 検定を行うよりも, データの差を考えた方が検出力があがることが多い. 対になっている確率変数の組 (Y_{1j}, Y_{2j}) $(j = 1, \cdots, m)$ に対して,

$$Y_j = Y_{1j} - Y_{2j} \tag{11.96}$$

を考える. Y_j が正規分布 $N(\mu, \sigma^2)$ に従うと仮定しよう. Y_{1j} と Y_{2j} の分散は等しくなくてよい.

2 つの母集団の平均に差がないという帰無仮説は $\mu = 0$ と表現できるから,

$$T = \frac{\sqrt{m}\,\overline{Y}_m}{\sqrt{\sum_j (Y_j - \overline{Y}_m)^2/(m-1)}} \sim t_{m-1} \tag{11.97}$$

を用いて検定を構成できる. 標本平均 $\overline{Y}_m$ は (5.18) に定義した. これを対応のある t 検定という.

他方, 等分散を仮定した上で, データの対応を無視して, 通常の t 検定を行う場合は次を利用する.

$$T^* = \frac{\sqrt{m}\,\overline{Y}_m}{\sqrt{\sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2/(m-1)}} \sim t_{2m-2} \tag{11.98}$$

ここで,

$$\overline{Y}_{..} = \frac{1}{2\,m} \sum_{i=1}^{2} \sum_{j=1}^{m} Y_{ij}\,, \quad \overline{Y}_{.j} = \frac{1}{2} \sum_{i=1}^{2} Y_{ij} \tag{11.99}$$

とおくと, 次が成り立つことに注意する.

$$2 \sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2 = \sum_j (Y_j - \overline{Y}_m)^2 + 4 \sum_j (\overline{Y}_{.j} - \overline{Y}_{..})^2$$

(11.97) と (11.98) を比較すると, 分散の評価が, 対応を無視した場合,

$$\frac{4 \sum_j (\overline{Y}_{.j} - \overline{Y}_{..})^2}{m-1} \tag{11.100}$$

だけ大きい. 実際, 対応のある場合は, σ^2 の不偏推定量は

$$\frac{\sum_j (Y_j - \overline{Y}_m)^2}{m-1} \tag{11.101}$$

であり, 対応を無視した場合, 仮定した等分散 $\sigma^2/2$ の不偏推定量が

$$\frac{\sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2}{2\,m-2} \tag{11.102}$$

であるから, (11.100) は, 両者における分散の評価の差にほかならない. これを第 j 組のブロック効果という.

対応のある t 検定では, 差を考えることによりブロック効果を消去しているわけである. 最初からブロック効果のない場合には, 自由度の差

$$(2m-2)-(m-1)=m-1$$

の分だけ検出力の損失を招くが, 通常はそのデメリットよりブロック効果を取り除くメリットの方が大きいと考えるのである.

平均の差 Δ の区間推定も (11.97) から導かれ, その $1-\alpha$ 信頼限界は次で与えられる.

$$\overline{Y}_m \pm t_{m-1}(\alpha/2)\left\{\frac{\sum_j (Y_j-\overline{Y}_m)^2}{m\,(m-1)}\right\}^{1/2} \tag{11.103}$$

ノンパラメトリックな比較

■分布が対称な場合　対をなすデータの差 Y_j $(j=1,\cdots,m)$ を μ に関して対称な分布 F に従う確率変数とする. $Z_j=Y_j-\mu$ とおき, $\nu=m(m+1)/2$ 個の $j \le j'$ なる (j,j') のすべての組に対して, 大きさの順に $(Z_j+Z_{j'})/2+\mu$ を並べて, $W_{(1)}<\cdots<W_{(\nu)}$ を得るとする. ここで,

$$W_{(i_{jj'})}-\mu=\frac{1}{2}\,(Z_j+Z_{j'}) \tag{11.104}$$

とおくと, $W_{(s_\alpha)} \le \mu$ は, $Z_j+Z_{j'} \le 0$ となる (j,j') の組の個数が s_α 以上になることと同値である. この組の個数は, Mann-Whitney 型の Wilcoxon の符号順位検定の統計量 S を用いて, $\nu-S$ と表される. したがって,

$$s_\alpha=\frac{m\,(m+1)+2}{4}+K_{\alpha/2}\sqrt{\frac{m\,(m+1)\,(2m+1)}{24}} \tag{11.105}$$

とおくと, 分布の対称性から, μ の $1-\alpha$ 信頼区間は

$$W_{(\nu+1-s_\alpha)}<\mu<W_{(s_\alpha)} \tag{11.106}$$

となる. s_α を求める際, 期待値に 1/2 を加えたのは連続補正である.

タイが存在する場合は s_α を求める際に, (8.47) を用いて計算すればよい.

■分布が対称でない場合　分布が対称とは限らない一般の場合は, 符号検定に基づいて中央値 μ の区間推定ができる.

確率変数 $Y_1, \cdots, Y_m$ が互いに独立で, 中央値より大きい値をとる確率が 1/2, 小さい値をとる確率が 1/2 であると仮定する. また, s_α を次の式を満たすように定める.

$$\left(\frac{1}{2}\right)^m \sum_{i=s_\alpha}^{m} {}_m\mathrm{C}_i = \left(\frac{1}{2}\right)^m \sum_{i=0}^{s_\alpha} {}_m\mathrm{C}_i = \frac{\alpha}{2} \tag{11.107}$$

任意の α に対して s_α が存在するとは限らないので, s_α が存在するように α を定めておく必要がある. このとき, s_α 個以上の Y_j が μ より大きな値をとる確率は,

$$P(Y_{(s_\alpha)} > \mu) = \left(\frac{1}{2}\right)^m \sum_{i=s_\alpha}^{m} {}_m\mathrm{C}_i = \frac{\alpha}{2} \tag{11.108}$$

である. 同様に,

$$Y_{(m-s_\alpha)} < \mu \tag{11.109}$$

は, $m - s_\alpha$ 個以上の Y_j が μ より小さな値をとることと同値であり, この事象の起こる確率は, 次で与えられる.

$$\left(\frac{1}{2}\right)^m \sum_{i=m-s_\alpha}^{m} {}_m\mathrm{C}_i = \left(\frac{1}{2}\right)^m \sum_{i'=0}^{s_\alpha} {}_m\mathrm{C}_{m-i'} = \frac{\alpha}{2} \tag{11.110}$$

以上から, 中央値 μ の信頼係数 $1 - \alpha$ の信頼区間は,

$$Y_{(m-s_\alpha)} < \mu < Y_{(s_\alpha)} \tag{11.111}$$

で与えられる. なお, 符号検定では, タイが存在しても構わない. 0 の値をとる観測値だけ除いておけばよい.

第 12 章

一元配置

本章では, 2 標本問題を拡張して, a 個の処理間の特性値の違いを統計学的に考察する. この実験デザインで最も一般的なのは一元配置の分散分析であるが, 実際のアクションに結び付けることは難しい. さらに, $a \geq 3$ では, どの水準間に差があるかを考察する多重比較の問題や水準に自然な順序がついている場合の解析など, 2 標本問題にはない特殊性が存在する.

12.1 オムニバスな検定

あるひとつの要因に注目して, a 個の母集団を考える. それぞれの母集団における要因の特性を水準という. このとき, 第 i 水準における実験数を n_i として,

$$\sum_{i=1}^{a} n_i = n \tag{12.1}$$

個の実験の順番を完全にランダム化する計画を一元配置という. Y_{ij} を第 i 母集団からの n_i 個の無作為標本とし, 線型モデル

$$Y_{ij} = \mu_i + \varepsilon_{ij} \quad (i = 1, \cdots, a\,;\ j = 1, \cdots, n_i) \tag{12.2}$$

を仮定する.

μ_i は第 i 母集団における特性値の平均を表し, ε_{ij} は互いに独立に正規分布 $N(0, \sigma^2)$ に従う確率変数であるとする. このとき,

帰無仮説 : $\mu_1 = \cdots = \mu_a$
対立仮説 : 少なくとも 1 組の $1 \leq k \neq l \leq a$ に対し $\mu_k \neq \mu_l$

なる検定方式を考える. 実際にはこの対立仮説から得られる情報は少ないが, 後に議論する多重比較の基礎にもなっていることから, この検定を構成することから議論を始める.

一元配置分散分析

Y_{ij}, ε_{ij} を辞書順に並べて得られる n 次列ベクトルをそれぞれ $\boldsymbol{Y}, \boldsymbol{\varepsilon}$ とし, μ_i を同様に並べて得られる a 次列ベクトルを $\boldsymbol{\mu}$ とすると, (12.2) は

$$\boldsymbol{Y} = X\boldsymbol{\mu} + \boldsymbol{\varepsilon} \tag{12.3}$$

と行列表現できる. ここに, X は次のデザイン行列である.

$$X = \begin{bmatrix} 1 & \cdots & 1 & & & & \\ & & & \ddots & & & \\ 0 & \cdots & 0 & & 1 & \cdots & 1 \end{bmatrix}' \tag{12.4}$$

のうち上段右は $0 \cdots 0$:

$$X = \begin{bmatrix} 1 & \cdots & 1 & & 0 & \cdots & 0 \\ & & & \ddots & & & \\ 0 & \cdots & 0 & & 1 & \cdots & 1 \end{bmatrix}'$$

つぎに, a 次直交行列

$$P' = \begin{bmatrix} \boldsymbol{n}' \\ Q_A' \end{bmatrix}, \quad \boldsymbol{n} = (\sqrt{n_1/n}, \cdots, \sqrt{n_a/n})' \tag{12.5}$$

を考える. また, X の各列ベクトルは直交するから,

$$Q'X = O, \quad Q'Q = I_{n-a} \tag{12.6}$$

なる $n-a$ 行 n 列行列 Q' が存在する.

ここで, Γ' を

$$\begin{bmatrix} P' \operatorname{diag}(1/\sqrt{n_i})\, X' \\ Q' \end{bmatrix} = \begin{bmatrix} (1/\sqrt{n})\, \boldsymbol{j}_a' \\ Q_A' \operatorname{diag}(1/\sqrt{n_i})\, X' \\ Q' \end{bmatrix} \tag{12.7}$$

で定めると, これは n 次直交行列となる.

そこで，

$$\boldsymbol{Z} = \Gamma' \boldsymbol{Y} \tag{12.8}$$

と直交変換すると，$\boldsymbol{Z}$ は n 変量正規分布に従い，

$$E[\boldsymbol{Z}] = \begin{bmatrix} E[Z_\mu] \\ E[\boldsymbol{Z}_A] \\ E[\boldsymbol{Z}_e] \end{bmatrix} = \begin{bmatrix} \sum_i n_i \mu_i / \sqrt{n} \\ Q'_A \operatorname{diag}(\sqrt{n_i}) \boldsymbol{\mu} \\ \boldsymbol{o} \end{bmatrix} \tag{12.9}$$

$$V[\boldsymbol{Z}] = \sigma^2 I \tag{12.10}$$

が成り立つ．$\boldsymbol{Z}$ は Γ' に応じてブロック分割した．

一般に，平均 $\boldsymbol{\mu}$ に対し，

$$C = \boldsymbol{\lambda}' \boldsymbol{\mu}, \quad (\boldsymbol{\lambda}' \boldsymbol{j} = 0) \tag{12.11}$$

となる C を μ_i $(i = 1, \cdots, a)$ に関する対比という．

対比に関する仮説 $C = 0$ を考えると，任意の水準間の平均の差の検定を構成できる．実際，

$$\boldsymbol{\lambda} = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0, \cdots, 0 \right)' \tag{12.12}$$

とすれば，C は μ_1 と μ_2 の差を表す．もっとも，対比の中には，

$$\frac{\mu_1}{2} + \frac{\mu_2}{2} - \frac{\mu_3}{\sqrt{2}}$$

などの意味不明のものも含まれ，仮説 $C = 0$ を機械的に検定すると検出力の低下を招く．

(12.5) は直交行列であったから，

$$\boldsymbol{\mu}_C = Q'_A \operatorname{diag}(\sqrt{n_i}) \boldsymbol{\mu} \tag{12.13}$$

は μ_i $(i = 1, \cdots, a)$ に関する対比であり，この最小二乗推定量は，

$$\boldsymbol{Z}_A = Q'_A \operatorname{diag}(\sqrt{n_i}) (X'X)^{-1} X' \boldsymbol{Y} = Q'_A \operatorname{diag}(1/\sqrt{n_i}) X'Y \tag{12.14}$$

と表される．$\boldsymbol{Z}_A$ の分布は $a - 1$ 変量正規分布であり，その平均は $\boldsymbol{\mu}_c$，分散は $\sigma^2 I_{a-1}$ である．

したがって，

$$S_A = \| \boldsymbol{Z}_A \|^2 \tag{12.15}$$

とおくと，S_A/σ^2 は非心度

$$\| \boldsymbol{\mu}_C \|^2 = \sum_{i=1}^{a} \frac{n_i\,(\mu_i - \overline{\mu})^2}{\sigma^2}\,, \quad \overline{\mu} = \sum_{i=1}^{a} \frac{n_i\,\mu_i}{n} \tag{12.16}$$

の非心 χ^2 分布に従う．帰無仮説のもとでは非心度は 0 となることに注意しよう．

ここで，

$$\overline{Y}_{..} = \sum_{i=1}^{a}\sum_{j=1}^{n_i} \frac{Y_{ij}}{n}\,, \quad \overline{Y}_{i.} = \sum_{j=1}^{n_i} \frac{Y_{ij}}{n_i} \tag{12.17}$$

とおくと，

$$S_A = \sum_{i=1}^{a} n_i\,(\overline{Y}_{i.} - \overline{Y}_{..})^2 \tag{12.18}$$

と書き下すことができる．これは要因の効果にほかならない．また，

$$\| \boldsymbol{Y} \|^2 - \| \boldsymbol{Z}_\mu \|^2 = \sum_{i=1}^{a}\sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{..})^2 \tag{12.19}$$

を S_T と表す．

これから要因効果 S_A を減じた

$$S_e = \| Q'\boldsymbol{Y} \|^2 = \sum_{i=1}^{a}\sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i.})^2 \tag{12.20}$$

が誤差である．この各成分は，互いに独立に正規分布 $N(0, \sigma^2)$ に従うから，S_e/σ^2 は χ^2_{n-a} に従う．$\mathit{\Gamma}'$ は直交行列であるから，S_e, S_A は互いに独立で，

$$W = \frac{S_A/(a-1)}{S_e/(n-a)} \sim F^{a-1}_{n-a} \tag{12.21}$$

が成り立つ．この検定を一元配置の分散分析 ANalysis Of VAriance (略して ANOVA) という．以上のことを分散分析表といわれる下の表のようにまとめることが多い．

要因	平方和	自由度	平方平均	F 統計量
処理	S_A	$a-1$	$S_A/(a-1)$	W
誤差	S_e	$n-a$	$S_e/(n-a)$	
計	S_T	$n-1$		

とくに, 2 標本の場合, t 統計量 (11.4) の二乗は, 帰無仮説のもとで

$$T^2 = \frac{\sum_i n_i(\overline{Y}_{i.} - \overline{Y}_{..})^2}{\{\sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2\}/(n-2)} \tag{12.22}$$

と変形できる. これは (12.21) において $a=2$ とおいたものにほかならず, 一元配置の分散分析は t 検定を拡張した解析と位置づけできる.

等分散性の検定

分散分析は, 正規性と等分散性を仮定した. t 検定と同様に, 漸近的にはノンパラメトリックな検定と同等であるが, これは検出力を保証するものではない. 2 標本の場合は分散比 σ_1^2/σ_2^2 の分布を直接比較できたが, 3 標本以上の場合は, 尤度比検定を用いる.

$$\frac{(n_i-1)\,\hat{\sigma}_i^2}{\sigma^2} \stackrel{\text{def}}{=} \sum_{j=1}^{n_i} \frac{(Y_{ij} - \overline{Y}_{i.})^2}{\sigma^2} \sim \chi^2_{n_i-1} \tag{12.23}$$

より, $\hat{\sigma}_1^2, \cdots, \hat{\sigma}_a^2$ の同時密度関数は, $n_i - 1 = m_i$ として,

$$\prod_i \frac{m_i/\sigma_i^2}{\Gamma(m_i/2)2^{m_i/2}} \left(\frac{m_i\,\hat{\sigma}_i^2}{\sigma_i^2}\right)^{(m_i-1)/2} \exp\left(-\frac{m_i\,\hat{\sigma}_i^2}{2\,\sigma_i^2}\right) \tag{12.24}$$

となる.

対数尤度

$$l(\sigma_1^2, \cdots, \sigma_a^2) = \text{const} - \sum_i \frac{m_i}{2} \log \sigma_i^2 - \sum_i \frac{m_i\,\hat{\sigma}_i^2}{2\sigma_i^2} \tag{12.25}$$

を σ_i^2 で偏微分して 0 とおくと, σ_i^2 の最尤推定量 $\hat{\sigma}_i^2$ を得る.

他方, (12.25) において $\sigma_1^2 = \cdots = \sigma_a^2 = \sigma^2$ とおいたときの最尤推定量は

$$\hat{\sigma}^2 = \sum_i m_i \hat{\sigma}_i^2 \Big/ \sum_i m_i \tag{12.26}$$

である. 以上から, 尤度比検定

$$-2\log\lambda = 2\Big\{ l(\hat{\sigma}_1^2, , \cdots, \hat{\sigma}_a^2) - l(\hat{\sigma}^2, \cdots, \hat{\sigma}^2) \Big\} \sim \chi^2_{a-1} \tag{12.27}$$

を得る. 実際には近似の精度を上げるため,

$$(-2\log\lambda) \Big/ \left\{ 1 + 3\left(\sum_i \frac{1}{m_i} - \frac{1}{\sum_i m_i} \right) \right\} \Big/ 3(a-1) \tag{12.28}$$

と統計量を補正する. これが Bartlett の検定である.

Kruskal-Wallis 検定

分布に正規性が仮定できないときは, a 個の水準すべてをプールしたときの Y_{ij} の順位を R_{ij} とし, そのスコア $s(R_{ij})$ の第 i 水準におけるスコアの平均

$$U_i = \frac{1}{n_i} \sum_{j=1}^{n_i} s(R_{ij}) \tag{12.29}$$

を考える. $\boldsymbol{U} = (U_1, \cdots, U_a)'$ とおくと, 漸近的に $\boldsymbol{U}$ はスコアに関する適当な条件のもとで多変量正規分布に従う. したがって,

$$W = \Big(\boldsymbol{U} - E[\boldsymbol{U}] \Big)' V[\boldsymbol{U}]^- \Big(\boldsymbol{U} - E[\boldsymbol{U}] \Big) \tag{12.30}$$

が漸近的に χ^2_{a-1} に従う. この検定を Q 検定という.

Y_{ij} を辞書順に並べて Y_l とし, c_i を標本が母集団 i からのものであれば 1 を, それ以外では 0 をとるダミー変数とすると系 8.2 より,

$$E[U_i] = \frac{n_i}{n} \sum_{l=1}^{n} s(l) \tag{12.31}$$

$$V[U_i] = \frac{n_i\,(n-n_i)}{n\,(n-1)} \sum_{l=1}^{n} \Big\{ s(l) - \bar{s} \Big\}^2 \tag{12.32}$$

を得る. 共分散は, (8.32) で与えられる.

$s(l) = l$, すなわち, スコアを順位とした場合, タイがなければ,

$$W = \frac{12}{n(n+1)} \sum_{i=1}^{a} n_i \left(\frac{1}{n_i} \sum_{j=1}^{n_i} R_{ij} - \frac{n+1}{2} \right)^2 \tag{12.33}$$

と書き下すことができる. これを Kruscal-Wallis 検定 という. タイがある場合は分散を (8.47) に応じて補正すればよい.

ほかにも, 中央値や van der Waerden スコアを利用した Q 検定も同様に構成できる. 母分布の違いにより検出力は各手法で違ってくるが, その特性は 2 標本問題と同様である.

12.2　多重比較

分散分析や Kruskal-Wallis 検定は, 平均の一様性に関するオムニバスな検定であった. しかし, 現実の問題として, どの水準とどの水準に差があるかが分からないと実際のアクションにつながりにくい.

そこで, ごく素朴に, 帰無仮説を

$$H_{ii'} : \mu_i = \mu_{i'}$$

として,

$$\nu = \frac{a\,(a-1)}{2} \tag{12.34}$$

個の (i, i') の組に対して有意水準 α で t 検定や Wilcoxon 検定を繰り返せばよいように思われる. しかし, これでは第一種の誤りの確率が α を大きく超えてしまうことが多い.

実際, $H_{ii'}$ に対する棄却域を $W_{ii'}$ とすると, 帰無仮説のもとで,

$$P(T \in \cup W_{ii'}) \geq P(T \in W_{ii'}) = \alpha \tag{12.35}$$

が成立する. T は検定統計量である. 等号は ν 個すべての $W_{ii'}$ が一致するときにのみ成立する. 以上のことを総括して検定の多重性の問題という.

そこで, 検定全体の有意水準を α に保つようにひとつひとつの検定を構成するように工夫しなければならない. このように多重性を考慮した検定を多重検定という.

Bonferroni の方法

いま, 帰無仮説 H が ν 個の仮説 H_l $(l = 1, \cdots, \nu)$ を用いて,

$$H_1 \cap \cdots \cap H_\nu$$

と表されるとする. H_l に対する棄却域を W_l とすると, H に対する第一種の誤りの確率を

$$P(W_1 \cup \cdots \cup W_\nu \,|\, H_1 \cap \cdots \cap H_\nu) \leq \alpha \tag{12.36}$$

となるように $W_1, \cdots, W_\nu$ を定めなければならない. このような $W_1, \cdots, W_\nu$ を有意水準 α の多重検定の棄却域という.

Bonferroni の方法は, すべての H_l に対して,

$$P\Big(T \in W_l \Big| H_l \Big) \leq \frac{\alpha}{\nu} \tag{12.37}$$

となるように棄却域を定めておけば, (12.36) が満たされることを利用した検定である. 実際, Bonferroni の不等式

$$P\Big(T \in \cup W_l \Big) \leq \sum_l P\Big(T \in W_l \Big) \tag{12.38}$$

より,

$$P\Big(T \in \cap W_l \Big| H \Big) \leq \sum_l P\Big(T \in W_l \Big| H \Big) \leq \sum_l \sup P\Big(T \in W_l \Big| H_l \Big)$$

が成り立つから, (12.37) を得る.

棄却域 W_l $(l = 1, \cdots, \nu)$ が排反でないときには, Bonferroni の方法は相当に保守的である. この点を改良した多重検定もいくつか提案されている.

ここで, 多重検定における有意確率を定義しておこう. 棄却域 W_l が, 検定統計量 T_l に対して, $T_l > c_l$ という形で与えられるときは, T_l の観測値 t_l に対して,

$$p_l = \sup P\Big(T_l > t_l \Big| H_l \Big) \tag{12.39}$$

を t_l の有意確率という.

Holm の方法

Holm の方法は最小の有意確率 $\min p_l$ のみ α/ν と比較し, 2 番目から先の有意確率は少し緩めの有意水準と比較する. $p_1, \cdots, p_\nu$ を小さい順に $p_{(1)}, \cdots, p_{(\nu)}$ とし, $p_{(l)}$ に対応する帰無仮説を $H_{(l)}$ とおく. このとき, $p_{(l)} > \alpha\Big/(\nu + 1 - l)$ となる最小の l を l^* として, $H_{(1)}, \cdots, H_{(l^*-1)}$ を棄却し, $H_{(l^*)}, \cdots, H_{(\nu)}$ を採択すると, 全体の有意水準が α 以下になる.

■証明　ν 個の仮説のうち, 真である仮説の集合を M, M の要素数を m とする. これらの真である仮説をすべて採択する確率

$$P\Big(p_l > \alpha/m, {}^\forall H_l \in M\Big) = 1 - P\Big(p_l \leq \alpha/m, {}^\exists H_l \in M\Big)$$

は, 明らかに,

$$1 - \sum_{H_l \in M} P\Big(p_l \leq \alpha/m\Big) \geq 1 - m \cdot \frac{\alpha}{m} = 1 - \alpha$$

より小さくない. ところが,

$$p_l > \frac{\alpha}{m}, {}^\forall H_l \in M \implies p_{(l)} > \frac{\alpha}{m}, {}^\forall l \geq \nu - m + 1$$

であり,

$$p_{(\nu)} \geq \cdots \geq p_{(\nu-m+1)} > \frac{\alpha}{m} = \frac{\alpha}{\nu - (\nu - m + 1) + 1}$$

より, $l^* \leq n - m + 1$ であるから, $H_l \in M$ はすべて採択される.

以上から, 真である仮説の少なくともひとつを棄却する確率は α 以下となる. ■

Shaffer の方法

Holm の方法にはまだ無駄がある. たとえば, 3 水準の場合, $\mu_1 \neq \mu_2$ であれば, $\mu_2 = \mu_3$ と $\mu_3 = \mu_1$ は同時には成り立たない.

したがって, 帰無仮説 $\mu_1 = \mu_2$ が最初に棄却されたときは, 次の検定の水準を $\alpha/2$ にする必要はなく, α でよいように思われる. Shaffer の方法はこの点において, Holm の方法にさらに改良を加えたものである.

$H_1, \cdots, H_\nu$ のうち, $H_{(1)}, \cdots, H_{(l-1)}$ が偽であるときに, 真でありうる仮説の最大個数を m_l とし, $p_{(l)} > \alpha/m_l$ となる最小の l を l^* とする. このとき, $H_{(1)}, \cdots, H_{(l^*-1)}$ を棄却し, $H_{(l^*)}, \cdots, H_{(\nu)}$ を採択すると, 真である仮説の少なくともひとつを棄却する確率は α 以下となる.

特性値が正規分布に従う場合

以上, Bonferroni の方法およびこれを改良した多重検定は統計量の分布に何の仮定も置かないものであった. 今度は, 特性値が正規分布に従う場合に多重検定を構成しよう.

一元配置分散分析モデル (12.3) を仮定し, 分散は未知であるとして, その不偏推定量

$$\hat{\sigma}^2 = \frac{S_e}{n-a} \tag{12.40}$$

で代用する. (12.40) は notation の違いを除いて (9.34) と一致するので, σ^2 の最良不偏推定量である.

Tukey 法

ν 個の $H_{ii'}$ すべてに興味がある場合, 第 i 群と第 i' 群の差を統計量

$$T_{ii'} = \frac{\overline{Y}_{i.} - \overline{Y}_{i'.}}{\sqrt{1/n_i + 1/n_{i'}}\hat{\sigma}} \tag{12.41}$$

で評価して,

$$T_{\max} = \max\left| T_{ii'} \right| \tag{12.42}$$

がその観測値 t より大きい値をとる確率を直接計算することができる. これを Tukey の方法という.

この有意確率は,

$$T_{\max} \le t \Longleftrightarrow \left|T_{12}\right| \le t, \cdots, \left|T_{(a-1)a}\right| \le t \tag{12.43}$$

となる確率を 1 から減じた値にほかならない.

帰無仮説で仮定した共通の平均は 0 とおいても一般性を失わないから, $\overline{Y}_{i.}$ が最大であると仮定すると, (12.43) は

$$\sqrt{\frac{n_{i'}}{n_i}}\,\frac{\sqrt{n_i}\,\overline{Y}_{i.}}{\sigma} - t\,\sqrt{1+\frac{n_{i'}}{n_i}}\,\frac{\hat{\sigma}}{\sigma} \le \frac{\sqrt{n_{i'}}\,\overline{Y}_{i'.}}{\sigma} \le \sqrt{\frac{n_{i'}}{n_i}}\,\frac{\sqrt{n_i}\,\overline{Y}_{i.}}{\sigma}$$

がすべての $i'(\neq i)$ について成立することと同値である.

$$U = \frac{\sqrt{n_i}\,\overline{Y}_{i.}}{\sigma} \sim N(0,1)\,, \qquad V = \frac{\hat{\sigma}}{\sigma} \sim \sqrt{\chi^2_{n-a}/(n-a)}$$

を考えると, $U = u, V = v$ を与えたときの上の不等式が成立する条件つき確率 $p_{ii'}(u,v)$ は,

$$\Phi\left(\sqrt{\frac{n_{i'}}{n_i}}\,u\right) - \Phi\left(\sqrt{\frac{n_{i'}}{n_i}}\,u - t\,\sqrt{1+\frac{n_{i'}}{n_i}}\,v\right) \tag{12.44}$$

で与えられる. ただし, $\Phi(y)$ は標準正規分布の分布関数である.

つぎに, u, v を動かすと, $p_{ii'}(u,v)$ の期待値は,

$$p_{ii'} = \int_0^{\infty}\int_{-\infty}^{\infty} p_{ii'}(u,v)\,\phi(u)\,f(v)\,du\,dv \tag{12.45}$$

となる. ただし, $\phi(u)$, $f(v)$ はそれぞれ U, V の密度関数である.

$\overline{Y}_{i.}$ が最大であるという a 個の事象は互いに排反であるから, 求める有意確率は

$$1 - \sum_{i=1}^{a}\int_0^{\infty}\int_{-\infty}^{\infty}\prod_{i'\neq i} p_{ii'}(u,v)\,\phi(u)\,f(v)\,du\,dv \tag{12.46}$$

となる. 実際の検定手順においては, 有意確率が有意水準 α と一致するような t_α を求め, $T_{ii'}$ がこの t_α より大きい場合, この水準間に差があると判断すればよい.

Scheffe 法

Scheffe の方法は, 任意の対比に関する帰無仮説

$$C = \sum_{i=1}^{a} \lambda_i \mu_i = 0 \quad \left(\sum_{i=1}^{a} \lambda_i = 0 \right) \tag{12.47}$$

に対する多重検定である. 既に述べたように, 対比の中には興味のないものまで含まれるので検出力は高くない.

まず, 統計量を

$$\sum_{i=1}^{a} \lambda_i \overline{Y}_{i.} = \sum_{i=1}^{a} \lambda_i (\overline{Y}_{i.} - \overline{Y}) \tag{12.48}$$

とする.

このとき, Cauchy-Schwarz の不等式

$$\left(\frac{1}{\hat{\sigma}} \sum_{i=1}^{a} \lambda_i \overline{Y}_{i.} \right)^2 \leq \sum_{i=1}^{a} \frac{\lambda_i^2}{n_i} \sum_{i=1}^{a} \frac{n_i (\overline{Y}_{i.} - \overline{Y})^2}{\hat{\sigma}^2} \tag{12.49}$$

および (12.21) より,

$$\sum_{i=1}^{a} n_i \frac{(\overline{Y}_{i.} - \overline{Y})^2}{(a-1)\,\hat{\sigma}^2} \sim F_{n-a}^{a-1} \tag{12.50}$$

を用いて, 棄却域を

$$\left| \sum_{i=1}^{a} \lambda_i \overline{Y}_{i.} \right| > \left\{ (a-1)\, f_{n-a}^{a-1}(\alpha) \sum_i \left(\lambda_i^2 \big/ n_i \right) \right\}^{1/2} \hat{\sigma} \tag{12.51}$$

とすれば, p 値を保守的に評価することができる. ここに, $f_{n-a}^{a-1}(\alpha)$ は F_{n-a}^{a-1} の上側 α 点である. これを Scheffe の方法という.

また, 以上のことから, C の $1-\alpha$ 同時信頼区間は

$$\sum_{i=1}^{a} \lambda_i \overline{Y}_{i.} \pm \left\{ (a-1)\, f_{n-a}^{a-1}(\alpha) \sum_i \left(\lambda_i^2 \big/ n_i \right) \right\}^{1/2} \hat{\sigma}^2 \tag{12.52}$$

で与えられる.

Dunnett 法

対照とその他すべての水準との差を検定したい場合も同様の考え方で検定を構成できる. これを Dunnett の方法という.

対照を第 1 水準として, 両側検定であれば,

$$T_{i1} = \left| \frac{\overline{Y}_{i.} - \overline{Y}_{1.}}{\sqrt{1/n_i + 1/n_1}\hat{\sigma}} \right| \leq t \tag{12.53}$$

がすべての $i(\neq 1)$ について成立する確率を 1 から減じると有意確率が得られる. t は統計量 $T = \max_i T_{i1}$ の観測値である.

この確率は,

$$1 - \int_0^{\infty}\int_{-\infty}^{\infty} \prod_{i=2}^{a} p(u,v)\,\phi(u)\,f(v)\,du\,dv \tag{12.54}$$

で与えられる. ただし, $p(u,v)$ は,

$$\Phi\left(\sqrt{\frac{n_i}{n_1}}\,u + t\,\sqrt{1+\frac{n_i}{n_1}}\,v\right) - \Phi\left(\sqrt{\frac{n_i}{n_1}}\,u - t\,\sqrt{1+\frac{n_i}{n_1}}\,v\right) \tag{12.55}$$

である.

右片側検定であれば, (12.53) のかわりに

$$\frac{\overline{Y}_{i.} - \overline{Y}_{1.}}{\sqrt{1/n_i + 1/n_1}\hat{\sigma}} \leq t \tag{12.56}$$

とすればよい. この有意確率は,

$$1 - \int_0^{\infty}\int_{-\infty}^{\infty} \prod_{i=2}^{a} \Phi\left(\sqrt{\frac{n_i}{n_1}}\,u + t\,\sqrt{1+\frac{n_i}{n_1}}\,v\right) \phi(u)\,f(v)\,du\,dv \tag{12.57}$$

で与えられる.

$\Phi(u), \phi(u), f(v)$ は Tukey の方法の項で定義した.

多重検定方式

Tukey や Dunett の多重比較で, 大きさが 2 番目以下の実測値に対しても, 最大のものと同じ基準を用いたのでは明らかに保守的に過ぎる. そこで, 真である帰無仮説を誤って棄却する確率を α 以下に押さえつつ, 大きさが 2 番目以下の実測値に対しては少しずつ基準を緩くしていくことを考える. このような検定方式は次の定理を利用して構成することができる.

■定理 12.1　帰無仮説 $H_1, H_2, \cdots$ に対する検定を考える. 任意の仮説 H_i に対し, 1 個以上の他の任意の仮説との積 $H_i \cap H_j \cap \cdots$ をそれぞれ有意水準 α で検定し, これらがすべて棄却されたときはじめて H_i を有意水準 α で検定するとすると, 真である帰無仮説の少なくともひとつを棄却する確率は α 以下である. (Mercus-Perits-Gabriel)

■証明　事象 A, B を次のように定める.

A : 任意の真である仮説 H_i が棄却される.
B : 積 $H_i \cap H_j \cap \cdots$ が棄却される.

このとき, 事象 A が起こるためには, まず B が起きなくてはならないから, $A = A \cap B$ である. したがって,

$$P(A) = P(A \cap B) = P(B) \cdot P\left(B \middle| A \right) \leq \alpha \cdot 1 = \alpha$$

が成り立つ. ■

Mercus-Perits-Gabriel の定理を用いて構成される検定を有意水準 α の多重検定方式という. 既に述べたように, Tukey や Dunnett の検定にもこの方式を応用することができる. たとえば, Dunnett の右片側検定であれば, 以下のように修正すればよい.

まず, 実測値 t_{i1} を大きさの順に大きい方から[*1]並べ, $t_{(1)}, \cdots, t_{(a-1)}$ とする. つぎに, $a-i$ 水準のときの Dunnett の右片側検定の有意確率が α となるような基準 $d_{a-i}(\alpha)$ と $t_{(i)}$ を比較し, $t_{(i)}$ の方が大きくなるような最大の i を i^* とする.

このとき, $\mu_{(1)}, \cdots, \mu_{(i^*)}$ は μ_1 より大きく, 他は有意差なしと判定する. 結局, T_{i1} を大きい順に検定を行い, はじめて有意でないと判定された時点で検定を中止すればよい. このとき, 水準数は検定を進めるごとに 1 ずつ減らしていくが, 分散の推定量は a 水準のときのまま (12.40) としてよい.

両側検定の場合はより複雑である. 3 水準の場合に Tukey の方法を用いるのであれば, t_{12}, t_{13}, t_{23} のうち, 絶対値が最大のものを Tukey 法で検定し, これが有意であれば, 残り 2 個を通常の両側 t 検定で比較すればよい. しかし, 4 水準以上の場合はこれほど単純ではない.

なお, 特性値が正規分布に従わない場合は順序カテゴリの場合と同一の議論となる.

12.3　特性値が名義変数の場合

特性値が名義変数の場合は一元配置のデータは分割表の形にまとめられる.

		反　応					計
		1	$\cdots$	j	$\cdots$	b	
水準	1	Y_{11}	$\cdots$	Y_{1j}	$\cdots$	Y_{1b}	$n_{1.}$
	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	i	Y_{i1}	$\cdots$	Y_{ij}	$\cdots$	Y_{ib}	$n_{i.}$
	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	a	Y_{a1}	$\cdots$	Y_{aj}	$\cdots$	Y_{ab}	$n_{a.}$
	計	$Y_{.1}$	$\cdots$	$Y_{.j}$	$\cdots$	$Y_{.b}$	n

[*1] 通常の順序統計量と異なり大きい方から並べていることに注意

水準に順序のない場合

要因を a 水準, 反応を b 水準として, 要因が第 i 水準, 反応が第 j 水準のときの頻度に対応する確率変数を Y_{ij} とする. 要因の第 i 水準の度数の周辺和 n_i は固定されているものとする. 当面, 要因側の水準には順序がついていないものとしよう. このとき, 分割表は上の表のようになる.

Pearson の適合度の検定

列に関する周辺度数も $Y_{.j} = n_{.j}$ と固定して考えると, Y_{ij} は多項超幾何分布に従い, 漸近的に

$$W = \left(\boldsymbol{Y} - E[\boldsymbol{Y}]\right)' V[\boldsymbol{Y}]^{-} \left(\boldsymbol{Y} - E[\boldsymbol{Y}]\right) \sim \chi^2_{(a-1)(b-1)} \tag{12.58}$$

が成り立つ. ここに, $\boldsymbol{Y} = (Y_{11}, \cdots, Y_{ab})'$ である. (12.58) に基づく検定を Pearson の適合度の検定という.

χ^2 分布の自由度は

$$n - a - b - 1 = (a-1)(b-1)$$

である. これは次の制約による.

$$\sum_{j=1}^{b} Y_{ij} = n_{i.}\,, \quad \sum_{i=1}^{a} Y_{ij} = n_{.j}\,, \quad \sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij} = n \tag{12.59}$$

(4.47) と

$$V[\boldsymbol{Y}]^{-} = (n-1)\,\mathrm{diag}\,(n_{i.}^{-1}) \otimes \mathrm{diag}\,(n_{.j}^{-1}) \tag{12.60}$$

より, (12.58) を書き下すと,

$$W \simeq \frac{n}{n-1}\,W = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{(Y_{ij} - n_{i.}\,n_{.j}/n)^2}{n_{i.}n_{.j}/n} \tag{12.61}$$

となる. これは (11.72) と同じ形式であることに注意しよう.

Fisher の直接確率法

2×2 表の場合と同様, セルの平均 $n_{i.}\,n_{.j}/n$ が 5 以下になるセルがある場合, 適合度の検定の χ^2 近似の精度は悪い. このような場合は, 多項超幾何分布の仮定のもとで分割表の得られる確率

$$\prod_i n_{i.}\,!\prod_j n_{.j}\,!\Big/ n\,!\prod_i\prod_j y_{ij}\,! \tag{12.62}$$

を求め, 実際のデータの得られる確率が (12.62) 以下となるようなすべての場合についての総和を有意確率とすればよい. この方法を Fisher の直接確率法という. この場合は両側確率のみが得られ, 片側確率を求めることはできない.

尤度比検定

確率分布として次のような a 個の多項分布を仮定する. 多項超幾何分布と多項分布は漸近的に同等であることに注意しよう.

		反　応					
		1	$\cdots$	j	$\cdots$	b	計
	1	p_{11}	$\cdots$	p_{1j}	$\cdots$	p_{1b}	1
水	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	i	p_{i1}	$\cdots$	p_{ij}	$\cdots$	p_{ib}	1
準	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	a	p_{a1}	$\cdots$	p_{aj}	$\cdots$	p_{ab}	1

ただし, 制約

$$\sum_{j=1}^{b} p_{ij} = 1 \quad (i = 1, \cdots, a) \tag{12.63}$$

があるので, 本質的なパラメタ数は $a\,(b-1)$ である.

このとき, 対数尤度は定数項を除いて次のように表される.

$$\sum_{i=1}^{a}\left\{\sum_{j=1}^{b-1} y_{ij}\ \log p_{ij} + y_{ib}\ \log(1 - p_{i1} - \cdots - p_{ib-1})\right\} \tag{12.64}$$

対立仮説のもとでの $p_{ij}\,(j = 1, \cdots, b-1)$ の最尤推定量 $\hat{p}_{ij}$ を求めよう. それには, 上の式を p_{ij} で偏微分して 0 とおくと,

$$\frac{y_{ij}}{p_{ij}} - \frac{y_{ib}}{1 - p_{i1} - \cdots - p_{ib-1}} = 0 \Longleftrightarrow \frac{y_{ij}}{p_{ij}} = \frac{y_{ib}}{p_{ib}} \tag{12.65}$$

を得る. これと

$$\sum_{j=1}^{b} y_{ij} = n_{i.} \tag{12.66}$$

より, p_{ij} の最尤推定量

$$\hat{p}_{ij} = y_{ij}/n_{i.} \tag{12.67}$$

を得る.

他方, 帰無仮説 $p_{1j} = \cdots = p_{aj} \stackrel{\text{def}}{=} p_j\ (j = 1, \cdots, b-1)$ のもとでの p_j の最尤推定量は, 同様にして, 次のようになる.

$$\hat{p}_j = n_{.j}/n \tag{12.68}$$

以上から, 尤度比検定の統計量は

$$-2\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}\left(\log\frac{Y_{.j}}{n} - \log\frac{Y_{ij}}{n_{i.}}\right) \tag{12.69}$$

となる. この検定量は, 両仮説のパラメタの次元の差

$$a\,(b-1) - (b-1) = (a-1)\,(b-1) \tag{12.70}$$

を自由度とする χ^2 分布に従う.

帰無仮説のもとで, Y_{ij} の平均は $n_{i.}\,p_j$ である. p_j を $\hat{p}_j$ で代用すると, 検定統計量は (11.83) と同じ形式になる. また, 尤度比検定と適合度の検定が漸近的に同等であるのも 2×2 表の場合と同じである.

なお, $b = 2$ の場合, すなわち特性値が二項分布に従う場合[*2]を除いては, 多重検定は解釈が難しい.

区間推定

2×2 表の場合と同様に, オッズ比

$$\phi_{ij} = \frac{p_{ij}\,p_{ab}}{p_{ib}\,p_{aj}} \quad (i = 1, \cdots, a-1\,;\ j = 1, \cdots, b-1)$$

の推定量として,

$$\hat{\phi}_{ij} = \frac{Y_{ab}\,Y_{ij}}{Y_{aj}\,Y_{ib}} \tag{12.71}$$

を考える. (11.88) と同様にして, ϕ_{ij} の $1-\alpha$ 信頼限界

$$\frac{y_{ab}\,y_{ij}}{y_{aj}\,y_{ib}} \exp\left\{\pm K_{\alpha/2}\left(\frac{1}{y_{ab}} + \frac{1}{y_{aj}} + \frac{1}{y_{ib}} + \frac{1}{y_{ij}}\right)^{1/2}\right\} \tag{12.72}$$

を得る.

12.4　特性値が順序カテゴリの場合

特性値が順序カテゴリの場合, 第 j カテゴリにスコア η_j $(\eta_1 \leq \cdots \leq \eta_b)$ を与え, 第 i 群における平均から構成される a 次元確率変数

$$\boldsymbol{U} = (U_1, \cdots, U_a)'\,, \quad U_i = \frac{1}{n_{i.}}\,\eta_j\,Y_{ij} \tag{12.73}$$

を考える.

Y_{ij} に多項超幾何分布 (4.44) を仮定すると, 漸近的に,

$$W = (\boldsymbol{U} - E[\boldsymbol{U}])'V[\boldsymbol{U}]^{-}(\boldsymbol{U} - E[\boldsymbol{U}]) \sim \chi_{a-1} \tag{12.74}$$

が成り立つ. χ^2 分布の自由度は, 次の制約からわかる.

$$\sum_i \Big(U_i - E[U_i]\Big) = 0 \tag{12.75}$$

[*2] 順序カテゴリの特別な場合として議論する.

(12.74) の統計量を成分で書き下そう. その準備として, $E[\boldsymbol{U}], V[\boldsymbol{U}]$ を求めておく.

まず, Y_{ij} の要素を辞書順に並べて得られるベクトルを $\boldsymbol{Y} = (Y_1, \cdots, Y_{ab})'$, スコアを並べたベクトルを $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_b)'$ とおくと,

$$\boldsymbol{U} = A\,\boldsymbol{Y}, \quad A = \mathrm{diag}\,(n_i^{-1}) \otimes \boldsymbol{\eta}' \tag{12.76}$$

と表せる. (4.51) により,

$$E[\boldsymbol{U}] = n^{-1}A\,(\boldsymbol{r} \otimes \boldsymbol{c}) = n^{-1}(\boldsymbol{\eta}'\boldsymbol{c})\,\boldsymbol{j} \tag{12.77}$$

を得る. つぎに, (4.52) より, 分散 $V[\boldsymbol{U}]$ は,

$$\frac{A\,\{(n\,\mathrm{diag}\,(n_{i.}) - \boldsymbol{r}\,\boldsymbol{r}') \otimes (n\,\mathrm{diag}\,(n_{.j}) - \boldsymbol{c}\,\boldsymbol{c}')\}\,A}{n^2\,(n-1)} \tag{12.78}$$

となる. ここで,

$$\sigma_\eta^2 = \frac{n}{n-1}\left\{\sum_j \frac{\eta_j^2\,n_{.j}}{n} - \left(\sum_j \frac{\eta_j\,n_{.j}}{n}\right)^2\right\} \tag{12.79}$$

とおくと,

$$V[\boldsymbol{U}] = \sigma_\eta^2\left\{\mathrm{diag}\,(n_{i.}^{-1}) - n^{-1}\,\boldsymbol{j}\,\boldsymbol{j}'\right\} \tag{12.80}$$

と表される.

したがって,

$$V[\boldsymbol{U}]^- = \frac{1}{\sigma_\eta^2}\,\mathrm{diag}\,(n_{i.}) \tag{12.81}$$

であり, (12.74) は

$$W = \frac{\sum_i \sum_j n_{i.}^{-1}\left\{\eta_j\,(Y_{ij} - n_{i.}n_{.j}/n)\right\}^2}{\sigma_\eta^2} \tag{12.82}$$

となる. この検定を一般化 Mantel 検定という. とくに, $a = 2$ のときを Mantel 検定という. 一般化 Mantel 検定は, スコアとしてカテゴリの平均順位を割り付ける場合は, Kruskal-Wallis の検定と同等になる.

多重検定

$\boldsymbol{U}$ の対比

$$C = \boldsymbol{\lambda}'\boldsymbol{U} \tag{12.83}$$

を考えると, (12.80) より

$$V[C] = \sigma_\eta^2 \boldsymbol{\lambda}' \operatorname{diag}\left(n_{i.}^{-1}\right) \boldsymbol{\lambda} \tag{12.84}$$

を得る. したがって, 対比に関する推測に限っては, 漸近的には通常の一元配置の分散分析において, 分散 σ_η^2, 繰り返し数 $n_{i.}$ の場合と一致する.

■**Tukey 法**　検定統計量を次のようにする.

$$T_{ii'} = \frac{1}{\sigma_\eta}\left(\frac{1}{n_{i.}} + \frac{1}{n_{i'.}}\right)^{-1/2} \sum_j \eta_j \left(\frac{Y_{ij}}{n_{i.}} - \frac{Y_{i'j}}{n_{i'.}}\right) \tag{12.85}$$

有意確率は

$$1 - \sum_{i=1}^{a} \int_{-\infty}^{\infty} \prod_{i' \neq i} \left\{ \Phi\left(\sqrt{\frac{n_{i'}}{n_i}}\, y\right) - \Phi\left(\sqrt{\frac{n_{i'}}{n_i}}\, y - t_{ii'} \sqrt{1 + \frac{n_{i'}}{n_i}}\right)\right\} \phi(y)\, dy \tag{12.86}$$

となる.

■**Scheffe 法**　いくつかの水準をプールして, 比較する 2 群を I, I' とし, 統計量 $T_{II'}$ を

$$\frac{1}{\sigma_\eta}\left(\frac{1}{N_I} + \frac{1}{N_{I'}}\right)^{-1/2} \sum_j \eta_j \left(\frac{\sum_{i\in I} Y_{ij}}{N_I} - \frac{\sum_{i'\in I'} Y_{i'j}}{N_{I'}}\right) \tag{12.87}$$

とする. ここに

$$N_I = \sum_{i\in I} n_{i.} \tag{12.88}$$

である. ところが, これは (12.82) より, 一般化 Mantel 検定の統計量の各成分の対比にほかならない. 他方,

$$\sum_i \lambda_i U_i \leq \left(\sum_i \frac{\lambda_i^2 \sigma_\eta^2}{n_i}\right) \cdot \frac{1}{\sigma_\eta^2} \sum_i n_i \left(U_i - \sum_j \frac{\eta_j\, n_{.j}}{n}\right)^2 \tag{12.89}$$

が成り立つから, χ_{a-1}^2 分布を用いて保守的な検定を構成できる.

■**Dunnett 法**　Tukey 法において, $i' = 1$ とすればよい. 右片側検定の有意確率は次で与えられる.

$$1 - \int_{-\infty}^{\infty} \prod_{i=2}^{a} \Phi\left(\sqrt{\frac{n_i}{n_1}}\, y + t_{i1} \sqrt{1 + \frac{n_i}{n_1}} \right) \phi(y)\, dy \tag{12.90}$$

両側検定の有意確率は

$$\begin{aligned} 1 - \int_{-\infty}^{\infty} \prod_{i=2}^{a} \Biggl\{ &\Phi\left(\sqrt{\frac{n_i}{n_1}}\, y + t_{i1} \sqrt{1 + \frac{n_i}{n_1}} \right) \\ &- \Phi\left(\sqrt{\frac{n_i}{n_1}}\, y - t_{i1} \sqrt{1 + \frac{n_i}{n_1}} \right) \Biggr\} \phi(y)\, dy \end{aligned} \tag{12.91}$$

で与えられる.

特性値が二項分布に従う場合

以上の各方法において $b = 2$ とおくと, 特性値が 2 項分布に従うときの多重検定が導かれる. 列の自由度は 1 になるので, 統計量はスコア η_j に依らない. 実際, (12.85) は任意の η_j に対して

$$\left(\frac{n-1}{n} \right)^{1/2} \frac{n}{\sqrt{n_{.1} n_{.2}}} \left(\frac{1}{n_{i.}} + \frac{1}{n_{i'.}} \right)^{-1/2} \left(\frac{Y_{i2}}{n_{i.}} - \frac{Y_{i'2}}{n_{i'.}} \right) \tag{12.92}$$

となる. $1 - 1/n \simeq 1$ とすれば, これは水準 i, i' における標本比率の差を

$$\left(\frac{Y_{i2}}{n_{i.}} - \frac{Y_{i'2}}{n_{i'.}} \right) \Bigg/ \left\{ \left(\frac{1}{n_{i.}} + \frac{1}{n_{i'.}} \right) \frac{n_{.1}}{n} \cdot \frac{n_{.2}}{n} \right\}^{1/2} \tag{12.93}$$

と標準化したものにほかならない.

12.5　水準に順序のついている場合

3 水準以上の特性値を比較するとき, 水準に自然な順序がついていることがある. たとえば, 正規標本 (12.2) であれば, 対立仮説 $\mu_1 \leq \cdots \leq \mu_a$ に興味がある場合がある.

この仮説は，

$$\varDelta' = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 \\ & & & \cdots & & & \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \tag{12.94}$$

とおくと，

$$\varDelta' \boldsymbol{\mu} \geq \boldsymbol{o} \tag{12.95}$$

と表現できる．ベクトルに関する不等式はすべての成分に関して不等号が成立することを意味するものとする．

$$\varDelta' \boldsymbol{Y} \sim N_{a-1}(\varDelta' \boldsymbol{\mu}, \sigma^2 \varDelta' \varDelta) \tag{12.96}$$

であるから，

$$(\varDelta' \varDelta)^{-1} \varDelta Y \geq \boldsymbol{c} \tag{12.97}$$

を棄却域とすれば，σ^2 が既知なら検定を構成できる．

Jonckheere の検定

特性値が順序カテゴリの場合には，$Y_{ij} < Y_{i'j'}$ となる (j, j') の組の総数

$$W_{ii'} = \sum_{j=1}^{n_i} \sum_{j'=1}^{n_{i'}} u(Y_{i'j'} - Y_{ij}) \tag{12.98}$$

を考える．ここに，$u(t)$ は指示関数 (8.44) である．総和

$$S = \sum_{i=1}^{a-1} \sum_{i'=i+1}^{a} W_{ii'} \tag{12.99}$$

は漸近的に正規分布する．

(12.99) を統計量とする検定を Jonckheere の検定という．S は離散値をとるので連続性の修正をすることがある．

$W_{ii'}$ は Mann-Whitney 型の Wilcoxon 統計量であるから, (11.59) より直ちに

$$E[W_{ii'}] = \frac{1}{2}\, n_i\, n_{i'} \tag{12.100}$$

$$V[W_{ii'}] = \frac{1}{12}\, n_i\, n_{i'}\,(n_i + n_{i'} + 1) \tag{12.101}$$

を得る. したがって, S の平均 $E[S]$ は

$$\frac{1}{2}\Big(\sum_{i=1}^{a}\sum_{i'=1}^{a}\frac{1}{2}n_i n_{i'} - \sum_{i=1}^{a}\frac{1}{2}n_i^2\Big) = \frac{1}{4}\Big(n^2 - \sum_{i=1}^{a} n_i^2\Big) \tag{12.102}$$

となる. 共分散 Cov $[\, W_{ii'}, W_{jj'}\,]$ については,

$$\sum_{k,k',l,l'} \mathrm{Cov}\left[\, u(Y_{i'k'} - Y_{ik}), u(Y_{j'l'} - Y_{jl})\right] \tag{12.103}$$

を実際に計算すればよい. たとえば, $i = j, i' \neq j'$ のとき,

$$\sum_{k,k',l'}\left\{\frac{2}{3\,!} - \left(\frac{1}{2}\right)^2\right\} = \frac{1}{12}\, n_i\, n_{i'}\, n_{j'}$$

であり, 他の場合についても同様に求めることができる. 結局, S の分散は, 簡単な計算の後,

$$V[S] = \frac{1}{72}\Big\{n^2(2n+3) - \sum_{i=1}^{a} n_i^2(2n_i+3)\Big\} \tag{12.104}$$

となる.

相関検定

特性値が順序カテゴリの場合, もうひとつのアプローチは, 要因と結果の両方にスコアを与えることである. 第 i 水準にスコア ξ_i を, 第 j カテゴリ j にスコア η_j を与え, 統計量

$$R = \sum_{i=1}^{a}\sum_{j=1}^{b}\xi_i\,\eta_j\,Y_{ij} = (\boldsymbol{\xi}' \otimes \boldsymbol{\eta}')\,\boldsymbol{Y} \tag{12.105}$$

を考える. ただし, $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_a)'$, $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_b)'$ である.

Y_{ij} が多項超幾何分布に従うことから, (4.51) (4.52) より

$$E[R] = \frac{1}{n}\sum_{i=1}^{a}\xi_i n_{i.} \sum_{j=1}^{b}\eta_j n_{.j} \tag{12.106}$$

$$V[R] = \frac{n^2}{n-1}\sum_i\left\{\frac{n_{i.}\xi_i^2}{n} - \bar{\xi}^2\right\}\sum_j\left\{\frac{n_{.j}\eta_j^2}{n} - \bar{\eta}^2\right\} \tag{12.107}$$

を得る. ここに,

$$\bar{\xi} = \sum_{i=1}^{a}\frac{\xi_i n_{i.}}{n}, \quad \bar{\eta} = \sum_{j=1}^{b}\frac{\eta_j n_{.j}}{n} \tag{12.108}$$

である. 結局, $(R - E[R])^2/V[R]$ は

$$(n-1)\frac{\left\{\sum_i\sum_j\left(\xi_i - \bar{\xi}\right)\left(\eta_j - \bar{\eta}\right)Y_{ij}/n\right\}^2}{\left\{\sum_i\left(\xi_i - \bar{\xi}\right)^2 n_{i.}/n\right\}\left\{\sum_j\left(\eta_j - \bar{\eta}\right)^2 n_{.j}/n\right\}} \tag{12.109}$$

と書き下すことができて, 漸近的に χ_1^2 に従う. この統計量は (ξ_i, η_j) のすべての組についての相関係数の二乗の $n-1$ 倍にほかならないから, この検定を相関検定という.

第 13 章

ブロック実験

13.1 乱塊法

a 個の処理の比較において, 各処理における繰り返し数 m が等しい場合を考える. 一元配置の場合, 全部で $n = am$ 個の実験の順序を完全にランダマイズした. これに対して, 実験室の状態を一定に保つことが難しいような場合は, ある状態において, a 個の処理に対する実験を 1 回ずつ行い, これを m 回繰り返すことによって, 状態の違いによる実験のばらつきを取り除くことができる. このような実験の方法を乱塊法といい, 実験の状態に当たるものをブロックという. とくに, $a = 2$ の場合は, 対応のあるデータの解析として既に議論した.

第 j ブロックにおける第 i 処理に対応するデータを Y_{ij} とし, 線型モデル

$$Y_{ij} = \mu_i + b_j + \varepsilon_{ij}\,, \quad (i = 1, \cdots, a;\ j = 1, \cdots, m) \tag{13.1}$$

を仮定する. μ_i は第 i 処理の効果, b_j は第 j ブロックの効果を表す. 誤差 ε_{ij} は互いに独立に正規分布 $N(0, \sigma^2)$ に従うとする.

このとき, (13.1) は, 次のように行列表現できる.

$$\boldsymbol{Y} = [\, X_\mu \quad X_b \,] \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{b} \end{bmatrix} + \boldsymbol{\varepsilon} \tag{13.2}$$

ただし, $\boldsymbol{Y}, \boldsymbol{\varepsilon}$ はそれぞれ Y_{ij}, ε_{ij} を辞書順に並べた n 次元ベクトルであり,

$$\boldsymbol{\mu} = (\mu_1, \cdots, \mu_a)', \quad \boldsymbol{b} = (b_1, \cdots, b_m)' \tag{13.3}$$

である.

デザイン行列は,

$$X_\mu = I_a \otimes \boldsymbol{j}_m, \quad X_b = \boldsymbol{j}_a \otimes I_m \tag{13.4}$$

である. ここで, $X = [\, X_\mu \quad X_b \,]$ とおくと, rank $X = a + m - 1$ であるから,

$$Q'X = 0, \quad Q'Q = I_{(a-1)(m-1)} \tag{13.5}$$

なる $(a-1)(m-1)$ 行 n 列行列 Q' が存在する. さらに, 2 個の直交行列

$$P'_a = \begin{bmatrix} a^{-1/2}\boldsymbol{j}'_a \\ Q'_a \end{bmatrix}, \quad P'_m = \begin{bmatrix} m^{-1/2}\boldsymbol{j}'_m \\ Q'_m \end{bmatrix} \tag{13.6}$$

を考えると,

$$\Gamma = \begin{bmatrix} n^{-1/2}\,\boldsymbol{j}'_n \\ m^{-1/2}\,Q'_a X'_\mu \\ a^{-1/2}\,Q'_m X'_b \\ Q' \end{bmatrix} \tag{13.7}$$

は n 次直交行列をなす.

$\boldsymbol{Z} = \Gamma'\boldsymbol{Y}$ とおき,

$$\boldsymbol{Z} = (Z_\mu, \boldsymbol{Z}'_A, \boldsymbol{Z}'_b, \boldsymbol{Z}'_e)'$$

と Γ' に応じてブロック分割すると, (13.2) は,

$$\boldsymbol{Z} = \begin{bmatrix} \sqrt{n}\,(\overline{\mu} + \overline{b}) \\ \sqrt{m}\,Q'_a\boldsymbol{\mu} \\ \sqrt{a}\,Q'_m\boldsymbol{b} \\ O \end{bmatrix} + \Gamma'\boldsymbol{\varepsilon} \tag{13.8}$$

と同値である. ここに,

$$\overline{\mu} = \frac{1}{a}\sum_i \mu_i, \quad \overline{b} = \frac{1}{m}\sum_j b_j \tag{13.9}$$

であり, $\Gamma'\boldsymbol{\varepsilon}$ は互いに独立に正規分布 $N(0, \sigma^2)$ に従う.

$\boldsymbol{Z}$ のノルムに注目して,

$$\| \boldsymbol{Z} \|^2 - \| Z_\mu \|^2 = \| \boldsymbol{Z}_A \|^2 + \| \boldsymbol{Z}_b \|^2 + \| \boldsymbol{Z}_e \|^2 \tag{13.10}$$

を成分で書き下すと,

$$\begin{aligned}\sum_i \sum_j (Y_{ij} - \overline{Y}_{..})^2 &= m \sum_i (\overline{Y}_{i.} - \overline{Y}_{..})^2 \\ &+ a \sum_j (\overline{Y}_{.j} - \overline{Y}_{..})^2 + \| Q'\boldsymbol{Y} \|^2\end{aligned} \tag{13.11}$$

を得る. ここに,

$$\overline{Y}_{..} = \frac{1}{am} \sum_i \sum_j Y_{ij}, \quad \overline{Y}_{i.} = \frac{1}{m} \sum_j Y_{ij}, \quad \overline{Y}_{.j} = \frac{1}{a} \sum_i Y_{ij} \tag{13.12}$$

である. (13.11) の右辺の第 1 項は処理の効果であり, S_A で表す. 第 2 項がブロック効果であり, S_b で表す. 第 3 項はモデルで説明できない残差平方和であり, S_e で表す. これから, 分散の最良不偏推定量

$$\hat{\sigma}^2 = \frac{S_e}{(a-1)(m-1)} \tag{13.13}$$

を得る.

ところで, 一元配置モデルにおける残差平方和は, 繰り返し数が等しい $(n_i = m)$ ときは, 次のようになる.

$$\sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2 = \sum_i m \overline{Y}_{i.}^2 - n \overline{Y}_{..}^2 \tag{13.14}$$

これは, (13.11) より $S_b + S_e$ と分解できる. したがって, 乱塊法では残差平方和がブロック効果 S_b だけ小さくなり, その分自由度が $m-1$ だけ損失を受けることがわかる.

対応のある t 検定の項で既に触れたが, ブロック効果がないときは, 自由度の損失の分だけ検出力が下がるが, 通常はそれよりブロック効果を取り除くメリットの方が大きい.

オムニバスな F 検定, Tukey , Scheffe , Dunnett の多重比較は, すべて一元配置において繰り返し数 m の場合と同様に構成できる. ただし, 分散の推定量として, (9.34) の代わりに (13.13) を用いなければならない.

13.2　ノンパラメトリック法

実際のデータ解析では, 因果関係の鎖に中に介在してこれに影響する因子が存在することがある. このように, 直接因果関係を考察する対象ではないが, それに影響を与える因子を交絡因子という. この値ごとに層別化して, それぞれの層をブロックとみなせば, ブロック効果を取り除くことができる.

2 値データに対する乱塊法

■マッチング　2 標本問題において, ブロックサイズ 2 の乱塊法を考える. これは疫学研究などでしばしば用いられる方法である. たとえば, 疾患群と対象群において, 過去の危険因子に対する暴露に差があったかどうか考察する際に, さまざまな交絡因子の似通った 2 人を対にして調査することがある. この方法をマッチングという.

母集団 $i = 1, 2$ において, 2 値反応を 1 (反応あり) , 0 (反応なし) で表す. このとき, 第 j ブロックにおける母集団 i に対応するデータを Y_{ij} とすると, 形式的には, 次のような 2×2 表を得る.

		$i=2$	
		0	1
$i=1$	0	a	b
	1	c	d

これは,

	反応	
	0	1
$i=1$	1	0
$i=2$	1	0

	反応	
	0	1
1	1	0
2	0	1

	反応	
	0	1
1	0	1
2	1	0

	反応	
	0	1
1	0	1
2	0	1

という対が, それぞれ a, b, c, d 個ずつであることを意味する.

母集団 $i = 1, 2$ ともに反応あり, またはともに反応なしの対は, 母集団の差に関して何の情報も与えないから, 母集団 $i = 1$ と 2 において反応に差が出る対 (unlike pair) だけを考える.

Y_{ij} に二項分布 $B(1, p_{ij})$ を仮定するとき, シフトモデルは

$$\log \frac{p_{ij}}{1 - p_{ij}} = \mu_j + \delta_i \varDelta \tag{13.15}$$

と表現できる. ここに, δ_i はサンプルが母集団 1 からのものであれば 1 を, 2 からのものであれば 0 をとるダミー変数であり, $\varDelta$ は母集団の差を表す正値のパラメタである.

帰無仮説 $\varDelta = 0$ のもとで,

$$Y_{1j} + Y_{2j} = y_{.j} \tag{13.16}$$

は十分統計量である.

実際, $y_{.j} = 0$ または $y_{.j} = 2$ のときは y_{1j} は一意に定まり, $y_{.j} = 1$ のときは, $Y_{1j} = y_{ij}$ となる条件つき確率は,

$$\frac{p_{1j}^{y_{1j}} (1 - p_{1j})^{1-y_{1j}}\, p_{2j}^{1-y_{1j}} (1 - p_{2j})^{y_{1j}}}{\sum_{y_{1j}} p_{1j}^{y_{1j}} (1 - p_{1j})^{1-y_{1j}}\, p_{2j}^{1-y_{1j}} (1 - p_{2j})^{y_{1j}}} = \frac{\exp(\varDelta\, y_{1j})}{1 + \exp \varDelta} \tag{13.17}$$

であるから, $\varDelta = 0$ のときは未知パラメタに依らない.

したがって, $y_{.j} = 1$ を与えたときの条件つき確率として有意確率を計算すればよい. 統計量を

$$T = \sum_j^{*} Y_{1j} \tag{13.18}$$

とすると, unlike pair の総数 ν を与えたときの分布は, 二項分布

$$B\Big(\nu, \exp \varDelta / (1 + \exp \varDelta)\Big)$$

になる. ただし, $\sum^*$ はすべての unlike pair に関する和を表す.

帰無仮説 $\varDelta = 0$ のもとでは,

$$T \sim B(\nu, 1/2) \tag{13.19}$$

となるので, これに基づいて検定を構成すればよい.

T , ν が十分大きい場合には, 正規近似を用いて, 漸近的に

$$\frac{(T-\nu/2)^2}{\nu\cdot(1/2)\cdot(1/2)} = \frac{(2T-\nu)^2}{\nu} \sim \chi_1^2 \tag{13.20}$$

が成立する. これは適合度の検定に相当する.

尤度比検定は

$$2T\log\frac{2T}{\nu} + 2(\nu-T)\log\frac{2(\nu-T)}{\nu} \sim \chi_1^2 \tag{13.21}$$

より構成すればよい.

ところで,

$$Y = \frac{(\sum_j^* Y_{1j} - \sum_j^* Y_{2j})^2}{4\nu} \tag{13.22}$$

を統計量とする検定を McNemar 検定という. 次章の Mantel-Haenszel 流に 2×2 表の効果を加えると統計量 Y (13.22) は自由度 1 の χ^2 分布に従うことがわかる. 実際,

$$Y = \frac{(2T-\nu)^2}{4\,\nu} \tag{13.23}$$

と変形できるので, (13.20) で両側検定とした場合と同等である.

Cochran の検定

今度は, a 通りの処理に対して乱塊法によるブロック実験を行い, 2 値の反応を得たとする. 第 j ブロックにおける分割表が下表のようになったとしよう. 列に関する周辺和がすべての行において 1 となっていることに注意されたい.

		反　応		
		0	1	計
	1	$1-Y_{1j}$	Y_{1j}	1
処	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	i	$1-Y_{ij}$	Y_{ij}	1
理	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	a	$1-Y_{aj}$	Y_{aj}	1
	計	$a-Y_{.j}$	$Y_{.j}$	a

$y_{.j}=0$ または $y_{.j}=a$ の場合は表のすべての要素が一意に定まる. そうでない場合をここでは unlike set と呼ぶことにする. $y_{.j}$ を与えたときの Y_{ij} の分布は多項超幾何分布である.

ここで,

$$\boldsymbol{T}=(T_1,\cdots,T_a)' \ , \quad T_i=\sum_j^{*} Y_{ij} \tag{13.24}$$

を統計量とすると, 処理の反応に対する均一性の検定を構成できる. ただし, $\sum^*$ は unlike set に関する和を表すとする.

異なるブロック間の独立性を仮定すると, (4.47) (4.50) より $\boldsymbol{T}$ の平均, 分散を得る：

$$E[\boldsymbol{T}]=(\sum_j^{*} y_{.j}/a)\boldsymbol{j}_a \ , \quad V[\boldsymbol{T}]=\frac{\sigma_c^2}{a-1}(aI_a-\boldsymbol{j}_a\boldsymbol{j}_a') \tag{13.25}$$

ここに,

$$\sigma_c^2=V[T_i]=\sum_j^{*} y_{.j}(a-y_{.j})/a^2 \tag{13.26}$$

である.

$$V[\boldsymbol{T}]^-=\frac{a-1}{a\sigma_c^2}I_a \tag{13.27}$$

が, $V[\boldsymbol{T}]$ の一般化逆行列のひとつを与えるから, $T_.=\sum_i T_i$ とおくと,

$$(\boldsymbol{T}-E[\boldsymbol{T}])'V[\boldsymbol{T}]^-(\boldsymbol{T}-E[\boldsymbol{T}])=\frac{(a-1)\sum_i T_i^2-T_.^2/a}{\sum_j^* y_{.j}-\sum_j^* y_{.j}^2/a} \tag{13.28}$$

は漸近的に χ^2_{a-1} に従う. (13.28) に基づく検定を Cochran の検定という.

とくに, $a = 2$ のときは, (13.28) は

$$\frac{T^2 + (\nu - T)^2 - \nu^2/2}{\nu - \nu/2} = \frac{(2T - \nu)^2}{\nu} \tag{13.29}$$

となり, McNemar 検定に一致する.

$\boldsymbol{T}$ の対比

$$C = \boldsymbol{\lambda}'\boldsymbol{T} \tag{13.30}$$

を考えると, (13.25) より

$$V[C] = \frac{a\sigma_c^2}{a-1} I_a \tag{13.31}$$

となるから, 対比に関する多重検定を考えるときは, データの数理構造は

$$V[\boldsymbol{T}] = \frac{a\sigma_c^2}{a-1} I_a \tag{13.32}$$

とした場合と同等になる.

したがって, まず, Tukey 法は, 検定統計量を

$$T_{ii'} = \frac{\sqrt{a-1}}{\sqrt{a}\sigma_c} \left| T_i - T_{i'} \right| \tag{13.33}$$

として, 有意確率を次で求めればよい.

$$1 - \sum_{i=1}^{a} \int_{-\infty}^{\infty} \prod_{i' \neq i} \left\{ \Phi\left(\frac{\sqrt{a}\sigma_c}{\sqrt{a-1}} y \right) - \Phi\left(\frac{\sqrt{a}\sigma_c}{\sqrt{a-1}} y - t_{ii'} \right) \right\} \phi(y)\, dy \tag{13.34}$$

第 1 の処理を対照とするとき, Dunnett 法は (13.33) において, つねに $i' = 1$ とすればよい. 右片側検定の有意確率は

$$1 - \int_{-\infty}^{\infty} \prod_{i=2}^{a} \Phi\left(\frac{\sqrt{a}\sigma_c}{\sqrt{a-1}} y + t_{ii'} \right) \phi(y)\, dy \tag{13.35}$$

で与えられる.

Scheffe 法は, 任意の対比

$$\sum_i \lambda_i T_i\,, \quad \sum_i \lambda_i = 0 \tag{13.36}$$

に対して

$$\frac{(a-1)(\sum_i \lambda_i\, T_i)^2}{a\,(\sum_i \lambda_i^2)\,\sigma_c^2} \tag{13.37}$$

を統計量とする. 保守的な有意確率は χ^2_{a-1} 分布から求められる.

順序カテゴリに対する乱塊法

Friedman の検定

		反　応					計
		1	$\cdots$	j	$\cdots$	b	
	1	Y_{11l}	$\cdots$	Y_{1jl}	$\cdots$	Y_{1bl}	1
処	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	i	Y_{i1l}	$\cdots$	Y_{ijl}	$\cdots$	Y_{ibl}	1
理	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	a	Y_{a1l}	$\cdots$	Y_{ajl}	$\cdots$	Y_{abl}	1
	計	$Y_{.1l}$	$\cdots$	$Y_{.jl}$	$\cdots$	$Y_{.bl}$	a

二項分布の場合の拡張として, 特性値が順序カテゴリのときの乱塊法を考えよう. a 通りの処理に対する反応または効果を表す順序カテゴリを $j = 1, \cdots, b$, ブロックサイズ m として, l 番目のブロックについて, 上の分割表を仮定する. 今度は j がブロックの番号ではなく, 反応を表す順序カテゴリであることに注意する.

すべてのブロックにおいて周辺和を所与としたとき, (4.51) (4.52) より,

$$E[Y_{ijl}] = \frac{y_{.jl}}{a} \tag{13.38}$$

$$\mathrm{Cov}\,[Y_{ijl}, Y_{i'j'l}] = \frac{(a\delta_{ii'} - 1)y_{.jl}(a\delta_{jj'} - y_{.j'l})}{a^2(a-1)} \tag{13.39}$$

を得る.

列の順序カテゴリ j にスコア η_j を与え, すべてのブロックについて加えた次の統計量を考える.

$$\boldsymbol{U} = (U_1, \cdots, U_a)', \quad U_i = \sum_{j=1}^{b} \sum_{l=1}^{m} \eta_j Y_{ijl} \tag{13.40}$$

いま,

$$y_{.j.} = \sum_{l} y_{.jl}, \quad (j = 1, \cdots, b) \tag{13.41}$$

とおくと, 異なるブロック間は相関がないとしているから, U_i の平均は

$$E[U_i] = \sum_{j} \sum_{l} \eta_j E[Y_{ijl}] = \sum_{j} \frac{\eta_j y_{.j.}}{a} \stackrel{\text{def}}{=} \overline{u} \tag{13.42}$$

であり, 共分散 Cov $[U_i, U_{i'}]$ は

$$\begin{aligned} &\frac{a\delta_{ii'} - 1}{a^2(a-1)} \sum_{l} \Big\{ \sum_{j} \eta_j^2 y_{.jl}(a - y_{.jl}) - \sum_{j} \sum_{j'} \eta_j \eta_{j'} y_{.jl} y_{.j'l} \Big\} \\ &\quad = \frac{a\delta_{ii'} - 1}{a-1} \Big\{ \sum_{j} \frac{\eta_j^2 y_{.j.}}{a} - \Big(\sum_{j} \frac{\eta_j y_{.j.}}{a} \Big)^2 \Big\} \end{aligned} \tag{13.43}$$

となる.

ここで,

$$\sigma_\eta^2 = \frac{a}{a-1} \sum_{j} \frac{\eta_j^2 y_{.j.}}{a} - \Big(\sum_{j} \frac{\eta_j y_{.j.}}{a} \Big)^2 \tag{13.44}$$

とおくと, $\boldsymbol{U}$ の分散は

$$V[\boldsymbol{U}] = \sigma_\eta^2 (I_a - a^{-1} \boldsymbol{j}_a \boldsymbol{j}_a') \tag{13.45}$$

と表現できる.

以上から, 一様性の検定の統計量

$$(\boldsymbol{U} - E[\boldsymbol{U}])' V[\boldsymbol{U}]^- (\boldsymbol{U} - E[\boldsymbol{U}]) = \frac{1}{\sigma_\eta^2} \sum_{i} (U_i - \overline{u})^2 \tag{13.46}$$

は漸近的に, χ^2_{a-1} に従う

とくに, 列のカテゴリをブロック内の順位とすると $b = a$ であり, タイがなければつねに $y_{.jl} = 1$ となる. このとき, スコアを $\eta_j = j$ $(j = 1, \cdots, a)$ としたものを Friedman の検定という.

(13.45) においても, 対比を考える限り, 分散の $\boldsymbol{jj'}$ の項を無視することができて, 次のように多重検定が構成される.

Tukey 法は

$$T_{ii'} = \frac{|\,U_i - U_{i'}\,|}{\sigma_\eta} \tag{13.47}$$

を統計量として, 有意確率が

$$1 - \sum_{i=1}^{a} \int_{-\infty}^{\infty} \prod_{i' \neq i} \Big\{ \Phi(y) - \Phi(y - t_{ii'}\sigma_\eta) \Big\}\, \phi(y)\, dy \tag{13.48}$$

で与えられる.

Scheffe 法は, λ_i $(i = 1, \cdots, a)$ を

$$\sum_i \lambda_i = 0 \tag{13.49}$$

を満たす任意の対比として

$$\frac{(\sum_i \lambda_i\, U_i)^2}{\sigma_\eta^2\, (\sum_i \lambda_i^2)} \tag{13.50}$$

を χ^2_{a-1} で評価すると, 保守的な有意確率が得られる.

Dunnett の右片側検定は, 対照を第 1 処理として,

$$T_{i1} = \frac{U_i - U_1}{\sigma_\eta} \tag{13.51}$$

を統計量にすればよい. 有意確率は次で与えられる.

$$1 - \int_{-\infty}^{\infty} \prod_{i=2}^{a} \Phi(y + t_{i1}\sigma_\eta)\, \phi(y)\, dy \tag{13.52}$$

処理にも順序のついている場合

処理にも自然な順序のついている場合は, 処理 i に対してもスコア ξ_i $(i = 1, \cdots, a)$ を与えて,

$$R = \sum_i \sum_j \sum_l \xi_i \eta_j Y_{ijl} \tag{13.53}$$

を統計量とすればよい. このとき, (13.42) (13.43) と同様にして,

$$E[R] = \frac{1}{a} \sum_i \sum_j \xi_i \eta_j y_{.j.} \tag{13.54}$$

$$V[R] = a \sigma_\eta^2 \left\{ \frac{\sum_i \xi_i^2}{a} - \left(\frac{\sum_i \xi_i}{a} \right) \right\} \stackrel{\text{def}}{=} \sigma^2 \tag{13.55}$$

を得る.

以上から, 漸近的に

$$\frac{R - E[R]}{\sqrt{V[R]}} \sim N(0, 1) \tag{13.56}$$

となることを利用した検定を構成できる. とくに, 列のカテゴリをブロック内の順位として, スコアを $\xi_i = i$, $\eta_j = j$ $(i, j = 1, \cdots, r)$ としたものを Page 検定という.

多重検定は, 処理に順序がついていない場合と同様にして構成できる.

分散分析を行う場合は, ブロックサイズが不均一であると解析がやや複雑になるが, ノンパラメトリック法は, このときも簡単に構成できる. ただし, この場合は分散共分散行列の一般化逆行列が対角行列になるとは限らないので成分で書き下すことは簡単ではない. 同じ理由から Tukey や Dunnett の多重検定の構成も困難になる. Scheffe 法はこの場合もデータを直交変換することにより簡単に構成できる.

第 14 章

二元配置

14.1　オムニバスな検定

交互作用と主効果

2 つの要因 A , B の反応に対する効果を調べたいとする．要因 A は A_i $(i = 1, \cdots, a)$ の a 値, 要因 B は B_j $(j = 1, \cdots, b)$ の b 値をとるものとする. 水準の組み合わせ (A_i, B_j) に対して, 測定を m 回ずつ行うとし, l 回目の測定値を Y_{ijl} とする. この組み合わせに対する平均を μ_{ij} として, 線型モデル

$$Y_{ijl} = \mu_{ij} + \varepsilon_{ijl} \tag{14.1}$$

を考える. ここに, Y_{ijl} は水準の組み合わせ (A_i, B_j) に対して, l 回目の測定で得られるデータに対応する確率変数である. 誤差 ε_{ijl} は互いに独立に, 正規分布 $N(0, \sigma^2)$ に従うものとする.

このとき, 形式的に, データを ab 水準の一元配置として解析することも考えられるが, この方法は誤差の推定が過大評価となり効率が悪い. 実際, 要因 A , B の間に交互作用がなければ, μ_{ij} は

$$\mu_{ij} - \overline{\mu} = (\overline{\mu}_{i.} - \overline{\mu}) + (\overline{\mu}_{.j} - \overline{\mu}) \Longleftrightarrow \mu_{ij} = \overline{\mu}_{i.} + \overline{\mu}_{.j} - \overline{\mu} \tag{14.2}$$

のように表せる.

ここに,

$$\overline{\mu}_{i.} = \frac{1}{b}\sum_{j=1}^{b}\mu_{ij}\ ,\ \overline{\mu}_{.j} = \frac{1}{a}\sum_{i=1}^{a}\mu_{ij}\ ,\ \overline{\mu} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\mu_{ij} \tag{14.3}$$

である.

この場合は, 既に何度かみたように, 実質的なパラメタ数は $a+b-1$ となるので, 繰り返し数が

$$\frac{n}{a+b-1} > m \tag{14.4}$$

となる. ただし, $n = abm$ は総実験数である.

交互作用の存在する場合は, モデルを

$$Y_{ijl} = \overline{\mu} + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijl} \tag{14.5}$$

とすればよい. γ_{ij} が相互作用を, $\overline{\mu}$ が総平均を表す. ただし, このとき, 未知パラメタは $1+a+b+ab$ 個あるので, ab 個の μ_{ij} と対応しない. そこで,

$$\overline{\mu}_{i.} = \sum_j \xi_j \mu_{ij}\ ,\quad \sum_j \xi_j = 1 \quad (i = 1, \cdots, a) \tag{14.6}$$

$$\overline{\mu}_{.j} = \sum_i \eta_i \mu_{ij}\ ,\quad \sum_i \eta_i = 1 \quad (j = 1, \cdots, b) \tag{14.7}$$

とおく. ξ_j , η_i は非負の荷重であり, $\overline{\mu}_{i.}$, $\overline{\mu}_{.j}$ はそれぞれ, 水準 A_i , B_j の加重平均と解釈できる. この定義から直ちに

$$\sum_i \eta_i \overline{\mu}_{i.} = \sum_j \xi_j \overline{\mu}_{.j} = \overline{\mu} \tag{14.8}$$

を得る. ここで,

$$\alpha_i = \overline{\mu}_{i.} - \overline{\mu}\ ,\ \beta_j = \overline{\mu}_{.j} - \overline{\mu} \tag{14.9}$$

とおくと, これらはそれぞれ A_i , B_j の効果を表すので, 主効果という. さらに, 交互作用を

$$\gamma_{ij} = (\mu_{ij} - \overline{\mu}_{.j}) - (\overline{\mu}_{i.} - \overline{\mu}) = \mu_{ij} - \overline{\mu}_{i.} - \overline{\mu}_{.j} + \overline{\mu} \tag{14.10}$$

で定義する.

これは水準 B_j における水準 A_i の効果から overall の A_i の効果を減じた差であるから, B_j に固有の A_i の効果を意味する. もちろん, i , j に関する対称性から A_i に固有の B_j の効果とも解釈できる.

結局, $\overline{\mu}$, α_i , β_j , γ_{ij} はすべて μ_{ij} のみの関数として表され, 実質的なパラメタの数は ab となる.

この代わりに, $\overline{\mu}$, α_i , β_j , γ_{ij} に, 以下の制約を設けてもよい.

$$\sum_i \eta_i \alpha_i = 0, \quad \sum_j \xi_j \beta_j = 0 \tag{14.11}$$

$$\sum_i \eta_i \gamma_{ij} = 0 \quad (j = 1, \cdots, b) \tag{14.12}$$

$$\sum_j \xi_j \gamma_{ij} = 0 \quad (i = 1, \cdots, a) \tag{14.13}$$

このとき, 制約の総数は $1 + 1 + b + a = a + b + 2$ 個であるが, 第 3 式と第 4 式の間には

$$\sum_j \xi_j \Big(\sum_i \eta_i \gamma_{ij} \Big) = \sum_i \eta_i \Big(\sum_j \xi_j \gamma_{ij} \Big) \tag{14.14}$$

が成立するので, 実質的に制約の個数は $a + b + 1$ である.

実は, $\overline{\mu}$, α_i , β_j , γ_{ij} を (14.6) から (14.10) のように表すことと, (14.5) に (14.11) から (14.13) の制約を課すことは同値である.

以上で定義した, 主効果および交互作用は, もちろん荷重 ξ_j , η_i に依存する. しかし, 交互作用は, ある荷重 $\{\xi_j^*\}$, $\{\eta_i^*\}$ に関して 0 であれば, 任意の荷重 $\{\xi_j\}$, $\{\eta_i\}$ に関しても 0 になる. すなわち, 交互作用が 0 であるという帰無仮説は荷重の取り方によらない. 以下にこのことを示そう.

$\{\xi_j^*\}$, $\{\eta_i^*\}$ を荷重としたときの $\overline{\mu}$, α_i , β_j をそれぞれ $\overline{\mu}^*$, α_i^* , β_j^* で表すと, このときの交互作用は 0 であるから,

$$\mu_{ij} = \overline{\mu}^* + \alpha_i^* + \beta_j^* \tag{14.15}$$

が成り立つ.

任意の荷重 $\{\xi_j\}$, $\{\eta_i\}$ に関しては (14.15) と

$$\sum_j \xi_j = \sum_i \eta_i = 1$$

より,

$$\sum_j \xi_j \mu_{ij} = \overline{\mu}^* + \alpha_i^* + \sum_j \xi_j \beta_j^* = \overline{\mu}^* + \alpha_i^*$$

$$\sum_i \eta_i \mu_{ij} = \overline{\mu}^* + \sum_i \eta_i \alpha_i^* + \beta_j^* = \overline{\mu}^* + \beta_j^*$$

$$\overline{\mu} = \overline{\mu}^* + \sum_i \eta_i \alpha_i^* + \sum_j \xi_j \beta_j^* = \overline{\mu}^*$$

が成り立つから, 任意の荷重 $\{\xi_j\}$, $\{\eta_i\}$ に関する交互作用は

$$\gamma_{ij} = \mu_{ij} - \sum_j \xi_j \mu_{ij} - \sum_i \eta_i \mu_{ij} + \overline{\mu} = 0$$

となる. ■

交互作用がない場合には, 主効果の対比もまた荷重の取り方によらない. 実際, 任意の α_i に関する対比は, 上の荷重 $\{\xi_j^*\}$, $\{\eta_i^*\}$ の場合を用いて,

$$\sum_i \lambda_i \alpha_i = \sum_i \lambda_i \Big(\sum_j \xi_j \mu_{ij} - \overline{\mu} \Big) = \sum_i \lambda_i \Big(\alpha_i^* - \sum_{i'} \eta_{i'} \alpha_{i'}^* \Big) = \sum_i \lambda_i \alpha_i^*$$

と表される.

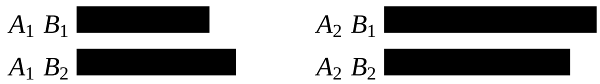
しかし, 交互作用のある場合には, 主効果に関する帰無仮説は荷重の取り方に依存する. 実際, 「交互作用あり」とは因子 A の水準により因子 B の効果が異なる (あるいは対称性により因子 B の水準により因子 A の効果が異なる) ことを表している.

簡単のため, $a = b = 2$ の場合を考えてみよう.

A_1 B_1 [redacted]　　A_2 B_1 [redacted]

A_1 B_2 [redacted]　　A_2 B_2 [redacted]

上の図の場合は交互作用がなく, 因子 A のどちらの水準においても因子 B の効果は一定である. (棒グラフが平行になっていることに注意) しかし, 下の図の場合はそうではない. 水準 A_1 においては, 水準 B_2 が B_1 に比してプラスに作用しているのに対し, 水準 A_2 においてはマイナスに作用していることがわかる. 繰り返すが, このような場合は因子 A の水準を固定しないと, 因子 B の効果が定まらないのである.



つぎに, 主効果, 交互作用の検定を構成しよう. まず,

$$\boldsymbol{Y} = (Y_{111}, \cdots, Y_{11m}, Y_{121}, \cdots, Y_{abm})' \tag{14.16}$$

$$\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_a)'\ ,\ \boldsymbol{\beta} = (\beta_1, \cdots, \beta_b)' \tag{14.17}$$

$$\boldsymbol{\gamma} = (\gamma_{11}, \cdots, \gamma_{1b}, \gamma_{21}, \cdots, \gamma_{ab})' \tag{14.18}$$

とおく. このとき, (14.5) は

$$\boldsymbol{Y} = \boldsymbol{j}\overline{\mu} + X_\alpha \boldsymbol{\alpha} + X_\beta \boldsymbol{\beta} + X_\gamma \boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{14.19}$$

と表現できる. ここに, $\boldsymbol{\varepsilon}$ は $N_n(\boldsymbol{o}, \sigma^2 I)$ に従う誤差ベクトルであり,

$$X_\alpha = I_a \otimes \boldsymbol{j}_{bm} \tag{14.20}$$

$$X_\beta = \boldsymbol{j}_a \otimes I_b \otimes \boldsymbol{j}_m \tag{14.21}$$

$$X_\gamma = \boldsymbol{j}_{ab} \otimes I_m \tag{14.22}$$

はデザイン行列である.

一元配置やブロック実験のときと同様に, データを直交変換すると見通しがよくなる.

まず, 2 つの直交行列

$$P'_a = \begin{bmatrix} a^{-1/2}\boldsymbol{j}'_a \\ Q'_a \end{bmatrix}, \quad P'_b = \begin{bmatrix} b^{-1/2}\boldsymbol{j}'_b \\ Q'_b \end{bmatrix} \tag{14.23}$$

を考える.

このとき,

$$Q' \,[\, \boldsymbol{j}_n \quad X_\alpha \quad X_\beta \quad X_\gamma \,] = O\,, \quad Q'Q = I_{n-ab} \tag{14.24}$$

を満たす $n-ab$ 行 n 列行列 Q' が存在して,

$$\Gamma' = \begin{bmatrix} n^{-1/2}\boldsymbol{j}_n' \\ (bm)^{-1/2} Q_a' X_\alpha' \\ (am)^{-1/2} Q_b' X_\beta' \\ m^{-1/2}(Q_a' \otimes Q_b') X_\gamma' \\ Q' \end{bmatrix} \tag{14.25}$$

が n 次直交行列となる.

$\boldsymbol{Z} = \Gamma'\boldsymbol{Y}$ と直交変換し, Γ' に応じて, $\boldsymbol{Z}$ をブロック分割すると, (14.19) は

$$\boldsymbol{Z} = \begin{bmatrix} Z_\mu \\ \boldsymbol{Z}_\alpha \\ \boldsymbol{Z}_\beta \\ \boldsymbol{Z}_\gamma \\ \boldsymbol{Z}_e \end{bmatrix} = \begin{bmatrix} \sqrt{n}(\overline{\mu} + \overline{\alpha} + \overline{\beta} + \overline{\gamma}) \\ \sqrt{m} Q_a' \{ \sqrt{b}\boldsymbol{\alpha} + (I_a \otimes \boldsymbol{j}_b')\boldsymbol{\gamma} \} \\ \sqrt{m} Q_b' \{ \sqrt{a}\boldsymbol{\beta} + (\boldsymbol{j}_a' \otimes I_b)\boldsymbol{\gamma} \} \\ \sqrt{m}(Q_a' \otimes Q_b')\boldsymbol{\gamma} \\ \boldsymbol{o} \end{bmatrix} + \Gamma'\boldsymbol{\varepsilon} \tag{14.26}$$

と書ける. ただし,

$$\overline{\alpha} = \frac{1}{a}\sum_{i=1}^{a} \alpha_i\,, \quad \overline{\beta} = \frac{1}{b}\sum_{j=1}^{b} \beta_j\,, \quad \overline{\gamma} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} \gamma_{ij} \tag{14.27}$$

である. ここで,

$$Y_{...} = n\overline{Y}_{...} = \sum_i \sum_j \sum_l Y_{ijl} \tag{14.28}$$

$$Y_{i..} = bm\overline{Y}_{i..} = \sum_j \sum_l Y_{ijl} \ \ (i = 1, \cdots, a) \tag{14.29}$$

$$Y_{.j.} = am\overline{Y}_{.j.} = \sum_i \sum_l Y_{ijl} \ \ (j = 1, \cdots, b) \tag{14.30}$$

$$Y_{ij.} = m\overline{Y}_{ij.} = \sum_l Y_{ijl} \ \ (i = 1, \cdots, a\,;\ j = 1, \cdots, b) \tag{14.31}$$

とおくと,

$$Z_\mu = \frac{1}{\sqrt{n}} Y_{...}\,, \quad \boldsymbol{Z}_\alpha = \frac{1}{\sqrt{bm}} Q_a'(Y_{1..}, \cdots, Y_{a..})'\,, \quad \boldsymbol{Z}_\beta = \frac{1}{\sqrt{am}} Q_b'(Y_{.1.}, \cdots, Y_{.b.})'$$

はすぐにわかる.

Q_a', Q_b' の定義の仕方から,

$$\| Z_\mu \|^2 = n^{-1} Y_{\cdots}^2 = n\overline{Y}_{\cdots}^2 \tag{14.32}$$

$$S_\alpha \stackrel{\text{def}}{=} \| \boldsymbol{Z}_\alpha \|^2 = (bm)^{-1} \sum_i Y_{i\cdot\cdot}^2 - n^{-1} Y_{\cdots}^2 \tag{14.33}$$

$$S_\beta \stackrel{\text{def}}{=} \| \boldsymbol{Z}_\beta \|^2 = (am)^{-1} \sum_j Y_{\cdot j\cdot}^2 - n^{-1} Y_{\cdots}^2 \tag{14.34}$$

は直ちに得られる. つぎに,

$$P_a' \otimes P_b' = \begin{bmatrix} (ab)^{-1/2} \boldsymbol{j}_{ab}' \\ a^{-1/2} \boldsymbol{j}_a' \otimes Q_b' \\ b^{-1/2} Q_a' \otimes \boldsymbol{j}_b' \\ Q_a' \otimes Q_b' \end{bmatrix} \tag{14.35}$$

は直交行列であるから,

$$\begin{aligned} m S_\gamma &\stackrel{\text{def}}{=} \| (Q_a' \otimes Q_b') X_\gamma' \boldsymbol{Y} \|^2 \\ &= \| X_\gamma' \boldsymbol{Y} \|^2 - (ab)^{-1} \| \boldsymbol{j}_{ab}' X_\gamma' \boldsymbol{Y} \|^2 - a^{-1} \| (\boldsymbol{j}_a' \otimes Q_b') X_\gamma' \boldsymbol{Y} \|^2 \\ &\qquad - b^{-1} \| (Q_a' \otimes \boldsymbol{j}_b') X_\gamma' \boldsymbol{Y} \|^2 \end{aligned} \tag{14.36}$$

が成り立つ. 他方, 簡単な計算により, 次が得られる.

$$\| \boldsymbol{j}_{ab}' X_\gamma' \boldsymbol{Y} \|^2 = Y_{\cdots}^2 \tag{14.37}$$

$$\| (\boldsymbol{j}_a' \otimes Q_b') X_\gamma' \boldsymbol{Y} \|^2 = \| Q_b' (Y_{\cdot 1 \cdot}, \cdots, Y_{\cdot b \cdot})' \|^2 = \sum_{j=1}^b Y_{\cdot j \cdot}^2 - a^{-1} Y_{\cdots}^2 \tag{14.38}$$

$$\| (Q_a' \otimes \boldsymbol{j}_b') X_\gamma' \boldsymbol{Y} \|^2 = \| Q_a' (Y_{1 \cdot\cdot}, \cdots, Y_{a \cdot\cdot})' \|^2 = \sum_{i=1}^a Y_{i \cdot\cdot}^2 - b^{-1} Y_{\cdots}^2 \tag{14.39}$$

(14.37) (14.38) (14.39) を (14.36) に代入, 整理すると, 結局,

$$\begin{aligned} S_\gamma &= m\Big(\sum_i \sum_j \overline{Y}_{ij\cdot}^2 - a \sum_j \overline{Y}_{\cdot j \cdot}^2 - b \sum_i \overline{Y}_{i\cdot\cdot}^2 + ab\overline{Y}_{\cdots}^2 \Big) \\ &= m \sum_i \sum_j (\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j \cdot} + \overline{Y}_{\cdots})^2 \end{aligned} \tag{14.40}$$

を得る.

以上から, 効果の変動は

$$S_T = \sum_i \sum_j \sum_l (Y_{ijl} - \overline{Y}_{...})^2 \tag{14.41}$$

$$S_\alpha = bm \sum_i (\overline{Y}_{i..} - \overline{Y}_{.j.})^2 \tag{14.42}$$

$$S_\beta = am \sum_j (\overline{Y}_{.j.} - \overline{Y}_{.j.})^2 \tag{14.43}$$

$$S_\gamma = m \sum_i \sum_j (\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...})^2 \tag{14.44}$$

と書き下すことができ, 結局, $\boldsymbol{Z}$ (14.26) のノルムの二乗は

$$S_T = S_\alpha + S_\beta + S_\gamma + S_e \tag{14.45}$$

と平方分解される.

誤差変動は $S_e = \| Q'\boldsymbol{Y} \|^2$ であり, 分散の最良不偏推定量は次であらわされる.

$$\hat{\sigma}^2 = S_e/(n - ab) \tag{14.46}$$

各変動の自由度は次の通りである.

$$\nu_\alpha = a - 1 \tag{14.47}$$

$$\nu_\beta = b - 1 \tag{14.48}$$

$$\nu_\gamma = (a - 1)(b - 1) \tag{14.49}$$

$$\nu_e = n - ab \tag{14.50}$$

結局, 二元配置の分散分析表は次のようになる.

変動	平方和	自由度	平方平均	F 統計量
行因子	S_α	ν_α	S_α/ν_α	$S_\alpha/\nu_\alpha\hat{\sigma}^2$
列因子	S_β	ν_β	S_β/ν_β	$S_\beta/\nu_\beta\hat{\sigma}^2$
交互作用	S_γ	ν_γ	S_γ/ν_γ	$S_\gamma/\nu_\gamma\hat{\sigma}^2$
誤差	S_e	ν_e	S_e/ν_e	
計	S_T	$n - 1$		

既に述べたように, 交互作用のある場合は, 主効果を検定することは無意味である. したがって, 二元配置の場合の解析手順は, 最初に, 交互作用を

$$\frac{S_\gamma/\nu_\gamma}{\hat{\sigma}^2} \sim F^{(a-1)(b-1)}_{n-ab} \tag{14.51}$$

で検定し, 有意でなければ, 行因子, 列因子の主効果を, それぞれ

$$\frac{S_\alpha/\nu_\alpha}{\hat{\sigma}^2} \sim F^{a-1}_{n-ab}, \quad \frac{S_\beta/\nu_\beta}{\hat{\sigma}^2} \sim F^{b-1}_{n-ab} \tag{14.52}$$

で検定する. 行因子, 列因子に関する多重検定は, それぞれ繰り返し数 bm , am の場合の一元配置とみなせばよい. ただし, 分散の推定量は (14.46) としなければならない.

繰り返し数が異なる場合

繰り返し数が異なる場合も, 線型モデルは, (14.19) は繰り返し数が均一な場合と同一でよい. ただし, デザイン行列は (14.21) などのように簡単には表せない.

この場合, (14.45) に相当する平方分解が一意に定まらず, モデルに取り入れる因子の順序の指定などが必要になる. ここでは, 行の因子の効果およびそれを除いた上での列の因子の効果を解析したいとしよう.

それにはまず, 水準 A_i の総実験数を $n_{i.}$, 水準 B_j の総実験数を $n_{.j}$ とする.

$$\boldsymbol{n}_A = n^{-1/2}(\sqrt{n_{1.}}, \cdots, \sqrt{n_{a.}})' \tag{14.53}$$
$$\boldsymbol{n}_B = n^{-1/2}(\sqrt{n_{.1}}, \cdots, \sqrt{n_{.b}})' \tag{14.54}$$

とおき, Q_A' , Q_B' をそれぞれ

$$Q_A'\boldsymbol{n}_A = \boldsymbol{o}, \quad Q_A'Q_A = I_{a-1} \tag{14.55}$$
$$Q_B'\boldsymbol{n}_B = \boldsymbol{o}, \quad Q_B'Q_B = I_{b-1} \tag{14.56}$$

を満たす $a-1$ 行 a 列, $b-1$ 行 b 列の行列とする.

このとき,

$$\Gamma' = \begin{bmatrix} n^{-1/2}\boldsymbol{j}_n' \\ Q_A'\mathrm{diag}\,(n_{i.}^{-1})X_\alpha' \\ Q_B'\{X_{\beta(\alpha)}'X_{\beta(\alpha)}\}^{-1/2}X_{\beta(\alpha)}' \\ (Q_A'\otimes Q_B')\{X_{\gamma(\beta,\alpha)}'X_{\gamma(\beta,\alpha)}\}^{-1/2}X_{\gamma(\beta,\alpha)}' \\ Q' \end{bmatrix} \tag{14.57}$$

は直交行列となる. ここに,

$$X_{\beta(\alpha)} = \{I_n - X_\alpha(X_\alpha'X_\alpha)^{-1}X_\alpha'\}X_\beta \tag{14.58}$$

$$X_{\gamma(\beta,\alpha)} = \{I_n - X_\alpha(X_\alpha'X_\alpha)^{-1}X_\alpha'\}\{I_n - X_{\beta(\alpha)}(X_{\beta(\alpha)}'X_{\beta(\alpha)})^{-1}X_{\beta(\alpha)}'\}X_\gamma \tag{14.59}$$

であり, Q' は

$$Q'\,[\,\boldsymbol{j}_n \quad X_\alpha \quad X_{\beta(\alpha)} \quad X_{\gamma(\beta,\alpha)}\,]\,, \quad Q'Q = I_{n-ab} \tag{14.60}$$

を満たす $n-ab$ 行 n 列行列である.

ここで, $\boldsymbol{Z} = \Gamma'\boldsymbol{Y}$ を Γ' に応じてブロック分割すれば, 繰り返し数が等しい場合と全く同様に要因の効果を分散分析できる.

このとき,

$$\Pi_{\alpha^\perp} = I_n - X_\alpha(X_\alpha'X_\alpha)^{-1}X_\alpha' \tag{14.61}$$

は Im X_α に沿った Ker X_α への直交射影行列であることに注意しよう. すなわち, Γ' (14.57) は n 次元ベクトル全体から成る集合 R^n を

$$R_n = V_0 \oplus V_\alpha \oplus V_{\beta-\alpha} \oplus V_{\gamma-\beta-\alpha} \oplus V_e \tag{14.62}$$

と直和分解していることになる. ここに,

$$\begin{aligned} V_0 &= \mathrm{Im}\ \boldsymbol{j}_n \\ V_\alpha &= (\mathrm{Im}\ X_\alpha) \cap V_0^\perp \\ V_{\beta-\alpha} &= (\mathrm{Im}\ X_\beta) \cap V_0^\perp \cap V_\alpha^\perp \\ V_{\gamma-\beta-\alpha} &= (\mathrm{Im}\ X_\gamma) \cap V_0^\perp \cap V_\alpha^\perp \cap V_{\beta-\alpha}^\perp \\ V_e &= R_n \cap V_0^\perp \cap V_\alpha^\perp \cap V_{\beta-\alpha}^\perp \cap V_{\gamma-\beta-\alpha}^\perp \end{aligned}$$

である.

14.2　繰り返しのない場合

交互作用の自由度は $m-1$ であったから, 繰り返しのない場合は, これを検定することはできない. 交互作用がないとわかっている場合は, $i=1,\cdots,a$; $j=1,\cdots,b$ に対し,

$$Y_{ij}=\mu_{ij}+\varepsilon_{ij}\,,\quad \mu_{ij}=\mu_{i.}+\mu_{.j}-\overline{\mu} \tag{14.63}$$

と表現できる. デザイン行列は

$$X_\alpha = I_a\otimes \boldsymbol{j}_b\,,\quad X_\beta=\boldsymbol{j}_a\otimes I_b$$

とすればよい.

ここで,

$$\boldsymbol{\alpha}=(\mu_{1.}-\overline{\mu},\cdots,\mu_{i.}-\overline{\mu})'\,,\quad \boldsymbol{\beta}=(\mu_{.1}-\overline{\mu},\cdots,\mu_{.j}-\overline{\mu})'$$
$$\boldsymbol{Y}=(Y_{11},\cdots,Y_{ab})'$$

とおくと, (14.63) は,

$$\boldsymbol{Y}=\boldsymbol{j}\overline{\mu}+X_\alpha\boldsymbol{\alpha}+X_\beta\boldsymbol{\beta}+\boldsymbol{\varepsilon} \tag{14.64}$$

と書くことができる. このとき, 行列

$$\varGamma'=\begin{bmatrix} n^{-1/2}\boldsymbol{j}_n' \\ b^{-1/2}Q_a'X_\alpha' \\ a^{-1/2}Q_b'X_\beta' \\ Q' \end{bmatrix} \tag{14.65}$$

が直交行列になるように行列 Q' を定めることができる. $\varGamma'$ による変換を考えると, 変動

$$S_T=\|\boldsymbol{Y}\|^2-\|n^{-1/2}\boldsymbol{j}_n'\boldsymbol{Y}\|^2=\sum_i\sum_j(Y_{ij}-\overline{Y}_{..})^2 \tag{14.66}$$

$$S_\alpha=\|b^{-1/2}Q_a'X_\alpha\boldsymbol{Y}\|^2=b\sum_i(\overline{Y}_{i.}-\overline{Y}_{..})^2 \tag{14.67}$$

$$S_\beta=\|a^{-1/2}Q_b'X_\beta\boldsymbol{Y}\|^2=a\sum_j(\overline{Y}_{.j}-\overline{Y}_{..})^2 \tag{14.68}$$

はそれぞれ (14.41) (14.42) (14.43) において $m=1$ とおいたものと一致する.

ただし, notation は多少異なる：

$$\overline{Y}_{i.} = \frac{1}{a}\sum_{i=1}^{a} Y_{ij}\,, \quad \overline{Y}_{.j} = \frac{1}{b}\sum_{j=1}^{b} Y_{ij}\,, \quad \overline{Y}_{..} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij} \tag{14.69}$$

誤差に関しては

$$S_{e'} = \| Q'\boldsymbol{Y} \|^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})^2 \tag{14.70}$$

となり, 繰り返しのある場合の交互作用項と同じ形式になる.

交互作用のないことが保証されていないにもかかわらず, 実験上の制約から繰り返し測定ができない場合がある. この場合は,

$$\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..} \quad \hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..} \tag{14.71}$$

なる条件のもとで, 統計量

$$S_\gamma = \frac{\{(\hat{\boldsymbol{\alpha}}' \otimes \hat{\boldsymbol{\beta}}')\boldsymbol{Y}\}^2}{\| \boldsymbol{\alpha} \|^2 \| \boldsymbol{\beta} \|^2} = \frac{(\sum_i \sum_j \hat{\alpha}_i \hat{\beta}_j Y_{ij})^2}{\sum_i \hat{\alpha}_i^2 \sum_j \hat{\beta}_j^2} \tag{14.72}$$

を考えればよい. ここに, $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \cdots, \hat{\alpha}_a)'$, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \cdots, \hat{\beta}_b)'$ である. このとき, S_α , S_β , S_γ は互いに独立である. 結局, S_γ と $S_{e'} - S\gamma$ は互いに独立にそれぞれ自由度 1 , $(a-1)(b-1)-1$ の χ^2 分布の σ^2 倍に従う. 以上から,

$$W = \frac{S_\gamma}{(S_{e'} - S_\gamma)/\{(a-1)(b-1)-1\}} \tag{14.73}$$

は $F^1_{(a-1)(b-1)-1}$ に従う. このことは, $S_{e'}$ から交互作用を表す S_γ を抽出したことを意味する.

14.3　交互作用解析

一般に, 交互作用が有意であるというオムニバスな検定だけでは情報が少なく, それがどのような性質であるのかを詳しく議論する必要が生じる. まずその準備として, 実験に関する因子を分類しておこう.

1. 制御因子 : 実験者が自由に水準を定めることのできる再現性のある因子. 多くの実験では, この因子の水準間の差を検討する.
2. ブロック因子 : 実験デザインには入れられるが, 再現性のない因子. 時間変化などがこれに当たるであろう.
3. 表示因子 : 再現性はあるが, 実験者が定められない因子. 癌のステージなどがこれに当たる. 以上の 3 種類を母数因子という.
4. 変動因子 : 実験の段階では再現性がある[*1]が, 結果を一般化する段階では誤差のように働く因子. 臨床研究における患者がしばしばこの因子の例として議論される.

二元配置の場合に, 因子の組み合わせによって解析方法がどのように変わってくるかを考えてみよう. 実験と名がつくからには, 少なくともひとつの因子は制御因子でなければならないであろう. そこで, 2 因子が次の組み合わせになる場合を考える.

■制御因子×制御因子の場合　最適な水準の組み合わせを求めることが目的であろう. 交互作用のない場合は, それぞれの因子の最適水準を組み合わせればよい. 一般の場合にはこの限りでないことは次の例を見ればすぐにわかる.

$$\alpha_1 = -\alpha_2 = 1, -\beta_1 = \beta_2 = 2, \gamma_{11} = \gamma_{22} = -\gamma_{12} = -\gamma_{21} = 3$$

主効果に注目した場合は α_1 , β_2 がそれぞれの最適水準であるが, 実際の最適組み合わせは (α_1, β_1) である.

すべての水準の組み合わせについて, あたかも一元配置のように比較して, 最適な組み合わせを求めるのが現実的であるが, 効率は悪い. 一般に交互作用の自由度は相当に大きいからである.

■表示因子×制御因子の場合　表示因子の水準ごとに, 最適な制御因子の水準を求めることが目的となるであろう. たとえば, 癌のステージごとに最適な治療方法を見つければよい.

[*1] この点で, 誤差とは異なる

この際, 表示因子間の多重比較を行えば, 解析結果がシンプルな行動に結びつく. 面倒な場合分けの各場合について述べた治療指針は実用的ではなかろう. 交互作用を視覚的に確認して, パタンの似通った表示因子をまとめ直すことは有用である. この場合は, 表示因子の値ごとに, 制御因子に関する一元配置データとして解析すればよい.

■変動因子×制御因子の場合　この場合は, 実験の目的がひとつに定まらない. 制御因子の最適水準を求めることが目的である場合は, 交互作用をも誤差に含めて解析するのが一般的である. 実際, 薬の有効性を検討する場合は, 個人ごとの誤差を超えて万人に有効な薬が推奨されるであろう. あるいは, 変動因子の水準が変わった場合に, 最も安定している制御因子を求めたい場合もある.

14.4　繰り返しがブロックをなす場合

要因のすべての組み合わせにおける繰り返しがブロックとみなせる場合は, ブロック効果を取り除くことによって効率があがる. 組み合わせごとに m 人の観測者が別々に観測を行う場合などがこれに当たる. このときは, ブロックに関するデザイン行列を

$$X_b = \boldsymbol{j}_{ab} \otimes I_m \tag{14.74}$$

として,

$$\Gamma' = \begin{bmatrix} n^{-1/2}\boldsymbol{j}_n' \\ (bm)^{-1/2}Q_a'X_\alpha' \\ (am)^{-1/2}Q_b'X_\beta' \\ m^{-1/2}(Q_a' \otimes Q_b')X_\gamma' \\ (ab)^{-1/2}Q_m'X_b' \\ Q' \end{bmatrix} \tag{14.75}$$

による直交変換を考えればよい. Q' は

$$Q'\,[\,\boldsymbol{j}_n \quad X_\alpha \quad X_\beta \quad X_\gamma \quad X_b\,] = O, \quad Q'Q = I \tag{14.76}$$

を満たす $(ab-1)(m-1)$ 行 n 列行列である.

ブロック因子に相当する平方和は

$$S_b = \| (ab)^{-1/2} Q_m' X_b' \boldsymbol{Y} \|^2 = ab \sum_{l=1}^{m} \left(\overline{Y}_{..l} - \overline{Y}_{...}\right)^2 \tag{14.77}$$

である. ただし,

$$\overline{Y}_{..l} = \sum_i \sum_j Y_{ijl} \Big/ ab \tag{14.78}$$

である.

残差平方和は

$$S_{e'} = S_e - S_b \tag{14.79}$$

であり, 再三指摘してきたように, σ^2 の推定量

$$\hat{\sigma}^2 = \frac{S_{e'}}{(ab-1)(m-1)} \tag{14.80}$$

は, (14.46) と比べ, 分母子ともブロック因子の分だけ小さくなる.

以上から, 下の分散分析表を得る.

変動	平方和	自由度	平方平均
行因子	S_α	$\nu_\alpha = a-1$	S_α/ν_α
列因子	S_β	$\nu_\beta = b-1$	S_β/ν_β
交互作用	S_γ	$\nu_\gamma = (a-1)(b-1)$	S_γ/ν_γ
ブロック	S_b	$\nu_b = m-1$	S_b/ν_b
誤差	$S_{e'}$	$\nu_{e'} = (ab-1)(m-1)$	$S_{e'}/\nu_{e'}$
計	S_T	$n-1$	

実際の手順は, まず交互作用を

$$W_\gamma = \frac{S_\gamma/(a-1)(b-1)}{S_{e'}/(ab-1)(m-1)} \tag{14.81}$$

で検定する. これが有意でなければ, つぎに, 行因子, 列因子をそれぞれ

$$W_\alpha = \frac{S_\alpha/(a-1)}{S_{e'}/(ab-1)(m-1)} \tag{14.82}$$

$$W_\beta = \frac{S_\beta/(b-1)}{S_{e'}/(ab-1)(m-1)} \tag{14.83}$$

に基づいて検定すればよい.

14.5　特性値が名義変数の場合

ロジスティックモデル

表示因子×制御因子 の二元配置データで, 反応が 2 値の場合を考える. 当面, 表示因子を層と考え, 層を i ($i = 1, \cdots, a$) , 要因を j ($j = 1, 2$) , 結果を k ($k = 0, 1$) で表す. このとき, 第 i 層において要因が j であるような観測数を n_{ij}, このうち結果が 1 である観測数に対応する確率変数を Y_{ij} とする. n_{ij} は本来は確率変数と考えられるが, ここでは定数と見て, n_{ij} を所与としたときの条件つきで議論を行う.

最も一般的な方法はロジスティック回帰である. これは, Y_{ij} が二項分布 $B(n_{ij}, p_{ij})$ に従うとして, 次のモデルを仮定する.

$$\log \frac{p_{ij}}{1 - p_{ij}} = \overline{\mu} + \alpha_i + \beta_j + \gamma_{ij} \tag{14.84}$$

もうひとつの方法は, すべての (i, j, k) の組み合わせに対するセル確率を p_{ijk} として, 対数線型モデルを仮定することである. このモデルは,

$$\log p_{ijk} = \overline{\mu} + \alpha_i + \beta_j + \gamma_k + \delta_{ij} + \zeta_{ik} + \eta_{jk} + \xi_{ijk} \tag{14.85}$$

を仮定することである. このモデルでは, 層と要因の交互作用は ξ_{ijk} で表されることに注意しよう. 同様に, 層の主効果は ζ_{ik} で, 要因の主効果は η_{jk} で表される. このとき, 反応の対数オッズは

$$\begin{aligned}\log \frac{p_{ij1}}{p_{ij0}} &= (\gamma_1 - \gamma_0) + (\zeta_{i1} - \zeta_{i0}) \\ &\quad + (\eta_{j1} - \eta_{j0}) + (\xi_{ij1} - \xi_{ij0})\end{aligned} \tag{14.86}$$

となる. ここで, 改めて,

$$\gamma_1 - \gamma_0 = \overline{\mu}, \ \ \zeta_{i1} - \zeta_{i0} = \alpha_i$$
$$\eta_{j1} - \eta_{j0} = \beta_j, \ \ \xi_{ij1} - \xi_{ij0} = \gamma_{ij}$$

とおくと, 結局, ロジスティックモデル (14.84) と一致する.

既に述べたように, ロジスティックモデルは要因が多値の場合, 反応が多値の場合, 反応が順序カテゴリの場合にも容易に拡張できる. それぞれの因子の有意性は Wald 検定により行えばよいが, 要因の効果は層との交互作用の存在する場合には一意に定まらない. この場合は, 層ごとに要因の効果を推論しなければならない.

Mantel-Haenszel 検定

層×要因×結果 の交互作用のない場合, さらに

$$\delta_{ij} = 0 : \text{層により例数に差がない.} \tag{14.87}$$

$$\zeta_{ik} = 0 : \text{層により効果に差がない.} \tag{14.88}$$

のどちらが成り立つ場合は層を併合して議論してよい.

3 因子交互作用はないが, (14.87) (14.88) のどちらも成り立たない場合は, 層そのものでなく層の効果を加え併せる. いま, 第 i 層 ($i = 1, \cdots, m$) において次の分割表を得たとしよう. 処理は添え字 j ($j = 1, 2$) で, 反応は添え字 k ($k = 1, 2$) で表す.

		反応 1	反応 2	計
処	1	Y_{i11}	Y_{i12}	$n_{i1.}$
理	2	Y_{i21}	Y_{i22}	$n_{i2.}$
	計	$n_{i.1}$	$n_{i.2}$	$n_{i..}$

すべての層において周辺度数を固定すると, この条件のもとで Y_{i11} の分布は超幾何分布になるから,

$$E[Y_{i11}] = \frac{n_{i1.}n_{i.1}}{n_{i..}}, \quad V[Y_{i11}] = \frac{n_{i1.}n_{i2.}n_{i.1}n_{i.2}}{n_{i..}^2(n_{i..} - 1)} \tag{14.89}$$

を得る. 異なる層のデータは独立であるとして, 統計量

$$\chi^2_{\mathrm{MH}} = \frac{\left\{\sum_i (Y_{i11} - E[Y_{i11}])\right\}^2}{\sum_i V[Y_{i11}]} \tag{14.90}$$

を考える.

すべての層において行因子と列因子の間に関係がないという仮説のもとで, χ^2_{MH} は漸近的に χ^2_1 分布に従う. これに基づく検定を Mantel-Haenszel 検定という. (14.90) は連続性の補正を行うことがある. (13.22) は (14.90) において, すべての j に対し $n_{i..} = 2$ とした場合である.

つぎに, 効果の指標を定義する. まず, 層 i において,

$$Y_{i11} \sim B(n_{i1.}, p_{i1}) \ , \quad Y_{i21} \sim B(n_{i2.}, p_{i2})$$

を仮定する. 超幾何分布と二項分布は漸近的に同等であったことを思い出そう.

第 i 層におけるリスク比 $\phi_i = p_{i1}/p_{i2}$ の推定量を

$$\hat{\phi}_i = \frac{Y_{i11} n_{i2.}}{Y_{i21} n_{i1.}} \tag{14.91}$$

とする. 層により効果に差がないという仮説のもとでは $\hat{\phi}_i$ は 一定の値 ϕ をとるから, (14.91) の分母を払った上ですべての i に関して加えると

$$\phi \sum_i Y_{i21} n_{i1.} = \sum_i Y_{i11} n_{i2.} \tag{14.92}$$

を得る.

以上から, 各層に共通なリスク比 $\phi = p_1/p_2$ の推定量として

$$\frac{\sum_i Y_{i11} n_{i2.}}{\sum_i Y_{i21} n_{i1.}} \tag{14.93}$$

が考えられる. 実際には (14.91) の分母子に共通の荷重 $1/n_{i..}$ を乗じて

$$\hat{\psi}_{\mathrm{MH}} = \frac{\sum_i (Y_{i11} n_{i2.}/n_{i..})}{\sum_i (Y_{i21} n_{i1.}/n_{i..})} \tag{14.94}$$

を ϕ の推定量とする. これを Mantel-Haenzsel リスク比という.

$\hat{\psi}_{\mathrm{MH}}$ の分散を求めよう. $\log \hat{\psi}_{\mathrm{MH}}$ が近似的に期待値 0 の正規分布に従うから, 漸近的には,

$$\chi^2_{\mathrm{MH}} = \frac{(\log \hat{\psi}_{\mathrm{MH}})^2}{V[\log \hat{\psi}_{\mathrm{MH}}]} \tag{14.95}$$

とおくことができる.

これは, Mantel-Haenzsel 検定の χ^2 値と Mantel-Haenzsel リスク比から得られる χ^2 値は同じ効果によるものであるから等しいという考えに基づいている. この式から直ちに

$$V[\log \hat{\psi}_{\mathrm{MH}}] = \frac{(\log \hat{\psi}_{\mathrm{MH}})^2}{\chi^2_{\mathrm{MH}}} \tag{14.96}$$

を得る. この方法を Miettinen の test-based method という.

以上から, ψ_{MH} の $1-\alpha$ 信頼限界は,

$$\frac{\sum_i (y_{i11} n_{i2.}/n_{i..})}{\sum_i (y_{i21} n_{i1.}/n_{i..})} \exp\left\{\pm K_{\alpha/2} \frac{\log \hat{\psi}_{\mathrm{MH}}}{\chi_{\mathrm{MH}}}\right\} \tag{14.97}$$

で与えられる. ここに, χ_{MH} は (14.90) の χ^2_{MH} の正の平方根である. 一般に, この方法で得られる信頼区間はやや狭めに得られることが知られている.

同様に, Mantel-Haenszel オッズ比を

$$\hat{\phi}_{\mathrm{MH}} = \frac{\sum_i (Y_{i11} Y_{i22}/n_{i..})}{\sum_i (Y_{i12} Y_{i21}/n_{i..})} \tag{14.98}$$

で推定する. 近似的な分散は, (14.96) と同様にして

$$V[\log \hat{\phi}_{\mathrm{MH}}] = \frac{(\log \hat{\phi}_{\mathrm{MH}})^2}{\chi^2_{\mathrm{MH}}} \tag{14.99}$$

とすることができ, これから ϕ_{MH} の $1-\alpha$ 信頼限界は

$$\frac{\sum_i (y_{i11} y_{i22}/n_{i..})}{\sum_i (y_{i12} y_{i21}/n_{i..})} \exp\left\{\pm K_{\alpha/2} \frac{\log \hat{\phi}_{\mathrm{MH}}}{\chi_{\mathrm{MH}}}\right\} \tag{14.100}$$

となる.

以上でみた 2×2 表の層別解析を拡張して, $a \times b$ 表の場合を考えよう. 第 i 層において, 次のような分割表を仮定する.

		反　応					計
		1	$\cdots$	k	$\cdots$	b	
処理	1	Y_{i11}	$\cdots$	Y_{i1k}	$\cdots$	Y_{i1b}	$n_{i1.}$
	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	j	Y_{ij1}	$\cdots$	Y_{ijk}	$\cdots$	Y_{ijb}	$n_{ij.}$
	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	a	Y_{ia1}	$\cdots$	Y_{iak}	$\cdots$	Y_{iab}	$n_{ia.}$
	計	$n_{i.1}$	$\cdots$	$n_{i.k}$	$\cdots$	$n_{i.b}$	$n_{i..}$

Y_{ijl} は, 第 i 層において多項超幾何分布に従い, 異なる層においては無相関となる. $E[Y_{ijk}]$, Cov $[Y_{ijk}, Y_{ij'k'}]$ は notation の違いを除き, (4.47) および (4.50) に与えた.

統計量を

$$\boldsymbol{T} = (T_{11}, \cdots, T_{1b}, T_{21}, \cdots, T_{ab})', \quad T_{jk} = \sum_{i=1}^{m} Y_{ijk} \tag{14.101}$$

とすると,

$$E[T_{jk}] = \sum_{i=1}^{m} E[Y_{ijk}] \tag{14.102}$$

$$\text{Cov}\,[T_{jk}, T_{j'k'}] = \sum_{i=1}^{m} \text{Cov}\,[Y_{ijk}, Y_{ij'k'}] \tag{14.103}$$

となるから, 漸近的に

$$(\boldsymbol{T} - E[\boldsymbol{T}])' V[\boldsymbol{T}]^{-} (\boldsymbol{T} - E[\boldsymbol{T}]) \sim \chi^2_{(a-1)(b-1)} \tag{14.104}$$

が成立し, これから検定を構成できる. χ^2 分布の自由度は一般の $a \times b$ 表の場合の適合度の検定と同様にして得られる.

■生存時間の比較　処理ごとに生存時間を比較したいときも, Mantel-Haenzsel 検定の特別な場合として検定を構成できる.

層 (i)	event (−)	event (+)	計
1	$n_{i1} - Y_{i1}$	Y_{i1}	n_{i1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
j	$n_{ij} - Y_{ij}$	Y_{ij}	n_{ij}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
a	$n_{ia} - Y_{ia}$	Y_{ia}	n_{ia}
計	$n_{i.} - d_{i.}$	$d_{i.}$	$n_{i.}$

イベントが起きた時点を t_i $(i = 1, \cdots, m)$ として, i を層の番号と考えればよい. 母集団を j $(j = 1, \cdots, a)$ で表し, 時点 t_i における母集団 j でのイベント数を Y_{ij}, リスク集合の大きさを n_{ij} とする. また, この時点におけるリスク集合全体の大きさを $n_{i.}$, 総イベント数を $d_{i.}$ とする :

$$n_{i.} = \sum_{j=1}^{a} n_{ij} \ , \quad d_{i.} = \sum_{j=1}^{a} Y_{ij} \tag{14.105}$$

w_i を時点 t_i における荷重として, 統計量

$$\boldsymbol{Y} = (Y_1, \cdots, Y_a)' \ , \quad Y_j = \sum_{i=1}^{m} w_i \, Y_{ij} \tag{14.106}$$

を考える. Y_{ij} が多項超幾何分布に従うことから,

$$E[Y_j] = \sum_{i=1}^{m} n_{ij} \, d_{i.}/n_{i.}$$

であり, 異時点におけるイベント数は無相関として, Cov $[Y_j, Y_{j'}]$ は,

$$\sum_{i=1}^{m} w_i^2 \, \frac{n_{ij} \, d_{i.}(n_{i.} - d_{i.})(\delta_{jj'} \, n_{i.} - n_{ij'})}{n_{i.}^2(n_{i.} - 1)} \tag{14.107}$$

となる.

以上から, $\boldsymbol{Y}' V[\boldsymbol{Y}]^{-} \boldsymbol{Y}$ が計算できる. これは漸近的に χ^2 分布に従うが, 制約

$$\sum_{j} Y_j = \sum_{j} E[Y_j] \tag{14.108}$$

より χ^2 分布自由度は $a - 1$ となる.

荷重 w_i をつねに $w_i = 1$ とした場合をログランク検定といい, $w_i = n_{i.}$ とした場合を一般化 Wilcoxon 検定という.

ログランクスコア $\boldsymbol{Y}$ に対して, 対比

$$C = \boldsymbol{\lambda}'\boldsymbol{Y} \tag{14.109}$$

を考えると, 多重検定を構成できる. 実際,

$$E[C] = \boldsymbol{\lambda}'E[\boldsymbol{Y}], \quad V[C] = \boldsymbol{\lambda}'V[\boldsymbol{Y}]\boldsymbol{\lambda} \tag{14.110}$$

より, 漸近的に

$$\frac{(C - E[C])^2}{V[C]} \sim \chi_1^2 \tag{14.111}$$

となる.

とくに, $\boldsymbol{\lambda}'$ として, 処理に使われた用量を定める場合を Tarone 検定という.

任意の対比を考えるときは, データを直交変換すればよい. $V[\boldsymbol{Y}]$ は非負定値対称行列なので, 適当な直交行列 Γ' によって,

$$\Gamma' V[\boldsymbol{Y}]\Gamma = \Sigma = \mathrm{diag}\ (\sigma_j^2) \tag{14.112}$$

と対角化できる. このとき, rank $V[\boldsymbol{Y}] = a - 1$ に注意して,

$$\sigma_1^2 \geq \cdots \geq \sigma_{a-1}^2 > 0 = \sigma_a^2 \tag{14.113}$$

となるように Γ' を定める. このとき,

$$\boldsymbol{Z} = \Gamma'\boldsymbol{Y} = (Z_1, \cdots, Z_a)' \tag{14.114}$$

とおくと,

$$V[\boldsymbol{Z}] = \mathrm{diag}\ (\sigma_j^2) \tag{14.115}$$

となる. したがって,

$$\boldsymbol{\kappa} = \Gamma'\boldsymbol{\lambda} = (\kappa_1, \cdots, \kappa_a)' \tag{14.116}$$

とおくと, C (14.109) の二乗は

$$(\boldsymbol{\lambda}'\boldsymbol{Y})^2 = (\boldsymbol{\lambda}'\Gamma\boldsymbol{Z})^2 = (\boldsymbol{\kappa}'\boldsymbol{Z})^2 \tag{14.117}$$

と書ける.

いま,

$$(\sigma_j^-)^2 = \begin{cases} 1/\sigma_j^2 & \cdots \quad i = 1, \cdots, a-1 \\ 0 & \cdots \quad i = a \end{cases} \tag{14.118}$$

とおき, Cauchy-Schwarz の不等式を用いると,

$$\begin{aligned} C^2 &\leq \left(\sum_j \sigma_j^2 \kappa_j^2\right)\left\{\sum_j (\sigma_j^-)^2 Z_j^2\right\} = (\boldsymbol{\kappa}'\Sigma\boldsymbol{\kappa})(\boldsymbol{Z}'\Sigma^-\boldsymbol{Z}) \\ &= \{\boldsymbol{\lambda}' V[\boldsymbol{Y}]\boldsymbol{\lambda}\}(\boldsymbol{Z}'\Sigma^-\boldsymbol{Z}) = V[C](\boldsymbol{Z}'\Sigma^-\boldsymbol{Z}) \end{aligned} \tag{14.119}$$

を得る. 結局, $C^2/V[C]$ を χ^2_{a-1} で評価すれば, Scheffe 法による保守的な有意確率が得られる.

14.6　特性値が順序カテゴリの場合

反応に順序のある場合は, カテゴリ k にスコア η_k を与え[*2], 処理 j に対するスコアの総和を統計量とする.

$$\boldsymbol{U} = (U_1, \cdots, U_a)', \quad U_j = \sum_{i=1}^{m}\sum_{k=1}^{b} \eta_k Y_{ijk} \tag{14.120}$$

このとき, U_j の平均, 分散は次で与えられる.

$$E[U_j] = \sum_{i=1}^{m}\sum_{j=1}^{a} \eta_k E[Y_{ijk}] \tag{14.121}$$

$$\mathrm{Cov}\,[U_j, U_{j'}] = \sum_{i=1}^{a}\sum_{k=1}^{b}\sum_{k'=1}^{b} \eta_k \eta_{k'} \mathrm{Cov}\,[Y_{ijk}, Y_{ij'k'}] \tag{14.122}$$

したがって, 漸近的に

$$(\boldsymbol{U} - E[\boldsymbol{U}])' V[\boldsymbol{U}]^- (\boldsymbol{U} - E[\boldsymbol{U}]) \sim \chi^2_{a-1} \tag{14.123}$$

が成立する. この検定が一般化拡張 Mantel 検定である. とくに $a = 2$ の場合を拡張 Mantel 検定という.

[*2] 一般には層ごとに異なるスコア η_{ik} を考えることも可能であるが, 各層で同じスコアを与える方が解釈が容易である.

ここでは, 反応に自然な順序を仮定したが, 処理に順序のついている場合も同様である.

処理にも順序のある場合, 処理 j にスコア ξ_j を, 反応 k にスコア η_k を与え, 次の統計量を考える.

$$R = \sum_{i=1}^{m} \sum_{j=1}^{a} \sum_{k=1}^{b} \xi_j \eta_k Y_{ijk} \tag{14.124}$$

Y_{ijk} は多項超幾何分布に従うから, $E[R]$, $V[R]$ は容易に求めることができる. 結局, 漸近的に

$$\frac{(R - E[R])^2}{V[R]} \sim \chi_1^2 \tag{14.125}$$

が成り立つことから検定を構成できる. これを拡張相関検定という.

14.7　水準に順序のついている場合

二元配置データでも, 水準に自然な順序のついている場合がある. このとき, 行と列両方に順序のついている場合と行と列片方に順序のついている場合が考えられる. 後者については, 列に順序がついているとしても一般性を失わない.

行の水準 i , 列の水準 j のときの平均を μ_{ij} とすると, 行列ともに水準を表す番号が大きくなるほど特性値が大きくなる傾向を示す場合は, 片側対立仮説を

$$\mu_{i+1j+1} - \mu_{i+1j} - \mu_{ij+1} + \mu_{ij} \geq 0 \tag{14.126}$$
$$(i = 1, \cdots, a-1;\ j = 1, \cdots, b-1)$$

とすればよい. 行, 列の番号が大きくなるほど特性値が小さくなる傾向がある場合は, 片側対立仮説を

$$\mu_{i+1j+1} - \mu_{i+1j} - \mu_{ij+1} + \mu_{ij} \leq 0 \tag{14.127}$$
$$(i = 1, \cdots, a-1;\ j = 1, \cdots, b-1)$$

とすればよい. 両側仮説は, (14.126) または (14.127) と表現できる.

列のみに順序のついている場合は, $i \neq i'$; $i, i' = 1, \cdots, a-1$; $j = 1, \cdots, b-1$ に対して, 両側対立仮説を

$$\mu_{i'j+1} - \mu_{i'j} - \mu_{ij+1} + \mu_{ij} \leq 0 \quad \text{or} \tag{14.128}$$
$$\mu_{i'j+1} - \mu_{i'j} - \mu_{ij+1} + \mu_{ij} \geq 0 \tag{14.129}$$

と表現できる. この場合は, i, i' に関する対称性から, 片側対立仮説を定義することはできない.

行列ともに順序のついている場合

行および列にそれぞれ単調非減少なスコア

$$\boldsymbol{\xi} = (\xi_1, \cdots, \xi_a)' \ , \ \xi_1 \leq \cdots \leq \xi_a \ , \ \boldsymbol{\xi}' \boldsymbol{j}_a = 0 \tag{14.130}$$
$$\boldsymbol{\eta} = (\eta_1, \cdots, \eta_b)' \ , \ \eta_1 \leq \cdots \leq \eta_b \ , \ \boldsymbol{\eta}' \boldsymbol{j}_b = 0 \tag{14.131}$$

を考える. このとき, $\overline{\boldsymbol{Y}} = (\overline{Y}_{11.}, \cdots, \overline{Y}_{ab.})'$ とすると, 帰無仮説

$$\mu_{ij} = \overline{\mu} \quad \text{for}\ ^{\forall} i, j \tag{14.132}$$

のもとで,

$$(\boldsymbol{\xi}' \otimes \boldsymbol{\eta}')\overline{\boldsymbol{Y}} \sim N(0, m^{-1} \parallel \boldsymbol{\xi} \parallel^2 \parallel \boldsymbol{\eta} \parallel^2 \sigma^2) \tag{14.133}$$

が成立する.

したがって, 帰無仮説のもとで,

$$T = \frac{\sqrt{m}(\boldsymbol{\xi}' \otimes \boldsymbol{\eta}')\overline{\boldsymbol{Y}}}{\{\parallel \boldsymbol{\xi} \parallel^2 \parallel \boldsymbol{\eta} \parallel^2 \hat{\sigma}^2\}^{1/2}} = \frac{\sqrt{m} \sum_i \sum_j \xi_i \eta_j \overline{Y}_{ij.}}{\{\sum_i \xi_i^2 \sum_j \eta_j^2 \hat{\sigma}^2\}^{1/2}} \tag{14.134}$$

は自由度 $n - ab$ の t 分布に従う. ここに, $\hat{\sigma}^2$ は (14.46) である.

この実験デザインの場合の多重比較は簡単ではないが, 行, 列をそれぞれどこで 2 分したときに交互作用が最大になるかを考えれば, 反応パタンの違いに対して答えることができる.

行を $i = 1, \cdots, k$ と $i = k+1, \cdots, a$ に, 列を $j = 1, \cdots, k'$ と $j = k'+1, \cdots, b$ に分けたとき,

$$\frac{1}{kk'} \sum_{i=1}^{k} \sum_{j=1}^{k'} \overline{Y}_{ij.} - \frac{1}{k} \sum_{i=1}^{k} \overline{Y}_{i..} - \frac{1}{k'} \sum_{j=1}^{k'} \overline{Y}_{.j.} + \overline{Y}_{...} \tag{14.135}$$

を考える.

この分散は

$$\frac{1}{m}\left(\frac{1}{k} - \frac{1}{a}\right)\left(\frac{1}{k'} - \frac{1}{b}\right)\sigma^2 \tag{14.136}$$

であるから,

$$\begin{aligned} T_{kk'} = \frac{nkk'}{(a-k)(b-k')\hat{\sigma}} \Bigg(& \frac{1}{kk'} \sum_{i=1}^{k} \sum_{j=1}^{k'} \overline{Y}_{ij.} \\ & - \frac{1}{k} \sum_{i=1}^{k} \overline{Y}_{i..} - \frac{1}{k'} \sum_{j=1}^{k'} \overline{Y}_{.j.} + \overline{Y}_{...} \Bigg) \end{aligned} \tag{14.137}$$

は自由度 $n - ab$ の t 分布に従う. そこで, $T_{kk'}$ または $|T_{kk'}|$ の k, k' に関する最大値を Bonferonni の方法によって調整した有意水準と比較すれば反応パタンの変化を検出できる.

すなわち, 両側検定, 右片側検定をそれぞれ

$$\max |T_{kk'}| > t_{n-ab}(\alpha/2(a-1)(b-1)) \tag{14.138}$$

$$\max T_{kk'} > t_{n-ab}(\alpha/(a-1)(b-1)) \tag{14.139}$$

により構成すればよい.

列のみに順序のついている場合

列にのみ, 単調非減少なスコア (14.131) を与え, 次の統計量を考える.

$$\boldsymbol{U} = (Q'_a \otimes \boldsymbol{\eta}')\,\overline{\boldsymbol{Y}} \tag{14.140}$$

$\boldsymbol{U}$ の分散 $V[\boldsymbol{U}]$ は次のようになる.

$$(Q'_a \otimes \boldsymbol{\eta}')\,\frac{\sigma^2}{m} I_{ab}\,(Q_a \otimes \boldsymbol{\eta}) = \frac{\sigma^2}{m} \sum_j \eta_j^2\, I_{a-1} \tag{14.141}$$

したがって,

$$\frac{m \parallel (Q'_a \otimes \boldsymbol{\eta}')\,\overline{\boldsymbol{Y}} \parallel^2}{(a-1)\hat{\sigma}^2 \sum_j \eta_j^2} \tag{14.142}$$

が F^{a-1}_{n-ab} に従うことから総括的な検定を構成できる.

多重比較は,

$$\boldsymbol{V} = \sqrt{m}(I_a \otimes \boldsymbol{\eta}')\,\overline{\boldsymbol{Y}} = (V_1, \cdots, V_a)' \tag{14.143}$$

が帰無仮説のもとで, 平均 $\boldsymbol{o}$, 分散 $\sigma^2 \sum_j \eta_j^2 I_a$ の a 変量正規分布に従うことから構成できる.

Tukey 法は

$$T_{ii'} = \frac{V_i - V_{i'}}{(\sum_j \eta_j^2)^{1/2}\,\hat{\sigma}} \tag{14.144}$$

を, (12.41) の代わりに用いれば一元配置の場合とまったく同じように有意確率を計算できる. Dunnett 法についても同様である.

Sheffe 法は, (14.143) の任意の対比

$$C = \boldsymbol{\lambda}'\boldsymbol{V} = (\boldsymbol{\lambda}' \otimes \boldsymbol{\eta}')\,\overline{\boldsymbol{Y}} \tag{14.145}$$

を考えればよい. このとき,

$$V[C] = m^{-1}\sigma^2(\boldsymbol{\lambda}'\boldsymbol{\lambda})(\boldsymbol{\eta}'\boldsymbol{\eta}) \tag{14.146}$$

より,

$$\frac{m\{(\boldsymbol{\lambda}' \otimes \boldsymbol{\eta}')\,\overline{\boldsymbol{Y}}\}^2}{(a-1)\sum_i \lambda_i^2 \sum_j \eta_j^2\,\hat{\sigma}^2} \tag{14.147}$$

を $F^{a-1}_{n-ab}(\alpha)$ と比較すると保守的な検定ができる.

第 15 章

分割実験

15.1　分割実験

要因 A の水準を $A_1, \cdots, A_a$, 要因 B の水準 $B_1, \cdots, B_b$ の二要因の実験を考える. このとき, ab 通りの水準の組み合わせに対して, 実験の順番をすべて無作為化するのが二元配置実験である. 一方, 任意の水準 A_i $(i = 1, \cdots, a)$ をブロックとして, b 通りの水準 B_j に対する実験を無作為に配置した実験を考えることがある. 水準の組み合わせ (A_i, B_j) における k 回目 $(k = 1, \cdots, m)$ の実験データに対応する確率変数を Y_{ijk} とする. このとき, モデルは

$$Y_{ijk} = \overline{\mu} + \alpha_i + \epsilon_{ik} + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \tag{15.1}$$

と書くことができる. 異なる j に関して, 共通の ϵ_{ik} を仮定しているので, Y_{ijk} はすべて独立というわけではない.

このように, ブロックに因子を重畳させる実験の方法を分割実験といい, ブロックに重畳させた因子を一次因子, もうひとつの因子を二次因子という. 一次因子に伴う誤差 ϵ_{ik} を一次誤差, 二次因子に伴う誤差 ε_{ijk} を二次誤差という. 分割実験は, 技術的に容易であるだけでなく, 二次因子に関しては, ブロック変動を除くことになるので, 解析上のメリットもある.

一次誤差 $\boldsymbol{\epsilon}$ は $N(\boldsymbol{o}, \sigma_1^2 I_{am})$ に, 二次誤差 $\boldsymbol{\varepsilon}$ は $N(\boldsymbol{o}, \sigma^2 I_n)$ に従うとする. 既に述べたように, これらは互いに独立ではない. この分散構造は後で示す.

モデル (15.1) を行列表現すると,

$$\boldsymbol{Y} = [\,\boldsymbol{j} \quad X_\alpha \quad X_\beta \quad X_\gamma\,]\begin{bmatrix}\mu\\ \boldsymbol{\alpha}\\ \boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{bmatrix} + X_\epsilon\boldsymbol{\epsilon} + \boldsymbol{\varepsilon} \tag{15.2}$$

となる. ここに,

$$\boldsymbol{\epsilon} = (\epsilon_{11}, \cdots, \epsilon_{1m}, \epsilon_{21}, \cdots, \epsilon_{am})' \tag{15.3}$$

$$X_\epsilon = I_a \otimes \boldsymbol{j}_b \otimes I_m \tag{15.4}$$

である. その他は二元配置のときと同一の notation である. ここで, $\boldsymbol{Y}$ の分散 $V[\boldsymbol{Y}]$ は,

$$\sigma^2 I_n + \sigma_1^2 X'_\epsilon X_\epsilon = \sigma^2 I_n + \sigma_1^2 (I_a \otimes \boldsymbol{j}_b \boldsymbol{j}'_b \otimes I_m) \tag{15.5}$$

であることに注意しよう.

今回も, 直交行列

$$\Gamma' = \begin{bmatrix} n^{-1/2}\boldsymbol{j}'_n \\ (bm)^{-1}Q'_a X'_\alpha \\ (am)^{-1}Q'_b X'_\beta \\ m^{-1/2}(Q'_a \otimes Q'_b)X'_\gamma \\ b^{-1/2}(I_a \otimes Q'_m)X_\epsilon \\ Q' \end{bmatrix} \tag{15.6}$$

を考えることができて, $\boldsymbol{Z} = \Gamma'\boldsymbol{Y}$ と変換すれば,

$$\boldsymbol{Z} = (Z_\mu, \boldsymbol{Z}'_\alpha, \boldsymbol{Z}'_\beta, \boldsymbol{Z}'_\gamma, \boldsymbol{Z}'_\epsilon, \boldsymbol{Z}'_\varepsilon)' \tag{15.7}$$

と Γ' に応じて分割できる.

一次誤差平方和は,

$$S_\epsilon = \| \boldsymbol{Z}'_\epsilon \|^2 = b\sum_i \sum_k \left(\overline{Y}_{i.k} - \overline{Y}_{i..}\right)^2 \tag{15.8}$$

である. 二次誤差平方和は二元配置の誤差平方和 S_e から S_ϵ を減じて求めればよい.

$$S_\varepsilon = S_e - S_\epsilon = \sum_i \sum_j \sum_k \left(Y_{ijk} - \overline{Y}_{ij.} - \overline{Y}_{i.k} + \overline{Y}_{i..}\right)^2 \tag{15.9}$$

ただし,

$$\overline{Y}_{i.k} = \frac{1}{b}\sum_{j=1}^{b} Y_{ijk} \quad (i = 1, \cdots, a\,;\ k = 1, \cdots m) \tag{15.10}$$

である.

分割実験においては, まず一次誤差が有意かどうかの検定を行う. 有意でなければ, 二次誤差平方和にプールして分散の推定値としてよい. もっとも, 二次誤差の自由度が十分大きいときは, このプーリングをしなくてもよい.

結局, 分散分析表は次のようになる. 一次因子と一次誤差は不偏分散の平均が σ^2 にならないことに注意しよう.

	平方和	自由度	不偏分散の期待値
一次因子	S_α	$a-1$	$\sigma^2 + b\sigma_1^2$
一次誤差	S_ϵ	$a(m-1)$	$\sigma^2 + b\sigma_1^2$
二次因子	S_β	$b-1$	σ^2
交互作用	S_γ	$(a-1)(b-1)$	σ^2
二次誤差	S_ε	$a(b-1)(m-1)$	σ^2
計	S_T	$n-1$	

一次因子, 一次誤差に関しては, 不偏分散の帰無仮説のもとでの期待値が上のようになる. このことは, (15.5) から簡単に得られる.

一次誤差が有意でない場合は, 通常の二元配置に帰着する. 有意であり, しかも交互作用がない場合は, 一次因子に関する推論を

$$\hat{\sigma}^2 + b\hat{\sigma}_1^2 = S_\epsilon \Big/ a(m-1) \tag{15.11}$$

に基づいて行う. これは分散の推定量が異なるだけで, 繰り返し数 bm の一元配置と同等である.

二次因子に関する推論は, 不偏分散の推定値として, (15.11) の代わりに

$$\hat{\sigma}^2 = S_\varepsilon \Big/ a(b-1)(m-1) \tag{15.12}$$

を用いればよい. これも, 繰り返し数 am の一元配置の分析と本質的に同等である.

一次誤差も交互作用も有意な場合は,

$$\overline{Y}_{ij.} = \frac{1}{m}\sum_k Y_{ijk} \tag{15.13}$$

を μ_{ij} の推定値として, その不偏分散 $(\sigma^2+\sigma_1^2)/m$ を (15.11) (15.12) を合成して求めればよい.

15.2　枝分れ実験

a 社の製品がそれぞれ品質の順に b ランクに分類されているようなときは, その特性を調べる際に次のモデルを考える.

$$Y_{ij} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk} \tag{15.14}$$

i 社の j 番目のランクと i' 社のそれとはなんら共通点がないので, β_j というパラメタは意味がない. むしろ $\beta_{j(i)}$ と書くべきかもしれない.

また, 会社の特性を表す α_i はむしろ誤差と考える方が適切であろう. つまり, 誤差構造を次のようにおく.

誤差	notation	分散
一次誤差	$\boldsymbol{\alpha}$	$\sigma_1^2 I_a$
二次誤差	$\boldsymbol{\epsilon}$	$\sigma_2^2 I_{ab}$
三次誤差	$\boldsymbol{\varepsilon}$	$\sigma^2 I_n$

モデルの行列表現は

$$\boldsymbol{Y} = \boldsymbol{j}\mu + X_1\boldsymbol{\alpha} + X_2\boldsymbol{\epsilon} + \boldsymbol{\varepsilon} \tag{15.15}$$

であり, 分散構造は

$$V[\boldsymbol{Y}] = \sigma_1^2 X_1 X_1' + \sigma_2^2 X_2 X_2' + \sigma^2 I_n \tag{15.16}$$

である. X_1, X_2 はそれぞれ前節の X_α, X_β と同一である.

直交行列

$$\Gamma' = \begin{bmatrix} n^{-1/2}\boldsymbol{j}' \\ (bm)^{-1/2}Q_a'X_1' \\ m^{-1/2}(I_a \otimes Q_b')X_2' \\ (I_{ab} \otimes Q_m') \end{bmatrix} \tag{15.17}$$

により変換し,

$$\boldsymbol{Z} = \Gamma'\boldsymbol{Y} = (Z_\mu, \boldsymbol{Z}_1', \boldsymbol{Z}_2', \boldsymbol{Z}_3')' \tag{15.18}$$

と Γ' に応じて分割する.

このとき, 各変動とその分布は次のようになる.

$$S_1 = \|\boldsymbol{Z}_1\|^2 = bm\sum_i(\overline{Y}_{i..} - \overline{Y}_{...})^2 \sim (bm\sigma_1^2 + m\sigma_2^2 + \sigma^2)\chi^2_{a-1} \tag{15.19}$$

$$S_2 = \|\boldsymbol{Z}_2\|^2 = m\sum_i\sum_j(\overline{Y}_{ij.} - \overline{Y}_{i..})^2 \sim (m\sigma_2^2 + \sigma^2)\chi^2_{a(b-1)} \tag{15.20}$$

$$S_3 = \|\boldsymbol{Z}_3\|^2 = \sum_i\sum_j\sum_k(Y_{ijk} - \overline{Y}_{ij.})^2 \sim \sigma^2\chi^2_{n-ab} \tag{15.21}$$

これから今までと同様に検定を構成できる.

15.3　反復測定データ

高脂血症の薬を投与されてから, 1 週後, 2 週後, $\cdots$ と血清コレステロール値を測定するような必要が生じる場合がある. しばしば見かけるのは, 時点ごとに t 検定を繰り返すことであるが, この方法はまず, 多重性を考慮していない. Bonferroni の方法を用いてこの問題をクリアしたとしても, 解釈の仕方が難しい. 実際, 時点 l において有意であり, $l-1$, $l+1$ において有意でないような場合はどのように解釈するのがよいのか疑問が生じる.

反応パタンがいくつかのパラメタに集約できるときは方法もある. たとえば, 薬剤が投与されてから, 血中濃度が最大になるまでの時間 $T_{\max}$ やそのときの濃度 $C_{\max}$ が代表的である. あるいは縦軸に濃度, 横軸に時間を指定したグラフをプロットし, プロットした曲線の下の面積 (AUC) を考える場合もある.

より一般的な場合, 処理を一次因子, 時間を二次因子, 処理と時間の交互作用 — これが有意だと処理の違いによって時間的な変化のパタンが異なることになる — を一次誤差と考えれば, モデルは本章の分割実験に形式的には帰着する. もっとも, 分割実験は, 一次因子の中で, 二次因子を無作為化できたが, 時間という概念は順序に意味があり, 無作為に並べ替えることは無理がある.

この解析方法は, 対象が人間である場合, もうひとつ大きな問題がある. 前章の交互作用解析の項で述べたが, 一般に処理に対する反応には個人差があり, これを無視して誤差とみなす分析は現実的といえない.

にもかかわらず, 実際にこのような解析が行われているのは, 個体差をモデルに取り込むとパラメタの数が飛躍的に大きくなるからであろう. 実際には処理と時間の交互作用を検定し, これが有意であれば, 変化パタンの似通った個体をグループ分けすることが有用であろう. これは主治医判定として日常行われていることである. グループ分けの際に恣意性が生じるが, 上のデメリットに比べればやむを得ないのでなあるまいか. 分類のための多重比較の方法も提案されている. グループごとに解析する場合は, 個人差と時間の交互作用を改めて誤差とみなして分割実験に対する分散分析を行えばよい.

付録 A

数学的補遺

A.1　線型代数学

行列の定義

　mn 個の複素数 a_{ij} $(i = 1, \cdots, m\,;\ j = 1, \cdots, n)$ を長方形の形に

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix}$$

と並べたものを m 行 n 列行列または $m \times n$ 型行列という. a_{ij} を行列 A の i 行 j 列成分または単に i, j 成分という. i 行 j 列成分が a_{ij} である行列 A を $[a_{ij}]$ と表すことがある. 成分がすべて 0 である行列 O を零行列という.

　$n \times n$ 型行列を n 次正方行列という. n 次正方行列 $A = [a_{ij}]$ の i 行 i 列成分 $(i = 1, \cdots, n)$ を対角成分という. すべての対角成分の和 $a_{11} + \cdots + a_{nn}$ を, 行列 A のトレースといい, $\mathrm{trace} A$ で表す. 対角成分以外 — 非対角成分という — がすべて 0 である n 次正方行列を対角行列という. i, i 成分が $a_{ii} = \alpha_i$ であるような対角行列を $\mathrm{diag}\,(\alpha_i)$ と書くことがある. とくに, $a_{ii} = 1$ $(i = 1, \cdots, n)$ である行列を単位行列といい, I_n または単に I と書く.

とくに, m 行 1 列行列を m 次列ベクトル, 1 行 n 列行列を n 次行ベクトルという. 列ベクトルを $\boldsymbol{a}$ などと書くことがある. すべての成分が 0 である列ベクトルを零ベクトルといい, $\boldsymbol{o}$ で表す. また, すべての成分が 1 である列ベクトルを $\boldsymbol{j}$ で表す. $\boldsymbol{o}$ や $\boldsymbol{j}$ の次数を強調して, $\boldsymbol{o}_n$ や $\boldsymbol{j}_n$ などと書くことがある.

行列の相等

ふたつの行列 A, B の型が一致し, しかもすべての成分が等しいとき, A と B は等しいといい, $A = B$ と表す.

行列の和

ふたつの m 行 n 列行列 $A = [a_{ij}]$, $B = [b_{ij}]$ に対し, i 行 j 列成分が $a_{ij} + b_{ij}$ である m 行 n 列行列を A と B の和といい, $A + B$ で表す.

行列のスカラ倍

m 行 n 列行列 $A = [a_{ij}]$ と複素数 α に対して, i 行 j 列成分が αa_{ij} である m 行 n 列行列を A の α 倍といい, αA で表す.

行列の積

l 行 m 列行列 $A = [a_{ij}]$, m 行 n 列行列 $B = [b_{ij}]$ に対し,

$$c_{ik} = \sum_{j=1}^{m} a_{ij} b_{jk}$$

を i 行 k 列成分にもつ l 行 n 列行列 C を A と B の積といい, AB で表す.

k 行 l 列行列 A, l 行 m 列行列 B, m 行 n 列行列 C に対し, $(AB)C = A(BC)$ が成り立つ. この積を ABC と書く.

正方行列 A の n 乗 $(n = 2, 3, \cdots)$ は $A^n = AA^{n-1}$ により帰納的に定義できる. ある n $(n = 2, 3, \cdots)$ に対して, $A^n = A$ となる行列をべき等行列という.

逆行列

正方行列 A に対し, $AX = XA = I$ となる正方行列 X が存在するとき, A は正則であるといい, X を A の逆行列という. これを $X = A^{-1}$ と書くことがある. ふたつの正方行列 A, B が正則であるとき, AB も正則で, 次が成り立つ.

$$(AB)^{-1} = B^{-1}A^{-1}$$

転置行列

m 行 n 列行列 $A = [a_{ij}]$ に対し, a_{ij} を j 行 i 列成分にもつ n 行 m 列行列を A の転置行列といい, A' で表す.

l 行 m 列行列 A, m 行 n 列行列 B に対し, $(AB)' = B'A'$ である. また, $A' = A$ となる行列を Hermite 行列, $A' = A^{-1}$ となる行列をユニタリ行列という. とくに, A の成分を実数に限れば, $A' = A$ となる行列を対称行列, $A' = A^{-1}$ となる行列を直交行列という.

ブロック分割

m 行 n 列行列 $A = [a_{ij}]$ を

$$A = \left[\begin{array}{c|c|c|c} A_{11} & A_{12} & \cdots & A_{1k} \\ \hline A_{21} & A_{22} & \cdots & A_{2k} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline A_{r1} & A_{r2} & \cdots & A_{rk} \end{array}\right]$$

のように rk 個の行列に分解して考えることをブロック分割という.

ブロック分割された行列の和, スカラ倍, 積に関して, あたかもブロックを成分にもつ行列のように計算することができる.

行列の直積

k 行 l 列行列 $A=(a_{ij})$, m 行 n 列行列 $B=(b_{ij})$ に対して, i, j ブロックが $a_{ij}B$ である km 行 ln 列行列を A と B の直積といい, $A\otimes B$ で表す. この定義から, 結合律 $(A\otimes B)\otimes C = A\otimes(B\otimes C)$ が成り立つ. これを $A\otimes B\otimes C$ と書く.

行列 A , B , C , D と複素数 α に対して, 以下が成り立つ.

$$\begin{aligned}(A\otimes B)' &= A'\otimes B'\\ (A+B)\otimes(C+D) &= A\otimes C + A\otimes D + B\otimes C + B\otimes D\\ (\alpha A)\otimes B &= A\otimes(\alpha B) = \alpha(A\otimes B)\\ (A\otimes C)(B\otimes D) &= (AB)\otimes(CD)\end{aligned}$$

ただし, 最後の式においては, A と B, C と D の積が定義できるものとする.

行列のランク

行列に対して,

1. ある行と他の行を交換する
2. ある行に他の行の定数倍を加える
3. ある行を k 倍する (k は 0 でない定数)

を行に関する基本変形という. 列に関する同様の変形を列に関する基本変形という. 行に関する基本変形と列に関する基本変形を行って, 任意の行列 A を,

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$$

の形に変形できる.

このとき r は変形の仕方によらない. この r を行列 A のランク (階数) といい, $\mathrm{rank}A$ で表す. A が n 次正方行列のとき,

$$A \text{ が正則} \iff \mathrm{rank}A = n$$

である.

ベクトル

ベクトルの定義, 相等, 加法, スカラ倍については既に述べた.

内積

n 次列ベクトル $\boldsymbol{a} = (a_1, \cdots, a_n)'$, $\boldsymbol{b} = (b_1, \cdots, b_n)'$ に対して,

$$\boldsymbol{a}'\overline{\boldsymbol{b}} = a_1\overline{b}_1 + \cdots + a_n\overline{b}_n$$

を $\boldsymbol{a}, \boldsymbol{b}$ の内積といい, $(\boldsymbol{a}, \boldsymbol{b})$ で表す. $\overline{b_i}$ は b_i の共役複素数を表す記号である.

ノルム

n 次列ベクトル $\boldsymbol{a} = (a_1, \cdots, a_n)'$ に対し,

$$\sqrt{(\boldsymbol{a}, \boldsymbol{a})} = \sqrt{a_1\overline{a}_1 + \cdots + a_n\overline{a}_n}$$

を, $\boldsymbol{a}$ のノルムといい, $\| \boldsymbol{a} \|$ で表す. ノルムが 1 であるベクトルを単位ベクトルという.

Cauchy-Schwarz の不等式

n 次列ベクトル $\boldsymbol{x}$, $\boldsymbol{y}$ に対し, 次の不等式が成立する. (Cauchy-Schwarz)

$$|(\boldsymbol{x}, \boldsymbol{y})| \leq \| \boldsymbol{x} \| \| \boldsymbol{y} \|$$

この定理から, $\boldsymbol{x}$, $\boldsymbol{y}$ のなす角 $\theta\ (0 \le \theta \le \pi)$ を

$$\cos\theta = \frac{(\boldsymbol{x}, \boldsymbol{y})}{\|\,\boldsymbol{x}\,\|\|\,\boldsymbol{y}\,\|}$$

で定義することができる.

とくに, $(\boldsymbol{a}, \boldsymbol{b}) = 0$ のとき, $\boldsymbol{a}, \boldsymbol{b}$ は直交するという.

行列式

n 個の自然数 $1, \cdots, n$ を自由に並べ換え, $i_1, \cdots, i_n$ を得たとする. この写像を置換といい, τ_n で表す. $\tau_n(k) = i_k\ (k = 1, \cdots, n)$ となる置換 τ_n を

$$\tau_n = \begin{pmatrix} 1 & \cdots & n \\ i_1 & \cdots & i_n \end{pmatrix}$$

などと書くことがある.

とくに, 互いに異なるある自然数 $i, j\ (1 \le i, j \le n)$ に対し,

$$\tau_n(i) = j, \quad \tau_n(j) = i$$

であり, $k \ne i, k \ne j$ なるすべての $k\ (k = 1, \cdots, n)$ に対しては

$$\tau_n(k) = k$$

となる置換を互換という.

任意の置換は互換のみの合成写像として表すことができる. このときの互換の個数は一意に定まらないが, その偶奇は一致する. 置換 τ_n が偶数個の互換の合成写像として表されるならば偶置換といい, 奇数個の互換の合成写像として表されるならば奇置換という.

置換 τ_n に対し,

$$\operatorname{sgn}\tau_n = \begin{cases} 1 & \cdots \quad \tau_n \text{ が偶置換} \\ -1 & \cdots \quad \tau_n \text{ が奇置換} \end{cases}$$

となる関数 $\operatorname{sgn}\tau_n$ を τ_n の符号という.

n 次正方行列 $A = (a_{ij})$ に対し, $n!$ 個の置換 τ_n すべてに関しての

$$(\mathrm{sgn}\,\tau_n)\ \prod_{k=1}^{n} a_{k\tau_n(k)}$$

の総和を行列 A の行列式といい, $\det A$ や $|A|$ で表す.

n 次正方行列 A, B に対し, 次が成立する.

$$\det AB = \det A \det B$$

線形独立

n 次列ベクトル $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_l$ と複素数 $\lambda_1, \cdots, \lambda_l$ に対して,

$$\boldsymbol{x} = \sum_{i=1}^{l} \lambda_i \boldsymbol{a}_i \tag{A.1}$$

を $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_l$ の線型結合という.

$\boldsymbol{x} = \boldsymbol{o} \Longrightarrow \lambda_1 = \cdots = \lambda_l = 0$ となるとき, $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_l$ は線型独立であるという.

線型空間とその次元

$\boldsymbol{a}_1, \cdots, \boldsymbol{a}_l$ を線型独立なベクトルとする. (A.1) と書けるような $\boldsymbol{x}$ 全体から成る集合 V を $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_l$ の張る線型空間といい, $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_l$ を V の基底という.

基底の取り方は一意ではないが, 基底を構成するベクトルの個数は一定である. たとえば, $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_m$ を V の基底とすると実は $m = l$ である. この l を線型空間 V の次元という.

線型空間 V の基底 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_l$ が, 互いに直交する単位ベクトルであるとき, この基底は正規直交基底をなすという. 任意の基底を用いて正規直交基底を作ることができる. (Gram-Schmidt の直交化)

線型空間 V の部分集合 W も線型空間で[*1], これを線形部分空間または単に部分空間という.

線型写像

線型空間 V から線型空間 W への写像 T が,

1. ${}^{\forall}\boldsymbol{x} \in V$, ${}^{\forall}\boldsymbol{y} \in V$ に対し, $T(\boldsymbol{x}+\boldsymbol{y}) = T(\boldsymbol{x}) + T(\boldsymbol{y})$
2. ${}^{\forall}\boldsymbol{x} \in V$ と任意の複素数 α について $T(\alpha\boldsymbol{x}) = \alpha T(\boldsymbol{x})$

を満たすとき, T を線形写像という. とくに, $V = W$ のときは T を線型変換という.

線形写像 T により V の任意の要素 $\boldsymbol{x}$ が $T(\boldsymbol{x})$ に写されるとき, $T(\boldsymbol{x})$ は行列 A を用いて $T(\boldsymbol{x}) = A\boldsymbol{x}$ と表される. この T を行列 A が表す線形写像という. とくに, A が直交行列のときを直交変換という.

直和, 補空間

V を線型空間とし, V_1, V_2 を V の部分空間とする. V の任意の要素 $\boldsymbol{x}$ が V_1 の要素 $\boldsymbol{x}_1$ と, V_2 の要素 $\boldsymbol{x}_2$ により, 一意に $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$ と表されるとき, V は V_1 と V_2 の直和であるといい,

$$V = V_1 \oplus V_2$$

と書く.

線型空間の部分空間 W に対し,

$$W \oplus W^* = V$$

となる W^* を W の補空間といい, W^C と表す.

[*1] 一般には, より広く, 演算に関する性質により, 公理的に線型空間を定義するので, 線型空間の部分集合は必ずしもそれ自身線型空間にはならない. しかし, 本書のように, 線型空間を基底ベクトルの張る空間と定義すれば, その部分集合は必ず線型空間である.

W を決めても W^* は一意に定まらない. とくに, W の任意の要素 $\boldsymbol{x}$ と W^* の任意の要素 $\boldsymbol{y}$ が直交するとき, W^C を W の直交補空間といい, $W^{\perp}$ と表す.

核と像

V, W を線型空間とし, V から W への線形写像を T とおく. このとき,

$$T^{-1} = \{\boldsymbol{x} \mid \boldsymbol{x} \in V,\ T(\boldsymbol{x}) = \boldsymbol{o}\}$$

は V の部分空間をなす. これを T の核といい, Ker T で表す.

また,

$$T(V) = \{T(\boldsymbol{x}) \mid \boldsymbol{x} \in V\}$$

は W の部分空間をなす. これを T の像といい, Im T で表す. このとき,

$$\dim V = \dim \mathrm{Ker}\ T + \dim \mathrm{Im}\ T$$

が成り立つ.

射影子, 射影行列

線型空間 V が, その部分空間 V_1 と V_2 の直和であるとき, V の要素 $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$ ($\boldsymbol{x}_1 \in V_1$, $\boldsymbol{x}_2 \in V_2$) を $\boldsymbol{x}_1$ に写す変換を V_2 に沿った V_1 への射影子という. この変換は明らかに線型変換で, 行列 $\Pi_{V_1 \cdot V_2}$ で表される. この行列 $\Pi_{V_1 \cdot V_2}$ を射影行列という.

n 次正方行列 Π の表す線型変換の核を Ker Π, 像を Im Π と表すとき, Π が Ker Π に沿った Im Π への射影行列であるための必要十分条件は,

$$\Pi^2 = \Pi$$

である.

とくに, Ker $\varPi$ が Im $\varPi$ の直交補空間となるとき, $\varPi$ を直交射影行列という. n 次正方行列 $\varPi$ が直交射影行列となるための必要十分条件は,

$$\varPi^2 = \varPi \wedge \varPi' = \varPi$$

である.

線型空間 V が, $V = V_1 \oplus W_1 = V_2 \oplus W_2$ の 2 通りの直和として表せるとする. このとき, W_1 に沿った V_1 への射影行列を P_1, W_2 に沿った V_2 への射影行列とすると, 次の 3 命題は同値である.

1. $P_2 - P_1$ は $V_1 \oplus W_2$ に沿った $V_2 \cap W_1$ への射影行列である.
2. $P_1P_2 = P_2P_1 = P_1$
3. $V_1 \subset V_2, W_2 \subset W_1$

固有値と対角化

A を n 次正方行列とする.

$$A\boldsymbol{a} = \lambda\boldsymbol{a}$$

なる複素数 λ と $\boldsymbol{o}$ でない n 次列ベクトル $\boldsymbol{a}$ が存在するとき, λ を行列 A の固有値, $\boldsymbol{a}$ を (λ に属する) 固有ベクトルという. このような λ は

$$\det(A - \lambda I_n) = 0 \tag{A.2}$$

の解として与えられる. (A.2) を (A の) 固有方程式という. (A.2) が相異なる n 個の固有値 $\lambda_1, \cdots, \lambda_n$ をもつならば,

$$P^{-1}AP = \text{diag}\ (\lambda_i)$$

となる行列 P が存在する. P は, 固有値 $\lambda_1, \cdots, \lambda_n$ に属する固有ベクトルをこの順に横に並べて得られる行列である. このとき A は対角化可能であるという[*2]. とくに A が対称 (Hermite) 行列のとき, P は直交 (ユニタリ) 行列となる.

[*2] 実は対角化可能である十分条件はもう少し緩い.

固有方程式 (A.2) の解と係数の関係から

$$\sum_{i=1}^{n} \lambda_i = \mathrm{trace} A, \qquad \prod_{i=1}^{n} \lambda_i = \det A$$

が成立する.

二次形式

n 個の変数 $x_1, \cdots, x_n$ に関する同次二次式

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}\, x_i\, x_j \tag{A.3}$$

を x_i $(i = 1, \cdots, n)$ の二次形式という. $a_{ij} = a_{ji}$ を仮定すると, 任意の二次形式 (A.3) に対し, a_{ij} は一意に定まる.

n 次対称行列 $A = (a_{ij})$, $\boldsymbol{x} = (x_1, \cdots, x_n)'$ を考えると, (A.3) の二次形式は

$$\boldsymbol{x}' A \boldsymbol{x}$$

と書くことができる.

非負定値行列, 正定値行列

以下, 成分が実数である実行列, 実ベクトルに話を限る. 任意の n 次列ベクトル $\boldsymbol{x}$ に対して, $\boldsymbol{x}'A\boldsymbol{x} \geq 0$ となるとき, A を非負定値 (対称) 行列という.

$$A \text{ が非負定値行列} \iff \text{行列 } A \text{ の固有値がすべて非負}$$

が成り立つ. とくに, $\boldsymbol{x} \neq \boldsymbol{o}$ なる任意の n 次列ベクトル $\boldsymbol{x}$ に対して $\boldsymbol{x}'A\boldsymbol{x} > 0$ となるとき, A を正定値 (対称) 行列という.

$$A \text{ が正定値行列} \iff \text{行列 } A \text{ の固有値がすべて正}$$

非負定値行列 A に対して, $X^2 = A$ となる行列 X が存在する. このような X を $A^{1/2}$ と書く. たとえば, A が直交行列 Γ によって対角化され,

$$\Gamma' A \Gamma = \Lambda = \mathrm{diag}\,(\lambda_i)$$

となるとき,

$$\Gamma\,\mathrm{diag}\,(\lambda_i^{1/2})\Gamma'$$

は $A^{1/2}$ である. A が正定値のときは, $A^{-1/2}$ を $(A^{1/2})^{-1}$ で定義する.

一般化逆行列

正方行列 A に対して, $AA^{-}A = A$ なる行列 A^{-} を A の一般化逆行列という.

非負定値対称行列 A に対しては,

$$\Gamma' A \Gamma = \Lambda = \mathrm{diag}\,(\lambda_i)$$

なる直交行列 Γ が存在し, $\lambda_i \geq 0\ (i = 1, 2, \cdots, n)$ である.

ここで,

$$\kappa_i = \begin{cases} 1/\lambda_i & \cdots \quad \lambda_i > 0 \\ 0 & \cdots \quad \lambda_i = 0 \end{cases} \qquad (i = 1, 2, \cdots, n)$$

とおき, $K = \mathrm{diag}\,(\kappa_i)$ とおくと,

$$\Gamma K \Gamma'$$

は A の一般化逆行列を与える. 一般に, 一般化逆行列は一意に定まらない.

A.2　微積分学

関数の連続

$f(x)$ を実数 $\boldsymbol{R}$ のある開区間 I で定義された関数とし, $a \in I$ とする. $\boldsymbol{R}$ は実数全体から成る集合を表す.

$$\lim_{x \to a} f(x) = f(a)$$

ならば, $f(x)$ は a において連続であるという. I 内のすべての点で連続であるとき, $f(x)$ は I において連続であるという.

区間 $\alpha < x < \beta$ で区分的に連続な関数とは, I 上に有限個の点 $\alpha < x_1 < \cdots < x_l < \beta$ をとると, $\alpha < x < x_1$, $x_l < x < \beta$ および各 $x_i < x < x_{i+1}$ $(i = 1, \cdots, l-1)$ において連続で, 各 x_i において左右の極限値が存在するものをいう.

微分係数と導関数

極限

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

が存在するならば, $f(x)$ は a において微分可能であるという. この極限を a における微分係数といい, $f'(a)$ で表す. $f(x)$ が I 内のすべての点で微分可能であるとき, $f(x)$ は I において微分可能であるといい, x の関数 $f'(x)$ を $f(x)$ の導関数という：

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

$y = f(x)$ とおくと, $f'(x)$ を y', dy/dx, $df(x)/dx$ などと表す. $f(x)$ の導関数を求めることを微分するという.

高次導関数

$f(x)$ の導関数 $f'(x)$ が I において微分可能であるとき, $f'(x)$ の導関数を第 2 次導関数といい, $f''(x)$, $d^2 f(x)/dx^2$ などの記号で表す. 第 n 次導関数は, 第 $n-1$ 次導関数の導関数として帰納的に定義でき, これを $f^{(n)}(x)$, $d^n f(x)/dx^n$ などと表す：

$$f^{(n)}(x) = \frac{d}{dx} f^{(n-1)}(x) \quad (n = 2, 3, \cdots)$$

関数 $f(x)$ の 第 k 次導関数が存在し, 第 k 次導関数が I 上で連続であるとき, $f(x)$ は C^k 級であるという. 任意の k に対し, C^k 級である関数を C^∞ 級という. とくに, C^1 級である関数を連続微分可能ということがある.

Taylor の定理

$f(x)$ が C^m 級であるとき,

$$f(x) = f(a) + f'(a)(x-a) + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k + \cdots + \frac{f^{(m-1)}(a)}{(m-1)!}(x-a)^{m-1} + R_m \tag{A.4}$$

と書ける. (Taylor) R_m は剰余項という：

$$R_m = \frac{f^{(m)}(\xi)}{m!}(x-a)^m$$

ただし, ξ はある θ $(0 < \theta < 1)$ に対し, $\xi = a + \theta(x-a)$ と表される.

とくに, $f(x)$ が C^∞ 級のとき, $R_m \to 0$ $(m \to \infty)$ となる x に対して,

$$f(x) = \sum_{m=0}^{\infty} \frac{f^{(m)}(a)}{m!}(x-a)^m$$

と書くことができる. これを Taylor 展開という.

(A.4) を a 近傍の $f(x)$ の近似とみるとき, R_m は $(x-a)^{m-1}$ より高次の無限小である. これを $o((x-a)^{m-1})$ と表すことがある. また, R_m は $(x-a)^m$ と同次以上の無限小であり, これを $O((x-a)^m)$ と表すことがある.

関数の極値

a に十分近い a 以外の任意の x に対して, $f(x) < f(a)$ となるとき, 関数 $f(x)$ は $x = a$ において極大になるといい, $f(a)$ を $f(x)$ の極大値という.

同様に極小を定義する. 関数 $f(x)$ が $x = a$ において極大または極小になるとき, $f(x)$ は $x = a$ において極値をとるといい, $f(a)$ を $f(x)$ の極値という.

$f(x)$ が $x = a$ において微分可能であるとすると, $f(x)$ が $x = a$ において極値をとるために, $f'(a) = 0$ が必要である.

n 変数関数の微分

U を $\boldsymbol{R}^n$ の連結な開集合, $f(x_1, \cdots, x_n)$ を U 上で定義された関数とし, U 内の 1 点を $(a_1, \cdots, a_n)$ とする. $\boldsymbol{R}^n$ は n 次ベクトル全体から成る集合である.

$\boldsymbol{R}^n$ の部分集合 D が連結であるとは, D 内の任意の 2 点 $(a_1, \cdots, a_n)$, $(b_1, \cdots, b_n)$ に対し,

$$a_i = p_i(0),\ \ b_i = p_i(1) \quad (i = 1, \cdots, n)$$

となる n 個の $0 \leq t \leq 1$ 上定義された関数 $p_i(t)$ が存在することをいう.

n 変数関数の連続

$$(x_1 - a_1)^2 + \cdots + (x_n - a_n)^2 \to 0$$

であるならば,

$$f(x_1, \cdots, x_n) \to f(a_1, \cdots, a_n)$$

となるとき, $f(x_1, \cdots, x_n)$ は $(a_1, \cdots, a_n)$ において連続であるという.

U 内のすべての点で連続であるとき, $f(x_1, \cdots, x_n)$ は U において連続であるという.

連続写像

$\boldsymbol{R}^n$ の連結な開集合 U 上で定義され, $\boldsymbol{R}^m$ に値をとる写像

$$F(x_1, \cdots, x_n) = (f_1(x_1, \cdots, x_n), \cdots, f_m(x_1, \cdots, x_n))$$

が U において連続であるとは, 各成分

$$f_k(x_1, \cdots, x_n) \quad (k = 1, \cdots, m)$$

が U において連続関数であることをいう.

偏導関数

$x_1 = a_1, \cdots, x_{i-1} = a_{i-1}, x_{i+1} = a_{i+1}, \cdots, x_n = a_n$ を固定すると,

$$f(a_1, \cdots, a_{i-1}, x_i, a_{i+1}, \cdots, a_n)$$

は x_i に関する 1 変数関数である.

この関数が $x_i = a_i$ において微分可能であるとき, 関数 $f(x_1, \cdots, x_n)$ は点 $(a_1, \cdots, a_n)$ において, x_i について偏微分可能であるという. この微分係数を関数 $f(x_1, \cdots, x_n)$ の $(a_1, \cdots, a_n)$ における x_i についての偏微分係数といい, $f_{x_i}(a_1, \cdots, a_n)$ で表す :

$$\begin{aligned} & f_{x_i}(a_1, \cdots, a_n) \\ & = \lim_{h_i \to 0} \frac{1}{h_i} \Big\{ f(a_1, \cdots, a_{i-1}, a_i + h_i, a_{i+1}, \cdots, a_n) \\ & \qquad\qquad - f(a_1, \cdots, a_{i-1}, a_i, a_{i+1}, \cdots, a_n) \Big\} \end{aligned}$$

関数 $f(x_1, \cdots, x_n)$ が U 内のすべての点 $(x_1, \cdots, x_n)$ において x_i について偏微分可能であるならば,

$$\begin{aligned} & \lim_{h_i \to 0} \frac{1}{h_i} \Big\{ f(x_1, \cdots, x_{i-1}, x_i + h_i, x_{i+1}, \cdots, x_n) \\ & \qquad\qquad - f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n) \Big\} \end{aligned}$$

は $x_1, \cdots, x_n$ の関数である.

これを $f(x_1, \cdots, x_n)$ の x_i についての偏導関数といい,

$$f_{x_i}(x_1, \cdots, x_n)\ ,\ \ \frac{\partial}{\partial x_i} f(x_1, \cdots, x_n)$$

などと表す. 偏導関数を求めることを偏微分するという.

高次偏導関数

$f(x_1,\cdots,x_n)$ の x_i についての偏導関数

$$f_{x_i}(x_1,\cdots,x_n) \quad (i=1,\cdots,n)$$

は $x_1,\cdots,x_n$ の関数である.

これが U 上で x_j $(j=1,\cdots,n)$ について偏微分可能であるとき,

$$\frac{\partial}{\partial x_j}\Big(\frac{\partial}{\partial x_i}f(x_1,\cdots,x_n)\Big)$$

を第 2 次偏導関数という. 以下同様に,

$$\frac{\partial}{\partial x_l}\Big\{\frac{\partial}{\partial x_j}\Big(\frac{\partial}{\partial x_i}f(x_1,\cdots,x_n)\Big)\Big\}$$

などのように高次偏導関数を定義する.

関数 $f(x_1,\cdots,x_n)$ のすべての第 k 次導関数が存在し, それらが定義域 U 上で連続であるとき, $f(x_1,\cdots,x_n)$ は C^k 級であるという. 任意の k に対し, C^k 級である関数を C^∞ 級という.

関数 $f(x_1,\cdots,x_n)$ が C^2 級であるとすると, 次が成り立つ.

$$\frac{\partial}{\partial x_j}\Big\{\frac{\partial}{\partial x_i}f(x_1,\cdots,x_n)\Big\}=\frac{\partial}{\partial x_i}\Big\{\frac{\partial}{\partial x_j}f(x_1,\cdots,x_n)\Big\}$$

この結果を用いると, 関数 $f(x_1,\cdots,x_n)$ が C^m 級であるとすると, 第 m 次偏導関数

$$\frac{\partial}{\partial x_l}\Big\{\frac{\partial}{\partial x_j}\Big(\cdots\frac{\partial}{\partial x_i}f(x_1,\cdots,x_n)\Big)\Big\}$$

は, 偏微分の順序によらないことがわかる. これを

$$\frac{\partial^m}{\partial x_i\partial x_j\cdots\partial x_l}f(x_1,\cdots,x_n)$$

などと書くことがある.

Taylor の定理 (n 変数)

$f(x_1, \cdots, x_n)$ が C^m 級であるとする. このとき,

$$\frac{\partial^{m_1+\cdots+m_n}}{\partial x_1^{m_1} \cdots \partial x_n^{m_n}} f(x_1, \cdots, x_n)$$

を $f^{(m_1,\cdots,m_n)}(x_1, \cdots, x_n)$ と書くと,

$$\begin{aligned} f(x_1, \cdots, x_n) = \sum \frac{f^{(m_1,\cdots,m_n)}(a_1, \cdots, a_n)}{m_1! \cdots m_n!} \\ \times (x_1 - a_1)^{m_1} \cdots (x_n - a_n)^{m_n} + R_m \end{aligned} \tag{A.5}$$

が成り立つ. $\sum$ は, $m_1 + \cdots + m_n \leq m - 1$ なる $(m_1, \cdots, m_n)$ のすべての組み合わせに関する和を表すとする.

剰余項 R_m は

$$R_m = \sum \frac{f^{(m_1,\cdots,m_n)}(\xi_1, \cdots, \xi_n)}{m_1! \cdots m_n!} (x_1 - a_1)^{m_1} \cdots (x_n - a_n)^{m_n}$$

と書ける. $\sum$ は, $m_1 + \cdots + m_n = m$ なる $(m_1, \cdots, m_n)$ のすべての組み合わせに関する和を表すとする. ξ_i $(i = 1, \cdots, n)$ はある θ $(0 < \theta < 1)$ に対し, $\xi_i = a_i + \theta(x_i - a_i)$ と表される.

$f(x_1, \cdots, x_n)$ が C^∞ 級であるならば,

$$\lim_{m \to \infty} R_m = 0$$

となる $(x_1, \cdots, x_n)$ に対して,

$$\begin{aligned} f(x_1, \cdots, x_n) = \sum \frac{f^{(m_1,\cdots,m_n)}(a_1, \cdots, a_n)}{m_1! \cdots m_n!} \\ \times (x_1 - a_1)^{m_1} \cdots (x_n - a_n)^{m_n} \end{aligned}$$

と書ける. $\sum$ は, $(m_1, \cdots, m_n)$ のすべての組み合わせに関する和を表すとする.

ベクトルを用いた表記

$y = f(x_1, \cdots, x_n) = f(\boldsymbol{x})$, $\boldsymbol{x} = (x_1, \cdots, x_n)'$ とおく. このとき, y の $\boldsymbol{x}$ による偏導関数, 第 2 次偏導関数をそれぞれ

$$\frac{\partial}{\partial \boldsymbol{x}} f(\boldsymbol{x}) = \left(\frac{\partial y}{\partial x_1}, \cdots, \frac{\partial y}{\partial x_n}\right)', \quad \frac{\partial^2}{\partial \boldsymbol{x}\,\partial \boldsymbol{x}'} f(\boldsymbol{x}) = \left(\frac{\partial^2 y}{\partial x_i \partial x_j}\right)$$

と書く. $\boldsymbol{\beta}$ を n 次定列ベクトル, A を n 次元定対称行列とすると,

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{\beta}'\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{\beta}) = \boldsymbol{\beta}, \quad \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'A\boldsymbol{x}) = 2A\boldsymbol{x}$$

である.

また, (A.5) の Taylor の定理は, $\boldsymbol{a} = (a_1, \cdots, a_n)'$ として,

$$f(\boldsymbol{x}) = f(\boldsymbol{a}) + (\boldsymbol{x} - \boldsymbol{a})'\frac{\partial f(\boldsymbol{a})}{\partial \boldsymbol{x}} + (\boldsymbol{x} - \boldsymbol{a})'\frac{\partial^2 f(\boldsymbol{a})}{\partial \boldsymbol{x}\,\partial \boldsymbol{x}'}(\boldsymbol{x} - \boldsymbol{a}) + R_3$$

と書くことができる.

n 変数関数の極値

$\boldsymbol{x} = (x_1, \cdots, x_n)'$, $\boldsymbol{a} = (a_1, \cdots, a_n)'$ とし, $f(x_1, \cdots, x_n) = f(\boldsymbol{x})$ とおく. $\|\boldsymbol{x} - \boldsymbol{a}\|$ が十分小さい $\boldsymbol{a}$ 以外の任意の $\boldsymbol{x}$ に対して, $f(\boldsymbol{x}) < f(\boldsymbol{a})$ となるとき, 関数 $f(\boldsymbol{x})$ は $\boldsymbol{a}$ において極大になるという. 同様に極小を定義する. $f(\boldsymbol{x})$ が $\boldsymbol{a}$ において極大または極小となるとき, $f(\boldsymbol{x})$ は $\boldsymbol{a}$ において極値をとるという.

$f(\boldsymbol{x})$ が x_i $(i = 1, \cdots, n)$ について偏微分可能であるとすると, n 変数関数 $f(\boldsymbol{x})$ が $\boldsymbol{a}$ において極値をとるために,

$$\left.\frac{\partial}{\partial \boldsymbol{x}} f(\boldsymbol{x})\right|_{\boldsymbol{x}=\boldsymbol{a}} = \boldsymbol{o}$$

が必要である.

条件つき極値

f, g を $\boldsymbol{R}^2$ の連結な開集合 U において定義された C^1 級関数とし, 条件 $g(x_1, x_2) = 0$ のもとで, $y = f(x_1, x_2)$ は U 内の点 (a_1, a_2) において極値をとるとする. このとき,

$$h(x_1, x_2) = f(x_1, x_2) + \lambda g(x_1, x_2) \quad (\lambda \text{ は } 0 \text{ でない定数})$$

とおくと, 極値を与える (a_1, a_2) および λ は

$$h_{x_1}(x_1, x_2) = h_{x_2}(x_1, x_2) = g(x_1, x_2) = 0$$

から求めることができる. (Lagrange の未定乗数法)

この方法は多変数の場合へ拡張できる. f, g_1, $\cdots$, g_m を $\boldsymbol{R}^n$ の連結な開集合 U において定義された C^1 級関数とする. 条件 $g_k(x_1, \cdots, x_n) = 0$ $(k = 1, \cdots, m)$ のもとで $f(x_1, \cdots, x_n)$ は U 内の点 $(a_1, \cdots, a_n)$ において極値をとるとする. このとき, $\lambda_1, \cdots, \lambda_m$ を 0 でない定数として,

$$h(x_1, \cdots, x_n) = f(x_1, \cdots, x_n) + \sum_{k=1}^{m} \lambda_k g_k(x_1, \cdots, x_n)$$

とおくと, 極値を与える $(a_1, \cdots, a_n)$ および $\lambda_1, \cdots, \lambda_m$ は

$$h_{x_i}(x_1, \cdots, x_n) = g_k(x_1, \cdots, x_n) = 0$$
$$(i = 1, \cdots, n;\ k = 1, \cdots, m)$$

から求めることができる.

原始関数

ある区間 I で定義された連続な関数 $f(x)$ に対して, $F'(x) = f(x)$ となる $F(x)$ があれば, それを $f(x)$ の原始関数という.

$f(x)$ の原始関数は一意ではないが, そのひとつを $F(x)$ とおくと, どの原始関数も $F(x)+C$ (C は定数) の形に書くことができる. 原始関数を求めることを (不定) 積分するという.

定積分

ある区間 I で定義された連続関数 $f(x)$ の原始関数のひとつを $F(x)$, a, b を I 内の点とするとき, $F(b)-F(a)$ を $f(x)$ の a から b までの定積分といい, 次で表す.

$$\int_a^b f(x)\,dx \tag{A.6}$$

広義積分

$f(x)$ が I 上で区分的に連続な場合, (A.6) を以下のように定義する.

1) 区間 $a<x<b$ において a 以外では有界とする. 十分小さい正数 ε に対し, $a+\varepsilon<x<b$ において $f(x)$ が有界で,

$$\lim_{\varepsilon\to 0}\int_{a+\varepsilon}^b f(x)\,dx$$

が存在する場合, これで (A.6) を定義する.

2) 区間 $a<x<b$ において b 以外では有界とする. 十分小さい正数 ε' に対し, $a<x<b-\varepsilon'$ において $f(x)$ が有界で,

$$\lim_{\varepsilon'\to 0}\int_a^{b-\varepsilon'} f(x)\,dx$$

が存在する場合, これで (A.6) を定義する.

3) $a<c<b$ なる c 以外で $f(x)$ が有界で, 十分小さい正数 ε , ε' に対し,

$$\lim_{\varepsilon\to 0}\int_a^{c-\varepsilon} f(x)\,dx+\lim_{\varepsilon'\to 0}\int_{c+\varepsilon'}^b f(x)\,dx$$

が存在する場合は, これで (A.6) を定義する.

このような点 c が有限個の場合も同様に定義する.

4) $-\infty < x < \infty$ において $f(x)$ が定義され, 上の意味で (A.6) が存在するとする. このとき,

$$\lim_{b\to\infty}\int_a^b f(x)\,dx$$

が存在すれば, これを

$$\int_a^\infty f(x)\,dx$$

とする. 同様に

$$\lim_{a\to-\infty}\int_a^b f(x)\,dx = \int_{-\infty}^b f(x)\,dx$$
$$\lim_{a\to-\infty,b\to\infty}\int_a^b f(x)\,dx = \int_{-\infty}^\infty f(x)\,dx$$

を定義する.

重積分

2 変数 x, y の関数 $f(x,y)$ が矩形

$$K = \{(x,y) \mid a \le x \le b\,,\ c \le y \le d\} \tag{A.7}$$

において連続であるとする.

このとき,

$$\int_a^b f(x,y)\,dx$$

は, 区間 $c \le y \le d$ において連続な y の関数であって, 次が定まる.

$$\int_c^d \left(\int_a^b f(x,y)\,dx\right) dy \tag{A.8}$$

これを

$$\int_a^b\int_c^d f(x,y)\,dx\,dy \tag{A.9}$$

と書く.

これを,

$$\int_K f(x,y)\,dx\,dy \tag{A.10}$$

のように略記することがある.[*3]

重積分は, 矩形以外の $\boldsymbol{R}^2$ の有界閉集合 V においても定義できる. たとえば, V が

$$V = \{(x,y) \mid a \le x \le b\,,\ \phi_1(x) \le y \le \phi_2(x)\}$$

と書けるとする. $\phi_1(x)$ および $\phi_2(x)$ は $\boldsymbol{R}$ の区間 $a \le x \le b$ における連続関数である. $f(x,y)$ が V において連続ならば,

$$\int_V f(x,y)\,dx\,dy = \int_a^b \int_{\phi_1(x)}^{\phi_2(x)} f(x,y)\,dy\,dx$$

である. 3 変数以上の場合も同様に重積分を定義する.

重積分における広義積分はより複雑であるが, 本書でたとえば

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

と書いた場合は, 矩形 (A.7) における (A.10) の積分範囲を xy 平面全体に拡げたものを表すとする.

重積分の変数変換

$\boldsymbol{R}^n$ の有界閉集合 V_1 をやはり $\boldsymbol{R}^n$ の有界閉集合 V_2 に写す 1 対 1 の連続微分可能な写像 (各成分が C^1 級関数)

$$(x_1, \cdots, x_n) \to (y_1, \quad \cdots, y_n)$$

を考える. このとき, 逆写像 $(y_1, \cdots, y_n) \to (x_1, \cdots, x_n)$ が存在する. $x_i = \phi_i(y_1, \cdots, y_n)$ $(i = 1, \cdots, n)$ とおくと, これらは V_2 上の n 変数 $y_1, \cdots, y_n$ の C^1 級関数である.

[*3] 数学的には, (A.8) の累次積分と (A.9) の重積分の定義は異なるが, 上の条件のもとで両者は一致する. 本書の範囲内では本文のように定義しても矛盾をきたさない.

いま, n 次正方行列 — Jacobian という—

$$J(y_1,\cdots,y_n)=\left(\frac{\partial x_i}{\partial y_j}\right)$$

を定義し, V_2 において, $J=\det J(y_1,\cdots,y_n)\neq 0$ を仮定する.

このとき, $f(x_1,\cdots,x_n)$ を V_1 で定義された連続関数とすると,

$$\int_{V_1} f(x_1,\cdots,x_n)\,dx_1\cdots dx_n$$

は次のように書き直せる. (重積分の変数変換公式)

$$\int_{V_2} f(\phi_1(y_1,\cdots,y_n),\cdots,\phi_n(y_1,\cdots,y_n))\,|J|\,dy_1\cdots dy_n$$

積分記号下の微分

$f(x,t)$ は矩形 $K=\{(x,t)\mid a\le x\le b\,,\ \alpha\le t\le\beta\}$ において定義された有界な関数で, t を固定したとき x について連続, x を固定したとき t について連続であるとする. このとき,

$$F(t)=\int_a^b f(x,t)\,dx$$

は $\alpha\le t\le\beta$ において t の連続関数である.

さらに, $f(x,t)$ が t について偏微分可能, 偏導関数 $f_t(x,t)$ が K において有界で, t を固定したとき x について連続ならば, $F(t)$ は微分可能で, 次が成り立つ.

$$F'(t)=\int_a^b f_t(x,t)\,dx$$

これは形式的には, 微積分の順序を交換することにほかならない.

A.3　複素解析

複素数

x についての方程式 $x^2 = -1$ は実数解をもたない．この解を $\pm i$ で表すと，二乗して $-2, -3, \cdots$ になる数は $\pm\sqrt{2}i, \pm\sqrt{3}i, \cdots$ と表すことができる．このように，二乗して負になる数を純虚数という．実数と純虚数の和からなる数

$$\alpha = a + bi \quad (a, b \in \boldsymbol{R}) \tag{A.11}$$

を複素数という．a, b をそれぞれ複素数 α の実部，虚部といい，$a = \mathrm{Re}\,\alpha$，$b = \mathrm{Im}\,\alpha$ で表す．(A.11) において，$b = 0$ であれば，α は実数を表し，$a = 0$ であれば，α は純虚数を表す．

複素数の相等

2 個の複素数 $\alpha = a + bi, \beta = c + di$ について

$$\alpha = \beta \Longleftrightarrow a = c\,,\ b = d$$

と定義する．

複素数の四則演算

複素数の四則演算は，実数と同じように行うことができる．

$$\begin{aligned}(a+bi) \pm (c+di) &= (a \pm c) + (b \pm d)i \quad (\text{複号同順})\\ (a+bi)(c+di) &= (ac - bd) + (ad + bc)i\\ \frac{a+bi}{c+di} &= \frac{ac+bd}{c^2+d^2} + \frac{bc-ad}{c^2+d^2}i\end{aligned}$$

共役複素数

複素数 $\alpha = a + bi$ に対して, $a - bi$ を α の共役複素数といい, $\overline{\alpha}$ と表す. α, $\overline{\alpha}$ の実部, 虚部について次が成り立つ.

$$\mathrm{Re}\ \alpha = \mathrm{Re}\ \overline{\alpha} = \frac{\alpha + \overline{\alpha}}{2}, \quad \mathrm{Im}\ \alpha = -\mathrm{Im}\ \overline{\alpha} = \frac{\alpha - \overline{\alpha}}{2}$$

絶対値

複素数 $\alpha = a + bi$ に対して, $\alpha\overline{\alpha} = a^2 + b^2$ の平方根を α の絶対値といい, $|\alpha|$ と表す. 2 個の複素数 $\alpha = a + bi, \beta = c + di$ について, 次が成り立つ.

$$|(a + bi)(c + di)| = |a + bi||c + di|$$
$$\left|\frac{a + bi}{c + di}\right| = \frac{|a + bi|}{|c + di|}$$

複素数平面

複素数 $\alpha = a + bi$ と xy 平面上の点 (a, b) は 1 対 1 に対応する. この平面を複素数平面という. x 軸, y 軸はそれぞれ実軸, 虚軸といわれる. この意味で, 複素数 α と複素数平面上の点 α を同一視することがある.

複素数 α の絶対値は, 複素数平面上, 原点 O と α の距離として表すことができる. より一般に, 2 点 $\alpha = a + bi, \beta = c + di$ の距離は, $\alpha - \beta$ の絶対値に等しい :

$$|\alpha - \beta| = \sqrt{(a - c)^2 + (b - d)^2}$$

複素数の極形式

複素数平面上の点 $\alpha = a + bi$ に対して,

$$r = \sqrt{a^2 + b^2} = |\alpha|, \quad \theta = \arctan(b/a) = \arg\alpha$$

をそれぞれ, α の絶対値, 偏角という. このとき,

$$\alpha = a + bi = r\,(\cos\theta + i\sin\theta)$$

である. この表示を複素数 α の極形式という.

$\exp(i\theta)$ を

$$\exp(i\theta) = \cos\theta + i\sin\theta$$

で定義する —Euler の公式という— と, 極形式は

$$\alpha = r\ \exp(i\theta)$$

と書ける.

複素数 $\alpha = r(\cos\theta + i\sin\theta), \beta = q(\cos\phi + i\sin\phi)$ に対し,

$$\begin{aligned}\alpha\beta &= rq\{(\cos\theta\cos\phi - \sin\theta\sin\phi) \\ &\qquad + i(\cos\theta\sin\phi + \sin\theta\cos\phi)\} \\ &= rq\{\cos(\theta+\phi) + i\sin(\theta+\phi)\}\end{aligned} \tag{A.12}$$

が成り立つ. 同様に次が得られる.

$$\frac{\alpha}{\beta} = \frac{r}{q}\{\cos(\theta-\phi) + i\sin(\theta-\phi)\}$$

(A.12) において, $r = q = 1$ とおき, これを繰り返し用いると

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta \tag{A.13}$$

を得る. これは, n が正の整数の場合だけでなく, 負の整数や有理数のときも成り立つ. これを de Moivre の公式という.

複素数の関数

複素数 z を変数とする関数を複素関数といい, $w = f(z)$ などと表す. z のとりうる値の集合を定義域, w のとりうる値の集合を値域という. 変数 z, w が動く複素数平面をそれぞれ z 平面, w 平面ということがある.

複素関数の連続

z 平面内の連結な開集合 D で定義された複素関数 $w = f(z)$ が D 内の点 α において連続であるとは次が成立することである.

$$\lim_{z\to\alpha} f(z) = f(\alpha)$$

ただし, $z \to \alpha$ とは, $z = x + yi$, $\alpha = a + bi$ とおいたときに, $x \to a$ かつ $y \to b$ となることをいう. D 内のすべての点において $w = f(z)$ が連続であるとき, $w = f(z)$ は D において連続であるという.

複素関数の微分

複素関数 $f(z)$ が D 内の点 α において微分可能であるとは

$$\lim_{\eta\to 0} \frac{f(\alpha+\eta) - f(\alpha)}{\eta}$$

が存在することをいう. この極限を α における微分係数といい, $f'(\alpha)$ で表す. $f(z)$ が D 内のすべての点で微分可能であるとき, $f(z)$ は D において微分可能であるという. このとき, z の関数 $f'(z)$ を $f(z)$ の導関数という :

$$f'(z) = \lim_{\eta\to 0} \frac{f(z+\eta) - f(z)}{\eta}$$

$w = f(z)$ とおくと, $f'(z)$ を w', dw/dz などと表す. $f(z)$ の導関数を求めることを微分するという. 微分可能な複素関数を正則関数という. また, $f(z)$ が $z = \beta$ で微分可能でないとき, この点 β を $f(z)$ の特異点という.

$f(z) = f(x + yi)$ の実部, 虚部をそれぞれ $u(x, y)$, $v(x, y)$ とおく：

$$f(z) = u(x, y) + v(x, y)\, i \tag{A.14}$$

このとき, 微分可能である必要十分条件は次である.

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} \tag{A.15}$$

(A.15) を Cauchy-Riemann の微分方程式という.

実数関数の場合と同様に, $f(z)$, $g(z)$ を D で定義された正則関数, α を定数とするとき, 次の公式が成り立つ.

$$\left\{\alpha f(z)\right\}' = \alpha f'(z)$$

$$\left\{f(z) \pm g(z)\right\}' = f'(z) \pm g'(z)$$

$$\left\{f(z)g(z)\right\}' = f'(z)g(z) + f(z)g'(z)$$

$$\left\{\frac{f(z)}{g(z)}\right\}' = \frac{f'(z)g(z) - f(z)g'(z)}{g(z)^2}$$

$w = f(\zeta)$, $\zeta = g(z)$ で, $f(\zeta)$ が ζ 平面の領域 $D^* = g(D)$ で正則であるとき,

$$\frac{dw}{dz} = f'(g(z))\, g'(z)$$

が成り立つ.

また, $w = f(z)$ の逆関数を $f^{-1}(w)$ で表すと次が成り立つ.

$$\frac{dw}{dz} = \frac{1}{d f^{-1}(w)/dw} \quad \left(\frac{d f^{-1}(w)}{dw} \neq 0\right)$$

代表的な正則関数

n を任意の正整数, α_i $(i = 0, 1, \cdots, n)$ を定数とするとき, z の整関数

$$P(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2 + \cdots + \alpha_n z^n$$

は正則関数で,

$$P'(z) = \alpha_1 + \alpha_2 z + \cdots + n\alpha_n z^{n-1}$$

である.

2 個の整関数 $P(z)$, $Q(z)$ を用いて, $P(z)/Q(z)$ と表される有理関数もまた正則である.

$z = x + iy$ に対して,

$$\exp z = \exp x\,(\cos y + i \sin y)$$

と定義される指数関数は正則で, 次が成り立つ.

$$\frac{d}{dz}\exp z = \exp z$$

三角関数は

$$\cos z = \frac{1}{2}\left\{\exp(iz) + \exp(-iz)\right\}$$
$$\sin z = \frac{1}{2i}\left\{\exp(iz) - \exp(-iz)\right\}$$

で定義する. これらは指数関数として表されるのでやはり正則で, 次が成り立つ.

$$\frac{d}{dz}\cos z = -\sin z\,, \quad \frac{d}{dz}\sin z = \cos z$$

他の三角関数は

$$\tan z = \frac{\sin z}{\cos z}$$

などのように定義すればよい. $\tan z$ も特異点を除いて正則である.

$$\frac{d}{dz}\tan z = \frac{1}{\cos^2 z}$$

複素積分

複素数平面上に, 2 点 P, Q をとり, この 2 点を結ぶ曲線を C で表す. C はパラメタ t を用いて次で表されるとする.

$$z = z(t) = x(t) + i\,y(t) \qquad (t_1 \le t \le t_2)$$

$x(t), y(t)$ はそれぞれ $z(t)$ の実部, 虚部であり, $z(t_1), z(t_2)$ はそれぞれ P, Q を表す.

複素関数 $f(z)$ の C に沿った積分 I を

$$I = \int_C f(z)\,dz = \int_{t_1}^{t_2} f(z(t))\,\frac{dz}{dt}\,dt$$

で定義する. この積分を複素積分といい, C を複素積分の積分路という.

複素積分の性質

複素積分は次の性質をもつ.

1) α, β を任意の定数として,

$$\int_C \{\alpha f(z) + \beta g(z)\}\,dz = \alpha \int_C f(z)\,dz + \beta \int_C g(z)\,dz$$

が成り立つ.

2) ある曲線に沿って P から Q まで積分するときの積分路を C とし, 同じ曲線に沿って Q から P まで積分するときの積分路を $-C$ で表すと,

$$\int_C f(z)\,dz = -\int_{-C} f(z)\,dz$$

が成り立つ.

3) 積分路 C を, 2 個の積分路 C_1 と C_2 に分割できるとき,

$$\int_C f(z)\,dz = \int_{C_1} f(z)\,dz + \int_{C_2} f(z)\,dz \qquad \text{(A.16)}$$

が成り立つ.

4) 積分路 C において, P, Q が一致するとき, C を閉曲線という. 自分自身と交わらない閉曲線を単一閉曲線というが, とくに断わらなくても, 以下閉曲線と書いたときは単一閉曲線を表す. 閉曲線 C に沿って 1 周する積分をとくに周回積分といい,

$$\oint_C f(z)\,dz$$

と書く.

周回積分は, 始点を閉曲線上のどこにとっても値は変わらないが, 閉曲線を逆の向きにまわると値が -1 倍になる. 閉曲線が囲む領域を左に見ながらまわるときを正の向きにまわるという. 周回積分においても (A.16) と同様に, 次が成り立つ.

$$\oint_C f(z)\,dz = \oint_{C_1} f(z)\,dz + \oint_{C_2} f(z)\,dz$$

$f(z)$ の周回積分は (A.14) のように

$$f(z) = u(x,y) + i\,v(x,y), \quad z = x + iy$$

とおいたときに,

$$\oint_C f(z)\,dz = i\int_D \left\{ \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \right) + i \left(\frac{\partial v}{\partial x} + i\,\frac{\partial u}{\partial y} \right) \right\} dx\,dy$$

と表すことができる. D は C が囲む領域である.

これと (A.15) から, 複素関数 $f(z)$ が閉曲線 C で囲まれる領域 D において正則で, 積分路 C 上で連続であるとき,

$$\oint_C f(z)\,dz = 0$$

が成り立つ. これを Cauchy の積分定理という.

この定理は, 1 個の閉曲線で囲まれた領域 —単連結領域という— のみでなく, 一般の連結な領域において成立する.

Cauchy の積分定理から, 複素積分における次の性質を得る.

1) $f(z)$ を単連結領域 D において正則な関数であるとする. D 内の 2 点 P, Q を結ぶ任意の曲線 C_1, C_2 に対して

$$\int_{C_1} f(z)\,dz = \int_{C_2} f(z)\,dz$$

が成立する. つまり, 複素積分は積分路によらない.

したがって, 正則関数 $f(z)$ に対して, 不定積分

$$F(z) = \int_{\alpha}^{z} f(\zeta)\, d\zeta$$

を定義することができる. 不定積分 $F(z)$ の微分は

$$\frac{d}{dz}F(z) = \lim_{\eta \to o} \frac{F(z+\eta) - F(z)}{\eta} = f(z)$$

である.

2) 互いに交わらない閉曲線 C_1 , C_2 があり, その囲む領域がそれぞれ D_1, D_2 ($D_1 \supset D_2$) であるとする. このとき, C_1, C_2 で囲まれる領域を $D = D_1 - D_2$ とし, $f(z)$ は D において正則で, C_1, C_2 上で連続であるとすると

$$\oint_{C_1} f(z)\, dz = -\oint_{C_2} f(z)\, dz$$

が成り立つ. 積分をともに反時計まわりに行うとすれば

$$\oint_{C_1} f(z)\, dz = \oint_{-C_2} f(z)\, dz$$

を得る.

索引

赤池の情報基準, 125
誤り
　第一種の—, 69
　第二種の—, 69

一次因子, 231
一次誤差, 231
一様最強力検定, 75
一様分布, 43
一致性, 87
一般化 Wilcoxon 検定, 224
一般化線型モデル, 120
一般化 Mantel 検定, 185
一般化 logit モデル, 120

Wilcoxon 検定, 146
Wilcoxon の符号順位検定, 106
上側 α 点, 75
Welch の検定, 138
後ろ向き研究, 129

枝分れ実験, 234
efficient score, 83
F 分布, 59

オッズ比, 119, 158

回帰係数, 120
χ^2 分布, 56
拡張相関検定, 226
拡張 Mantel 検定, 225
撹乱母数, 77
確率
　の公理, 4
確率関数, 8
確率空間, 4
確率測度, 4
確率分布, 7
確率変数, 7
片側検定, 70
Kaplan-Meier 推定量, 127
頑健, 98
観測値, 7
Γ 分布, 39

棄却域, 68
危険率, 69
基準ハザード関数, 126
期待値, 9
帰無仮説, 68
強度, 34
共分散, 21
共変量, 120

Kruscal-Wallis 検定, 172

決定係数, 114
検出力, 69
検出力関数, 74
検定, 68
検定関数, 68

交互作用, 204
交絡因子, 194
Cochran-Armitage の検定, 160
Cochran の検定, 197

最強力検定, 72
最小二乗推定量, 110
再生性
　χ^2 分布の—, 56
　正規分布の—, 50
最尤推定法, 86
最尤推定量, 86

Savage 検定, 149
残差, 110
残差平方和, 115

Shaffer の方法, 175
Scheffe の方法, 177
施行, 3
事後確率, 6
事象, 3
指数分布, 41
指数分布族, 92
事前確率, 6
四分偏差 (四分位差), 98
自由度, 56
十分統計量, 91
周辺確率関数, 17
周辺分布関数, 17
主効果, 109, 204
順位検定, 102
順位相関係数
    Kendall の—, 99
    Spearman の—, 99
順序統計量, 98
条件つき確率, 5
条件つき分布, 19
信頼域, 89
信頼区間, 89
信頼係数, 89
信頼限界, 89

水準, 165
推定値, 79
推定量, 79
スコア, 102, 149
スコア検定, 123

正規分布, 37
正規方程式, 110
制御因子, 215
生存関数, 126
積率, 10
積率母関数, 12, 31
漸近的, 36
線型順位統計量, 102
線型モデル, 110
尖度, 11

相関行列, 30
相関係数, 22
相関検定, 190
相似検定, 77

タイ, 98, 106
対応のある t 検定, 161
対数線型モデル, 126
大数の法則, 61, 62
対比, 167
対立仮説, 68
多項超幾何分布, 46
多項分布, 44
多重検定, 172
Dunnett の方法, 178
多変量正規分布, 48
単純仮説, 72

中央値, 98
中央値検定, 147
中心極限定理, 62
超幾何分布, 35

t 検定, 78
t 分布, 58
適合度の検定, 155, 181
デザイン行列, 110
Tukey の方法, 175
点推定, 79

統計量, 53
同時確率関数, 17
同時分布関数, 17
同時密度関数, 18, 24
等分散性の検定, 136
特性関数, 12

並べ替え検定, 96, 139

二項分布, 33
二次因子, 231
二次誤差, 231
二重対数モデル, 120
Newton-Raphson 法, 121

ノンパラメトリック, 95

パーセント点, 98
Bartlett の検定, 170

ハザード関数, 126
外れ値, 98
反復測定, 235

p 値, 70
非心度, 57, 58
非線型モデル, 121
表示因子, 215
標準化, 38
標準誤差, 10
標準正規分布, 37
標準偏差, 10
標本, 60
—の大きさ, 60
標本空間, 3
標本点, 3
比例オッズモデル, 121
比例ハザードモデル, 126

van der Waerden 検定, 145
Fisher-Yates 検定, 144
Fisher の情報行列, 83
Fisher の情報量, 81
Fisher の直接確率法, 155
符号検定, 105
部分尤度関数, 128
不偏推定量, 80
不偏性
検定の—, 76
Friedman の検定, 201
ブロック因子, 215
ブロック効果, 193
probit モデル, 120
分割実験, 231
分割表, 154
分散, 10
分散共分散行列, 29
分散分析
一元配置の—, 168
二元配置の—, 210
分布関数, 8

平均, 9
Page 検定, 202
B 分布, 41
Bernoulli 試行, 33
変数減少法, 125
変数選択, 125
変数増加法, 125
変動因子, 215

Poisson 分布, 34
母集団, 60
母数空間, 78
Holm の方法, 174
Bonferroni の方法, 173

McNemar 検定, 196
マッチング, 194
Mantel 検定, 160
Mantel-Haenszel 検定, 220
Mantel-Haenzsel リスク比, 221
Mann-Whitney 型, 106

Miettinen の test-based method, 221
密度関数, 8

無作為化, 60

有意確率, 70
有意水準, 69
有効推定量, 82
尤度関数, 78
尤度比検定, 78
尤度方程式, 86

Jonckheere の検定, 188

乱塊法, 191
ランダム検定, 71

離散分布, 8
リスク集合, 127
両側検定, 70
両側指数分布, 41
臨界値, 70

連続分布, 8
連続補正, 97

ログランク検定, 224
ロジスティック分布, 43
logistic モデル, 119
logit モデル, 119

歪度, 11
Wald 検定, 122

■著者略歴

医学博士　野 口 千 明 (のぐち ちあき)

1963 年東京生まれ

東京大学医学部医学科卒業

東京大学大学院医学博士課程修了

現在, 都内診療所所長のほか老人ホーム顧問,

イベント会社経営など幅広く活動中

著書

「ケーススタディでみえる医学統計学」 中外医学社

高校数学参考書 (共著), 論文など多数

数理統計の理論と応用

発　行　　2003 年 7 月 29 日 初版 1 刷

著　者　　野 口 千 明

発行者　　デュナミス出版

東京都渋谷区本町 1 - 53 - 8

電話　03 - 3320 - 9751

印刷·製本　中西印刷株式会社

本書の一部または全部を著作権法の定める範囲を

超え, 無断で複写, 複製, 転載することを禁じます.

© 2003 野口千明

ISBN 4-9901788-0-7