目 次

第 1 章　検定の前提となる事柄 **1**

1.1　確率変数 1

1.1.1　標本点、標本空間 1

1.1.2　$\sigma-$ 集合体 1

1.1.3　確率、確率空間 2

1.1.4　確率変数 4

1.1.5　確率分布 5

1.1.6　度数関数と密度関数 5

1.1.7　積率、平均および分散 7

1.1.8　積率母関数 10

1.2　多変数分布 13

1.2.1　離散的二変数分布 14

1.2.2　連続的二変数分布 15

1.2.3　二変数分布の積率 17

1.2.4　多変数分布の積率母関数 19

1.2.5　二変数分布の変数変換 20

1.3　標本分布 25

1.3.1　標本、確率標本 25

1.3.2　標本分布 27

1.3.3　統計量、統計量の分布 27

1.3.4　標本積率 28

1.4　正規分布 31

1.4.1　正規分布 31

1.4.2　正規分布を母集団とする標本集団の性質 34

1.5　ガンマ分布 35

1.6　カイ二乗分布 37

1.7　t 分布 41

第 2 章　統計的決定理論 **46**

2.1　決定理論 46

2.1.1　決定理論の一般的手順 46

2.1.2　損失関数、危険関数 48

第 3 章　母数の推定 **51**
3.1 点推定 51
3.1.1 点推定量の条件 53
3.1.2 点推定の方法 56
3.1.3 積率法 56
3.1.4 最大尤度法 58
3.2 区間推定 61
3.2.1 正規母集団の平均値の区間推定 63
3.2.2 正規母集団の分散の区間推定 64

第 4 章　検定 **67**
4.1 検定の基礎事項 67
4.1.1 検定とは 67
4.1.2 検定で注意すべき点 68
4.1.3 検出力関数 70
4.1.4 尤度比検定 71
4.2 母平均の検定 78
4.2.1 正規母集団の母平均の検定の一般的手順 78
4.2.2 母分散 σ^2 が既知の場合 79
4.2.3 母分散 σ^2 が未知の場合 79
4.2.4 応用例 79
4.3 2 つの正規母集団の母平均の差の検定 80
4.3.1 2 つの母分散 σ_x^2, σ_y^2 が既知の場合 80
4.3.2 2 つの母分散 σ_x^2, σ_y^2 は未知だが、$\sigma_1^2 = \sigma_2^2$ と仮定できる場合 . 81
4.3.3 母分散 σ_x^2, σ_y^2 が共に未知の場合 83
4.4 無相関の検定 83

第 5 章　未解決な問題 **84**

Appendix A. 正規分布の重要性とその応用 **85**

Appendix B. ガンマ関数とその性質 **86**

Appendix C. ベータ関数とその性質 **88**

Appendix D. 用語 **90**
参考文献 90

第1章　検定の前提となる事柄

1.1　確率変数

1.1.1　標本点、標本空間

同一条件下で繰り返される実験あるいは観察を試行といい、試行の結果として起こる事柄を事象という。また、ある試行について、考え得る限りの可能な個々の結果を標本点 (素事象、根源事象)[1]といい、このような結果の全体 (集合[2]) を標本空間といい、Ω で表す。例を挙げると、1 つのサイコロを 1 回投げるとき、偶数の目がでるという事象 A は、$A = \{2, 4, 6\}$、標本空間は、$\Omega = \{1, 2, 3, 4, 5, 6\}$ と表される。

いま、ある事象に対応する部分集合[3]を A とし、試行の結果 ω が A に属する 1 点を表すとき、すなわち

$$\omega \in A$$

であるとき、この試行で事象 A が起こったということにすれば、普通用いられる集合演算に次のような具体的な事象的な意味が対応つけられる。

$A \cup B \cdots\cdots$ 事象 A, B のうち少なくとも 1 つが起こるという事象、
$A \cap B \cdots\cdots$ 事象 A および B がともに起こるという事象、
$A - B \cdots\cdots$ 事象 A が起こり、B が起こらないという事象、
$A^c \quad \cdots\cdots$ 事象 A が起こらないという事象。

1.1.2　$\sigma-$ 集合体

次にこの事象 A に確率という数値を対応させることを考える。確率はある事象が起こる「確からしさ」を表す実数値である。確率を定義するとき、その定義域にあたる集合族をきちんと決める必要がある。この集合族としては、次のような $\mathcal{U}$ (これを $\sigma-$ 集合体ということにする) を考える。$\mathcal{U}$ は次の性質をもっている。 Ω の部分集合 (A など。ここで $A \in \Omega$) からなる、空でない集合族である。

---

[1]根源事象：それ以上細かく分解して考えられない事象。例えば、サイコロを投げたとき、1 の目が出ること。

[2]ある明確に定まった条件に適するものの集まり。

[3]部分集合：2 つの集合 A、B において、B の要素がすべて A の要素であるとき、B は A の部分集合、あるいは B は A に含まれるといい $B \in A$ と書く。

$\mathcal{U}$ の条件 **1** Ω の部分集合 (集合 A など。ここで $A \in \Omega$) からなる、空でない集合族である。

$\mathcal{U}$ の条件 **2** 集合 A が $\mathcal{U}$ に属するとき、補集合 A^c もまた $\mathcal{U}$ に属する。

$\mathcal{U}$ の条件 **3** 集合 $A_i (i = 1, 2, \cdots)$ が $\mathcal{U}$ に属するとき、
和集合 $\cup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup A_3 \cdots$ もまた $\mathcal{U}$ に属する。

これらの条件を満たすから、任意の $\sigma-$ 集合体 $\mathcal{U}$ は次のような性質をもつ。

$\mathcal{U}$ の性質 **1** 常に元の空間 Ω (確実に起こる事象の意味にとれる) を含む。
なぜなら、$\mathcal{U}$ に属している一つの集合を A とおくと、$\mathcal{U}$ の条件 2 より $A^c \in \mathcal{U}$.

$\mathcal{U}$ の性質 **2** 常に空集合[4] ϕ(起こり得ない事象の意味) を含む。
なぜなら、上と同様に $\mathcal{U}$ に属している一つの集合を A とおくと、$\mathcal{U}$ の条件 2 より $A^c \in \mathcal{U}$. また $\mathcal{U}$ の条件 3 より

$$\Omega = A \cup A^c \in \mathcal{U}, \quad \phi = \Omega^c \in \mathcal{U}.$$

$\mathcal{U}$ の性質 **3** A_1, A_2 が $\mathcal{U}$ に属するときは $A_1 \cap A_2$ もまた $\mathcal{U}$ に属す。

なぜなら $A_1 \cap A_2 = (A_1^c \cup A_2^c)^c \in \mathcal{U}$.

$\mathcal{U}$ の性質 **4** A_1, A_2 が $\mathcal{U}$ に属するとき、$A_1 - A_2$ もまた $\mathcal{U}$ に属す。

なぜなら $A_1 - A_2 = (A_1^c \cup A_2)^c \in \mathcal{U}$.

$\mathcal{U}$ の性質 **5** 一般的に、$\mathcal{U}$ に属する集合の間に $\cup, \cap, -, c$ の集合演算を可付番無限回ほどこして得られる集合もまた $\mathcal{U}$ に属する。

ここで、Ω と $\mathcal{U}$ の例を挙げておく。1 つのサイコロを 1 回投げる場合を考えると、$\Omega = \{1, 2, 3, 4, 5, 6\}$ に対し、
$\mathcal{U} = \{1\}, \{2, 4\}, \{1, 2, 3, 4, 5, 6\}, \{\phi\}, \cdots$
となる。

1.1.3　確率、確率空間

$\sigma-$ 集合体 $\mathcal{U}$ のすべての集合 A に対し、一意的に定められた関数 $P(A)$ が次の条件を満たすとき、$P(A)$ を事象 A が起こる (または事象 A の) 確率という。

確率の条件 **1** $P(A) \geq 0, \quad (A \in \mathcal{U})$,

確率の条件 **2** $P(\Omega) = 1$,

[4] 要素 (集合を構成する個々のもの、元ともいう) を全く含まない集合のこと。

確率の条件 3 $A_i \in \mathcal{U}, A_i \cap A_j = \phi, (i \neq j)$ であるとき、

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \tag{1.1}$$

注意

1. 確率の条件 1 の中に $P(A) \leq 1$ が欠けているようにみえるが、これは他の 2 条件が満たされれば成り立つ。
   証明

   $$\Omega = A \cup A^c, A \cap A^c = \phi \tag{1.2}$$

   よって確率の条件 3 および確率の条件 2 より

   $$P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1.$$

   $$P(A) \geq 0, \quad P(A^c) \geq 0$$

   なので、

   $$0 \leq P(A) \leq 1, \quad 0 \leq P(A^c) \leq 1. \quad \text{(証明終わり)}$$

2. 確率の条件 3 は確率事象 $A_1, A_2, \cdots$ のどの 2 つをとっても排反事象 (同時に起こることがあり得ない事象) であるとき、$A_1, A_2, \cdots$ のうち少なくとも 1 つの事象が起こる確率は、各事象の起こる確率の和で表されるという意味である。これは $A_1, A_2, \cdots$ が有限個の場合も含んでいる。すなわち $n < i$ に対し $A_i = \phi$ とおけば確率の条件 3 より

   $$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i).$$

3. 上の結果は通常確率の加法定理と呼ばれるもので、ここでは証明なしに公理的に与えられた性質であるが、確率の条件 1 ~ 3 があればいろいろな確率の結果を導くことができる。例えば次の一般化された加法定理も示される

   $$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

   確率の加法定理の証明
   集合 $A \cup B$ は次のように独立な 3 つの集合に分けられる。

   $$A \cup B = (A - B) \cup (B - A) \cup (A \cap B).$$

   よって

   $$P(A \cup B) = P(A - B) + P(B - A) + P(A \cap B).$$

   ここで

   $$P(A) = P(A - B) + P(A \cap B), \quad P(B) = P(B - A) + P(A \cap B).$$

変形して上式に代入し、整理すると

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \qquad (\text{証明終わり})$$

以上により今後、確率について考えるときには、理論的にはまずそれに対し空間 Ω を定め、その部分集合よりなる $\sigma-$ 集合体 $\mathcal{U}$ を決め、その集合体に属する任意の集合に対し上のような確率 P が定義されなければならないことになる。この $\Omega, \mathcal{U}, P$ をまとめて $(\Omega, \mathcal{U}, P)$ と書き、これを確率空間という[5]。

例として、ある試行の起こり得る結果の数が有限個または可付番無限個の場合[6]の確率空間は次のように定められる。$\Omega = \{\omega_1, \omega_2, \cdots, \omega_n, \cdots\}$ とおく。$\sigma-$ 集合体 $\mathcal{U}$ としては Ω のすべての部分集合 (空集合 ϕ を含む) よりなる集合属を考える[7]。次に確率 P は次のような数列 $\{p_i\}$ で定義する。

$$P(\omega_i) = p_i, \quad p_i \geq 0, \quad \sum_{i=1}^{\infty} p_i = 1.$$

$\mathcal{U}$ に属する任意の集合

$$A_i = \{\omega_{i1}, \omega_{i2}, \cdots, \omega_{ik}\}$$

に対しては $P(A_i) = \sum_{l=1}^{k} p_{il}$ とし、特に $P(\phi) = 0$ とする。このように定義された $P(A_i)$ は、確率の条件 $1. \sim 3.$ を明らかに満たす[8]。

1.1.4　確率変数

標本空間を構成する標本点 (試行の結果) に対応して定められる実数値を確率変数という[9]。確率変数自身は確率を表すものではなく、サイコロ投げの例では「$1 \sim 6$ までのいずれかの目が出る」という可能性を表すものに過ぎない。以下では確率変数 X の数学的な定義を示す。

まず確率空間を $(\Omega, \mathcal{U}, P)$ とおく。Ω の上で定義された実数値関数[10] $X(\omega), (\omega \in \Omega)$ が次の条件、「任意の実数 x に対し、$X(\omega) \leq x$ を満たす ω の集合が常に $\sigma-$ 集合体 $\mathcal{U}$ に属する」、即ち

$$\{\omega; X(\omega) \leq x\} \in \mathcal{U}$$

[5] 標本空間 Ω の元 ω は標本点である。P は $\mathcal{U}$ 上で定義される集合関数である。

[6] 連続だと可付番にはならない。ex. 実数値には番号が付けられない。

[7] $\mathcal{U}$ を定義するのは、空集合 ϕ を含むため、より厳密に確率を定義できるからと考えられる。$P(A) = 0$ があり得るのは $\mathcal{U}$ があるため。確率 $P(\mathcal{U})$ は定義できないことに注意!

[8] $A_1 = \{\omega_{11}, \omega_{12}\}, A_2 = \{\omega_{21}, \omega_{22}\}, B = \{\omega_{11}, \omega_{12}, \omega_{21}, \omega_{22}\}, C = \{\omega_{11}, \omega_{22}\}$ とする。このとき $P(B) = P(A_1) + P(A_2), P(C) = P(A_1) + P(C) - p_{11}$. つまり、独立な事象の確率のみ足されていくことになるので、条件 3 を満たしていると言える。

[9] 例 1) サイコロ投げの場合。　標本点：1 から 6 までの目が出るという結果の各々。確率変数：1,2,3,4,5,6。
例 2)1 枚のコインを 2 回投げたときの表の数。　標本点：s_1 ={ 表、表 }, s_2 ={ 表、裏 },s_3 ={ 裏、表 },s_4 ={ 裏、裏 } の各々。確率変数 (表の数)：$x_1 = 2, x_2 = 1, x_3 = 1, x_4 = 0$。

[10] ☆確率変数 X が $\mathcal{U}$ に属することで、各々の X についてではなく、まとめて議論することができる。

を満たすとき、この $X(\omega)$ を確率変数という。

1.1.5　確率分布

集合 $\{\omega; X(\omega) \leq x\}$ は $\mathcal{U}$ に属しているから、この集合の $\mathcal{U}$ に対する確率が与えられ、次のように表される。

$$P(X \leq x) = P\{\omega; X(\omega) \leq x\} = F(x).$$

これを確率変数 X の分布関数という。

確率変数 X がとる値を支配している偶然性は、ある確率法則により規定される。その確率法則を確率分布ということにする。この確率分布は上の分布関数で表されていることになる。分布関数 $F(x)$ には次の性質がある。

分布関数の性質 1 $0 \leq F(x) \leq 1.$

分布関数の性質 2 $\lim_{x\to-\infty} F(x) = F(-\infty) = 0, \quad \lim_{x\to\infty} F(x) = F(+\infty) = 1.$

分布関数の性質 3 $x < y$ ならば $F(x) \leq F(y)$.[11]

分布関数の性質 3 から明らかなように、分布関数 $F(x)$ は x の単調増加関数である。

$P(a < X \leq b)$ と $F(a), F(b)$ の関係は以下のようになる。ここで $a, b(a < b)$ は X の変域内の二つの値とする。

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b)$$

であるから、X が a より大で b を超えない確率 $P(a < X \leq b)$ は明らかに

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) \tag{1.3}$$

である。

1.1.6　度数関数と密度関数

この節では、確率変数の分布関数上のある点またはある微小部分における確率を定義することを考える。確率変数が離散的な場合と連続的な場合について分けて考える方が分かりやすい。

[11] $x < y$ ならば $(-\infty, x) \subset (-\infty, y)$ である。

離散的確率変数の場合

確率変数 X が離散量 $x_1 < x_2 < x_3 < \cdots$ をとるとき、個々の離散量に対応する確率が定まる。離散的確率変数 X が特定の値 x_i をとる確率を

$$f(x_i) = P(X = x_i)$$

で示し、これを離散的確率変数の度数関数という。$f(x)$ を X の変域内のすべての値に対して考えるとき、この $f(x)$ を変数 X の離散的確率分布あるいは簡単に確率分布という。度数関数は明らかに次の性質を持つ。

1. $0 \leq f(x_i) \leq 1$.
2. $\sum_i f(x_i) = 1$.

逆に任意の関数 $f(x)$ が x のある変域に対してこの二つの条件をみたすとき、その関数は度数関数と考えることができる。

度数関数 $f(x)$ と分布関数 $F(x)$ との関係は次式で示される。確率変数 X が離散量 $x_1 < x_2 < x_3 < \cdots$ をとるとき、その分布関数 $F(x)$ は離散量 $x_1, x_2, x_3, \cdots$ ごとに上昇するが、x_i と x_{i+1} の中間においては不変である。故に離散的確率変数の分布は個々の離散量に対応する確率によって定まる。

$$F(x) = P(X \leq x) = \sum_{i; x_i \leq x} f(x_i). \tag{1.4}$$

この 式の右辺は x を超えない x_i のすべての値に対する $f(x_i)$ の値の総和を意味する。この関係式により分布関数 $F(x)$ を度数関数 $f(x_i)$ から計算することができ、逆に $f(x_i)$ を $F(x)$ から計算することもできる。

連続的確率変数の場合

確率変数が連続量の場合、変数の特定の値に対して non ゼロ の確率を定めることはできない。しかし分布関数が変域内で微分可能な場合 (以下このような分布関数のみを考える)、連続的確率変数に対する確率を次のような形で定めることができる。次の式で定義される $f(x)$ を確率変数 X の確率密度あるいは密度関数という。

$$f(x) = \lim_{h \to 0} \frac{P(X \leq x + h) - P(X \leq x)}{h}$$

密度関数 $f(x)$ の定義式から直ちに密度関数と分布関数 $F(x)$ との関係を示す次の式が得られる。

$$f(x) = \frac{dF(x)}{dx} \tag{1.5}$$

$f(x)dx = dF(x)$ をしばしば確率素分という。確率素分 $f(x)dx$ は点 x の付近における確率である。厳密にいえば連続変数の場合 x の特定値に対して確率はゼロで

ある。確率素分は変数 X のきわめて小さい区間に対する確率 (次式で表される) と考えればよい。

$$\int_x^{x+\Delta x} f(x)dx \cong f(x)\Delta x$$

(1.5) 式から直ちに明らかなように

$$F(x) = \int_{-\infty}^{x} f(t)dt \tag{1.6}$$

である。したがって $a < b$ のとき、X が区間 (a, b) の値をとる確率は

$$\begin{aligned} P(a < X \leq b) &= \int_a^b f(x)dx \\ &= F(b) - F(a) \end{aligned} \tag{1.7}$$

である[12]。

連続関数の場合、確率密度関数が次の性質をもつことは明らかである。

1. $f(x) \geq 0$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.

1 は分布関数 $F(x)$ が非減少であることの当然の結果であり、2 は分布関数の性質 $F(\infty) = 1$ と同値である。密度関数 $f(x)$ はこの二つの条件をみたさなければならないし、反対に任意の関数 $f(x)$ が一定の区間でこの二条件をみたすとき、それは密度関数と考えることができる。

1.1.7　積率、平均および分散

ここでは確率変数 X が連続変数である場合について書くが、離散変数の場合も同様である。
X が確率変数で $f(x)$ がその密度関数のとき、任意の実数値 a を原点とする (k 次の) 積率を次の式で定義する。

$$E\left[(X-a)^k\right] = \int (x-a)^k f(x)dx / \int f(x)dx$$

ここで、$a = 0$ とおいた場合を原点に対する X の k 次積率、あるいは原点積率といい、これを記号 μ'_k で示す。また、$a = \mu$ とおくと、平均に対する積率 μ_k が定義され、これを中心積率という。

一次の原点積率を期待値もしくは平均値という。すなわち、

$$\mu'_1 = E\left[(X^1)\right] = \int_{-\infty}^{\infty} xf(x)dx / \int_{-\infty}^{\infty} f(x)dx. \tag{1.8}$$

[12] 点 a, b において確率はゼロであるから、区間 $(a,b)[a,b)(a,b][a,b]$ の確率はいずれも同じ。(記号 ( ) は開区間 (端の値を含む)、[ ] は閉区間を示す。)

$E(X)$ が発散するときは期待値は存在しない[13]。以下、平均値 μ_1' を記号 μ で示すことにする。平均値 μ は分布の中心 (正確には重心) を示す。

二次の中心積率を分布の分散 (あるいは単に分散) という。記号 σ^2 または $\sigma^2(X)$、σ_X^2, あるいは $var(X)$ で示す。すなわち[14]

$$\begin{aligned}\sigma^2 &= \mu_2 = E\left[(X-E(X))^2\right] = E\left[(X-\mu)^2\right] \\ &= \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx \Big/ \int_{-\infty}^{\infty} f(x)dx. \end{aligned} \tag{1.9}$$

分散の正の平方根を標準偏差という。記号 σ で示す。すなわち

$$+\sqrt{\sigma^2} = \sigma = +\sqrt{\mu_2}.$$

分散 $\sigma^2 = \mu_2$ は分布のばらつきの大小を示す。ただし変数 X と同次の量でばらつきの大小を示すためには分散の平方根である標準偏差 σ が便利である。

任意の a を原点とする二次積率と二次の中心積率との間には次の関係が成り立つ[15]。

$$\begin{aligned} E\left[(X-a)^2\right] &= E\left[\{(X-\mu)+(\mu-a)\}^2\right] \\ &= E\left[(X-\mu)^2 + 2(X-\mu)(\mu-a) + (\mu-a)^2\right] \\ &= \int\{(x-\mu)^2 + 2(x-\mu)(\mu-a) + (\mu-a)^2\}f(x)dx \Big/ \int f(x)dx \\ &= \int(x-\mu)^2 f(x)dx + 2(\mu-a)\int(x-\mu)f(x)dx + (\mu-a)^2 \\ &= E\left[(X-\mu)^2\right] + 2(\mu-a)E(X-\mu) + (\mu-a)^2 \\ &= \mu_2 + (\mu-a)^2 \end{aligned} \tag{1.10}$$

(なぜならば　　$E(X-\mu) = 0$) [16]

$(\mu-a)^2 \geq 0$ であるから次の関係が成り立つ。

$$E(X-a)^2 \geq \mu_2$$

故に二次積率の中で最も小さいものが分散である。

[13] というらしい。もしくは「考えない」のだと思われる
[14] 離散変数の場合は $\sigma^2 = \sum_i (x_i-\mu)^2 f(x_i) / \sum_i f(x_i)$
[15] 3 行目から 4 行目で、密度関数の性質より $\int f(x)dx = 1$ である。
[16]

$$\begin{aligned} E(X-\mu) &= \int(x-\mu)f(x)dx \Big/ \int f(x)dx \\ &= (\int xf(x)dx - \mu\int f(x)dx) \Big/ \int f(x)dx \\ &= (\mu - \mu\cdot 1)/1 = 0 \end{aligned}$$

次に変数 X の一次変換 $Y = aX + b$(a は目盛の変更、b は座標原点の変更を示す) を考える。$E(Y) = aE(X) + b$ であるから[17]、

$$\begin{aligned}\sigma^2(Y) &= E\left[Y - E(Y)\right]^2 \\ &= E\left[(aX+b) - \{aE(X)+b\}\right]^2 \\ &= E\left[a\{X - E(X)\}\right]^2 = a^2 E[X - E(X)]^2 \\ &= a^2\sigma^2(X) \end{aligned} \tag{1.11}$$

故に目盛の変更は分散の大きさに関係するが、座標の原点の変更は分散に無関係である[18]。

分散 (二次の積率) について知られる性質をまとめる。

分散の性質 1　c を定数とすれば

$$\sigma^2(c) = 0.^{19} \tag{1.12}$$

分散の性質 2

$$\sigma^2(X + c) = \sigma^2(X).^{20} \tag{1.13}$$

分散の性質 3

$$\sigma^2(cX) = c^2\sigma^2(X). \tag{1.14}$$

分散の性質 2 と 3 は (1.11) の直接の結果である。

---

[17]
$$\begin{aligned}\sigma^2(Y) &= \int \{y - E(Y)\}^2 f(x)dx = \int \{ax + b - E(aX+b)\}^2 f(x)dx \\ &= \int \{ax + b - aE(X) - b\}^2 f(x)dx = a^2 \int \{x - E(X)\}^2 f(x)dx = a^2\sigma^2(X)\end{aligned}$$

[18]例えば $y = 100x$ のとき $\int f(y)dy = 1$ は成り立たない。即ち変換後は確率密度関数とは言えない。$\int f'(y)/dy/100$ を考える。

[19]
$$\sigma^2[g(x)] = \int_{-\infty}^{\infty} \{g(x) - E[g(x)]\}^2 f(x)dx$$

であるから、

$$\sigma^2(c) = \int_{-\infty}^{\infty} \{c - c\}^2 f(x)dx = 0.$$

[20]$X + c$ の平均は $\mu + c$ であるから、

$$\sigma^2(X + c) = \int_{\infty}^{\infty} \{(x + c) - (\mu + c)\}^2 f(x)dx = \sigma^2(X).$$

分散の性質 4 二次の原点積率と中心積率との間には次の関係が成り立つ[21][22]。

$$\sigma^2(X) = \sigma_X^2 = \mu_2 = \mu_2' - \mu^2 \tag{1.17}$$

以上では中心積率、すなわち平均値に対する積率 $E(X-\mu)^k$ や、任意の原点 a に対する積率 $E(X-a)^k$ を考えた。もっと一般的には、X の任意の関数 $g(X)$ についての積率を次のように定義できる。確率変数 X の任意の関数を $g(X)$ とするとき、$g(X)$ の k 次の積率は[23]

$$E[\{g(X)\}^k] = \int_{-\infty}^{\infty} [g(x)]^k f(x)dx \qquad (\text{連続関数}). \tag{1.18}$$

なお参考までに、3 次の中心積率 μ_3 は分布のゆがみの大小を示し、4 次の中心積率 μ_4 は同じく尖り方の大小を示すことを加えておく。

X が離散変数の場合について付け加えておく。

$$E\left[(X-a)^k\right] = \sum_i (x_i - a)^k f(x_i) / \sum_i f(x_i). \tag{1.19}$$

$$E(X) = \sum x_i p_i = \mu, \quad Var(X) = \sum (x_i - \mu)^2 p_i. \tag{1.20}$$

1.1.8　積率母関数

確率変数の積率は統計の理論および実際で重要な役割をもっている。前節のように積率は確率変数の分布の種々の性質を定めることができ、普通の場合すべての積率が与えられればそれで確率分布の形が決定する[24]。従ってすべての積率を包括的に代表するような関数を見出すことができれば大変好都合である。その一つが積率母関数である。確率分布の形によっては、積率の定義式に基づいて積率を計算するときに、特別の工夫をしなければならない場合や、計算の面倒な場合があるが、積率母関数を使うと積率を一定の方法でたいてい簡単に計算することができる。X を確率変数、t を実変数とするとき、次式で定義される関数 $M_x(t)$ を X の積率母関数という。

$$M_x(t) = E(e^{tX}).$$

[21] 同様に三次の中心積率も次のように原点積率の項で示すことができる。

$$\mu_3 = \mu_3' - 3\mu\mu_2' + 2\mu^3 \tag{1.15}$$

一般に次式が成り立つ。

$$\mu_k = \mu_k' - \binom{k}{1}\mu_{k-1}'\mu + \cdots + (-1)^i \binom{k}{i}\mu_{k-i}'\mu^i + \cdots + (-1)^{k-1}(k-1)\mu^k \tag{1.16}$$

ただし、$\binom{k}{i} = {}_kC_i = \frac{k\cdot(k-1)\cdot\cdots\cdot(k-i+1)}{i!}$. これらの関係式を使って中心積率を原点積率から計算することができる。逆に原点積率を中心積率から計算する式を作ることもできる。

[22] (1.10) 式で $a=0$ とおくと、$\mu_2' = \mu_2 + \mu^2 = \sigma^2 + \mu^2$.

[23] 離散関数の場合は $\sum_i [\{g(x_i)\}^k] f(x_i)$.

[24] 級数の考え方と似ているのだろうが、実際にグラフが描けるかはよく分からない。

X が離散変数ならば

$$M_X(t) = \sum_x e^{tx} f(x), \tag{1.21}$$

X が連続変数ならば

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx. \tag{1.22}$$

指数関数 e^{tX} をベキ級数に展開[25]すると、積率母関数を次のような級数の形で示すことができる。

$$\begin{aligned} M_X(t) = E(e^{tX}) &= E(1 + tX + \frac{1}{2!}t^2X^2 + \cdots + \frac{1}{k!}t^kX^k + \cdots) \\ &= 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \cdots + \frac{t^k}{k!}E(X^k) + \cdots \\ &= 1 + \mu'_1 t + \mu'_2 \frac{t^2}{2!} + \cdots + \mu'_k \frac{t^k}{k!} + \cdots \end{aligned} \tag{1.24}$$

故に積率母関数をこのようにベキ級数の形に展開することができれば、その展開式の $t^k/k!$ の項の係数は k 次の原点積率 μ'_k を与える。これが $M_X(t)$ を積率母関数 (積率を生み出す関数) と呼ぶ理由である。

積率母関数から積率を計算するためには次の方法を用いることもできる。積率母関数を t について k 回微分すると

$$\begin{aligned} \frac{d^k}{dt^k} M_X(t) &= E\left[\frac{d^k}{dt^k} e^{tX}\right] \\ &= E(X^k e^{tX}) \end{aligned}$$

となり[26]、$t = 0$ とおけば

$$M_X^{(k)}(0) = E(X^k) = \mu'_k \tag{1.25}$$

となる。故に母関数を k 回微分して $t = 0$ とおくことにより k 次の原点積率が得られる (この結果は 1.24 式を項ごとに微分して $t = 0$ とおいても得られる[27])。
X を確率変数、$g(X)$ をその関数とするとき、$g(X)$ の積率母関数を次の式で定義する。

$$M_{g(X)}(t) = E(e^{tg(X)}) \tag{1.26}$$

[25] 一般的なベキ級数展開は

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + \cdots. \tag{1.23}$$

この公式の x に対応するのが tX である。

[26] $\frac{d^k}{dt^k} M_X(t) = \int (\frac{d^k}{dt^k} e^{tx}) f(x) dx.$ これは ( ) 内についての平均、即ち $E\left[\frac{d^k}{dt^k} e^{tx}\right]$ である。また e^{tX} を t について k 回微分すれば $X^k e^{tX}$ となる。

[27] t について k 回微分したときの $k - n$ 番目の項は $\frac{1}{(k-n)!} t^{k-n} \cdot \mu'_{k-n}$.$k$ 番目より前では定数の微分になるため 0、k 番目より後では $t = 0$ を入れたときに 0 となり、結局 μ'_k しか残らない。

X の積率母関数 $M_X(t)$ をベキ級数に展開した (1.24) 式で、e^{tx} を $e^{tg(X)}$ に置き換えると次の式が得られる。

$$E(e^{tg(X)}) = 1 + tE\left[g(X)\right] + \frac{t^2}{2!}E\left[g(X)\right]^2 + \cdots + \frac{t^k}{k!}E\left[g(X)\right]^k + \cdots$$

積率母関数についての定理を示す。

積率母関数の定理 1　X を確率変数、$g(X)$ を任意の関数、c を任意の定数とすれば

$$M_{cg(X)}(t) = M_{g(X)}(ct). \tag{1.27}$$

<証明>　(1.26) 式で $g(X) = cg(X)$ として、

$$M_{cg(X)}(t) = \int_{-\infty}^{\infty} e^{tcg(X)} f(x)dx.$$

(1.26) 式で t を ct で置き換えると、

$$M_{cg(X)}(t) = M_{g(X)}(ct).$$

(証明終わり)

(1.27) 式で $g(X) = X$ のときは

$$M_{cX}(t) = M_X(ct) \tag{1.28}$$

積率母関数の定理 2　X を確率変数、$g(X)$ をその関数、c を任意の定数とすれば

$$M_{g(X)+c} = e^{ct} M_{g(X)}(t). \tag{1.29}$$

<証明>

$$\begin{aligned} M_{g(X)+c}(t) &= \int_{-\infty}^{\infty} e^{t(g(x)+c)} f(x)dx \\ &= \int_{-\infty}^{\infty} e^{tg(x)} \cdot e^{ct} f(x)dx \\ &= e^{ct} \cdot M_{g(X)}(t). \end{aligned}$$

(証明終わり)

式 (1.29) で $g(X) = X$ のとき

$$M_{X+c}(t) = e^{ct} M_X(t) \tag{1.30}$$

式 (1.29) で $g(X) = X, c = -\mu$ のとき

$$M_{X-\mu}(t) = e^{-\mu t} M_X(t) \tag{1.31}$$

$M_{X-\mu}(t)$ は中心積率母関数と呼ばれ、確率変数の平均値に対する偏差の積率母関数となっている。この中心積率母関数は原点積率母関数に $e^{-\mu t}$ をかけたものに等しい。$M_{X-\mu}(t)$ をベキ級数に展開するか[28]、または $M_{X-\mu}(t)$ を k 回微分して $t=0$ とおけば k 次の中心積率が得られる。

$$\frac{d^k}{dt^k}M_{X-\mu}(t)|_{t=0} = \mu_k.^{29} \tag{1.34}$$

1.2　多変数分布

[30]二つまたは二つ以上の確率変数を同時に考える必要がしばしば起こる。これは標本空間の各標本点に二つまたは二つ以上の実数値が相伴って対応している場合といえる。例えば標本点 ω を任意に抽出された世帯とし、これに三つの確率変数 $X(\omega), Y(\omega), Z(\omega)$ を対応させる。X は収入、Y は家族の人数、Z は食費と考えてよい。この三変数は相互に関連しながら確率的に変化する。ある場合には標本空間そのものが n 次元空間であって、標本点は n 個の座標によって決定されるその空間内の一点を定めることも考えられる。例えば 2 枚のコインを投げる場合、表を 1、裏を 0 とすると、標本点は $(1,1),(1,0),(0,1),(0,0)$ であり、標本空間は 2 次元となる。

このような n 個の座標によって決定される標本点についても確率が定まる。この章では二つ (以上) の確率変数の同時分布について考える。

---

[28](1.31) 式に (1.24) 式を代入すると大変。つまり、

$$M_{X-\mu}(t) = e^{-\mu t}\cdot E(e^{tX}) = e^{-\mu t}\cdot E(1+tX+\frac{1}{2!}t^2X^2+\cdots) \tag{1.32}$$

とすると、とっても面倒になる。下のようにすればいい。

$$\begin{aligned} M_{X-\mu}(t) = E(e^{t(X-\mu)}) &= E\{1+t(X-\mu)+\frac{1}{2!}t^2(X-\mu)^2+\cdots+\frac{1}{k!}t^k(X-\mu)^k+\cdots\} \\ &= 1+tE(X-\mu)+\frac{1}{2!}t^2E\{(X-\mu)^2\}+\cdots \end{aligned} \tag{1.33}$$

[29]

$$\begin{aligned} M_{X-\mu}(t) &= \int_{-\infty}^{\infty} e^{t(x-\mu)}f(x)dx\cdot\frac{d^k}{dt^k}M_{X-\mu}(t) \\ &= \int_{-\infty}^{\infty} e^{t(x-\mu)}(x-\mu)^k f(x)dx\cdot\frac{d^k}{dt^k}M_{X-\mu}(t)|_{t=0} = \int_{-\infty}^{\infty}(x-\mu)^k f(x)dx = \mu_k. \end{aligned}$$

[30]カイ二乗分布、t 分布を議論する際、本当は多変数 (三変数以上) についての、確率変数の和の分布を知る必要がある。このノートでは、積率母関数を除き、二変数について扱うものとする。

1.2.1　離散的二変数分布

確率変数の定義された標本空間を S とし、X, Y を Ω の上で定義された二つの離散的確率変数とする。任意の標本点が $X = x_i,\ Y = y_j$ という実現値をとる確率は、このような値をとる標本点の確率の和に等しい。この確率を

$$f(x_i, y_j) \equiv P(X = x_i, Y = y_j) \tag{1.35}$$

で示す。上の式を $i = 1, 2, \cdots, m,\ j = 1, 2, \cdots, n$ の全てについて考えるとき、これを二変数 X, Y の同時度数関数という。明らかに

$$f(x_i, y_j) \geq 0,$$

$$\sum_i \sum_j f(x_i, y_j) = 1$$

である。X, Y の同時分布関数は次の式で定義される。

$$F(x_i, y_j) = P(X \leq x_i, Y \leq y_j) = \sum_{x \leq x_i} \sum_{y \leq y_j} f(x, y). \tag{1.36}$$

いま Y の値は何でもいいからとにかく x_i がでる確率 $P(X = x_i)$(これを $g(x_i)$ とおく) は

$$g(x_i) = \sum_j f(x_i, y_j)$$

である。この計算は x_i のすべての値に対してすることができるから、

$$g(x) = \sum_y f(x, y) \tag{1.37}$$

は X の確率分布を示す。$g(x)$ を X の周辺度数 (関数) という。同様にして

$$h(y) = \sum_x f(x, y) \tag{1.38}$$

は Y の周辺度数 (関数) を与える。X, Y の同時密度関数が与えられるとき、その周辺度数は一意的に定まる。

Y の値が一定値 (例えば y) に定められたときの X の条件付き確率を $f(x|y)$ とすれば[31]

[31] 確率の教科書では、条件つき確率は以下の様に求められている。事象 A が起こったときに事象 B が起こるという条件つき確率を $P_A(B)$ で表す。事象 A, B を表す集合をそれぞれ $\mathbf{A}, \mathbf{B}$ とし、事象 A の起こる場合の数を $n(\mathbf{A})$ で表す。ただし $n(\mathbf{A}) \neq 0$ とする。また標本空間を $\mathbf{U}$ とする。事象 A も B も起こる場合は $\mathbf{A} \cap \mathbf{B}$ で表される。この $\mathbf{A} \cap \mathbf{B}$ は集合 $\mathbf{A}$ の部分集合である。従って、条件つき確率は

$$P_A(B) = \frac{n(\mathbf{A} \cap \mathbf{B})}{n(\mathbf{A})} \tag{1.39}$$

$$f(x|y) = \frac{f(x,y)}{h(y)} \quad h(y) \neq 0 \tag{1.41}$$

である。いまもし X と Y が互いに独立ならば

$$f(x|y) = g(x), \quad f(y|x) = h(y)$$

であるから

$$f(x,y) = g(x)h(y) \tag{1.42}$$

である。

1.2.2　連続的二変数分布

X, Y を標本空間 Ω の上で定義された二つの連続的確率変数とし、A を xy 平面上の点 (x, y) の任意の部分集合とする。このとき点 (x, y) がこの部分集合に属することに対して一定の確率が定まるならば、二変数 X, Y の同時分布を定めることができる。いま変域 $-\infty < x < \infty, -\infty < y < \infty$ のすべての x, y に対して非負の関数 $f(x, y)$ が存在し、xy 平面上の任意の領域 A に対して次の関係が成り立つとき、$f(x, y)$ を二変数 X, Y の同時密度関数という。

$$P\left[(X = x, Y = y) \in A\right] = \int\int_A f(x,y)dxdy.^{32} \tag{1.43}$$

明らかに

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

である。同時分布関数は次の式で定義される。

$$\begin{aligned} F(x,y) &= P(X \leq x, Y \leq y) \\ &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dydx. \end{aligned} \tag{1.44}$$

故に一変数の場合と同様にこの場合も密度関数は分布関数を微分して求めることができる。

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y). \tag{1.45}$$

となる。条件つき確率の場合、分母が $n(\mathbf{A})$ となることに注意しなければならない。単に事象 A も B も起こるという確率は

$$P(\mathbf{A} \cap \mathbf{B}) = \frac{n(\mathbf{A} \cap \mathbf{B})}{n(\mathbf{U})}$$

である。(1.39) 式の右辺の分母、分子を $n(\mathbf{U})$ で割ると、条件つき確率は次の式で表される。

$$P_A(B) = \frac{P(\mathbf{A} \cap \mathbf{B})}{P(\mathbf{A})}. \tag{1.40}$$

[32] 分布関数なので、右辺の値は 0 から 1 である。

離散変数の場合と同様に、変数 X, Y の単独の密度関数 (周辺分布という) は次の式によって与えられる[33]。

$$\begin{aligned} g(x) &= \int_{-\infty}^{\infty} f(x,y)dy, \\ h(y) &= \int_{-\infty}^{\infty} f(x,y)dx. \end{aligned} \tag{1.46}$$

即ち一変数の分布は二変数の同時分布をもう一方の変数に関して積分することによって導かれる。

条件付き確率の場合と同様、周辺分布についても

$$f(x|y) = \frac{f(x,y)}{h(y)}, \quad f(y|x) = \frac{f(x,y)}{g(x)} \tag{1.47}$$

が成り立つ。これらの式で定められる分布を Y(または X) に対する、あるいは Y(または X) を条件とする X(または Y) の条件付き分布という。上式の $f(x|y), f(y|x)$ がそれぞれ密度関数を定めることは次の通り明らかである。即ち $f(x|y), f(y|x)$ はいずれも二つの確率密度の比であるから非負である。また明らかに

$$\begin{aligned} \int_{-\infty}^{\infty} f(x|y)dx &= \int_{-\infty}^{\infty} \frac{f(x,y)}{h(y)}dx = \frac{1}{h(y)} \int_{-\infty}^{\infty} f(x,y)dx \\ &= \frac{1}{h(y)} h(y) = 1 \end{aligned}$$

である。

いまもし $f(x|y) = f(x,y)/h(y)$ が Y に無関係ならば、即ちこの式の右辺を整理して y を含まない式になるときは、$f(x|y) = g(x)$ であり、従って

$$f(x,y) = g(x)h(y) \tag{1.48}$$

となる。$f(y|x) = h(y)$ のときも同様である。このことは次のことを意味する。

二変数分布の定理 **1**　$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$、即ち X, Y が互いに独立ならば、X, Y の同時分布はそれぞれ X, Y のみを含む二関数の積の形で表される。逆に同時分布がそれぞれ X または Y だけの関数の積として示すことができ、かつそれぞれの変数の変域が互いに無関係のときは、二変数 X, Y は互いに独立である。

X, Y が互いに独立ならば、その分布関数についても明らかに次の関係が成り立つ。

$$F(x,y) = F(x)F(y). \tag{1.49}$$

[33] $g(x)$ は、まず $x = x$ と x の値を固定し、y の値について足し合わせていく。その後、x の値は動いても構わない。

1.2.3　二変数分布の積率

一変数分布の場合と同様にして二変数分布に対しても変数の期待値が定義される。X, Y を二つの連続確率変数、その同時密度関数を $f(x, y)$ とし、X, Y の任意の関数を $\phi(X, Y)$ とすれば、関数 $\phi(X, Y)$ の期待値は次式で定義される。

$$E\left[\phi(X, Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(x, y) f(x, y) dxdy \tag{1.50}$$

X, Y が離散変数の場合も期待値をこれに準じて定義することができる。

二つの確率変数が互いに独立のとき、その積の期待値について次の定理が成り立つ。

二変数分布の定理 **2** 二つの確率変数が互いに独立のとき、その積の期待値は各変数の期待値の積に等しい。

< 証明 >
いま $\phi(X, Y) = XY$ とすれば、二変数の積 XY の期待値は

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x, y) dxdy$$

で定められる。もし X, Y が互いに独立ならば、$f(x, y) = g(x)h(y)$ であるから

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty} xg(x)dx \int_{-\infty}^{\infty} yh(y)dy \\ &= E(X)E(Y) \end{aligned} \tag{1.51}$$

である。　　(証明終わり)
この定理は三変数以上の場合に拡張される。

二変数分布の場合も X, Y のそれぞれの周辺分布に対してそれぞれの積率が定まるが、この他に X, Y の同時分布に対しても次のような同時積率が定義される。X, Y が確率変数のとき次の式で定まる値

$$E\left[(X - a)^r (Y - b)^s\right]$$

を X, Y の $X = a,\ Y = b$ に対する r, s 次の積率という。いま $a = b = 0$ とおけば X, Y の原点積率

$$\mu'_{rs} = E(X^r Y^s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^r y^s f(x, y) dxdy \tag{1.52}$$

が定義され、$a = E(X) = \mu_X,\ b = E(Y) = \mu_Y$ とおけば中心積率

$$\begin{aligned} \mu_{rs} &= E\left[(X - \mu_X)^r (Y - \mu_Y)^s\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)^r (y - \mu_Y)^s f(x, y) dxdy \end{aligned} \tag{1.53}$$

が定義される。特に

$$\mu_{11} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f(x,y)dxdy \tag{1.54}$$

を X,Y の共分散と称し、記号 $cov(X,Y)$ あるいは σ_{XY} で示す。次の関係式は重要である。

$$\begin{aligned} cov(X,Y) &= \mu_{11} = E\left[(X-\mu_X)(Y-\mu_Y)\right] \\ &= E\left[XY-\mu_X Y-\mu_Y X+\mu_X\mu_Y\right] \\ &= E(XY)-\mu_X E(Y)-\mu_Y E(X)+\mu_X\mu_Y \\ &= \mu'_{11}-\mu_X\mu_Y \end{aligned} \tag{1.55}$$

$E(XY)=\mu'_{11}$ を (原点に対する)X,Y の乗積率という。X,Y が互いに独立なときは二変数分布の定理 2 により $E(XY)=E(X)E(Y)$、即ち $\mu'_{11}=\mu_X\mu_Y$ であるから、(1.55) 式は 0 となる。即ち二変数が互いに独立のとき、その共分散は 0 である。ただしこの逆は必ずしも真ではない。

なお、同時積率の記号を用いて X,Y の平均値や分散を次のように表記することができる。(以下、確率変数表示の添字には便宜上小文字を用いる。)

$$\begin{aligned} E(X) &= \mu_x = E(X^1Y^0) = \mu'_{10} \\ E(Y) &= \mu_y = \mu'_{01} \\ var(X) &= \sigma_x^2 = E\left[(X-\mu_x)^2\right] = \mu_{20} \\ var(Y) &= \sigma_y^2 = \mu_{02} \\ cov(X,Y) &= \sigma_{xy} = \mu_{11} \end{aligned}$$

例として二変数 X,Y の和および差の分散を求めると次のようになる。

$$\begin{aligned} var(X\pm Y) &= E\left[X\pm Y-E(X\pm Y)\right]^2 \\ &= E\left[(X-\mu_x)\pm(Y-\mu_y)\right]^2 \\ &= E\left[(X-\mu_x)^2\right]+E\left[(Y-\mu_y)^2\right]\pm 2E\left[(X-\mu_x)(Y-\mu_y)\right] \\ &= \sigma_x^2+\sigma_y^2\pm 2\sigma_{xy} \end{aligned} \tag{1.56}$$

いまもし X,Y が互いに独立ならば共分散は $\sigma_{xy}=0$ である。よってこのときは

$$var(X\pm Y)=\sigma_x^2+\sigma_y^2 \tag{1.57}$$

となる。すなわち次の定理が成り立つ。

二変数分布の定理 **3** 二変数 X,Y が互いに独立ならば、その和 (または差) の分散は各変数の分散の和に等しい。

この定理は三変数以上の場合にも拡張される。

1.2.4　多変数分布の積率母関数

ここではまず、二変数分布について述べる。二変数 X, Y の同時分布の積率母関数は次式によって定義される。

$$M_{X,Y}(t_x, t_y) = E(e^{t_x X + t_y Y}) \tag{1.58}$$

いま X, Y を共に連続変数とすれば、積率母関数は次式で計算される。

$$M(t_x, t_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_x x + t_y y} f(x, y)dxdy \tag{1.59}$$

一変数の場合と同様に、この母関数は次のような形にベキ級数展開される。

$$M(t_x, t_y) = \sum_{r,s=0}^{\infty} \mu'_{rs} \frac{t_x^r t_y^s}{r!s!}.^{34} \tag{1.60}$$

従って任意の次数の同時積率が次の計算によって得られる。

$$\mu'_{rs} = \frac{\partial^{r+s}}{\partial t_x^r \partial t_y^s} M(t_x, t_y)|_{\substack{t_x=0 \\ t_y=0}}.^{35} \tag{1.61}$$

同様にして平均値に対する中心積率母関数が次のように定義される。

$$M_{X-\mu_x, Y-\mu_y}(t_x, t_y) = \left[e^{t_x(X-\mu_x)+t_y(Y-\mu_y)}\right] \tag{1.62}$$

X, Y が互いに独立のときは $f(x, y) = g(x)h(y)$ であるから

$$\begin{aligned} M_{X,Y}(t_x, t_y) &= \int_{-\infty}^{\infty} e^{t_x x} g(x)dx \int_{-\infty}^{\infty} e^{t_y y} h(y)dy \\ &= M_X(t_x) \cdot M_Y(t_y) \end{aligned} \tag{1.63}$$

となる。よって次の定理が得られる。

二変数分布の定理 4　二変数 X, Y が互いに独立のときは、その同時分布の積率母関数は各変数の積率母関数の積に等しい。

二変数分布の定理 4 は、多変数分布について次のように拡張できる。

---

[34]

$$M(t_x, t_y) = E(e^{t_x X \cdot e^{t_y Y}}) = (\sum (t_x x)^r \cdot \frac{1}{r!})(\sum (t_y y)^s \cdot \frac{1}{s!}) = \sum ((t_x x)^r \cdot \frac{1}{r!})((t_y y)^s \cdot \frac{1}{s!}).$$

$E(1 + t_x X + t_y Y + \frac{1}{2!}(t_x X + t_y Y)^2 \cdots)$ として計算してもいいが、大変。また、

$$\mu'_{rs} = E(X^r Y^s) = \int\int x^r y^s f(x, y)dxdy.$$

[35] $r-1, s-1$ までの項は微分されて落ち、$r+1, s+1$ 以上の項は $t_x = 0, t_y = 0$ で消える。

独立変数の和の積率母関数の定理　確率変数 $X_1, X_2, \cdots$ が互いに独立でかつ積率母関数 $M_{X_1}(t), M_{X_2}(t), \cdots$ を有するとき、$Y = X_1 + X_2 + \cdots$ の積率母関数は

$$M_Y(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdots \tag{1.64}$$

で与えられる。

< 証明 > 連続変数のばあいについて考える。定義により

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = E\left[e^{t(X_1+X_2+\cdots)}\right] \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{t(x_1+x_2+\cdots)} f(x_1, x_2, \cdots) dx_1 dx_2 \cdots \\ &= \int_{-\infty}^{\infty} e^{tx_1} f_{X_1}(x_1) dx_1 \int_{-\infty}^{\infty} e^{tx_2} f_{X_2}(x_2) dx_2 \cdots \\ &= M_{X_1}(t) M_{X_2}(t) \cdots \end{aligned}$$

(証明終わり)

1.2.5　二変数分布の変数変換

[36]この節では連続変数について考える。またこの節においては、確率変数 X の密度関数を f_X と表記する。確率変数 X の密度関数 $f_X(x)$ が与えられたとき、X を X の関数 $\mathbf{Y} = \mathrm{g}(\mathbf{X})$ で変数変換することを考える。このとき Y の密度関数は次式で定められる。

$$f_Y(y) = f_X\{g^{-1}(y)\} \left|\frac{dx}{dy}\right| \tag{1.65}$$

ただし $g^{-1}(Y)$ は $g(X)$ の逆関数 $(X = g^{-1}(Y))$ であり、$g(X)$ の逆関数が存在するとき (一変数分布の場合は $g(X)$ が単調増加もしくは単調減少の時) のみを考える。すなわち Y の密度関数は $f_X(x)$ の x を $x = g^{-1}(y)$ で y におきかえてそれに dx/dy (の絶対値) をかければよいのであって、このとき dx/dy は X の確率素分 $f(x)dx$ を Y の確率素分 $f(y)dy$ に変換するための換算因子の働きをしている。

< (1.65) 式の証明 >

まず $Y = g(X)$ が増加関数の場合について考える。$g(X)$ は単調増加であるから逆関数 $X = g^{-1}(Y)$ も単調増加である。いま X の分布関数を $F_X(x), a$ を X の任意の値とすれば

$$F_X(a) = P(X \leq a) = \int_{-\infty}^{a} f_X(x) dx.$$

Y の分布関数を $F_Y(y)$ とすれば

$$F_Y\{g(a)\} = P[Y \leq g(a)] = P(X \leq a) = F_X(a)$$

である。よって

$$\int_{-\infty}^{g(a)} f_Y(y) dy = \int_{-\infty}^{a} f_X(x) dx$$

[36]一変数の場合の変数変換は 1.7 の t 分布の節で用いる。二変数分布の変数変換は、直接使う訳ではなかった。X, Y の同時分布が与えられており、$U = X + Y$ などの分布を求める場合、変数変換によって解くこともできる。

右辺を $x = g^{-1}(y),\ f_X(x) = f_X\{g^{-1}(y)\},\ dx = \frac{d}{dy}g^{-1}(y)dy$ で置き換えると

$$\int_{-\infty}^{g(a)} f_Y(y)dy = \int_{-\infty}^{g(a)} f_X\{g^{-1}(y)\}\frac{d}{dy}g^{-1}(y)dy$$

となり、従って

$$f_Y(y) = f_X\{g^{-1}(y)\}\frac{d}{dy}g^{-1}(y)$$

となる。
$Y = g(X)$ が X の減少関数であるときは、

$$F_Y\{g(a)\} = P[Y \leq g(a)] = P(X > a)$$

であるから

$$\begin{aligned}\int_{-\infty}^{g(a)} f_Y(y)dy &= \int_a^{\infty} f_X(x)dx = \int_{\infty}^{a} -f_X(x)dx \\ &= \int_{-\infty}^{g(a)} -f_X\{g^{-1}(y)\}\frac{d}{dy}g^{-1}(y)dy\end{aligned}$$

となる。従って

$$f_Y(y) = -f_X\{g^{-1}(y)\}\frac{d}{dy}g^{-1}(y).$$

このとき減少関数 $g^{-1}(y)$ の導関数は負であるから、$f_Y(y)$ は正である。以上の結果を合わせると、(1.65) 式となる。　　(証明終わり)

次にこの考え方を拡張して二変数分布の場合の変数変換の方法を考える。問題は確率変数 X, Y の同時分布 $f_{X,Y}(x, y)$ が与えられているとき、X, Y の与えられた関数 $U = g(X, Y),\ V = h(X, Y)$ の同時分布 $f_{U,V}(u, v)$ を求めることである。ただし関数 g, h は X, Y の変域内で連続かつ微分可能とする。このとき X, Y の同時分布の確率素分 $f(u, v)dudv$ に変換するための換算因子 (一変数のときの dx/dy に相当するもの) は次式で与えられる。

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \left|\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}\right| \tag{1.66}$$

この換算因子を (U, V に関する X, Y の) ヤコービアンという。
<ヤコービアンに関する説明>
一組の変数 U, V が他の一組の変数 X, Y の関数として次の式

$$U = g(X, Y), \quad V = h(X, Y) \tag{1.67}$$

で表されていて、g および h は連続でかつ微分できるものとする。もし X に一定の値を与えれば、これら二つの式から Y を消去して U および V を結び付けることができる。幾何学的には、$X =$const は UV 平面上の一つの曲線を表す。同様に $Y =$const もまた一つの曲線を表す。即ち $X = c_1, X = c_2, \cdots$ 等は一群の曲線を、

$Y = k_1, Y = k_2, \cdots$ 等は他の一群の曲線を表す。もし式 (1.67) の逆関数があるときはこれを

$$X = G(U,V), \quad Y = H(U,V) \tag{1.68}$$

とし、G および H は連続でかつ微分できるものとする。ある平面上において X を横軸、Y を縦軸にとれば、$U = a_1, U = a_2, \cdots$ および $V = l_1, V = l_2, \cdots$ 等はそれぞれ曲線群を表すことになる。(1.67) を (1.68) に代入すれば

$$X = G(g(X,Y), h(X,Y)), \quad Y = H(g(X,Y), h(X,Y))$$

が得られる。これらをそれぞれ X および Y に関して微分すれば、

$$\begin{aligned} 1 &= \left.\frac{\partial G}{\partial U}\right|_V \left.\frac{\partial g}{\partial X}\right|_Y + \left.\frac{\partial G}{\partial V}\right|_U \left.\frac{\partial h}{\partial X}\right|_Y, \quad 0 = \left.\frac{\partial G}{\partial U}\right|_V \left.\frac{\partial g}{\partial Y}\right|_X + \left.\frac{\partial G}{\partial V}\right|_U \left.\frac{\partial h}{\partial Y}\right|_X, \\ 0 &= \left.\frac{\partial H}{\partial U}\right|_V \left.\frac{\partial g}{\partial X}\right|_Y + \left.\frac{\partial H}{\partial V}\right|_U \left.\frac{\partial h}{\partial X}\right|_Y, \quad 1 = \left.\frac{\partial H}{\partial U}\right|_V \left.\frac{\partial g}{\partial Y}\right|_X + \left.\frac{\partial H}{\partial V}\right|_U \left.\frac{\partial h}{\partial Y}\right|_X. \end{aligned}$$

ゆえに

$$J = \frac{\partial G}{\partial U}\frac{\partial H}{\partial V} - \frac{\partial G}{\partial V}\frac{\partial H}{\partial U} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u} \tag{1.69}$$

が 0 でなければ[37]、

$$\frac{\partial g}{\partial X} = \frac{1}{J}\frac{\partial H}{\partial V}, \quad \frac{\partial g}{\partial Y} = -\frac{1}{J}\frac{\partial G}{\partial V}, \quad \frac{\partial h}{\partial X} = -\frac{1}{J}\frac{\partial H}{\partial U}, \quad \frac{\partial h}{\partial Y} = \frac{1}{J}\frac{\partial G}{\partial U}$$

が得られる。以上より、

$$J = \left| \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} \right| = \frac{\partial(x,y)}{\partial(u,v)}.$$

(説明終わり)

(1.65) 式に対応し、二変数分布の変数変換については次の定理が成り立つ。

二変数分布の定理 5　X, Y の同時密度関数が $f_{X,Y}(x,y)$ で与えられ、関数 $U = g(X,Y),\ V = h(X,Y)$ が X, Y の変域内で連続かつ微分可能なとき、U, V の同時密度関数は次式で与えられる。

$$f_{U,V}(u,v) = f_{X,Y}\{G(u,v)H(u,v)\} \left| \frac{\partial(x,y)}{\partial(u,v)} \right| \tag{1.72}$$

ただし $G(u,v), H(u,v)$ は $g(x,y), h(x,y)$ の逆関数である。

[37]

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \tag{1.70}$$

であるとすると、

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{1.71}$$

結局、$J = ad - bc \neq 0$ は逆行列が存在するための条件。

< 証明 >
ヤコービアンが X,Y の変域内で 0 でなければ X,Y の逆関数

$$X = G(U,V), \quad Y = H(U,V)$$

が一価的に定まる。よって xy 平面上の任意の領域 R に含まれる点 (x,y) は uv 平面上のこれに対応する領域 S 内の一点 (u,v) と一対一で対応する。このとき点 (x,y) が集合 R に属する確率は点 (u,v) が集合 S に属する確率に等しい。即ち

$$P[(u,v) \in S] = P[(x,y) \in R]$$

である。この関係から

$$\int\int_S f_{U,V}(u,v)dudv = \int\int_R f_{X,Y}(x,y)dxdy$$

が成り立つ。ここで

$$X = G(U,V), \quad Y = H(U,V)$$

による変数変換を考える。XY 座標系において、ベクトル $(dx,0),(0,dy)$ 間の面積は

$$|(dX,0)\times(0,dY)| = \begin{vmatrix} dX & 0 \\ 0 & dY \end{vmatrix} = dXdY$$

となる。ところで、

$$\begin{aligned} dX &= \frac{\partial G}{\partial U}dU + \frac{\partial G}{\partial V}dV, \\ dY &= \frac{\partial H}{\partial U}dU + \frac{\partial H}{\partial V}dV \end{aligned}$$

であるから、次に UV 座標系におけるベクトル、$(\frac{\partial G}{\partial U}dU, \frac{\partial G}{\partial V}dV), (\frac{\partial H}{\partial U}dU, \frac{\partial H}{\partial V}dV)$ 間の面積は次のようになる。

$$\left|(\frac{\partial G}{\partial U}dU, \frac{\partial G}{\partial V}dV)\times(\frac{\partial H}{\partial U}dU, \frac{\partial H}{\partial V}dV)\right| = \begin{vmatrix} \frac{\partial G}{\partial U}dU & \frac{\partial H}{\partial U}dU \\ \frac{\partial G}{\partial V}dV & \frac{\partial H}{\partial V}dV \end{vmatrix} = JdUdV.$$

よって

$$\begin{aligned} \int\int_S f_{U,V}(u,v)dudv &= \int\int_R f_{X,Y}(x,y)dxdy \\ &= \int\int_S f_{X,Y}\{G(u,v)H(u,v)\}\left|\frac{\partial(x,y)}{\partial(u,v)}\right|dudv \end{aligned}$$

という関係が導かれる。この関係は uv 平面上のどのような集合 S についても成り立つから、二変数分布の定理 5 が成り立つ。　　(証明終わり)

即ち U,V の同時密度関数を求めるためには X,Y の変換式 $U = g(X,Y),\ V = h(X,Y)$ を X,Y について解いてこれを X,Y の同時密度関数 $f_{X,Y}(x,y)$ に代入し、

U, V に関する X, Y のヤコービアンの絶対値を確率素分の変換因子としてかければいい。新しい変数 U, V の変換は X, Y の変域と $U = g(X, Y),\ V = h(X, Y)$ の関係からそれぞれの場合につき決定される。なおヤコービアンの式 (1.66) の計算については

$$\frac{\partial(x,y)}{\partial(u,v)} = \left[\frac{\partial(u,v)}{\partial(x,y)}\right]^{-1} = \left|\begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}\right|^{-1}$$

であるから、左辺よりも右辺の方が計算しやすいときにはこの方法で計算すればよい。

問題例 **1**　X, Y の同時分布が $f(x, y) = e^{-(x+y)},\ x > 0,\ y > 0$ で与えられ、$U = X + Y,\ V = X/Y$ と定めたとき、$f(u, v)$ を求めよ。

U, V の逆関数は

$$X = \frac{UV}{1+V}, \quad Y = \frac{U}{1+V}$$

であるから変数変換のヤコービアンは

$$J = \left|\begin{pmatrix} \frac{v}{1+v} & \frac{u}{(1+v)^2} \\ \frac{1}{1+v} & \frac{-u}{(1+v)^2} \end{pmatrix}\right| = \left|\frac{-u}{(1+v)^2}\right|$$

となる。(ただしこの場合は変換式をそのまま使ってヤコービアンの逆数を求める方が計算は簡単である。即ち

$$J^{-1} = \left|\begin{pmatrix} 1 & 1 \\ \frac{1}{y} & -\frac{x}{y^2} \end{pmatrix}\right| = \left|\frac{-(x+y)}{y^2}\right| = \left|\frac{-(1+v)^2}{u}\right|$$

となる。) よって U, V の同時分布は

$$f_{U,V}(u, v) = e^{-u}\left|\frac{-u}{(1+v)^2}\right|$$

で定められる。u, v の変域は $x > 0,\ y > 0$ であるから $u = x + y > 0,\ v = x/y > 0$ である。よって

$$f_{U,V}(u, v) = \frac{ue^{-u}}{(1+v)^2}$$

X, Y の同時分布 $f(x, y)$ が与えられていてこの二変数の与えられた一関数 $g(X, Y)$ の分布を求めるときも以上の方法を応用して計算することができる。即ち $U = g(X, Y),\ V = X$ または $V = Y$ とおいて $f_{X,Y}(x, y) \to f_{U,V}(u, v)$ の変換を行ない、$f_U(u) = \int f_{U,V}(u, v)dv$ であるから $f_{U,V}$ を $V = X$ または $V = Y$ の変域について積分すればよい。具体例は次の通りである。

問題例 **2**　$f(x, y) = e^{-(x+y)},\ x > 0,\ y > 0$ のとき $U = X + Y$ の分布を求めよ。

$$U = X + Y, \quad V = Y$$

とおけば、逆関数は

$$X = U - V, \quad Y = V$$

である。よって

$$J = \left| \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \right| = 1,$$

$$f_{U,V}(u,v) = e^{-u} \cdot 1 = e^{-u}$$

である。$y > 0$ であるから $v > 0$ であるが、$x = u - v > 0$ であるから $u > v$ である。よって V の変域は $0 < v < u$ となる。従って $U = X + Y$ の分布は

$$\begin{aligned} f_U(u) &= \int_0^u f_{U,V}(u,v)dv = \int_0^u e^{-u}dv \\ &= ue^{-u} \quad u > 0. \end{aligned}$$

積率母関数による変数変換

[38]一変数分布の場合と同様に、二変数分布の場合にも積率母関数を使って変数変換のできる場合がある。いま X, Y の同時分布を $f(x,y)$ とし、$U = g(X,Y)$ を X, Y の任意の関数とする。このとき U の積率母関数が存在するならば、それは

$$M_U(t) = E(e^{tU}) = \int\int e^{tg(x,y)} f(x,y)dxdy$$

で与えられる[39]。もしこの関数が既知の確率変数の積率母関数と同形になるならば、U はそれと同じ形の分布をもつことになる。この方法は X, Y の異なった二つの関数 $g(X,Y),\ h(X,Y)$ の同時分布を求める場合 (一般に X, Y の関数のいくつかの同時分布を求める場合) にも拡張することができる。

1.3　標本分布

1.3.1　標本、確率標本

われわれが経験世界で現実に当面する確率変数が本来どのような分布を有するかは通常明らかではない。このようなときわれわれはその分布についての情報を得るために、その確率変数についての試験的観察を何回か試みる。このように確率変数について実際に観察を行なうことを標本観察といい、この標本観察の対象として選ばれる個体の集合を標本という。これに対して標本観察によって情報を得ようとする確率変数が定義されている標本空間を、その標本の母集団という。統計解析の中心問題の一つは、標本観察の結果によって母集団に関する情報をさぐ

[38]後で直接使うわけではない。

[39]

$$M_U(t) = E(e^{tU}) = \int e^{tU} f(u)Jdu = \int\int e^{tg(x,y)} f(x,y)dxdy$$

ということ。

ることである。ここで母集団に関する情報とは、母集団が確率変数に関してどのような分布を有し、その分布がどのような値の母数によって定められているかということである。われわれはこのような母数の推定値を標本観察の結果から計算しようとする。

母集団から確率的に選ばれた個々の観察対象を ω とし、i 番目に選ばれた ω_i について観察された結果を $X_i = X(\omega_i)$ とする。X_i は確率変数である。n 回の観察結果は n 個の確率変数の集合

$$\{X_1, X_2, \cdots, X_n\}$$

を作る。これを母集団 X からの大きさ n 個の標本と呼ぶことにする。ここで $\{X_1, X_2, \cdots, X_n\}$ は確率変数 (個々の試行の結果にある実数値を与えるルール、あるいはどのような値が考えられるかを集めた可能性一覧のようなもの。直接、確率を表す訳ではなく、確率が定義されるという前提がある場合に用いられる、一種の変数。) の集合としての標本を意味し、実際に標本観察の結果として得られた数値の集まり $\{x_1, x_2, \cdots, x_n\}$ とは区別されている。

標本観察は対象の持つ量的属性または質的属性について行なわれる。量的属性の観察結果は直ちに確率変数として取り扱われる。質的属性の場合も特定の性質を持つか否かによって $(0, 1)$ 変数で表現することができる。例えば製品の合格不合格の観察で、合格を 1、不合格を 0 とすれば、n 個の標本観察の結果は例えば

$$\{1, 1, 0, 1, 0, 1, \cdots\}$$

で示される。この場合、確率標本 X の具体的な数値が $x_1 = 1, x_2 = 1, x_3 = 0, \cdots$ となる。

母集団 X の分布が $f(x)$ で与えられ、n 個の対象が同一かつ不変のこの母集団から互いに独立に抜き取られるとき、一組の標本観測値 $\{x_1, x_2, \cdots, x_n\}$ が得られる確率あるいは確率密度は

$$f(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n)$$

である。

標本 $\{X_1, X_2, \cdots, X_n\}$ の確率または確率密度が

$$f(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n) \tag{1.73}$$

で与えられるとき、$\{X_1, X_2, \cdots, X_n\}$ を度数関数あるいは密度関数 $f(x)$ を有する母集団 X からの大きさ n 個の確率標本 (無作為標本、標本) という。条件 (1.73) が成り立つためには $X_1, X_2, \cdots$ は互いに独立でなければならない。

確率標本についての注意を挙げておく。確率標本 $\{X_1, X_2, \cdots, X_n\}$ は、一定の分布 F をもつ互いに独立な確率変数の組であり、結局、確率標本とは母集団から

互いに独立に抜き取られた (無作為抽出された) 標本であるといえる。さらに加えると、確率標本 X は集合ではない。また $f(x_1)$ は確率もしくは確率密度なので、$X_1 \neq f(x_1)$.

1.3.2　標本分布

母集団 $f(x)$ から[40]抜き取られた大きさ n の確率標本を $\{X_1, X_2, \cdots, X_n\}$ とすれば

$$f(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n) \tag{1.74}$$

はこのような標本の得られる確率あるいは密度関数である。このような度数関数あるいは密度関数の定める分布を標本分布という[41]。

1.3.3　統計量、統計量の分布

$\{X_1, X_2, \cdots, X_n\}$ が確率変数を示すとき、そのおのおのの観察値を独立変数とする関数

$$Y = g(X_1, X_2, \cdots, X_n)$$

を統計量という。統計量は明らかに確率変数である。

統計量の定義式には母集団のパラメーター (母数) は含まれていない。故にそれは母集団の知識なしに標本観察値のみから計算できる。例えば標本について観察された平均値や分散、

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n),$$

$$s^2 = \frac{1}{n}[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2]$$

はいずれも統計量である。

確率標本 $\{X_1, X_2, \cdots, X_n\}$ はその母集団分布によって定まる一定の標本分布 $f(x_1, x_2, \cdots, x_n)$ を持つ。そして統計量 $Y = g(X_1, X_2, \cdots, X_n)$ が一価関数であれば、Y の分布はこの定義式を変換式として、与えられた標本分布 $f(x_1, x_2, \cdots, x_n)$ から導くことができる。統計量 Y の分布を標本分布ということもある。

統計量の中には、順序統計量というものがあり、以下のようなものである。標本観察値を観察された順序ではなく大きさの順に配列し直し、

$$X_{(1)}, X_{(2)}, \cdots, X_{(n)}$$

[40]確率標本の各々が、同一の分布に従う (あるいはある一つの母集団から n 個の確率標本が順番に抜き取られる) というのがポイントのようである。

[41]ただし $f(x_1, x_2, \cdots, x_n)$ は大きさ n の標本がこの順序で得られる確率あるいは確率密度である。一般に標本は観察単位の順序づけられた集合である[42]。

(ただし $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$) とする。このように並べられたベクトル、あるいはこのベクトルに基づいて定義された量が順序統計量である。例えばメジアン (中央値)、レンジ (範囲) $R = X_{(n)} - X_{(1)}$ などは順序統計量である。これらの標本統計量に対しても同様に標本分布が考えられる。

統計量の分布 (標本分布) は母集団分布を推定するために重要である。母集団分布の形 $f(x;\theta)$(ただし x は確率変数、θ は母数 (母集団分布の形を決めるパラメータ) を示す) が与えられているが、母数 θ の値が未知のとき、この母集団から選ばれた標本 $\{X_1, X_2, \cdots, X_n\}$ に基づいて θ の値を推定することである。標本 $\{X_1, X_2, \cdots, X_n\}$ は母集団分布 $f(x;\theta)$ の母数 θ に関する情報を含んでいるはずであるから、統計量 $Y = g(X_1, X_2, \cdots, X_n)$ も同様に θ に関する情報を含んでおり、その形を適当に選べば母数 θ の値をこの統計量によって推定することができるであろう。このとき統計量 Y の標本分布が母数 θ とどのような関係にあるかを知ることができれば、Y を推定量として用いるときの推定の正確さについて明確な判断を下すことができる。故に、母数推定の問題は、推定量として用いられる統計量の標本分布の問題に帰着する。

1.3.4　標本積率

$X_1, X_2, \cdots, X_n$ を大きさ n の確率標本とするとき、k 次の原点標本積率は次の式で定義される。

$$m'_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \tag{1.75}$$

標本平均値、標本分散はこの積率記号でそれぞれ次のとおりに示される。

$$\bar{X} = m'_1, \qquad s^2 = m'_2 - m'^2_1. \tag{1.76}$$

原点標本積率と同様に中心標本積率も次のとおり定義される。

$$m_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k \tag{1.77}$$

標本積率は標本観察値で定められるから統計量であって、母集団に関する情報を含む。この点に関して次の諸定理は重要である[43]。

[43]期待値という用語の意味は、「最も期待されそうな値」であり、以下の定理では記号 $E(\ )$ と共に使われる。平均値と同義で使うこともあるが、注意が必要である。標本については、平均値と期待値は等しく、$(x_1 + x_2 + \cdots + x_n)/n$ である。これに対し、母集団については確率が分かっていなければならない。即ち、母集団分布が離散的な場合、母平均の期待値は $E(X) = \sum^n x_i p_i$ であり、母平均の分散は $var(X) = \sum^n (x_i - \mu)^2 p_i$ である。 母平均値は分からない場合が普通であるので、標本平均値の n が大きくなると母平均の期待値に近付くといった性質を利用して、母集団について期待値を求めるのである。

定理 1 k 次の原点標本積率の期待値は母集団の原点積率に等しい。

<証明>

$$\begin{aligned} E(m'_k) = E(\frac{1}{n}\sum_{i=1}^{n} X_i^k) &= \frac{1}{n}E(\sum X^k) \\ &= \frac{1}{n}\sum \Big[E(X^k)\Big] \end{aligned}$$

然るに積率の一般定義 (母集団の積率の定義) により、$E(X^k) = \mu'_k$ であるから[44]、

$$E(m'_k) = \frac{1}{n}\sum \mu'_k = \frac{1}{n}n\mu'_k = \mu'_k \tag{1.78}$$

(証明終わり)

ただし、中心積率 m_k については必ずしもこれと同じ関係は成り立たない。

標本平均値の期待値は母集団平均値に等しい。すなわち

$$E(\bar{X}) = \mu'_1 = \mu \tag{1.79}$$

ただし分散についてはこの関係は成り立たない。

定理 2 標本平均値の積率母関数は母集団分布の積率母関数で次のとおり示される。

$$M_{\bar{X}}(t) = \Big[M_X(\frac{t}{n})\Big]^n \tag{1.80}$$

<証明> $X_1, X_2, \cdots$ は互いに独立であるから

$$\begin{aligned} M_{\bar{X}}(t) &= E(e^{t\bar{X}}) = E\left[e^{t(X_1+X_2+\cdots+X_n)/n}\right] \\ &= E(e^{\frac{t}{n}X_1})E(e^{\frac{t}{n}X_2})\cdots E(e^{\frac{t}{n}X_n}) \\ &= \Big[M_X(\frac{t}{n})\Big]^n \end{aligned} \tag{1.81}$$

(証明終わり)

定理 3 母集団分散 (母分散)σ^2 が有限値のとき、標本平均値の分散は

$$var(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n} \tag{1.82}$$

で与えられる。

<証明>

$$\begin{aligned} var(\bar{X}) &= E\Big[\bar{X} - E(\bar{X})\Big]^2 = E\Big[(\bar{X}-\mu)^2\Big] = E\left[(\frac{1}{n}\sum X_i - \mu)^2\right] \\ &= E\left[(\frac{1}{n}\sum X_i - \frac{1}{n}\sum \mu)^2\right] = E\left[\Big(\frac{1}{n}\sum (X_i-\mu)\Big)^2\right] \end{aligned}$$

[44] 確率変数 $X_1, X_2, \cdots, X_n$ はそれぞれ独立に同じ分布をしている母集団から取り出すので、期待値などは等しくなる。

$$
\begin{aligned}
&= \frac{1}{n^2}E\left[\left(\sum(X_i-\mu)\right)^2\right] \\
&= \frac{1}{n^2}E\left[\{(X_1-\mu)+\cdots+(X_n-\mu)\}\{(X_1-\mu)+\cdots+(X_n-\mu)\}\right] \\
&= \frac{1}{n^2}E\left[\sum_i(X_i-\mu)^2+\sum_{i\neq j}(X_i-\mu)(X_j-\mu)\right] \\
&= \frac{1}{n^2}\left(E\left[\sum_i(X_i-\mu)^2\right]+E\left[\sum_{i\neq j}(X_i-\mu)(X_j-\mu)\right]\right)
\end{aligned}
$$

然るに

$$E\left[\sum(X_i-\mu)^2\right]=\sum^n\left[E(X_i-\mu)^2\right]=n\sigma^2.^{45}$$

$$
\begin{aligned}
E\left[\sum_{i\neq j}(X_i-\mu)(X_j-\mu)\right] &= \sum\left[E(X_i-\mu)(X_j-\mu)\right] \\
&= \sum\left[E(X_i-\mu)E(X_j-\mu)\right]
\end{aligned}
$$

(X_i, X_j は互いに独立であり、二変数分布の定理 4 (1.63 式を参照)「二つの確率変数が互いに独立のとき、その積の期待値は各変数の期待値の積に等しい」が成り立つので。)
ここで、

$$E(X_i-\mu)=\sum(x_i-\mu)p_i=\sum x_ip_i-\mu\sum p_i=\mu-\mu\cdot 1=0.$$

よって、

$$E\left[\sum_{i\neq j}(X_i-\mu)(X_j-\mu)\right]=\sum\left[E(X_i-\mu)E(X_j-\mu)\right]=\sum(0\cdot 0)=0$$

故に

$$var(\bar{X})=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n} \tag{1.83}$$

(証明終わり)

定理 4 標本分散 $s^2=\frac{\sum(X-\bar{X})^2}{n}$ の期待値は $\frac{n-1}{n}\sigma^2$ である[46]。

< 証明 >

$$
\begin{aligned}
E(s^2) &= E\left[\frac{1}{n}\sum(X_i-\bar{X})^2\right]=E\left[\frac{1}{n}\sum(X_i-\mu-(\bar{X}-\mu))^2\right] \\
&= E\left[\frac{1}{n}\{\sum(X_i-\mu)^2-2\sum(X_i-\mu)(\bar{X}-\mu)+\sum(\bar{X}-\mu)^2)\}\right]
\end{aligned}
$$

[45] X_i の i は何番目の試行かということ。各確率変数は同じ母集団分布から独立に抜き出したものなので、どの i についても $\left[E(X_i-\mu)^2\right]=\sigma^2$ となる。

[46] $s^2=\frac{\sum(X-\bar{X})^2}{n-1}$?

$$
\begin{aligned}
&= E\left[\frac{1}{n}\{\sum(X_i-\mu)^2 - 2(n\bar{X}-n\mu)(\bar{X}-\mu)+n(\bar{X}-\mu)^2)\}\right] \\
&= E\left[\frac{1}{n}\{\sum(X_i-\mu)^2 - 2n(\bar{X}-\mu)^2+n(\bar{X}-\mu)^2)\}\right] \\
&= E\left[\frac{1}{n}\sum(X_i-\mu)^2 - (\bar{X}-\mu)^2\right] \\
&= \frac{1}{n}\sum^{n}\left[E(X_i-\mu)^2\right] - E(\bar{X}-\mu)^2 \\
&= \frac{1}{n}n\sigma^2 - \sigma^2_{\bar{X}} = \sigma^2 - \frac{\sigma^2}{n} \\
&= \frac{n-1}{n}\sigma^2. \qquad (1.84)
\end{aligned}
$$

(証明終わり)

故に標本分散は平均すると一般に母平均よりも小さい。分散は二次の中心積率であるが、このように中心積率の場合、標本積率の期待値は必ずしも母集団積率に等しくならない。

1.4　正規分布

1.4.1　正規分布

確率密度関数が

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < +\infty,\ \ \sigma > 0$$

となる分布を正規分布と呼ぶ。以下に示すように正規分布の平均値は μ、分散は σ^2 となる。平均値が μ 、 分散が σ^2 の正規分布を $\mathrm{N}(\mu,\sigma^2)$ と表すことがある。特に、$\mathrm{N}(0,1^2)$ の場合を標準正規分布と呼ぶ。
正規分布する母集団では、$\mu\pm\sigma$ の範囲に 68%、$\mu\pm1.65\sigma$ の範囲に 90%、$\mu\pm1.96\sigma$ の範囲に 95% の構成要素が含まれている。

$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \mu$ の証明

正規分布の確率密度関数は

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < +\infty$$

である。確率変数 X の平均は (1.8) 式で表されるから、

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} xf(x)dx \\
&= \int_{-\infty}^{\infty} x\frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx. \qquad (1.85)
\end{aligned}
$$

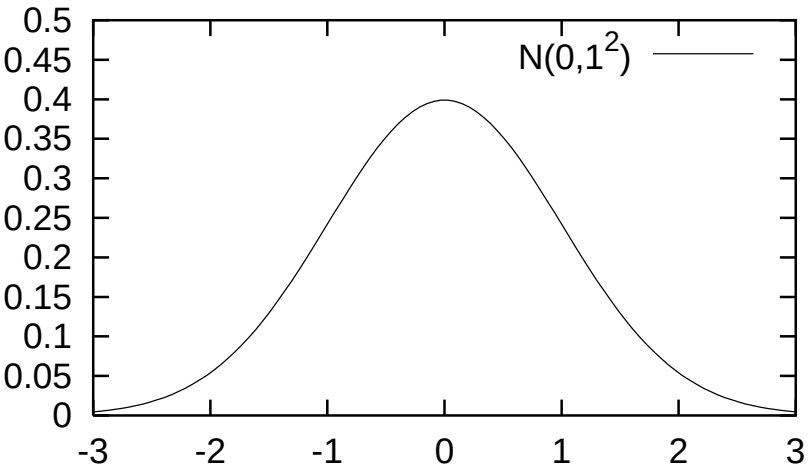

図 1.1: 標準正規分布のグラフ

ここで $y = x - \mu$ とおくと、

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} (y+\mu)\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{y^2}{2\sigma^2}} dy \\ &= \int_{-\infty}^{\infty} y\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{y^2}{2\sigma^2}} dy + \int_{-\infty}^{\infty} \mu\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{y^2}{2\sigma^2}} dy. \end{aligned} \tag{1.86}$$

ここで第 1 項は、$z = \frac{y}{\sqrt{2}\sigma}$ とおいて、

$$\begin{aligned} (1.86)\text{ の第 1 項} &= \int_{-\infty}^{\infty} \sqrt{2}\sigma z\frac{1}{\sigma\sqrt{2\pi}}\, e^{-z^2}\ \sqrt{2}\sigma dz = \frac{\sqrt{2}\sigma}{\sqrt{\pi}}\int_{-\infty}^{\infty}\ ze^{-z^2}dz \\ &= \frac{\sqrt{2}\sigma}{\sqrt{\pi}}\int_{-\infty}^{0}\ ze^{-z^2}dz + \frac{\sqrt{2}\sigma}{\sqrt{\pi}}\int_{0}^{\infty}\ ze^{-z^2}dz. \end{aligned}$$

さらに $z^2 = a$ とおくと $2zdz = da$ であるから、

$$\begin{aligned} (1.86)\text{ の第 1 項} &= \frac{\sqrt{2}\sigma}{\sqrt{\pi}}\int_{\infty}^{0}\ \frac{1}{2}e^{-a}da + \frac{\sqrt{2}\sigma}{\sqrt{\pi}}\int_{0}^{\infty}\ \frac{1}{2}e^{-a^2}da \\ &= \frac{1}{2}\frac{\sqrt{2}\sigma}{\sqrt{\pi}}\left[-e^{-a}\right]_{\infty}^{0} + \frac{1}{2}\frac{\sqrt{2}\sigma}{\sqrt{\pi}}\left[-e^{-a}\right]_{0}^{\infty} \\ &= \frac{1}{2}\frac{\sqrt{2}\sigma}{\sqrt{\pi}}(-1+0+0+1) = 0. \end{aligned}$$

第 2 項については、係数の μ はとりあえず別にして (1.86) 式の第 2 項$=\mu\cdot A$ とおく。ただし

$$A = \int_{-\infty}^{\infty}\ \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{y^2}{2\sigma^2}} dy.$$

$z = y/\sigma$ で変数変換すると

$$A = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\ e^{-\frac{1}{2}z^2} dz$$

となる。積分変数の選択は任意であるから、

$$A^2 = A \cdot A = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}a^2} da$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2+a^2)} dz da$$

$z = r\sin\theta, a = r\cos\theta$ で極座標に変換すれば、

$$\begin{aligned} A^2 &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{1}{2}r^2} \, r dr d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} d\theta \int_0^{\infty} r e^{-\frac{1}{2}r^2} \, dr \\ &= \frac{1}{2\pi} \left[\theta\right]_0^{2\pi} \cdot \left[-e^{-\frac{1}{2}r^2}\right]_0^{\infty} \\ &= \frac{1}{2\pi} 2\pi(-0+1) = 1 \end{aligned}$$

ここで A^2 について計算したので、A の符号を考える。$\sigma > 0$ のとき明らかに $A > 0$ であるから、$A = 1$ である。以上をまとめれば、

$$E(X) = 0 + \mu \cdot 1 = \mu \qquad \text{(証明終わり)}$$

$var(X) = \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx = \sigma^2$ の証明

正規分布をする確率変数 X の分散は (1.9) 式で表されるから、

$$var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} \, e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

ここで $y = \frac{x-\mu}{\sigma}$ とおくと、

$$var(X) = \int_{-\infty}^{\infty} \sigma^2 y^2 \frac{1}{\sigma\sqrt{2\pi}} \, e^{-\frac{1}{2}y^2} \cdot \sigma dy = \int_{-\infty}^{\infty} \sigma^2 \frac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}y^2} y^2 dy.$$

積分する関数は偶関数なので、

$$var(X) = 2\int_0^{\infty} \sigma^2 \frac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}y^2} y^2 dy.$$

ここで、定積分の公式

$$\int_0^{\infty} e^{-ax^2} x^{2n} dx = \frac{(2n-1)(2n-3)\cdots 3\cdot 1}{2^{n+1}} \sqrt{\frac{\pi}{a^{2n+1}}}$$

を用いて、$a = 1/2$, $n = 1$ の時を考えると、

$$var(X) = 2 \cdot \sigma^2 \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{4} \cdot \sqrt{8\pi} = \sigma^2.$$

(証明終わり)

1.4.2　正規分布を母集団とする標本集団の性質

平均値 μ 、分散 σ^2 の正規分布をしている母集団から大きさ n のサンプルをランダムに抜き取った場合、標本集団の性質は以下のようになる。

定理 1 母集団が平均値 μ 、分散 σ^2 の正規分布をするとき、大きさ n の確率標本を $\{X_1, X_2, \cdots, X_n\}$ とすれば、標本平均値 $\bar{X}$ は平均値 μ 、分散 σ^2/n で正規分布する。

<証明> X_i は正規分布変数であるから、その積率母関数は以下のようになる。

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx - \frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x^2 - 2(\sigma^2 t+\mu)x + \mu^2}{2\sigma^2}} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{\{x-(\sigma^2 t+\mu)\}^2 - (\sigma^4 t^2 + 2\sigma^2 t\mu)}{2\sigma^2}} dx \\ &= \frac{e^{\frac{1}{2}\sigma^2 t^2 + t\mu}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{\{x-(\sigma^2 t+\mu)\}^2}{2\sigma^2}} dx. \end{aligned}$$

ここで $y = x - (\sigma^2 t + \mu)$ とおくと、

$$M_X(t) = \frac{e^{\frac{1}{2}\sigma^2 t^2 + t\mu}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma^2}} dy.$$

また、$A = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma^2}} dy$ とおく。$z = y/\sigma$ で変数変換すると、

$$A = \sigma \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz.$$

積分変数の選択は任意であるから、

$$\begin{aligned} A^2 = A \cdot A &= \sigma^2 \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \cdot \int_{-\infty}^{\infty} e^{-\frac{a^2}{2}} da \\ &= \sigma^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2+a^2)} dz da. \end{aligned}$$

$z = r\sin\theta, a = r\cos\theta$ で極座標変換すると、

$$\begin{aligned} A^2 &= \sigma^2 \int_0^{2\pi} \int_0^{\infty} e^{-\frac{1}{2}r^2} r dr d\theta \\ &= \sigma^2 \left[\theta\right]_0^{2\pi} \cdot \left[-e^{-\frac{1}{2}r^2}\right]_0^{\infty} = \sigma^2 \cdot 2\pi(-0+1) = \sigma^2 2\pi \end{aligned}$$

ここで A^2 について計算したので A の符号を考える。$f(x) > 0$ となるために $\sigma > 0$ なので、$A > 0$. よって、$A = \sigma\sqrt{2\pi}$ である。以上より、

$$M_X(t) = \frac{e^{\mu t + \frac{1}{2}\sigma^2 t^2}}{\sigma\sqrt{2\pi}} \cdot \sigma\sqrt{2\pi} = e^{\mu t + \frac{1}{2}\sigma^2 t^2}.$$

標本平均値は

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1}{n} + \frac{X_2}{n} + \cdots + \frac{X_n}{n}$$

であるから、まず X/n の積率母関数を求めると (1.27) 式より

$$M_{X/n}(t) = M_X(\frac{t}{n}) = e^{\mu\frac{t}{n}+\frac{1}{2}\frac{\sigma^2 t^2}{n^2}}$$

となる。X_i は互いに独立であるから (1.64) より、

$$\begin{aligned} M_{\bar{X}}(t) &= M_{X_1/n}(t)\cdot M_{X_2/n}(t)\cdots M_{X_n/n}(t) \\ &= \left[M_X(\frac{t}{n})\right]^n = e^{\mu t+\frac{1}{2}(\frac{\sigma^2}{n})t^2} \end{aligned} \tag{1.87}$$

となる。これは平均値 μ 、分散 σ^2/n の正規分布の積率母関数である。(証明終わり)

定理 2 範囲 R[47] は、平均値がもとの母集団の標準偏差の定数倍、つまり $d_{2\sigma}$ で、標準偏差がもとの母集団の標準偏差の定数倍、つまり $d_{3\sigma}$ の正規分布をする。定数 d_2 と d_3 はサンプルの大きさ n によって決まった値をとる (数値表にのっている)。(証明は省略)

定理 3 偏差平方和 S の平均値は、もとの母集団の分散 σ^2 の $(n-1)$ 倍となる。(証明は省略)

1.5　ガンマ分布

確率変数 X の確率密度が次の式で定められるとき、X をガンマ変数、その分布をガンマ分布という。

$$f(x) = \frac{\lambda^n}{\Gamma(n)}e^{-\lambda x}x^{n-1} \qquad x > 0 \tag{1.88}$$

ただし、λ, n は分布の母数であって、$\lambda > 0,\ n > 0$ である。$\Gamma(n)$ はガンマ関数。詳しくは Appendix A を参照。

$n = 1$ のとき、この分布は指数分布とよばれる。$n \leq 1$ ならばガンマ分布の形は逆 J 字型になる (図 1.2 参照)[48]。

[47] 範囲 R は標本値の最大値と最小値との差、すなわち

$$R = x_n - x_1.$$

[48] $n > 1$ のときは $X = (n-1)/\lambda$ がモードである。モードとは度数分布において、度数の最大値に対応する値、つまり最頻値である。この場合では分布関数の極大値に対応する (即ち $f'(x) = 0$ のときの)X の値である。

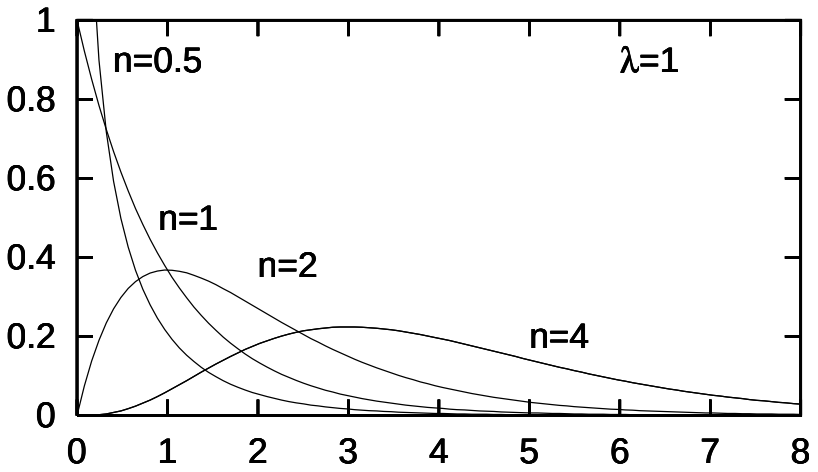

図 1.2: ガンマ分布のグラフ

(1.88) 式が密度関数の条件 $\int_{-\infty}^{\infty} f(x)dx = 1$ をみたすことは次の通り証明される。

$$\int_0^{\infty} f(x)dx = \frac{\lambda}{\Gamma(n)} \int_0^{\infty} e^{-\lambda x}(\lambda x)^{n-1}dx.$$

$y = \lambda x$ とおいて変換すれば、

$$\frac{\lambda}{\Gamma(n)} \cdot \frac{1}{\lambda} \int_0^{\infty} e^{-y} y^{n-1} dy = \frac{\lambda}{\Gamma(n)} \frac{1}{\lambda} \Gamma(n) = 1.$$

(証明終わり)

ガンマ分布の分布関数は次式で与えられる。

$$F(x) = \frac{\lambda^n}{\Gamma(n)} \int_0^{x} e^{-\lambda x} x^{n-1} dx. \tag{1.89}$$

ガンマ関数の積率母関数は次のように計算される。

$$\begin{aligned} M(t) &= \int_0^{\infty} e^{tx} \frac{\lambda^n}{\Gamma(n)} e^{-\lambda x} x^{n-1} dx \\ &= \frac{\lambda^n}{(\lambda - t)^n} \int_0^{\infty} \frac{(\lambda - t)^n}{\Gamma(n)} e^{-(\lambda - t)x} x^{n-1} dx, \end{aligned}$$

積分を $y = (\lambda - t)x$ で変換して

$$= \frac{\lambda^n}{(\lambda - t)^n} \frac{1}{\Gamma(n)} \int_0^{\infty} e^{-y} y^{n-1} dy = (\frac{\lambda}{\lambda - t})^n.$$

よってガンマ分布の積率母関数は

$$M(t) = (1 - \frac{t}{\lambda})^{-n} \tag{1.90}$$

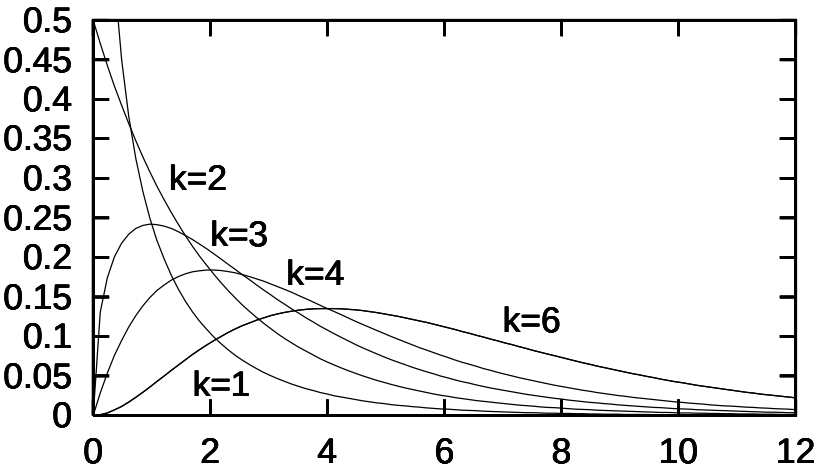

図 1.3: カイ二乗分布のグラフ

で与えられる。この母関数は次のように級数展開される[49]。

$$M(t) = 1 + \frac{n}{\lambda}t + \frac{n(n+1)}{\lambda^2}\frac{t^2}{2!} + \frac{n(n+1)(n+2)}{\lambda^3}\frac{t^3}{3!} + \cdots$$

この展開式からガンマ分布の平均値、分散は直ちに次の通り定まる。

$$E(X) = \frac{n}{\lambda} \tag{1.91}$$

$$E(X^2) = \frac{n^2+n}{\lambda^2}$$

$$var(X) = E(X^2) - [E(X)]^2 = \frac{n}{\lambda^2} \tag{1.92}$$

1.6　カイ二乗分布

自由度が k のカイ二乗分布の確率密度関数は

$$f(x) = \begin{cases} \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} & (x > 0) \\ 0 & (x \leq 0) \end{cases} \tag{1.93}$$

と表される。カイ二乗分布は標本分布の理論で重要な役割をもつ分布の一つである。以下に示すように、カイ二乗分布の平均は k 、分散は $2k$ となる。またカイ二乗分布はガンマ分布の特殊な場合である。(1.88) 式に $n = k/2, \lambda = 1/2$ を代入すればよい。
よって積率母関数は

$$M(t) = (1-2t)^{-\frac{k}{2}} \tag{1.94}$$

[49] (1.23) 式の x に t を対応させればよい。

となる (1.90 式を参照)。この積率母関数は次のように級数展開することができる。

$$M(t) = 1 + kt + k(k+2)\frac{t^2}{2!} + k(k+2)(k+4)\frac{t^3}{3!} + \cdots$$

よって

$$\begin{aligned} \mu_1' &= k. \\ \mu_2' &= k(k+2). \\ \mu_3' &= k(k+2)(k+4). \end{aligned}$$

一般に

$$\mu_i' = k(k+2)(k+4)\cdots(k+2i-2) \tag{1.95}$$

である。この結果からカイ二乗分布の平均値と分散は次のように定められる。

$$E(\chi^2) = \mu_1' = k. \tag{1.96}$$

$$var(\chi^2) = \mu_2' - \mu_1'^2 = 2k. \tag{1.97}$$

正規変数の二乗の和の分布がカイ二乗分布になること

いま $X_i(i = 1, 2, \cdots)$ が平均値 μ_i、分散 σ_i^2 の正規変数であって、互いに独立のとき、

$$Z_i = \frac{X_i - \mu_i}{\sigma_i} \tag{1.98}$$

は標準正規分布変数である。ここで

$$\chi^2 = Z_1^2 + Z_2^2 + \cdots + Z_k^2 \tag{1.99}$$

と定める。明らかに χ^2 は X_i の関数である。この χ^2 の分布は積率母関数によって次の通り導かれる。$Z_1, Z_2, \cdots, Z_k$ の同時密度関数は[50]

$$\begin{aligned} f(z_1, z_2, \cdots, z_k) &= \prod_{i=1}^{k}\left[\frac{1}{\sqrt{2\pi}}e^{-z_i^2/2}\right] \\ &= (\frac{1}{2\pi})^{\frac{k}{2}} e^{-\sum z_i^2/2} \end{aligned}$$

である。従って $\chi^2 = Z_1^2 + Z_2^2 + \cdots + Z_k^2$ の積率母関数は次の通りになる。

$$\begin{aligned} M_{\chi^2}(t) = E(e^{t\chi^2}) &= \int\cdots\int e^{t\chi^2} f(z_1, z_2, \cdots, z_k) dz_1 dz_2 \cdots dz_k \\ &= (\frac{1}{2\pi})^{\frac{k}{2}} \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{t\sum z^2} e^{-\sum z^2/2} dz_1 \cdots dz_k \\ &= \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(1-2t)z_i^2} dz_i \end{aligned}$$

[50] $\prod$ は n 個の数についての相乗積 (2 個以上の数を掛け合わせて得た積)。

ところで $\frac{\sqrt{1-2t}}{\sqrt{2\pi}}e^{-\frac{1}{2}(1-2t)z^2}$ は分散 $\frac{1}{1-2t}$ の正規分布の密度関数であるから

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(1-2t)z^2}dz=\frac{1}{\sqrt{1-2t}}$$

である。よって

$$M_{\chi^2}(t)=(\frac{1}{1-2t})^{\frac{k}{2}}\quad t<\frac{1}{2}$$

である[51]。従って、k 個の正規変数の二乗和の分布は自由度 k のカイ二乗分布となる。

以上の結果を使って $k\to\infty$ のときカイ二乗分布が正規分布に近付くことを次のように証明することができる。

$k\to\infty$ のときカイ二乗分布が正規分布に近付くことの証明

上記のようにカイ二乗変数の平均値は k、分散は $2k$ であるから、$K=(\chi^2-k)/\sqrt{2k}$ は標準正規変数 $Z=(X-\mu)/\sigma$ に見合う値である。χ^2 の積率母関数は $M_{\chi^2}(t)=(1-2t)^{-k/2}$ であるから、上の変換式で K の積率母関数を求めると次の結果が得られる[52]。

$$\begin{aligned}
M_K(t) &= E(e^{\frac{\chi^2-k}{\sqrt{2k}}t})=e^{-\frac{kt}{\sqrt{2k}}}(1-\frac{2t}{\sqrt{2k}})^{-\frac{k}{2}}\\
&= \left[e^{t\sqrt{2/k}}(1-\sqrt{\frac{2}{k}}t)\right]^{-\frac{k}{2}}\\
&= \left[(1+\sqrt{\frac{2}{k}}t+\frac{1}{2!}\left(\sqrt{\frac{2}{k}}t\right)^2+\cdots)(1-\sqrt{\frac{2}{k}}t)\right]^{-\frac{k}{2}}\\
&= \left[1-\frac{t^2}{k}-\left(\frac{2}{k}\right)^{\frac{3}{2}}\frac{t^3}{3}+\cdots\right]^{-\frac{k}{2}}
\end{aligned}$$

$k\to\infty$ のとき、括弧内の第 3 項は第 2 項よりも高次の無限小となるから、この式は近似的に次の形で示される。

$$M_K(t)\cong\left[1-\frac{t^2}{k}\right]^{-\frac{k}{2}}.$$

$k\to\infty$ のとき、右辺は $e^{\frac{t^2}{2}}$ に収束する[53]。これは標準正規分布の積率母関数である。よってカイ二乗分布は $k\to\infty$ のとき正規分布に近付く。　(証明終わり)

次で標本分散の場合について考える。基本的に似た話である。

[51] χ^2 はこれで 1 つの文字と同じ扱いをする。

[52] p.12 の積率母関数についての定理 1、2 を用いる。

[53]
$$\lim_{n\to\infty}(1+\frac{a}{n})^n=e^a$$

母集団分布が正規分布の時、その標本分散はカイ二乗分布をすること

母集団分布を平均値 μ、分散 σ^2 の正規分布とし、大きさ n の確率標本から計算した標本平均値、標本分散[54]をそれぞれ $\bar{X}, s^2$ とする。

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

s^2 の式を次のように変形する。

$$s^2 = \frac{1}{n}\sum(X-\mu)^2 - (\bar{X}-\mu)^2.$$

両辺に n/σ^2 をかけて移項すれば

$$\sum(\frac{X-\mu}{\sigma})^2 = \frac{ns^2}{\sigma^2} + \frac{(\bar{X}-\mu)^2}{\sigma^2/n}. \tag{1.100}$$

即ち、$\sum(\frac{X-\mu}{\sigma})^2$ が s^2 の項と $\bar{X}$ の項に分解されたことになる。X が正規分布変数のとき、この $\bar{X}$ と s^2 は互いに独立であることが証明される[55]。今、(1.100) 式を次の記号で示すことにする。

$$R = P + Q$$

両辺の積率母関数を考えると

$$M_R(t) = M_P(t) \cdot M_Q(t)$$

となる。ところで $R = \sum^n(\frac{X-\mu}{\sigma})^2$ は自由度 n のカイ二乗変数である。従って

$$M_R(t) = (1-2t)^{-\frac{n}{2}}$$

である。また $\sqrt{Q} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ は標準正規変数であるから、「X が平均値 0、分散 1 の標準正規変数のとき、X^2 は自由度 1 のカイ二乗分布をする」という定理より、Q は自由度 1 のカイ二乗変数である。よって

$$M_Q(t) = (1-2t)^{-\frac{1}{2}}$$

となる。従って $P = ns^2/\sigma^2$ の積率母関数は次の通りになる。

$$M_P(t) = (1-2t)^{-\frac{n-1}{2}}$$

これは自由度 $n-1$ のカイ二乗分布の積率母関数である。よって次の定理が成立する。

であるから、

$$\lim_{k\to\infty}\left\{(1+\frac{-t^2}{k})^k\right\}^{-\frac{1}{2}} = (e^{-t^2})^{-\frac{1}{2}} = e^{\frac{t^2}{2}}.$$

[54] 普通は $s^2 = (\sum_{i=1}^{n}(X_i-\bar{X})^2)/(n-1)$ の場合が多いと思う。確認して、必要だったら直すこと。

[55] 『統計数理入門』では証明が省略されている。

定理 母集団が分散 σ^2 の正規分布をするとき、s^2 を大きさ n の確率標本の標本分散とすれば、ns^2/σ^2 は自由度 $n-1$ のカイ二乗分布をする。

$s^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/n$ であるから

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

である。この式は $R = \sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2$ の母平均 μ の代りに標本平均 $\bar{X}$ を使ったものである。いずれもカイ二乗変数であるが、後者の自由度は n であるのに対し、前者の自由度は $n-1$ である。このとき $\sum(\frac{X-\bar{X}}{\sigma})^2$ は、平均値からの偏差の二乗和の計算に、母平均の代りに標本平均を使用したことによって自由度を一つ失った、あるいは平均値の推定のために自由度を 1 つ使用したという。

今、ns^2/σ^2 の自由度が $n-1$ であることに注意して、カイ二乗分布の密度関数 (1.93) 式を $x = ns^2/\sigma^2$ で変数変換すれば、標本分散 s^2 の分布が次の式で定められる。

$$f(s^2) = \left(\frac{n}{2\sigma^2}\right)^{\frac{n-1}{2}}(s^2)^{\frac{n-3}{2}}e^{-\frac{ns^2}{2\sigma^2}}/\Gamma\left(\frac{n-1}{2}\right). \tag{1.101}$$

1.7 t 分布

自由度が n の t 分布の確率密度関数は

$$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})(1+\frac{t^2}{n})^{\frac{n+1}{2}}} \quad , -\infty < t < \infty \tag{1.102}$$

と表される。t 分布の平均は 0、分散は $n/(n-2)$ (ただし $n \geq 2$ [56]) である。

1.4.2 節の定理 1 によれば、正規母集団の確率標本から計算された $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ は平均値 0、分散 1 の標準正規分布をする。このことは標本平均値 $\bar{X}$ の誤差を考えるときの基本原理として重要な意味を持つが、この式に含まれる母数 σ は一般に未知である。よって σ の値を何らかの方法で推定して使うしかない。

---

[56] $n = 1$ のときは

平均 $\int_{-\infty}^{\infty} tf(t)dt = \int_{-\infty}^{\infty}\frac{t}{\pi(1+t^2)}dt.$

奇関数なので平均 $= 0$。

$$\begin{aligned} \text{分散}\ \int_{-\infty}^{\infty} t^2 f(t)dt &= \int_{-\infty}^{\infty}\frac{t^2}{\pi(1+t^2)}dt = [t - 1\times\arctan t]_{-\infty}^{\infty} \\ &= \infty - \arctan\infty - ((-\infty) - \arctan(-\infty)) \\ &= \infty - (n\pi + \frac{\pi}{2}) + (n\pi - \frac{\pi}{2}) = \infty \end{aligned}$$

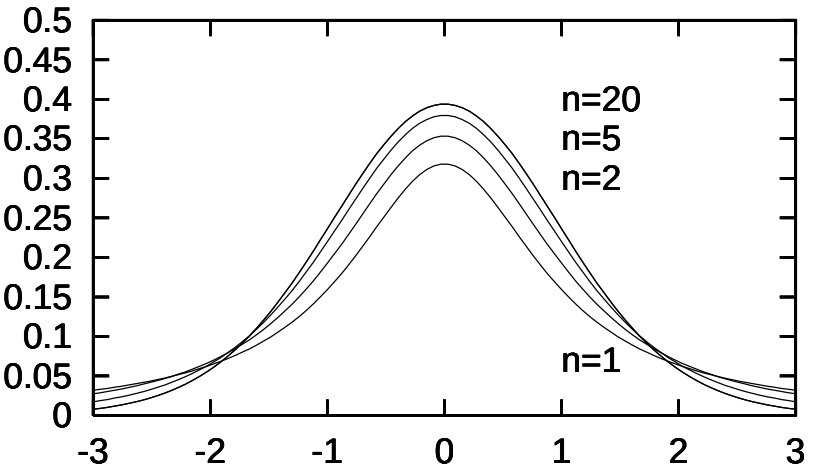

図 1.4: t 分布のグラフ

しかし σ をその推定値、例えば標本標準偏差 s で置き換えると、n が極めて大でない限り Z はもはや正規変数ではなくなる。そこで σ の推定値として何が適当であるか、そのときその推定値を使った新しい統計量がどのような分布をするかが問題となる。ここで登場するのが t 分布である。

t 分布は次のような分布である。X を標準正規分布する確率変数、Y を自由度 n のカイ二乗分布をするそれぞれ独立な確率変数とする。このとき、$t = X/\sqrt{Y/n}$ は自由度 n の t 分布をする。

< 証明 >

X、Y の密度関数は分かっているので、t の密度関数は次のように導かれる。X、Y は独立であるから、その同時密度関数は

$$f(x,y) = f_X(x) \cdot f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \cdot y^{\frac{n}{2}-1} e^{-\frac{1}{2}y} / 2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)$$

である。この式の変数 X を $t = X/\sqrt{Y/n}$ で t に変換すると

$$\begin{aligned} f(t,y) &= f\{x(t),y\}\left|\frac{dx}{dt}\right| \\ &= \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(t^2)\frac{y}{n}} y^{\frac{n}{2}-1} e^{-\frac{1}{2}y}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \left(\frac{y}{n}\right)^{\frac{1}{2}} \\ &= \frac{y^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}y(1+\frac{t^2}{n})}}{(2\pi n)^{\frac{1}{2}} 2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \end{aligned}$$

となる。さらに変数 Y を $Z = \frac{Y}{2}(1 + \frac{t^2}{n})$ で Z に変換すると

$$f(t,z) = f\{t,y(z)\}\left|\frac{dy}{dz}\right|$$

$$
\begin{aligned}
&= \frac{2^{\frac{1}{2}(n-1)} e^{-z} z^{\frac{1}{2}(n-1)}}{(1+\frac{t^2}{n})^{\frac{1}{2}(n-1)} (2\pi n)^{\frac{1}{2}} 2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \cdot 2\left(1+\frac{t^2}{n}\right)^{-1} \\
&= \frac{z^{\frac{1}{2}(n-1)} e^{-z}}{\sqrt{\pi n}(1+\frac{t^2}{n})^{\frac{1}{2}(n+1)} \Gamma(\frac{n}{2})}
\end{aligned}
$$

となり、これを Z の区間 $(0,\infty)$ について積分すれば、t の分布が次の通り計算される。

$$
\begin{aligned}
f(t) &= \int_0^\infty f(t,z)dz \\
&= \frac{1}{\sqrt{\pi n}(1+\frac{t^2}{n})^{\frac{1}{2}(n+1)} \Gamma(\frac{n}{2})} \int_0^\infty z^{\frac{n+1}{2}-1} e^{-z} dz \\
&= \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}(1+\frac{t^2}{n})^{\frac{1}{2}(n+1)} \Gamma(\frac{n}{2})}.
\end{aligned}
$$

t 分布の平均と分散の証明

t 分布は $t=0$ に対して対称であるから、奇数次の積率はすべて 0 である。よって平均値は 0 である。以下、分散について証明する。
t 分布は $t=0$ について対称であるから、

$$
var(t) = 2\int_0^\infty t^2 f(t)dt = \frac{2\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \int_0^\infty t^2 \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} dt
$$

$u = \frac{t^2}{n+t^2}$ とおけば、$dt = \frac{n^{\frac{1}{2}} u^{-\frac{1}{2}}}{2(1-u)^{\frac{3}{2}}} du$ であるから、

$$
\begin{aligned}
var(t) &= \frac{2\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \int_0^1 \frac{nu}{(1-u)^{\frac{1-n}{2}}} \frac{n^{\frac{1}{2}} u^{-\frac{1}{2}}}{2(1-u)^{\frac{3}{2}}} du \\
&= \frac{n\Gamma(\frac{n+1}{2})}{\sqrt{\pi}\Gamma(\frac{n}{2})} \int_0^1 u^{\frac{1}{2}} (1-u)^{\frac{n-2}{2}-1} du \\
&= \frac{n\Gamma(\frac{n+1}{2})}{\sqrt{\pi}\Gamma(\frac{n}{2})} B\left(\frac{3}{2}, \frac{n-2}{2}\right) \\
&= \frac{n\Gamma(\frac{n+1}{2})}{\sqrt{\pi}\Gamma(\frac{n}{2})} \frac{\Gamma(\frac{3}{2})\Gamma(\frac{n-2}{2})}{\Gamma(\frac{n+1}{2})} = \frac{\frac{1}{2}n\sqrt{\pi}}{\frac{n-2}{2}\sqrt{\pi}} = \frac{n}{n-2}
\end{aligned}
\tag{1.103}
$$

(証明終わり)[57]

[57] $B(m,n)$ はベータ関数である。ベータ関数には次の性質がある。ただし、$m>0, n>0$ である。

1. $B(m,n) = \int_0^1 x^{m-1}(1-x)^{n-1}dx = \int_0^1 z^{n-1}(1-z)^{m-1}dz = B(n,m)$
2. $B(m,n) = (\Gamma(m)\Gamma(n))/\Gamma(m+n)$

ベータ関数の性質の証明は、Appendix B を参照のこと。

$n \to \infty$ の時、t 分布が正規分布に近付くことの証明

今、(1.102) 式を

$$f(t) = C\left(1+\frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)} \quad \text{ただし} \quad C = \Gamma\left(\frac{n+1}{2}\right)/\sqrt{\pi n}\Gamma\left(\frac{n}{2}\right) \tag{1.104}$$

とおく。階乗計算のスターリングの公式は一般にガンマ関数 $\Gamma(n+1)$ に対して (n が整数であるとないとに関わらず)n が大きいときに成り立つ近似式[58]である。よって C をこの近似式で計算すると

$$\begin{aligned} C &= \frac{\sqrt{2\pi}\left(\frac{n-1}{2}\right)^{\frac{n}{2}} e^{-\frac{1}{2}(n-1)}}{\sqrt{2\pi}\left(\frac{n}{2}-1\right)^{\frac{1}{2}(n-1)} e^{-(\frac{n}{2}-1)}\sqrt{\pi n}} = \frac{(n-1)^{\frac{n}{2}}(n-2)^{-\frac{1}{2}(n-1)}}{2^{\frac{1}{2}}e^{\frac{1}{2}}\sqrt{\pi n}} \\ &= e^{-\frac{1}{2}}(\pi n)^{-\frac{1}{2}}\left(\frac{n-1}{n-2}\right)^{\frac{1}{2}(n-2)}\left(\frac{n-1}{n-2}\right)^{\frac{1}{2}}\left(\frac{n-1}{2}\right)^{\frac{1}{2}} \end{aligned} \tag{1.105}$$

となる。$n \to \infty$ のとき[59]

$$\left(\frac{n-1}{n-2}\right)^{\frac{1}{2}(n-2)} = \left(1+\frac{1}{n-2}\right)^{(n-2)\cdot\frac{1}{2}} \to e^{\frac{1}{2}}$$

$$\left(\frac{n-1}{n-2}\right)^{\frac{1}{2}} \to 1, \qquad \left(\frac{n-1}{n}\right)^{\frac{1}{2}} \to 1$$

であるから、

$$C \to (2\pi)^{-\frac{1}{2}}$$

となる。また $f(t)$ の残りの部分は

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

で、$n \to \infty$ のとき[60]

$$\left(1+\frac{t^2}{n}\right)^{n} \to e^{t^2}, \left(1+\frac{t^2}{n}\right) \to 1$$

---

[58]
$$\Gamma(n+1) \cong \sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}.$$
n が正の整数ならば $\Gamma(n+1) = n!$

[59]
$$\lim_{n\to\infty}\left(1+\frac{a}{n}\right)^n = e^a$$

[60] $h = n/t^2$ とおけば、$n \to \infty$ のとき $h \to \infty$ であるから、

$$\lim_{n\to\infty}(1+\frac{t^2}{n})^n = \lim_{h\to\infty}(1+\frac{1}{h})^{ht^2} = e^{t^2}$$

であるから

$$\left(1+\frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)} \to e^{-\frac{1}{2}t^2}$$

となる。よって $n \to \infty$ のとき

$$f(t) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$$

となる。右辺は標準正規分布の密度関数である。(証明終わり)

標本平均値の分布

1.4.2 節で見たように、正規分布を母集団とする標本についての統計量

$$Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$$

の分布は標準正規分布となる[61]。しかし母分散 σ^2 は未知な場合が多いので、σ^2 の代りに標本分散 $s^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/n$ を用いた[62]

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n-1}}} \tag{1.106}$$

がどのような分布に従うか調べておくと、 μ の区間推定に都合がよい。(1.106) 式は次のように書き改めることができる。

$$t = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \Big/ \sqrt{\frac{\frac{ns^2}{\sigma^2}}{(n-1)}}$$

この式の分子は標準正規分布変数であり、分母の $\frac{ns^2}{\sigma^2}$ は自由度 $(n-1)$ のカイ二乗変数である[63]。よって (1.106) 式の t は自由度 $(n-1)$ の t 分布をする[64]。この分布は σ の代りに s を用い、平均値を 0、標準偏差を 1 に調節したものとも言える。

[61] ただし $\bar{X}$ は正規母集団から取り出した大きさ n の確率標本の平均値、μ は母集団平均値である。

[62] $\sigma^2 = (n/n-1)s^2$ という関係を用いる。

[63] p.40 の 母集団分布が正規分布の時、$\cdots$ を参照のこと

[64] p.42 の t 分布についての証明を参照のこと。

第2章　統計的決定理論

2.1　決定理論

[1]前章まで、確率の理論から、推測統計の問題を処理するのに必要な各種の道具立てについて述べてきた。この章では、これらの道具を推測統計の諸問題にどう応用するかを考える。従来、推測統計の理論は ＜母数の推定＞ と ＜仮説の検定＞ の二つの問題に重点がおかれていた。ここでも同じように説明を進めるが、その前に 1940 年頃から提示された統計的決定の理論について簡単に説明する。

2.1.1　決定理論の一般的手順

母集団の母数 (統計学で、母集団の特性を表す定数。θ で表すことにする) は一般的に不明である。そこで我々は、母数 θ に代わる観察可能なデータ、即ち標本を使用し、できるだけ適切な決定 (母数の推定・仮説の検定) を行おうとする。このとき統計的決定理論では、次のように判断を進める。

1. 母数 θ がとりうるあらゆる可能な値の集合を定める。この集合を母数空間といい、記号 Φ で示す。
2. 母数 θ の各値に対してそれぞれ適当な一つの行動 a が対応する。a の集合を行動空間といい、記号 A で示す。
3. 決定関数の定義 母集団 $f(x;\theta)$ から任意の確率標本 $\{X_1, X_2, \cdots, X_n\}$ を選び、a が上記の行動空間の要素として含まれるように次のような関数を定める。

$$a = d(X_1, X_2, \cdots, X_n)$$

そして標本 $\{X_1, X_2, \cdots, X_n\}$ が観察されたとき、この関数によって定まる行動 a を決定するものとする。このとき関数 d を決定関数という (即ち決定関数は標本として得られた情報 $\{X_1, X_2, \cdots, X_n\}$ を行動空間の要素に誘導する関係式である)。決定関数はいろいろ考えられるが、この際できるだけよい推定値を与えるものでなくてはならない[2]。

---

[1]疑問点。1。母数空間とは何か? 2。「行動空間」は必要ないのでは。 3。母集団分布を推定することはあるのだろうか。 4。「検定」ではほとんどの場合、母集団分布を正規分布と仮定するのだろうか。

[2]この決定関数、後で登場する損失関数といった概念により、いろいろな問題を一般的に扱うことができる。なるべく損失の少ない決定関数を選べば、推定値は好ましいものとなるはずである。ただ実際には損失関数を考えるというより、よい推定値の条件として、不偏性・一致性・不変性などを考え、決定関数が選ばれるようである。

統計的推理は以上のような順序をふんで行われる。従来推測統計で取り扱われてきた二つの問題をこの決定理論の推理方式に当てはめてみると、母数推定の問題は、行動空間の要素が無限 (あるいは連続) の場合の決定問題であり、仮説検定の問題は、結論が仮説を容認するかしないかのどちらかであるから、行動空間が要素二個より成る特別の場合である。いまこれを平均値に関する推定および検定を例にとって説明すると次の通りである。

(1) 推定の問題

母集団分布 $f(x;\theta)$ を平均値 μ の正規分布とする。平均値 μ を推定する場合、母数 $\theta = \mu$ である。統計的決定理論に従うと、母数推定の手順は次のようになる。

1. まず、母数空間を決める。実数範囲と定めれば、

   $$\Phi = \{\mu; -\infty < \mu < \infty\}$$

   となる。
2. このとき行動は、μ の推定量 $\hat{\mu}$ を決定することである。推定量 $\hat{\mu}$ もまた実数範囲で考えられるから、行動空間は

   $$A = \{\hat{\mu}(\text{を決定する}); -\infty < \hat{\mu} < \infty\}$$

   となる。
3. 決定関数を定義する。$\hat{\mu}$ を決定するために、大きさ n の確率標本 $\{X_1, X_2, \cdots, X_n\}$ を観察して、例えばその平均値

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

   を推定値として選んだとする。これは決定関数を次の形で決めたことを意味する。

   $$a = d(X_1, X_2, \cdots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

   ここで、決定関数としてはこの他に例えば

   $$a' = d'(X_1, X_2, \cdots, X_n) = \left(\prod_{i=1}^{n} X_i\right)^{\frac{1}{n}} \quad (\text{幾何平均})$$

   あるいは

   $$a'' = d''(X_1, X_2, \cdots, X_n) = X_{\left(\frac{n+1}{2}\right)} \quad (\text{メジアン})$$

   を選ぶことも考えられる。この際できるだけよい推定値を与えるような決定関数を選ぶべきことはいうまでもない。

(2) 検定の問題

前と同様に母集団 $f(x;\theta)$ を平均値 μ の正規分布とする。検定とは例えば次のような二つの行動のうちどちらをとるかという選択を行うことである。二つの行動とは即ち、$\mu < k$ (k は所与の定数) と判断するか、$\mu \geq k$ と判断するか (あるいは仮説 $\mu < k$ をとるか、反対の仮説 $\mu \geq k$ をとるか) である。推定の問題と同様、統計的決定理論では、以下の手順で判断を行う。

1. 母数空間を $\Phi = \{\mu; -\infty < \mu < \infty\}$ と定める。
2. 行動空間を定義する。その前に「行動」は、$\mu < k$ と判断すること (a_1 で示す)、$\mu \geq k$ と判断すること (a_2 で示す) とする。このとき行動空間は、推定の場合と異なり、ただ二点 a_1, a_2 の集合

$$A = \{a; a = a_1, a_2\}$$

   である。
3. 決定関数を決める。大きさ n 個の確率標本を観察し、その平均値から次のように決定するものとする。即ち

$$\begin{aligned} \bar{X} &< k \quad \text{ならば } a_1 \text{をとり、} \\ \bar{X} &\geq k \quad \text{ならば } a_2 \text{をとる。} \end{aligned}$$

   このとき決定関数は次式で示される。

$$\begin{aligned} a = d(X_1, X_2, \cdots, X_n) &= a_1 \qquad \bar{X} < k, \\ &= a_2 \qquad \bar{X} \geq k. \end{aligned}$$

もちろんこれ以外の決定関数の決め方も考えられる。検定に関していえば、標本から計算される「検定統計量」によることが多い。どのように決定関数を決めるのがよいかを考えるのが検定理論の問題である。

推測統計の問題では以上の二つの類型に当てはめ得ないような問題も少なくない[3]。しかし、統計的決定の理論は、いろいろな問題を幅広く統一的に扱えるのである。

2.1.2　損失関数、危険関数

決定理論についての以上の説明で明らかなように、決定関数の決め方はいくつか考えられる。この複数の決定方式の中から最も適切と思われるものを選び出すためには、各行動の結果についての評価が必要となる。そこで決定理論では、損失関数という概念が導入される。

[3]即ち行動空間が有限個の多数の要素を含むとき。例えば実験で三つ以上の処理方法を試みてその効果の順位を決定する問題であるとか、観察の結果に基づいて三つ以上の可能な行動のうちの一つを選択するような問題。

損失関数の定義　母数が θ のとき、a という行動をとることによって生ずる損失を $l(a;\theta)$ で示し、これを損失関数という。損失関数は非負の実関数で、a が θ にとって最適である場合、$l(a;\theta)=0$ であると定める。

われわれは与えられた不完全な情報、即ち観察の結果 $\{x_1, x_2, \cdots, x_n\}$ に基づいて行動を次のように決定する。

$$a_0 = d(x_1, x_2, \cdots, x_n).$$

X_i は確率変数であるから、それに基づいて決定された行動 a もまた確率変数である。従って行動 a_0 に伴う損失

$$l(a_0;\theta) = l\{d(x_1, x_2, \cdots, x_n);\theta\}$$

もまた確率変数である。従って損失を小さくするということは、この損失関数の期待値を小さくするということ以外に考えようがない。

危険 (関数) の定義　損失の期待値を行動の危険 (危険関数) といい、$R(d;\theta)$ で示す。即ち

$$\begin{aligned} R(d;\theta) &= E[l(a;\theta)] \\ &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} l\{d(x_1, x_2, \cdots, x_n);\theta\} f(x_1;\theta) f(x_2;\theta) \\ &\qquad \cdots f(x_n;\theta) dx_1 dx_2 \cdots dx_n. \end{aligned}$$

この $R(d;\theta)$ は決定関数 d、損失関数 l、および母数 θ に依存しているが、実際にどのような情報 (標本) が観察されたかには無関係である。良好な決定関数は母数空間内のすべての θ に対してこの危険関数 R が最小であることが望ましい。しかし危険関数は行動 a の関数であると同時に母数 θ の関数でもある。従ってすべての θ に対して R が一様に最小になるような決定 d が存在するとは限らない。決定理論の問題はこのようなとき、どのような原理に基づいて決定方式を選択するかということである。

この点に関しては二つの原理が知られている。一つは Bayes の原理として知られているもので、もう一つは Wald によって提唱されたミニマクス原理である[4]。

ベイズの原理は母数 θ に関する事前確率、即ち事の真相についてどのような状態がどの程度可能であるかということがあらかじめ知られているか、あるいは推定可能であることを前提として、危険関数、即ち損失の期待値を最小にするような決定関数を求める。しかしこの原理は事前確率についてまったく情報が得られない場合には使えない。この困難を避けるためにワルドは、事の真相のいかんに関わらず、予想される危険 R の中の最大のものが最小であるような決定方式を選ぶ

[4]決定関数 (損失関数) を考える (複数個ある) → 決定原理 →ある決定関数 (なるべく小さい損失関数) に決まる、ということか。

こと (ミニマクス原理) を提唱した。言い換えると、事の真相が決定者にとって最も不利な場合を予想して、そのときの危険を最小にとどめようという考え方であって、消極的ではあるが安全第一の考え方である。

決定関数の選択原理はこの他にも幾通りも考えることができる。問題によって適当な原理を選ぶことが大切である。

第3章　母数の推定

この章では母数の推定を扱う。母数の推定は上記のように統計的決定問題の一つとして性格づけられるが、推定理論自体は決定理論が提唱される以前から、推測統計学の重要な問題の一つとして展開されてきた長い歴史を持っている。

既に述べたように、決定理論の一問題としての推定問題の場合、目的とする行動は母数の推定値を決定することである。以下本章で述べる推定の理論は、いわゆるパラメトリックの理論と呼ばれるものであって、ここで我々は母集団分布が一つまたは幾つかの母数によって定められる特定の形で与えられている場合に限定する。問題はこの母数の推定値を実験あるいは観察の結果に基づいて決定することである。

伝統的な母数の推定方法には点推定の方法と区間推定の方法とがある。点推定の方法は最もよいと思われるただ一つの推定値をもって母数を推定する方法で、区間推定の方法は母数の真価が高い確率をもって含まれていると考えられる実数値の区間によって母数を推定する方法である。まず初めに点推定の方法を説明する。

3.1　点推定

母集団分布を定める母数は必ずしも一個に限られていない。そして母数の推定は同時に二個以上の母数の推定を目的とする場合もあるが、以下では簡単な単一母数の場合について考える。問題の構造は複数母数の場合についても同様である。

点推定の定義　点推定の決定関数を

$$\hat{\theta} = d(X_1, X_2, \cdots, X_n)$$

とするとき、$\hat{\theta}$ を母数 θ の推定量といい、$\{X_1, X_2, \cdots, X_n\}$ の実際の観察値から計算された $\hat{\theta}$ の値を θ の推定値という[1]。

ここで $\hat{\theta}$ は推定という行動を意味する[2]と同時に、推定量という決定関数を意味する。$\hat{\theta} = \theta$ ならば損失関数は $l(\hat{\theta}, \theta) = 0$ である。点推定の問題はこのような損失関数を考慮して、危険 R の小さい決定関数、即ち推定量を求めることである。

[1]推定量：$\hat{\theta}$ は $\{X_1, X_2, \cdots, X_n\}$ の関数。推定値：$\hat{\theta}$ は $\{X_1, X_2, \cdots, X_n\}$ から計算されたある値。

[2]行動を考える必要はあるのだろうか?

推定の場合の損失関数としては次の形の関数が使われることが多い[3]。

$$l(\hat{\theta};\theta) = c(\theta)(\hat{\theta}-\theta)^2 \tag{3.1}$$

ここで $c(\theta)$ は θ のすべての値に対して正であるとする。この損失関数は母数の推定量 $\hat{\theta}$ と真の母数 θ との差が大になるほど損失が大となることを意味する。実際にはこの $c(\theta)$ を定数、特に $c(\theta)=1$ として、即ち損失関数を簡単に

$$l(\hat{\theta};\theta) = (\hat{\theta}-\theta)^2 \tag{3.2}$$

の形で使う場合が多い。以下これを平方誤差損失関数と呼ぶことにする。

そこで問題は母集団 $f(x;\theta)$ からの観察標本 $\{X_1, X_2, \cdots, X_n\}$ に対して、次の危険関数を最小にするような決定関数を見出すことである。

$$\begin{aligned} R(d;\theta) &= E[c(\theta)(\hat{\theta}-\theta)^2] \\ &= c(\theta)E[(\hat{\theta}-\theta)^2]. \end{aligned}$$

いま推定量を二つの決定関数

$$\hat{\theta}_1 = d_1(X_i) \qquad \hat{\theta}_2 = d_2(X_i)$$

の中から選び出すものとして、θ の一定の値に対して $R(d_1;\theta)$ と $R(d_2;\theta)$ とを比較する。

$$R(d_1;\theta) > R(d_2;\theta)$$

ならば明らかに $d_2 = \hat{\theta}_2$ が選ばれる。このとき上の式は次のように書ける。

$$E[l(\hat{\theta}_1;\theta)] > E[l(\hat{\theta}_2;\theta)].$$

いま上記の損失関数 (3.1) を使用すれば、この不等式は次の通りになる。

$$c(\theta)E[(\hat{\theta}_1-\theta)^2] > c(\theta)E[(\hat{\theta}_2-\theta)^2].$$

$c(\theta) > 0$ であるから

$$E[(\hat{\theta}_1-\theta)^2] > E[(\hat{\theta}_2-\theta)^2].$$

よって損失関数 (3.1) の場合、上の不等式が成立するときは、$\hat{\theta}_1$ の危険は $\hat{\theta}_2$ のそれよりも大である。よって危険の最小な推定量を求める問題は、$E[(\hat{\theta}-\theta)^2]$ 即ち平均平方誤差の最小な推定量を求める問題に帰着する。ただし残念ながら多くの $f(x;\theta)$ の場合、すべての θ に対して平均平方誤差が一様に最小になるような推定量は存在しない。つまり θ の値によって $E[(\hat{\theta}-\theta)^2]$ を最小にするような $\hat{\theta}$ が異なるのが普通である。よって平均平方誤差の最小は推定量選択の基本的な標準ではあるが、この標準だけで推定量の最良の形を推定することは困難であって、この他に種々の性質が推定量の選択標準として考えられることになるのである。

[3] 一般的には k 乗について。そういう意味では分散が重要なときに二乗したものを用いるべきだが、損失関数としては最もよく使われているようである。

3.1.1　点推定量の条件

推定量の条件として平均平方誤差の最小が基本的であることは上記の通りであるが、平均平方誤差が最小となること以外で推定量として望ましい性質としては、いくつかのものが考えられる。ここでは不偏性、一致性、効率の三つについて説明するが、これらは推定量の性質として最も重要、あるいは最も望ましいというよりは、むしろ推定の理論に関して伝統的に最もしばしば取り上げられてきたものというのが正しいであろう。

不偏性

平均平方誤差は次のように表すことができる。

$$\begin{aligned} E[(\hat{\theta}-\theta)^2] &= E[\{\hat{\theta}-E(\hat{\theta})\}-\theta-E(\hat{\theta})]^2 \\ &= E[\{\hat{\theta}-E(\hat{\theta})\}^2]+[\theta-E(\hat{\theta})]^2 \\ &= var(\hat{\theta})+[\theta-E(\hat{\theta})] \end{aligned} \tag{3.3}$$

よって平均平方誤差は二つの非負値の和である。$\theta - E(\hat{\theta})$ を推定量 θ の偏り (バイアス) という。推定量はその分散が小であって、同時に偏りも小であるほど良好である。従って推定量の偏りが 0 であること、即ち推定量が不偏であることは、それ自体望ましい性質である。

不偏推定量の定義　母数空間 Φ に属するすべての母数 θ に対して

$$E(\hat{\theta}) = \theta$$

が成り立つとき、$\hat{\theta}$ を θ の不偏推定量という。

推定量が不偏であることは望ましいことであるが、それは推定量の不可欠の条件ではない。幾分偏りがあっても、他の性質が優れていれば、推定量として優先することも十分考えられる。推定量の分布の重心が母数に一致することと、推定量の分布が母数の付近に集中していることとは別問題で、重要なのは後の方の性質である。

推定量が母数に対して偏りを有するとき、その偏りが一定であれば、その偏りの大きさだけ推定量を修正すれば不偏推定量が得られる。例えば $E(\hat{\theta}) = k\theta$ ならば、$E(\hat{\theta}/k) = \theta$ である。

例 分散の不偏推定量

標本分散 $s^2 = \sum(X-\bar{X})^2/n$ は母分散 σ^2 の不偏推定量ではない。

$$E(s^2) = \frac{n-1}{n}\sigma^2$$

だからである。ただしこの式から直ちに明らかなように

$$E(\frac{n}{n-1}s^2) = \sigma^2$$

であるから、分散の不偏推定量は標本分散 s^2 を次のように修正することによって簡単に得られる。

$$\frac{n}{n-1}s^2 = \frac{n}{n-1}\frac{\sum(X-\bar{X})^2}{n} = \frac{\sum(X-\bar{X})^2}{n-1} \tag{3.4}$$

一致性

よい推定量は標本の大きさが大となるほど危険の大きさが小さくなるようなものでなければならない。いま

$$\hat{\theta_n} = d_n(X_1, X_2, \cdots, X_n)$$

を大きさ n の標本 $\{X_1, X_2, \cdots, X_n\}$ に基づく推定量を表すものとする。このとき母数空間内のすべての θ に対して

$$\lim_{n\to\infty} R(d_n;\theta) = 0$$

が成り立てば、推定量 $\hat{\theta}$ は母数 θ の一致推定量であるという。損失関数に平方誤差形式 (3.2) を仮定すれば、この一致推定量は次の通り定義される。

一致推定量の定義 **1**　母数空間 Φ に属するすべての母数 θ に対して

$$\lim_{n\to\infty} E[(\hat{\theta}_n - \theta)^2] = 0 \tag{3.5}$$

ならば、$\hat{\theta}$ は θ の (平均誤差) 一致推定量である。

(3.3) 式で示したように、$E[(\hat{\theta}_n - \theta)^2] = var(\hat{\theta}) + [\theta - E(\hat{\theta}_n)]^2$ であるから、上の定義中の (3.5) 式は

$$\lim_{n\to\infty}[var(\hat{\theta}_n) + \{\theta - E(\hat{\theta}_n)\}^2] = 0$$

となり、一致推定量では n が大となるとき、推定量の分散も偏りも共に 0 に近付くことになる。従って上の定義はまた次のように表現することもできる。

一致推定量の定義 **2**　$n \to \infty$ のとき $E(\hat{\theta}) \to \theta$、$var(\hat{\theta}) \to 0$ ならば $\hat{\theta}$ は母数 θ の一致推定量である。

例　標本平均値 $\bar{X}$ は母平均 μ の一致推定量である。(1.73) 式の確率標本についての定義の系より $E(\bar{X}) = \mu$。母分散が σ^2 のとき、標本平均値 $\bar{X}$ の分散は $var(\bar{X}) = \sigma^2/n$ である。σ^2 が有限ならば、$n \to \theta$ のとき $\sigma^2/n \to 0$ となる。よって $\bar{X}$ は母平均 μ の一致推定量である。

例　母平均が正規分布のとき、標本分散 $S^2 = \sum(X-\bar{X})^2/n$ は母分散 σ^2 の一致推定量である。p.30 の定理 4 により、$E(s^2) = ((n-1)/n)\sigma^2$ であるから、$n \to \infty$ のとき $E(s^2) \to \sigma^2$ である。また母集団が正規分布のとき、41 の定

理より ns^2/σ^2 は自由度 $n-1$ のカイ二乗分布をする。そしてカイ二乗変数の分散は自由度の 2 倍である。よって

$$\begin{aligned} var(ns^2/\sigma^2) &= (n/\sigma^2)^2 var(s^2) = 2(n-1) \\ var(s^2) &= 2(\frac{n-1}{n^2})\sigma^4 \end{aligned} \tag{3.6}$$

となる。従って $n \to \infty$ のとき $var(s^2) \to 0$ であり。即ち母集団が正規分布のとき、標本分散 s^2 は母分散 σ^2 の一致推定量である。なお分散の不偏推定量 $(n/(n-1))s^2$ も同様に一致推定量であることが同じようにして容易に証明される。

一致推定量の定義 2 の条件、 $n \to \infty$ のとき $E(\hat{\theta}) \to \theta$、$var(\hat{\theta}) \to 0$ が成り立てば、次の条件式の成り立つことが証明される (しかし証明は省略)。ただし ϵ は任意に小さい正の値とする。

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \geq \epsilon) = 0 \tag{3.7}$$

条件 (3.7) 式は一致推定量の定義 1 の平方誤差の条件 (3.5) 式よりもゆるい。条件 (3.5) が成り立つためには二次積率の存在が必要であるが、条件 (3.7) はそれを必要としないから、適用の範囲が広く、従ってより一般的な一致性の条件として使われてきた。この条件によると一致推定量は次の通り定義される。

一致推定量の定義 3　任意の正の値 ϵ に対して

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \geq \epsilon) = 0$$

が成り立つとき、$\hat{\theta}_n$ を θ の一致推定量という。

簡単にいえば、標本の大きさが限りなく大となる時、ますます正確に母数に近付くような推定量が一致推定量である。

定理　標本積率は母集団積率の一致推定量である。 (証明は省略)

効率 (有効性)

効率の定義 1　母数 θ の二つの推定量を

$$\hat{\theta}_1 = d_1(X_i), \quad \hat{\theta}_2 = d_2(X_i) \qquad i = 1, 2, \cdots, n$$

として、その危険関数を次の通りとする。

$$R(d_1;\theta) \leq R(d_2;\theta)$$

このとき推定量 $\hat{\theta}_1$ を $\hat{\theta}_2$ よりも有効であるという。

$$e(d_1, d_2) = e(\hat{\theta}_1, \hat{\theta}_2) = \frac{R(d_2;\theta)}{R(d_1;\theta)} \tag{3.8}$$

を $\hat{\theta}_2$ に対する $\hat{\theta}_1$(あるいはより正確には d_2 に対する d_1) の相対効率という。

効率の定義 2　推定量が不偏で、かつその分散が他の不偏推定量の分散よりも小であるとき、この推定量を最良不偏推定量、あるいは最有効推定量という。

3.1.2　点推定の方法

母数推定の問題を前記のように統計的決定の問題として取り上げるならば、問題に応じて適当な形の損失関数を設定し、適当な選択原理、例えばミニマクス原理やベイズ理論等に基づいて推定方式を定める決定関数を求めるという順序で問題を処理するのが正しい考え方である。ただし母数推定の基本原理としては、ミニマクス原理にもベイズ原理にも種々の困難な問題が残されている (例えば既述のように損失関数に平方誤差形式を仮定するときは、ミニマクス原理を適用できないから損失関数になんらかの変更を加えることが必要である)。以下では決定理論的な問題から離れて、点推定の伝統的な方法の主なものを伝統的な接近方法で説明することにする。この際、推定方式の選択基準は、推定量が以上で述べた諸標準をできるだけよく満足することである。以下ではこれらの伝統的な方法のうち、積率法と最大尤度法について説明する[4]。

3.1.3　積率法

点推定の方法のうち、母数推定の最も歴史の古い方法が積率法である。積率法では、次の仮定がなされている。

1. 母集団分布は k 個の母数をもつ密度関数 (離散変数のときは度数関数) $f(x;\theta_1,\theta_2,\cdots,\theta_k)$ で定められている。
2. この母集団からの確率標本は、無作為抽出されたものとする。従って確率標本は、母集団の性質をよく表していると考える。

これらを仮定し、母数 $\theta_1,\theta_2,\cdots$ の推定量を求めることを考える。積率は一般にこれらの母数の関数として示すことができる。即ち、

$$\mu_i' = \mu_i'(\theta_1,\theta_2,\cdots,\theta_k)$$

である。いまこの母集団から大きさ n の確率標本 $\{X_1,X_2,\cdots,X_n\}$ をとり、その最初の k 次までの標本積率を $m_1',m_2',\cdots,m_k'$ とすると

$$m_i' = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

である。積率法では次の k 個の方程式

$$m_i' = \mu_i'(\theta_1,\theta_2,\cdots,\theta_k) \qquad i=1,2,\cdots,k \tag{3.9}$$

を θ_i について解いて得られる解を母数の推定量 $\hat{\theta}_i$ とする。母集団積率が直ちに分布の母数であるときは、標本積率は直接にその推定量となる。例えば平均値が分布の母数であるとき、標本の一次積率は常に母平均の推定量である。

[4]最小二乗法もまた、多用されている伝統的方法である。

例 1 積率法によって正規分布の母数、平均値と分散の推定量を求めよ。
平均値と分散は分布と無関係に原点積率によって次の通り表現できる。

$$\begin{aligned}\mu &= \mu_1' \\ \sigma^2 &= \mu_2' - \mu_1'^2\end{aligned}$$

従って最初の二次の原点積率を母数 $\mu,\ \sigma^2$ で表現すれば

$$\begin{aligned}\mu_1' &= \mu \\ \mu_2' &= \sigma^2 + \mu^2\end{aligned}$$

となる。そこで $m_1' = \mu_1',\ m_2' = \mu_2'$ とおいて

$$\begin{aligned}m_1' &= \mu \\ m_2' &= \sigma^2 + \mu^2\end{aligned}$$

この二式を $\mu,\ \sigma^2$ について解いて推定量

$$\begin{aligned}\hat{\mu} &= m_1' = \frac{1}{n}\sum X \\ \hat{\sigma}^2 &= m_2' - m_1'^2 = \frac{1}{n}\sum(X^2) - \left[\frac{1}{n}\sum(X)\right]^2 \\ &= \frac{1}{n}\sum(X - \bar{X})^2\end{aligned}$$

を得る。ただし平均値と分散の推定についての以上の結果は、正規分布以外でも平均値と分散をもつ全ての分布について成り立つことは明らかである。

例 2 ガンマ分布 $f(x) = \lambda^n e^{-\lambda x} x^{n-1}/\Gamma(n)$ の母数 $n,\ \lambda$ の推定量を求めよ。
ガンマ分布では $\mu_1' = n/\lambda$,$\mu_2' = (n(n+1))/\lambda^2$ である。よって

$$\begin{aligned}m_1' &= \frac{n}{\lambda} \\ m_2' &= \frac{n(n+1)}{\lambda^2}\end{aligned}$$

とおき、この二式を $n,\ \lambda$ について解いて次の結果を得る。

$$\begin{aligned}\hat{n} &= \frac{(m_1')^2}{m_2' - (m_1')^2} \\ \hat{\lambda} &= \frac{m_1'}{m_2' - (m_1')^2}\end{aligned}$$

以上の二例で明らかなように、積率法で得られる母数の推定量は標本積率 m'_i の関数の形で与えられる。標本積率は母集団積率の一致推定量[5]であるが[6]、このような標本積率の関数もまた通常の場合可能な条件のもとで一致推定量であることが証明される。積率法による母数推定量は必ずしも有効推定量ではないが、これを一次近似として更に効率の高い推定量を求める出発点とすることができる。

3.1.4　最大尤度法

母数の推定方法として最も広く使用されているのが、最大尤度法である。積率法と同様に、最大尤度法においても以下が仮定される。

1. 母集団分布は k 個の母数をもつ密度関数 (離散変数のときは度数関数) $f(x;\theta_1,\theta_2,\cdots,\theta_k)$ で定められている。
2. この母集団からの確率標本は、無作為抽出されたものとする。従って確率標本は、母集団の性質をよく表していると考える。

ここでは簡単のために母集団分布が一関数の密度関数 (または度数関数) $f(x;\theta)$ で定められているものとし、関数の形 f は既知であるが、母数 θ は未知で、その推定方法を考えるものとする。大きさ n の確率標本 $\{X_1,X_2,\cdots,X_n\}$ が n 個の一組の数値 $\{x_1,x_2,\cdots,x_n\}$ の内容で選ばれる確率密度 (離散変数のときは確率) は

$$L(\theta) \quad = \quad f(x_1,x_2,\cdots,x_n;\theta)$$

[5]いま

$$\hat{\theta}_n = d_n(X_1,X_2,\cdots,X_n)$$

を大きさ n の標本 $\{X_1,X_2,\cdots,X_n\}$ に基づく推定量を示すものとする。このとき母数空間内のすべての θ に対して危険関数 R が

$$\lim_{n\to\infty} R(d_n;\theta) = E[c(\theta)(\hat{\theta}-\theta)^2] = c(\theta)E[(\hat{\theta}-\theta)^2] = 0$$

となるならば、推定量 $\hat{\theta}$ は母数 θ の一致推定量であるという。

[6]定理 1　標本積率は母集団積率の一致推定量である。
< 証明 > 標本観察値を $\{X_1,X_2,\cdots,X_n\}$ とし、$Y_i = X_i^k$ とする。母集団分布が積率を有するものと仮定すれば、X は分散をもつから Y もまた分散をもつ。このとき大数の法則により、$n\to\infty$ のとき、任意に小さい正の値 ϵ に対して

$$P(|\bar{Y}_n - E(Y)| > \epsilon) \to 0.$$

従って $n\to\infty$ のとき $\bar{Y}_n \to E(Y)$ である。ところで $\bar{Y}_n = \frac{1}{n}\sum(X^k) = m'_k$、$E(Y) = E(X^k) = \mu_k$ であるから、$n\to\infty$ のとき

$$m'_k = \mu_k$$

である。(証明終わり)
大数の法則 X を平均値 μ、分散 σ^2 の母集団からの大きさ n の確率標本の標本平均値とすれば、$n\to\infty$ のとき、任意の正の値 ϵ に対して

$$P(|\bar{X}-\mu| \geq \epsilon) \to 0$$

あるいは

$$P(|\bar{X}-\mu| < \epsilon) \to 1$$

である。いいかえると $n\to\infty$ のとき X は母平均 μ に確率的に収束する。

$$= \quad f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta) \tag{3.10}$$

で与えられる。この確率密度は標本観察値 $\{x_1, x_2, \cdots, x_n\}$ にも依存しているが、同時に母集団を規定する母数 θ にも依存している。標本観察値 $\{x_1, x_2, \cdots, x_n\}$ を与えられたものとしてみるとき、この関数は母数 θ の関数となり、このとき $L(\theta)$ を確率標本 $\{X_1, X_2, \cdots, X_n\}$ の尤度関数という[7]。最大尤度法は母数 θ の変化に対して尤度関数 $L(\theta)$ の値を最大にするような統計量 $\hat{\theta}$ (そのような統計量が存在するならば) を母数 θ の推定量として決定する方法である。即ち最大尤度法では、実際に観察された結果としての標本の成立を最もよく説明するような母数 θ の値 $\hat{\theta}$ を見出して、それを θ の推定量と定めるのである。いいかえると標本観察値 $\{x_1, x_2, \cdots, x_n\}$ を既に与えられたものとして、θ を可能な範囲で変化させて考え、実際に観察されたような結果 (標本) が得られる確率密度 $L(\theta)$ が最大となるような θ の値 $\hat{\theta}$ を求める。そして観察された標本の実現を最もよく説明する $\hat{\theta}$ が定まれば、それが θ の真値であるときに最善と考えられる行動をとるのである。この行動は一定の損失関数に基づき、損失を最小にするような方法で決定される。

観察標本 $\{x_1, x_2, \cdots, x_n\}$ の最良の説明は標本そのものの内容に依存するから、損失を最小にする $\hat{\theta}$ はまたこの標本観察値 $\{x_1, x_2, \cdots, x_n\}$ に依存する。従ってそれは観察標本の関数 (従って統計量) であり、標本観察の可能な種々の結果に一定の行動 (推定値の決定) を対応させる一つの決定関数である。

最大尤度法についての定義　$L(\theta)$ が母集団 $f(x;\theta)$ の確率標本 $\{X_1, X_2, \cdots, X_n\}$ の尤度関数のとき、$L(\theta)$ を最大にする θ の値 $\hat{\theta}$ を母数 θ の最尤推定量という。

確率標本の場合、尤度関数 (3.10) 式は標本観察値のおのおのに対する確率密度関数の積である。従ってその最大値を求めるとき、そのままの形で取り扱うよりも、次のような対数の形で取り扱った方がしばしば簡単である。

$$\begin{aligned} \log L(\theta) &= \log\left[\prod_{i=1}^{n} f(x_i;\theta)\right] \\ &= \sum_{i=1}^{n} \log f(x_i;\theta) \end{aligned} \tag{3.11}$$

対数関数は単調増加であるから、$L(\theta)$ とその対数 $\log L(\theta)$ とは θ の同値で最大値をとる。よって $L(\theta)$ が微分可能のとき、次の方程式 (尤度方程式という)

$$\frac{d}{d\theta}\log L(\theta) = 0 \tag{3.12}$$

ただし

$$\frac{d^2}{d\theta^2}\log L(\theta) < 0$$

[7]確率密度関数と尤度関数とは、式の上では両者同じものにみえるが ((3.10) 式)、変数値と固定値との関係がちょうど逆になっている。

を解いて得られる $\theta = \hat{\theta}$ が最尤推定量である。

母集団分布が二つ以上の母数を有し、それらが同時に推定の対象となるとき[8]、尤度関数は

$$L(\theta_1, \theta_2, \cdots, \theta_k) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2, \cdots, \theta_k)$$

であって、母数 θ の最尤推定量 $\hat{\theta_1}, \hat{\theta_2}, \cdots, \hat{\theta_k}$ は次の連立方程式を解くことによって決定される。

$$\begin{aligned} \frac{\partial}{\partial \theta_1} \log L(\theta_1, \theta_2, \cdots, \theta_k) &= 0 \\ \vdots & \quad \vdots \\ \frac{\partial}{\partial \theta_k} \log L(\theta_1, \theta_2, \cdots, \theta_k) &= 0 \end{aligned} \tag{3.13}$$

例 正規分布の母数 μ,σ^2 の推定 (連続分布、二母数の例)

分布の密度関数は

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

であるから、尤度関数は次の形で示される。

$$L(\mu, \sigma^2) = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum (X-\mu)^2},$$

$$\log L(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum (X-\mu)^2. \tag{3.14}$$

よって μ,σ^2 の最大値は次の二つの方程式を解いて求められる。

$$\begin{aligned} \frac{\partial}{\partial \mu} \log L &= \frac{1}{\sigma^2}\sum (X-\mu) = 0, \\ \frac{\partial}{\partial \sigma^2} \log L &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum (X-\mu)^2 = 0. \end{aligned}$$

第一式より

$$\hat{\mu} = \frac{1}{n}\sum X = \bar{X}.$$

これを第二式に代入して

$$\hat{\sigma^2} = \frac{1}{n}\sum (X - \bar{X})^2 = s^2$$

が得られる。この結果で明らかなように分散の最尤推定量は不偏推定量ではない。なお以上の計算の過程で明らかなように、平均値 μ の推定値は分散 σ^2 の知識なしに定まるが、σ^2 の推定値は $\hat{\mu}$ の値が定まった後でなければ定まらない。

[8]母数が二つ以上あっても、そのうちの一つが推定の対象で、他の母数が全て既知として取り扱えるときは、既述の一母数の場合の方法で処理することができる。

3.2　区間推定

点推定の方法では最もよいと思われる推定値が一つ示されるだけで、その正確さについての表示は含まれていない。推定値の計算の基礎になっている標本の大きさが分かれば、ある程度推定値の信頼度について判断することもできるが、正確な判断をするためには誤差の理論をわきまえている必要がある。従って一般的にいって推定値を示す場合、その信頼度に関して直接何らかの表示が望まれることになる。

一つの方法は推定値の標準偏差を信頼度の表示として (このときこれを標準誤差という) 推定値に付記することである。この推定値の標準誤差は、理想的には母数 θ と推定値 $\hat{\theta}$ の差の二乗の期待値の平方根として計算される。そして損失関数に平方誤差形式を考えるとき、それは推定値の信頼度の最も適切な示し方である。しかしこの標準誤差の計算式は通常未知の母数 θ に依存しているため計算不能である。従って多くの場合大標本を仮定して、この未知の母数の代わりにその標本からの推定値を用いて計算し、なおかつそれを推定値の標準誤差と称している。標準誤差が信頼度の尺度として使用されるのは、母数推定値の多くが (少なくとも近似的に) 正規分布をすること、そしてこのとき標準偏差は分布のばらつきの範囲を最も簡明に示し得ることに基づいていると考えられる。

例　平均値の推定の信頼度
標本平均値 $\bar{X} = \frac{1}{n}\sum(X)$ は母平均 μ の最尤推定量である。母集団が分散 σ^2 の正規分布のとき、$\bar{X}$ は平均値 μ、分散 σ^2/n で正規分布をする。母集団分布が正規でなくても、大標本の場合 $\bar{X}$ は近似的に同様の正規分布をする。従って母分散 σ^2 が既知ならば $\sigma/\sqrt{n}$ を推定値 $\bar{X}$ の標準誤差として使うことができる。σ^2 が未知でも n が大であれば標本分散 s^2 を σ^2 の近似値として使って $\bar{X}$ の標準偏差が計算される。いずれにしても標本平均値 $\bar{X}$ は母平均 μ を中心にして (少なくとも近似的に) 正規分布をするから、$\bar{X}\pm$ [標準誤差の 2 倍または 3 倍] の区間が母平均 μ を含む確率はかなり大きいと考えてよい。

このように標準誤差は推定値のばらつきの範囲を示すという意味で信頼度の表示として使用されているのであって、本来それによって示そうと意図されていることは、母数がその間に含まれることがほとんど確実であると考え得る数値の範囲である。そしてこのことは直ちに次に述べる区間推定の考え方に結びつく。

区間推定の考え方は、例えば母平均 μ の推定値を次のような形で示す。

$$P(\bar{X} - a < \mu < \bar{X} + b) = \gamma \quad a > 0, b > 0$$

このとき区間 $(\bar{X} - a < \mu < X\bar{} + b)$ を母平均 μ の信頼区間、$\bar{X} - a, \bar{X} + b$ をそれぞれ信頼限界、γ をその信頼係数という。次にこの区間推定の方法を一般的な母数の推定に関して説明する。

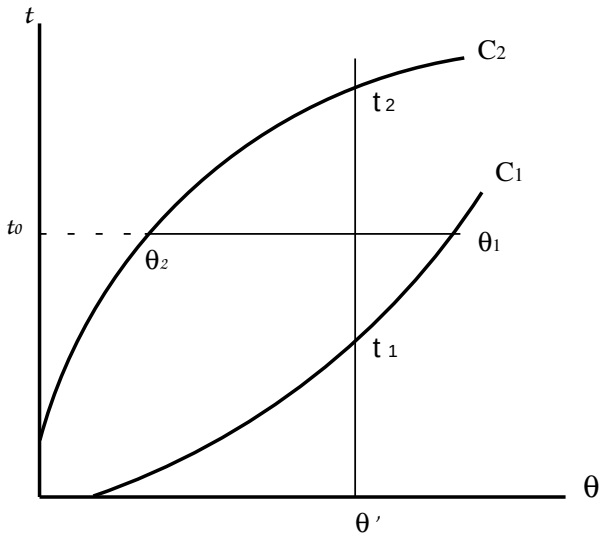

図 3.1: 信頼区間について

信頼区間の意味

いま分布の与えられた母集団 $f(x;\theta)$ からの大きさ n の確率標本を $X_1, X_2, \cdots, X_n$ とし、この標本から計算される統計量を $T = t(X_1, X_2, \cdots, X_n)$ とする。母集団分布 $f(x;\theta)$ が分かっているから、θ が与えられたとき、統計量 T の分布 $g(t;\theta)$ もまた定まる。このとき母数空間内の任意の θ に対して統計量 T が一定の区間 (t_1, t_2) の中に含まれる確率 γ を計算することができ、逆に一定の確率 γ に対して条件

$$P(t_1 < T < t_2) = \gamma \tag{3.15}$$

を満たす t_1, t_2 を計算することも可能である。

一定の確率 γ に対して条件 (3.15) を満たす区間 (t_1, t_2) の決め方は幾通りも考えられる。区間推定の考え方からいうと間隔 $t_2 - t_1$ の最も小さいものが望ましい。T の分布が対称のときは、t_1, t_2 を T の期待値に対して対称にとるとき $t_2 - t_1$ が最小となることは明らかである。

いま母数 θ がある任意の値 θ' のとき、条件 (3.15) を満たす統計量 T の値 t_1, t_2 を

$$P(T < t_1) = P(t > t_2) = \frac{1}{2}(1 - \gamma)$$

となるように定め、横軸に θ、縦軸に t をとってこの二点 t_1, t_2 を θ' の上の垂線上にとる。θ' は仮に定めた量であるから、母数空間内の全ての θ に対して上記の条件を満たす t_1, t_2 を考えるとき、明らかにそれは図 3.1 の C_1, C_2 で示すような二本の曲線を作るであろう。そして $(C_1, C_2$ がいずれも単調で θ の変域内で交わらないとすれば$)$ θ がどのような値をとるにしても統計量 T がこの二本の曲線ではさま

れる区間 (t_1, t_2) の中に含まれる確率が γ であることは明らかである。幾何学的にいえば統計量 $T = t$ の高さに引かれた水平線が線分 t_1t_2 と交わる確率は γ である。

さて観察標本 $\{x_1, x_2, \cdots, x_n\}$ に基づいて統計量 T の値 t_0 が得られたとする。t_0 を通る水平線を引き、曲線 C_1, C_2 との交点をそれぞれ θ_1, θ_2 とする。線分 $\theta_1\theta_2$ が線分 t_1t_2 と交われば、明らかに母数 $\theta = \theta'$ は区間 (θ_1, θ_2) に含まれる。我々は母数 θ の真値を知らないが、線分 $\theta_1\theta_2$ が線分 t_1t_2 と交わる確率が γ であることを知っている。これは母数 θ の値 θ' が変域内のどこにあっても真である。よって、θ_1, θ_2 は信頼係数 γ に対する母数 θ の信頼限界を与える[9]。

次に区間推定の方法を解析的に考えてみる。図 3.1 で明らかなように、$t = C_1(\theta), C_2(\theta)$ の逆関数 $\theta = C_1^{-1}(t), C_2^{-1}(t)$ を求めることができれば、統計量 T の観察値から θ の信頼限界を計算することができる。これを整理すると次のようになる。統計量 T の分布 $g(t;\theta)$ が知られているとき、与えられた信頼係数 γ に対して次の条件

$$P[t_1(\theta) < T < t_2(\theta)] = \gamma \tag{3.16}$$

を満たす二数 $t_1(\theta), t_2(\theta)$ が定まる。ここに $t_1(\theta), t_2(\theta)$ は未知の母数 θ に依存する値である。統計量 T が後に述べるような条件を満たすとき、不等式 $t_1(\theta) < T < t_2(\theta)$ を θ について解いて次の形の不等式を定めることができる。

$$\theta_2(t) < \theta < \theta_1(t) \tag{3.17}$$

(3.16) 式が成り立つとき、この不等式も同じく確率 γ で成り立つことは明らかである。統計量 $T = t$ を標本観察値に基づいて計算し $\theta_1(t), \theta_2(t)$ に代入すれば、信頼限界の値が計算される。この限界値は統計量 T に依存しており、従って確率変数である。そしてこのような計算によって定まる区間 (θ_1, θ_2) が母数 θ を含む確率が信頼係数 γ である。いいかえると、信頼係数はこのような方法によって計算される信頼区間が、真の母数 θ を含むことに成功する確率である。

3.2.1　正規母集団の平均値の区間推定

母集団を平均値 μ、分散 σ^2 の正規変数とし、分散 σ^2 が与えられているとして、未知の平均値 μ の推定を考える。大きさ n の確率標本からの標本平均値を $\bar{X}$ とするとき

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

が標準正規分布をすることは既知である。いま信頼係数を $\gamma = 0.95$ とすれば

$$P(-1.96 < Z < 1.96) = 0.95$$

[9]以上では母集団分布を連続的と仮定して説明したが、離散変数の場合も以上に準じて同じように考えることができる。このとき信頼限界を定める曲線は分布曲線の場合と同様の階段状の線になる。以下では引続き連続変数の場合について説明を進める。

である[10]。括弧内の不等式の Z を上式で置き換えれば

$$-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96$$

となる。この不等式を母数 μ について解いて次の不等式を得る。

$$\bar{X} - 1.96\sigma/\sqrt{n} < \mu < \bar{X} + 1.96\sigma/\sqrt{n} \tag{3.18}$$

この式は信頼係数 95%に対する母平均 μ の信頼限界を与える。

次に母集団を平均値 μ、分散 σ^2 が未知の正規変数とし、未知の平均値 μ の推定を考える。先の例では母分散 σ^2 を既知としたが、実際の場合 σ^2 は母平均 μ と同様におおむね未知である。このとき上の信頼限界は母数 σ を含んでいるから計算不能である。n が大ならば標本標準偏差 s を σ の近似値として代用できるが、n が小であるときは誤差が大きくなる。しかしこのような場合、我々は統計量

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n-1}}$$

が自由度 $n-1$ の t 分布変数であることを知っている。従って次の条件

$$P(t_1 < t < t_2) = 0.95$$

を満足する t_1, t_2 を計算することは可能である。このとき括弧内の不等式は次のような形になる。

$$-t_{0.95} < \frac{\bar{X} - \mu}{s/\sqrt{n-1}} < +t_{0.95}.$$

ただし $t_{0.95}$ は t 分布表から信頼係数 $\gamma = 0.95$、自由度 $n-1$ に対して得られる定数である。この不等式を μ について解いて母平均 μ の信頼限界が次の通り計算される。

$$\bar{X} - t_{0.95}\frac{s}{\sqrt{n-1}} < \mu < \bar{X} + t_{0.95}\frac{s}{\sqrt{n-1}}. \tag{3.19}$$

3.2.2　正規母集団の分散の区間推定

ここでは、母集団を分散 σ^2 の正規変数とし、未知の分散 σ^2 の推定を考える。1.6 節のカイ二乗分布に関する「定理」でみたように、分散 σ^2 の正規母集団からの大きさ n の標本から計算された標本分散を s^2 とすれば、ns^2/σ^2 は自由度 $n-1$ のカイ二乗分布をする。よって信頼係数を γ として、条件

$$P\left(\chi_a^2 < \frac{ns^2}{\sigma^2} < \chi_b^2\right) = \gamma$$

[10]ここで ± 1.96 は信頼係数 $\gamma = 0.95$ のときの標準正規分布の信頼限界である。この値は計算や数表によって求めることができる。

を満たす値 χ_a^2, χ_b^2 を定めることができる。括弧の中の不等式を母数 σ^2 について解けば、σ^2 の信頼限界が

$$\frac{ns^2}{\chi_b^2} < \sigma^2 < \frac{ns^2}{\chi_a^2} \tag{3.20}$$

の形で与えられる。ただし χ_a^2, χ_b^2 の値の決め方によって信頼区間の幅

$$\frac{ns^2}{\chi_a^2} - \frac{ns^2}{\chi_b^2}$$

は一様ではない。この幅が最小であるためには $1/\chi_a^2 - 1/\chi_b^2$ が最小でなければならない。本来、カイ二乗分布は非対称であるが、$n \to \infty$ のとき正規分布となるので、簡単のために対称と考えて以下のようにするとよい。n が極めて小さい値でなければ

$$P\left(\frac{ns^2}{\sigma^2} < \chi_a^2\right) = P\left(\frac{ns^2}{\sigma^2} > \chi_b^2\right) = \frac{1}{2}(1-\gamma)$$

となるように χ_a^2, χ_b^2 を定める。これで最適解に近い信頼区間が容易に得られる。

以下は細かい話である。区間推定の上記の方法がどのような母数の推定に対しても常に可能であるとは限らない。例えば先の例で、母分散 σ^2 が所与の場合、母平均 μ の信頼限界を計算するために我々は統計量 $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ を利用した。ここで σ は既知であるから、未知の母数は推定の対象である μ だけである。しかもこの統計量 Z は母数 μ に関係なく平均値 0、分散 1 の標準正規分布をする。なおこの統計量を定める $\bar{X}$ は充足統計量[11]である。このように区間推定が上記の方法で解析的に可能であるためには、次のような性質を持つ統計量を見出すことが必要である。

(1) それが充足統計量を通じて標本観察値と関係していること。

(2) 推定の対象である母数以外に (未知の) 母数を含まないこと。

(3) 母数の値に関係なく一定の分布を有すること。

統計量がこの三条件を備えているとき、この統計量をかなめ量という。母数の区間推定で σ^2 が未知の場合に使用した統計量 $t = (\bar{X} - \mu)/(s/\sqrt{n-1})$ も自由度 $n-1$ の t 分布変数であるから上の諸条件を満たしている。このようなかなめ量が存在するとき、不等式 (3.16) から不等式 (3.17) への転換が解析的に可能である。

[11]充足統計量の定義　統計量 T が母数 θ に関して標本 $X_1, X_2, \cdots, X_n$ のもつ情報のすべてを含むとき、この統計量を充足統計量という。
統計量 T が母数 θ に対して充足であるか否かは次の標準で判定することができる。
充足統計量に関する定理 標本 $\{X_1, X_2, \cdots, X_n\}$ の確率密度を $f(x_1, x_2, \cdots, x_n)$ とするとき、この密度関数が次の形

$$f(x_1, x_2, \cdots, x_n; \theta) = g(t; \theta)h(x_1, x_2, \cdots, x_n) \tag{3.21}$$

に因数分解することができれば、統計量 $T = t(x_1, x_2, \cdots, x_n)$ は母数 θ に対して充足である。

このような条件を満たすかなめ量が得られないときは、信頼区間の決定は別の方法によって行わなければならない。いま (3.16) の条件式

$$P[t_1(\theta) < T < t_2(\theta)] = \gamma$$

を次の形で定める。

$$P[T < t_1(\theta)] = P[T > t_2(\theta)] = \frac{1}{2}(1-\gamma). \tag{3.22}$$

母数 θ が例えば θ' という値をとるとき、この条件は次のことを意味する。

$$\begin{aligned} \int_{-\infty}^{t_1(\theta')} g(t;\theta')dt &= \frac{1}{2}(1-\gamma) \\ \int_{t_2(\theta')}^{\infty} g(t;\theta')dt &= \frac{1}{2}(1-\gamma) \end{aligned} \tag{3.23}$$

この式は θ' が θ の変域内にある限り常に当てはまる。ところで図 3.1 で明らかなように、

$$\begin{aligned} \theta &= \theta_1\text{のとき} \quad t_1(\theta_1) = t_0 \\ \theta &= \theta_2\text{のとき} \quad t_2(\theta_2) = t_0 \end{aligned}$$

である。そして (3.22) で定めたように

$$P[T < t_1(\theta_1)] = P[T > t_2(\theta_2)] = \frac{1}{2}(1-\gamma)$$

であるから、次の式の成り立つことも明らかである。

$$\begin{aligned} \int_{-\infty}^{t_0} g(t;\theta_1)dt &= \frac{1}{2}(1-\gamma) \\ \int_{t_0}^{\infty} g(t;\theta_2)dt &= \frac{1}{2}(1-\gamma) \end{aligned} \tag{3.24}$$

この式を θ_1, θ_2 について解くことができれば、信頼係数 γ に対する信頼区間 $(\theta_1 > \theta > \theta_2)$ を定めることができる。

以上では連続変数の場合について述べたが、離散変数の場合についても同様にこの方法が可能であって、離散変数の場合には特にこの方法が有用である。

第4章　検定

4.1　検定の基礎事項

4.1.1　検定とは

部分的な情報から、正確には知られていない母集団に関して 2 つのうちどちらの行動を選ぶか、という決定理論の 1 つの応用として仮説の検定というものがある。いま母数 (パラメータ) 空間を Φ 、その部分集合を ϕ とし、未知の母数を θ とする。母数に関する統計的仮説 H は母数 θ が母数空間の特定の部分集合 ϕ に含まれていること、即ち $\theta \in \phi$ で表現される。

検定に関する定義 **1**　統計的仮説 $H:\ \theta \in \phi$ において、集合 Φ がただ一つの要素から成るときこの仮説を単純仮説、また Φ が二つまたはそれ以上の要素を持つときこの仮説を複合仮説という。

検定において、母数空間 Φ は二つの部分集合 ϕ_0, ϕ_1 に分割される。この部分集合に対応し仮説検定では、命題 $\theta \in \phi_0$ を検定仮説、$\theta \in \phi_1$ を対立仮説という。前者を H_0、後者を H_1 で示す。一般に $H_0 : \theta \in \phi_0$ は現状の不変あるいは無差別を意味する命題の場合が多い。なお、検定仮説 H_0 を選ぶことを検定仮説を採択するといい、対立仮説を取ることを検定仮説を棄却するという。仮説検定では、検定仮説 H_0 と対立仮説 H_1 のいずれかを正しいと決定しなければならない。なお、対立仮説を想定しない検定仮説は存在しない。

さて、仮説の採否を決定するためには、母集団に関する情報収集 (データの観察) を行わなければならない。この観察データは確率標本を構成する。いま母集団 $f(x;\theta)$ より求められた大きさ n の確率標本を $\{X_1, X_2, \cdots, X_n\}$ とし、行動 (仮説の採否) を決める決定関数を

$$a = d(X_1, X_2, \cdots, X_n) \tag{4.1}$$

とする。行動 $a = \{a : a = a_0, a_1\}$ に対応して標本空間 S も同様に二つの部分集合 $S_0, S_1 (= S - S_0)$ に分割される。標本点 $s = (x_1, x_2, \cdots, x_n)$ が集合 S_0 に属するとき、

$$d(x_1, x_2, \cdots, x_n) = a_0$$

とし、検定仮説 H_0 を採択する。また標本点 s が集合 S_1 に属するとき

$$d(x_1, x_2, \cdots, x_n) = a_1$$

とし、仮説 H_0 を棄却する。部分集合 S_0 を仮説の採択域、S_1 をその棄却域という。

行動の決定は標本観察値、すなわち標本点によって直接行うのではなく、その標本点が定める特定の統計量 T によることが多い。例えば母平均 μ に関する検定を標本平均値 $\bar{X}$ によって行うなどである。このとき決定関数は

$$a = d(T) \quad T = T(X_1, X_2, \cdots, X_n) \tag{4.2}$$

で与えられる。このときの統計量 T を検定統計量という。各標本点 s はそれぞれ対応する T の値を有するから、標本空間内の採択域と棄却域に対応して検定統計量も仮説 H_0 を採択する領域と棄却する領域とに分かれる。これらの領域に対しても同様に採択域、棄却域の名称が用いられる。

4.1.2　検定で注意すべき点

以上の行動は未知の母数に関する不完全な情報に基づいているのであるから、これらの行動には常に損失の危険が伴っている。真の状態、即ち母数が θ のとき、行動 a を取ることによって生ずる損失は、損失関数 $l(a;\theta) \geq 0$ で与えられる。そして

$$\begin{aligned} &H_0\text{が真ならば (即ち}\theta \in \omega_0\text{ならば)} \quad l(a_0;\theta) &= 0 \\ &H_1\text{が真ならば (即ち}\theta \in \omega_1\text{ならば)} \quad l(a_1;\theta) &= 0 \end{aligned} \tag{4.3}$$

である。これらのとき行動 a_0, a_1 はいずれも正しかったことになる。決定 d に伴う危険は危険 (関数) の定義により次式で定められる。

$$R(d;\theta) = l(a_0;\theta)P(s \in S_0|\theta) + l(a_1;\theta)P(s \in S_1|\theta) \tag{4.4}$$

ただし $P(s \in S_0|\theta)$ は母数が θ のとき標本点 s が標本空間の採択域 S_0 に属する確率である。このとき行動 a_0 (仮説 H_0 の採択) が取られるから、$P(s \in S_0|\theta)$ を a_0 の行動確率 (H_0 の採択確率) という。同様にして $P(s \in S_1|\theta)$ は a_1 の行動確率 (H_0 の棄却確率) である。

いま母数 θ が集合 ϕ_i に属するとき、行動 a_j に伴う損失を $l(a_j;\theta \in \phi_i)$ で示すことにすれば、θ が ϕ_0 に属するときの危険は (4.4) 式により

$$R(d;\theta \in \phi_0) = l(a_0;\theta \in \phi_0)P(s \in S_0|\theta \in \phi_0) + l(a_1;\theta \in \phi_0)P(s \in S_1|\theta \in \phi_0)$$

である。(4.3) 式より $l(a_0;\theta \in \phi_0) = 0$ であるから

$$R(d;\theta \in \phi_0) = l(a_1;\theta \in \phi_0)P(s \in S_1|\theta \in \phi_0) \tag{4.5}$$

であり、同様にして

$$R(d;\theta \in \phi_1) = l(a_0;\theta \in \phi_1)P(s \in S_0|\theta \in \phi_1) \tag{4.6}$$

である。(4.5) 式右辺の $P(s \in S_1|\theta \in \phi_0)$ は母数 θ が集合 ϕ_0 に属し、従って行動 a_0 を取るべきであるにも関わらず、標本点 s が棄却域 S_1 に属していたため、誤った行動を取る確率であって、つまり検定仮説 $H_0 : \theta \in \phi_0$ が正しいに関わらず、これを採択しないで棄却する誤りをおかす確率である。この誤りを検定仮説の < 第一種の誤り > という。これに対して (4.6) 式の $P(s \in S_0|\theta \in \phi_1)$ は θ が ϕ_1 に属し、従って行動 a_1 を取るべきに関わらず a_0 を取る誤り、即ち仮説 H_0 が誤っているのに関わらずこれを採択する誤りをおかす確率であって、この誤りを < 第二種の誤り > という。(4.5) 式、(4.6) 式はそれぞれ第一種、第二種の誤りをおかす危険の大きさを示すことになる。

(4.4) 式と (4.5)(4.6) 式とで明らかなように、(4.4) 式の中の損失関数 $l(a_0;\theta), l(a_1;\theta)$ のいずれか一つは 0 であるから、(4.4) 式は次の形で表すことができる。

$$R(d;\theta) = l(\theta)\epsilon(d;\theta) \tag{4.7}$$

ただし $l(\theta)$ は損失関数、$\epsilon(d;\theta)$ は誤った行動をとる確率 (誤りの確率) で、(4.5)(4.6) 式で明らかなように次の通り定められる。

$$\begin{aligned} \theta \in \phi_0 \text{ならば} \quad l(\theta) &= l(a_1;\theta) \\ \epsilon_I &= P(s \in S_1|\theta \in \phi_0) \\ \theta \in \phi_1 \text{ならば} \quad l(\theta) &= l(a_0;\theta) \\ \epsilon_{II} &= P(s \in S_0|\theta \in \phi_1) \end{aligned}$$

ϵ_I は第一種の誤りの確率、ϵ_{II} は第二種の誤りの確率である。検定理論は母数空間 Φ に属するすべての値に対してこの危険関数が最小になるような検定方式、即ち決定関数を見出すことである。しかし一般にこの危険を最小にするような決定関数 (検定方式) は母数 θ の値によって一様ではない。そして検定仮説の場合母数 θ は未知であるから、与えられた問題の場合にこのような条件を満たす決定方式を発見することは容易な事柄ではない。

もう一つ、仮説検定の困難な問題は、危険関数を定める損失関数 l の形が一般に不明であるか、あるいは少なくとも正確に知られていないということである。もし損失関数の形が分からないとしたら、危険関数を小さくするために (4.7) 式のもう一つの因子である誤りの確率 ϵ を小さくすること、従ってそのような条件を満たす決定関数を求めることは合理的である。しかし母数 θ が ϕ_0 に属するか不明であるから、誤りの確率として ϵ_I と ϵ_{II} のいずれかを考えるべきかも不明である。そしてこの二つ、即ち第一種の誤りの確率と第二種の誤りの確率とは、一方を小さくすれば他方が大きくなるという背反的な性質を持っているから、二つの ϵ を同時に小さくすることは不可能である。そこで伝統的な仮説検定方式では、比較的小さな確率 α (例えば 0.01,0.05,0.10 等) を第一種の誤りの確率 ϵ_I の許容限度として定め、このような条件、即ち

$$\epsilon_I = P(s \in S_1|\theta \in \phi_0) \leq \alpha \tag{4.8}$$

を満たす検定方式、即ち棄却域 S_1 の集合を考え、このような集合の中で第二種の誤りの確率、即ち

$$\epsilon_{II} = P(s \in S_0 | \theta \in \phi_1) \tag{4.9}$$

を最小にするような検定方式、あるいはその余事象の確率

$$1 - \epsilon_{II} = P(s \in S_1 | \theta \in \phi_1) \tag{4.10}$$

を最大にするような検定方式を求める。

検定に関する定義 **2**　定められた一定の確率 α に対して検定方式が次の条件

$$\begin{aligned} &\phi_0\text{に属するすべての}\theta\text{に対して} \quad \epsilon_I \leq \alpha \\ &\phi_1\text{に属するすべての}\theta\text{に対して} \quad 1 - \epsilon_{II} \quad \text{が最大} \end{aligned}$$

を満たすとき、この検定方式を最良あるいは最強力であるという。

この検定方式では損失関数 $l(\theta)$ は明示的には考慮されていないが、事実は条件式 (4.8) の α の値の決定のときにそれが考えられている。もし第一種の誤りをしたときの損失 $l(a_1; \theta \in \phi_0)$ が第二種の誤りをしたときの損失 $l(a_0; \theta \in \phi_1)$ に比べて非常に重大であれば、検定方式の危険を小さくするためにそのような誤りをおかす確率の限度 α をなるべく小さな値に定めておかねばならない。故にこの検定方式は第一種の誤りが第二種の誤りに比べて特に重要であるという考え方に基づいているといえる。なぜならばこの場合は第一種の誤りをする確率を許容できる最小限度に抑えておくことが第一の条件だからである。

検定に関する定義 **3**　条件 $(4.8\ )$ 式の α 、即ち第一種の誤りの確率の許容される限度を検定の水準、有意水準あるいは棄却域の大きさという。

4.1.3　検出力関数

以下、統計的仮説の検定方式、即ち検定のための決定関数の求め方について説明する。ここでは検定仮説、対立仮説がともに単純仮説の場合について考える。この場合、検定に関する定義 2 の意味での最強力検定方式を求めることは容易である。いま、検定仮説を $H_0 : \theta = \theta_0$、対立仮説を $H_1 : \theta = \theta_1 (\neq \theta_0)$ とする。θ_0、θ_1 はいずれも一定値である。また第一種の誤りの確率を ϵ_I、第二種の誤りの確率を ϵ_{II} とする。この場合、ϵ_{II} の大きさは θ_1 次第で変わってくるから、ϵ_{II} は θ_1 の関数である。一般にある検定について

$$\pi(\theta_1) = 1 - \epsilon_{II}(\theta_1) \tag{4.11}$$

を、その検定の検出力という。また $\pi(\theta_1)$ を θ_1 の関数とみて検出力関数という。

一般によい検定とは仮説 $H_0 : \theta = \theta_0$、$H_1 : \theta = \theta_1$ に対し、α が小さく、かつ $\pi(\theta_1)$ が全ての θ_1 についてなるべく大きいような検定をいう。指定された α を有する検定方式のうち、全ての θ_1 について他のものよりも検出力の大きいものがあるならば、それは一様最強力な検定と呼ばれる。また $\theta = \theta_0$ において $\pi(\theta_1)$ が最小になる検定を不偏な検定という。

ここで、最強力検定に関連して有名なネイマン-ピアソンの定理を挙げておく。

ネイマン-ピアソンの定理　仮説 $H_0 : \theta = \theta_0$、対立仮説 $H_1 : \theta = \theta_1$ が与えられたとき、H_0 を検定する最強力な棄却域は、不等式

$$\frac{\prod_{i=1}^{n} f(x_i|\theta_1)}{\prod_{i=1}^{n} f(x_i|\theta_0)} \geq c \tag{4.12}$$

を満たす点 $x_1, x_2, \cdots, x_n$ の集合で与えられる (証明は省略)[1]。

検定仮説が複合仮説の形で与えられているとき、一様に最強力な検定方式を見出すことは困難であるから、この場合には母数 θ の全ての領域に対して最強力な検定方式を求めるという考え方をあきらめて、別の標準で検出力の比較的大きい検定方式を求めることを工夫しなければならない。次に述べる尤度比検定はその一つの方法である。

4.1.4　尤度比検定

以下の尤度比検定の方法は、必ずしも常によい検定方式を与えるという訳ではないが、多くの場合、特に大標本の場合はよい結果を導く方法である。いま母集団分布を簡単のために一母数関数の密度関数 $f(x;\theta)$ で定められるものとし、検定仮説を $H_0 : \theta \in \phi_0$ とする。ただし ϕ は母数空間 Φ の中の特定の部分集合とする。対立仮説は $H_1 : \theta \in \phi_1$、ただし $\phi_1 = \Phi - \phi_0$ である。標本の大きさを n とし、確率標本 $\{X_1, X_2, \cdots, X_n\}$ によって H_1 に対する H_0 の検定を考える[2]。

まず確率標本の尤度関数は $L = \prod_{i=1}^{n} f(X_i;\theta)$ で与えられる。いま母数 θ の関数としての尤度関数 L が母数空間 Φ 全体の上でとりうる最大値を

$$L(\hat{\Phi}) = \prod f(X_i;\hat{\theta}) \tag{4.13}$$

と定める。このとき $\hat{\theta}$ は明らかに母数 θ の最尤推定値である。これに対して仮説 $H_0 : \theta \in \phi_0$ が真のとき、即ち部分集合 ϕ の範囲内で尤度関数 L がとりうる最大値を

$$L(\hat{\phi}) = \prod f(X_i;\hat{\theta}') \tag{4.14}$$

[1] この定理の意味、使い方がよく分からない。分布の形を変えるのだろうか?
[2] こういう言い方をするのだろうか?

と定める。このとき ϕ は Φ の部分集合であって、部分集合では母数空間全体におけるよりも母数の範囲が制限されているから、尤度関数 L が大となる機会も制限される。従っていま尤度比を

$$\lambda = \frac{L(\hat{\phi})}{L(\hat{\Phi})} \tag{4.15}$$

と定義すれば明らかに $\lambda \leq 1$ である。この尤度比 λ は標本観察値のみの関数であって母数を含まないから統計量である。

さて尤度関数 L は確率標本 $\{X_1, X_2, \cdots, X_n\}$ の確率密度である。従って $\lambda = L(\hat{\phi})/L(\hat{\Phi})$ が 1 に近いほど、仮説 $H_0 : \theta \in \phi_0$ の真実性は大であるといえる。逆に λ が 0 に近いほど仮説 $H_0 : \theta \in \phi_0$ の真実性は乏しい。従って尤度比 λ は仮説 H_0 の検定のための検定統計量として利用することができる。

いま仮説 H_0 が真であって、そのときの尤度比 λ の確率密度 $g(\lambda|H_0)$ が知られているとする。$g(\lambda|H_0)$ は母集団分布 $f(x;\theta)$ が与えられていればそれから導くことができる訳である。もし $g(\lambda|H_0)$ が未知の母数を含まないならば、与えられた有意水準 α に対して次式の定める c の値を計算することができる。

$$\alpha = \int_0^c g(\lambda|H_0)d\lambda \tag{4.16}$$

このとき仮説 $H_0 : \theta \in \phi_0$ の棄却域は $0 < \lambda < c$ で与えられる。

尤度比検定に関する定理 **1**

[3]検定の有意水準を α とするとき、尤度比 λ による仮説 H_0 の棄却域は $0 < \lambda < c$ で定められる。ただし c は次式で定められる。

$$\alpha = \int_0^c g(\lambda|H_0)d\lambda$$

ここで $g(\lambda|H_0)$ は仮説 H_0 を条件とする尤度比 λ の確率密度である。

この定理に基づく仮説検定の方法を一般に尤度比検定という。時として統計量 λ の分布よりもその関数 $u(\lambda)$ の分布の方が計算しやすい場合がある。このときは次の定理によって $U = u(\lambda)$ を検定統計量として使用することができる。

尤度比検定に関する定理 **2**

$U = u(\lambda)$ が尤度比 λ の単調関数のとき、統計量 U に基づく仮説検定は λ による仮説検定と同値である。このとき $0 < \lambda < c$ に対応する棄却域は $u(0) < U < u(c)$ で与えられる。ただし U が λ の単調減少関数のときは $u(c) < U < u(0)$ で与えられる (証明は省略)。

[3]証明はしない。意味については前の段落を参照のこと。

以上では簡単のために母集団分布の母数を一個と仮定したが、母集団分布が二個以上の母数を有するときも、尤度比検定の方法は全く同一である。このとき尤度関数は $L = \prod_{i=1}^{n} f(X_i; \theta_1, \theta_2, \cdots)$ で与えられ、検定仮説は $H_0 : (\theta_1, \theta_2, \cdots) \in \phi_0$ である。

例 1 正規母集団の分散 $\sigma^2 = 1$ が与えられているとき、$H_0 : \mu = \mu_0$、$H_1 : \mu \neq \mu_0$ の検定方式を尤度比検定の方法によって求めよ。

$\sigma^2 = 1$ であるから尤度関数は

$$L(\mu) = (\frac{1}{\sqrt{2\pi}})^n e^{-\frac{1}{2}\sum(X-\mu)^2}$$

である。μ の最尤推定量は $\hat{\mu} = \bar{X}$ であるから母数空間 Φ での L の最大値は

$$L(\hat{\Phi}) = L(\hat{\mu}) = (\frac{1}{\sqrt{2\pi}})^n e^{-\frac{1}{2}\sum(X-\bar{X})^2}$$

である[4]。次に仮説 H_0 のもとでは $\mu = \mu_0$ は一定であって他に推定すべき母数はこの場合存在しないから、尤度関数 L の値はこの $\mu = \mu_0$ を代入した

$$L(\hat{\phi}) = L(\mu_0) = (\frac{1}{\sqrt{2\pi}})^n e^{-\frac{1}{2}\sum(X-\mu_0)^2}$$

以外にはない。よって尤度比は

$$\begin{aligned} \lambda = \frac{L(\hat{\phi})}{L(\hat{\Phi})} &= \exp\left[-\frac{1}{2}\{\sum(X-\mu_0)^2 - \sum(X-\bar{X})^2\}\right] \\ &= \exp\left[-\frac{n}{2}(\bar{X}-\mu_0)^2\right] \end{aligned}$$

となる。ここで n、μ_0 は与えられており、$\bar{X}$ の分布は正規であるから、上の式から $g(\lambda|H_0)$ を求めることは容易ではない。ただしこの場合 λ の対数を考えた方が処理がいっそう容易である。即ち上式から

$$\log\lambda = -n(\bar{X}-\mu_0)^2/2$$

を得る。$\mu = \mu_0$ のとき $\bar{X}$ は平均値 μ_0、分散 $1/n$ の正規分布変数であるから

$$-2\log\lambda = n(\bar{X}-\mu_0)^2$$

は自由度 1 のカイ二乗変数である。いま

$$Y = -2\log\lambda$$

とおけば、Y は λ の単調減少関数で、$\lambda = 0$ のとき $Y = \infty$ であるから、有意水準を α とすれば

$$\alpha = \int_0^c g(\lambda|H_0)d\lambda = \int_{-2\log c}^{\infty} h(y)dy$$

[4] $L(\hat{\Phi})$ は必ず ϕ_0 の方にあるのだろうか?

ただし $h(y)$ は自由度 1 のカイ二乗分布の密度関数である。$\alpha = 0.05$ とすればカイ二乗分布の数値表より

$$-2\log c = 3.84$$

を得るから検定の棄却域は

$$n(\bar{X} - \mu_0)^2 > 3.84$$

あるいは

$$\bar{X} < \mu_0 - \frac{1.96}{\sqrt{n}}, \qquad \bar{X} > \mu_0 + \frac{1.96}{\sqrt{n}}$$

で与えられる。このように棄却域が検定統計量 $\bar{X}$ の分布の両側にできるから、この検定方式は両側検定である。

例 2 例 1 では母分散 $\sigma^2 = 1$ が与えられていたが、同じく正規母集団で母分散が未知の場合について $H_0 : \mu = \mu_0$、$H_1 : \mu \neq \mu_0$ の検定方式を求めよ。
μ、σ^2 の最尤推定量は

$$\hat{\mu} = \bar{X} \qquad \hat{\sigma}^2 = \sum (X - \bar{X})^2/n$$

で与えられる[5]。よって

$$L(\hat{\Phi}) = L(\hat{\mu}, \hat{\sigma}^2) = \left[\frac{1}{(2\pi/n)\sum(X - \bar{X})^2}\right]^{\frac{n}{2}} e^{-\frac{n}{2}}$$

である。次に仮説 $H_0 : \mu = \mu_0$ のもとで尤度関数を最大にするための条件を求める。μ は仮説で定めたので、もう一つの未知の母数 σ^2 を L を最大にするように決定すればよい。σ^2 も仮説によって規定され、

$$\hat{\sigma}_0^2 = \sum (X - \mu_0)^2/n$$

である。よって

$$L(\hat{\phi}) = L(\mu_0, \hat{\sigma}_0^2) = \left[\frac{1}{(2\pi/n)\sum(X - \mu_0)^2}\right]^{\frac{n}{2}} e^{-\frac{n}{2}}$$

である。従って尤度比は

$$\lambda = \left[\frac{\sum(X - \bar{X})^2}{\sum(X - \mu_0)^2}\right]^{\frac{n}{2}}$$

となる。次の問題はこの λ の分布を求めることである。$X, \bar{X}$ の分布はともに正規であるからこのことは可能であるが、前の例ほど簡単ではない。まず λ の分母は

$$\sum (X - \mu_0)^2 = \sum (X - \bar{X})^2 + n(\bar{X} - \mu_0)^2$$

[5] p.60 の例「正規分布の母数 μ,σ^2 の推定」を参照のこと。

とかけるから、λ は次のような形で示すことができる。

$$\lambda = \left[\frac{1}{1 + n(\bar{X} - \mu_0)^2} \cdot \frac{1}{\sum (X - \bar{X})^2} \right]^{\frac{n}{2}}$$

ところで $\mu = \mu_0$ のときは

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n-1}} = \frac{\bar{X} - \mu_0}{\sqrt{\sum (X - \bar{X})^2 / n(n-1)}}$$

は自由度 $n-1$ の t 分布変数であるから

$$\lambda = \left[\frac{1}{1 + t^2/(n-1)} \right]$$

となる。λ の分布は t 分布をこの式によって変数変換して求めることができる。ただしこの場合、実際に分布 $g(\lambda)$ の形を計算する必要はない。λ は t^2 の単調減少関数になっているから、λ の棄却域に対応する t^2 の棄却域を定め、t を検定統計量として使用すれば問題は解決される。$\lambda = 1$ のとき、$t^2 = 0$ で、λ が 0 に近づくに従って t^2 は大となるから、λ の棄却域 $0 < \lambda < c$ に対応する t^2 の棄却域の形は $t^2 > A$ となる。よって任意の有意水準 α に対する t の棄却域の限界を $\pm t_{\frac{\alpha}{2}}$ とすれば[6]、仮説 $\mu = \mu_0$ の棄却域は $t = 0$ の両側で

$$t < -t_{\frac{\alpha}{2}} \qquad t > t_{\frac{\alpha}{2}}$$

で与えられ、$t_{\frac{\alpha}{2}}$ の値は

$$\int_{t_{\frac{\alpha}{2}}}^{\infty} f(t; n-1) dt = \frac{\alpha}{2}$$

から計算される。ただし $f(t; n-1)$ は自由度 $n-1$ の t 分布を示す。従って検定方式はこの場合、次のようにまとめることができる。

統計量 $t = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n-1}}$ (ただし $s^2 = \dfrac{\sum (X - \bar{X})^2}{n}$) を計算し、$-t_{\frac{\alpha}{2}} < t < t_{\frac{\alpha}{2}}$ ならば

仮説 $\mu = \mu_0$ を採択、その他の場合は仮説を棄却する。

この例で仮説 $\mu = \mu_0$ の採択域は次の不等式で与えられている。

$$-t_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu_0}{s/\sqrt{n-1}} < t_{\frac{\alpha}{2}}$$

この不等式は次の形で示すことができる。

$$\bar{X} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}} < \mu_0 < \bar{X} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}}$$

[6] t_α を $t_n(\alpha), t_{n-1}(\frac{\alpha}{2})$ などと書く場合もある。このとき、t_n, t_{n-1} は自由度が n もしくは $n-1$ の t 分布であることを示している。

この式は信頼係数 $\gamma = 1 - \alpha$ に対する母平均 μ の信頼限界である。故に尤度比検定による仮説 H_0 の検定は結局次の方法と同値である。即ち母平均の信頼係数 $\gamma = 1 - \alpha$ の信頼区間を計算し、もし仮説 $\mu = \mu_0$ がこの区間内にあれば仮説を採択し、区間内になければ仮説を棄却する。

多くの場合、母数に関する仮説の採択域はこのようにその母数の信頼区間の形で表すことができる。一般に仮説 $\theta = \theta_0$ に対する有意水準 α の検定 (採択域) は、母数 θ に対する信頼係数 $1 - \alpha$ の信頼区間で与えられる。

以上の 2 例では λ の関数を媒介として λ の分布に代わる既知の分布が容易に見出されたが、このようなことが可能でないとき、一般には λ の分布は複雑で、従って λ を検定量とするときの棄却域 $\lambda > c$ の計算は必ずしも容易ではない。このような場合、次の定理は特に有用である。

尤度比検定についての定理 **3**

極めて一般的な条件のもとで統計量 $-2\log\lambda$ の分布は、検定仮説 H_0 が真であれば、標本の大きさ n が大であるとき、カイ二乗分布で近似され、その自由度は仮説 H_0 に含まれる未知の母数の数に等しい。

例 2 では $-2\log\lambda = n(\bar{X} - \mu_0)^2$ はまさしく自由度 1 のカイ二乗変数であるから、この例は $-2\log\lambda$ が正確にカイ二乗変数で与えられている特別の例である。この例の場合と同様に $Y = -2\log\lambda$ は λ の減少関数であるから、Y の棄却域はカイ二乗分布の右側で定められる。従って大標本の場合に有意水準 α で尤度比検定を行うためには $Y = -2\log\lambda$ の棄却域 $Y > A$ をカイ二乗分布の右側で $\alpha = P(Y > A)$ となるように定めればよい。

例 3 平均値の差の検定

二つの正規母集団 X, Y で分散は未知であるが同一であることが知られている。仮説 $H_0 : \mu_x - \mu_y = 0$、$H_1 : \mu_x - \mu_y \neq 0$ の検定方式を求めよ。

今、母集団 X からの標本の大きさを m 、その標本平均値を $\bar{X}$ 、Y からの標本の大きさを n 、標本平均値を $\bar{Y}$ とし、両母集団に共通の母分散を σ^2 とする。母集団 X からの標本を $\{X_1, X_2, \cdots, X_m\}$ 、母集団 Y からの標本を $\{Y_1, Y_2, \cdots, Y_n\}$ とすれば、このとき標本の尤度関数は

$$L = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{m+n}{2}} e^{-\frac{1}{2\sigma^2}\left[\sum^m (X-\mu_x)^2 + \sum^n (Y-\mu_y)^2\right]} \tag{4.17}$$

である。これは p.60 の (3.14) 式から x についての式と y についての式をかけ合わせることで得られる。次に平均値 μ_x、μ_y および分散 σ^2 の最尤推定値についても p.60 の例のようにログをとって、

$$\frac{\partial}{\partial\mu_x}\log L = 0, \quad \frac{\partial}{\partial\mu_y}\log L = 0, \quad \frac{\partial}{\partial\sigma^2}\log L = 0$$

を解くことで以下のように得られる。

$$\hat{\mu_x} = \bar{X}, \quad \hat{\mu_y} = \bar{Y}$$

$$\hat{\sigma}^2 = \frac{1}{m+n}\left[\sum^m (X-\bar{X})^2 + \sum^n (Y-\bar{Y})^2\right]$$

従って

$$L(\hat{\Phi}) = \left[\frac{m+n}{2\pi\left[\sum(X-\bar{X})^2 + \sum(Y-\bar{Y})^2\right]}\right]^{\frac{m+n}{2}} e^{-\frac{m+n}{2}}$$

である。次に仮説 $H_0 : \mu_x = \mu_y = \mu$ が真のとき、平均値、分散の最尤推定値は (4.17) 式において $\mu_x = \mu_y = \mu$ とし、

$$\frac{\partial}{\partial \mu}\log L = 0, \quad \frac{\partial}{\partial \sigma^2}\log L = 0$$

を解いて次のようになる。

$$\begin{aligned}
\hat{\mu} &= \frac{1}{m+n}(\sum^m X + \sum^n Y) = \frac{m\bar{X} + n\bar{Y}}{m+n} \\
\hat{\sigma}^2 &= \frac{1}{m+n}\left[\sum^m (X-\hat{\mu})^2 + \sum^n (Y-\hat{\mu})^2\right] \\
&= \frac{1}{m+n}\left[\sum^m (X-\bar{X})^2 + \sum^n (Y-\bar{Y})^2 + \frac{mn}{m+n}(\bar{X}-\bar{Y})^2\right]
\end{aligned}$$

となる。従って、

$$L(\hat{\phi}) = \left[\frac{m+n}{2\pi\{\sum(X-\bar{X})^2 + \sum(Y-\bar{Y})^2 + \frac{mn}{m+n}(\bar{X}-\bar{Y})^2\}}\right]^{\frac{m+n}{2}} e^{-\frac{m+n}{2}}$$

である。よって尤度比は次の形で与えられる。

$$\lambda = \left[1 + \frac{\frac{mn}{m+n}(\bar{X}-\bar{Y})^2}{\sum(X-\bar{X})^2 + \sum(Y-\bar{Y})^2}\right]^{-\frac{m+n}{2}} \tag{4.18}$$

この検定は t 分布を使って表すことができる。その理由は以下の通りである。まず、標本平均値 $\bar{X}$、$\bar{Y}$ はそれぞれ平均値 μ_x、μ_y、分散 σ^2/m、σ^2/n の正規分布変数で[7]、互いに独立であるから、その差 $\bar{X}-\bar{Y}$ は平均値 $\mu_x - \mu_y$、分散 $\sigma^2(\frac{1}{m}+\frac{1}{n})$ の正規分布をする[8]。仮説 H_0 が真ならば $\mu_x - \mu_y = 0$ であるから、$\bar{X}-\bar{Y}$ の期待値も 0 である。次に $\sum^m (X-\bar{X})^2/\sigma^2$、$\sum^n (Y-\bar{Y})^2/\sigma^2$ はそれぞれ自由度 $m-1$、$n-1$ のカイ二乗変数であるから[9]、その和は自由度 $m+n-2$ のカイ二乗変数である。さらに仮説 H_0 が真のとき、

$$Z = \frac{\bar{X}-\bar{Y}}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}}$$

[7] p.34 の「正規分布を母集団とする標本集団の性質」の「定理 1」を参照のこと。

[8] 分散については、p.18 の「二変数分布の定理 3」を参照のこと。

[9] p.41 の「母集団分布が正規分布の時、その標本分散はカイ二乗分布をすること」の定理を参照のこと。$\sum^m (X-\bar{X})^2/\sigma^2$ は定理の ns^2/σ^2 にあたる。

は平均値 0、分散 1 の正規分布をするから[10]、

$$t = \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}} \Bigg/ \sqrt{\frac{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2}{\sigma^2(m + n - 2)}} \tag{4.19}$$

は自由度 $m + n - 2$ の t 分布変数である[11]。(4.18) 式と (4.19) 式とから

$$\lambda = \left[1 + \frac{t^2}{m + n - 2}\right]^{-\frac{m+n}{2}}$$

を得る。よって λ の代りに統計量 t を用いて仮説の検定を行なうことができる。即ち対立仮説 $H_1 : \mu_x \neq \mu_y$ に対して棄却域は

$$t < -t_{\frac{\alpha}{2}} \quad t > t_{\frac{\alpha}{2}}$$

$$\int_{t_{\frac{\alpha}{2}}}^{\infty} f(t; m + n - 2)dt = \frac{\alpha}{2}$$

で与えられる。またもし対立仮説が例えば $H_1 : \mu_x < \mu_y$ (片側検定) のときは

$$t < -t_\alpha$$

$$\int_{t_\alpha}^{\infty} f(t; m + n - 2)dt = \alpha$$

で与えられる。

4.2　母平均の検定

4.2.1　正規母集団の母平均の検定の一般的手順

正規母集団 $N(\mu, \sigma^2)$ の母平均 μ を検定するとき、母分散 σ^2 が既知か未知かで 2 通りに場合分けされるが、手順としては共に次のようになる。

1. 検定統計量 t を計算する。
2. 有意水準 α に対する t の棄却域の限界値 t_α を求める。
3. t が 棄却域に含まれるならば 仮説 $\mu = \mu_0$ を採択、そうでないときには仮説を棄却する。

ここで、対立仮説 H_1 の取り方により、棄却域の限界値 t_α は変わってくる。即ち、$H_1 : \mu \neq \mu_0$(両側検定) のときには仮説 $\mu = \mu_0$ の棄却域が $t = 0$ の両側で

$$t < -t_{\frac{\alpha}{2}}, \quad t > t_{\frac{\alpha}{2}}$$

[10] p.38 の (1.98) 式で、X を $X - Y$、μ を $\mu_x - \mu_y$、σ を $\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}$ に置き換えて考えればよい。
[11] p.42 の t 分布についての < 証明 > を参照のこと。

で与えられ、$t_{\frac{\alpha}{2}}$ の値は

$$\int_{t_{\frac{\alpha}{2}}}^{\infty} f(t;n-1)dt = \frac{\alpha}{2}$$

から計算される。ただし $f(t;n-1)$ は自由度 $n-1$ の t 分布を示す。これに対し $H_1 : \mu < \mu_0$ もしくは $H_1 : \mu > \mu_0$(片側検定) のときには、仮説 $\mu = \mu_0$ の棄却域が $t = 0$ の片側で

$$t < -t_\alpha \text{もしくは} t > t_\alpha$$

となり、t_α の値は

$$\int_{t_\alpha}^{\infty} f(t;n-1)dt = \alpha$$

から計算される。

4.2.2　母分散 σ^2 が既知の場合

このとき大きさ n の標本 $\{X_1, X_2, \cdots, X_n\}$ による母平均 μ の検定統計量 $T(\bar{X})$ は

$$T(\bar{X}) = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

である。この場合の検定統計量が上の様になる理由は p.73 の尤度比検定の例 1 を参照のこと。$T(\bar{X})$ の分布は標準正規分布 $N(0, 1^2)$ に従う。証明は p.34 の定理 1 を参照のこと。

4.2.3　母分散 σ^2 が未知の場合

普通、母分散 σ^2 は分からないことが多い。この場合、母分散 σ^2 の代りに標本分散 s^2 を用いて検定を行う。このとき大きさ n の標本 $\{X_1, X_2, \cdots, X_n\}$ による母平均 μ の検定統計量 $T(\bar{X}, s^2)$ は

$$T(\bar{X}, s^2) = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n-1}}}, \quad \text{ただし } s^2 = \frac{\sum(X - \bar{X})^2}{n}$$

である。この場合の検定統計量が上の様になる理由は p.74 の尤度比検定の例 2 を参照のこと。$T(\bar{X}, s^2)$ の分布は自由度 $n-1$ の t 分布に従う。証明は 1.7、t 分布の節の「標本平均値の分布」を参照のこと[12]。

4.2.4　応用例

ある領域における東西風データのうち、特定の年が以上であることを検定したい。具体的な問題は以下の通りである。インド付近のある領域における春の東西風の

[12]テキストによって、標本分散を $s^2 = (\Sigma(X - \bar{X})^2)/n$ とするものと $s^2 = (\Sigma(X - \bar{X})^2)/(n-1)$ とするものとがある。標本の数が少ない場合は後者の方が好ましいように思われる。$s^2 = (\Sigma(X - \bar{X})^2)/(n-1)$ としたときは、検定統計量は $T(\bar{X}, s^2) = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$ となる。

年々変動を考える。使用するデータは 1948 年から 1999 年までの 52 年間。領域平均した東西風の時系列データを作ると、1948、54、70、81、85、94 年は特に強いように見える。これらの年は普段の年に比べて異常であることを検定したい。以上の検定は「母分散が未知の場合の母平均の検定」を行えばよい。母集団は解析期間 (52 年間) の東西風とした。[13]この場合、母平均 $\mu_0 = 26.95$ は解析期間で平均した東西風となる。標本は循環の強い年 $(n = 6)$ の東西風であるとする。このとき標本平均は $\mu = 29.14$ で、標本分散は $s^2 = 0.3597$ である。また、仮説を $H_0 : \mu = \mu_0$ とする。これに対し、対立仮説は $H_1 : \mu < \mu_0$ となるので、片側検定を行えばよい。有意水準を 0.05 とするとき、自由度 5 の t 分布における棄却域の限界値は、$t_5(0.025) = 2.015$ となる。ここで検定統計量を計算すると $T(\bar{X}, s^2) = 8.151$ となる。T の値は棄却域に入るので、仮説 H_0 は棄却される。こうして、選び出した 6 年間の東西風は、普段の年に比べて異常に (有意に) 強いことが分かった。

4.3　2 つの正規母集団の母平均の差の検定

2 つの正規母集団 $N(\mu_x, \sigma_x^2)$、$N(\mu_y, \sigma_y^2)$ に母平均の差があることを示したい場合に用いる。差の検定についても 2 つの母分散 σ_x^2、σ_y^2 が既知か未知かによって、3 通りの方法が考えられる。

4.3.1　2 つの母分散 σ_x^2, σ_y^2 が既知の場合

このとき、大きさ m と n の二つの標本 $\{X_1, X_2, \cdots, X_m\}$、$\{Y_1, Y_2, \cdots, Y_n\}$ による母平均の差 $\mu_x - \mu_y$ の検定統計量 $T(\bar{x_1}, \bar{x_2})$ は

$$T(\bar{X}, \bar{Y}) = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}}$$

で、$T(\bar{X}, \bar{Y})$ の分布は標準正規分布 $N(0, 1^2)$ に従う[14]。

< 証明 >

今、母集団 X からの標本の大きさを m 、その標本平均値を $\bar{X}$ 、Y からの標本の大きさを n 、標本平均値を $\bar{Y}$ とし、母分散をそれぞれ σ_x^2、σ_y^2 とする。母集団 X からの標本を $\{X_1, X_2, \cdots, X_m\}$ 、母集団 Y からの標本を $\{Y_1, Y_2, \cdots, Y_n\}$ とすれば、このとき標本の尤度関数は

$$L = \left(\frac{1}{2\pi}\right)^{\frac{m+n}{2}} \left(\frac{1}{\sigma_x^2}\right)^{\frac{m}{2}} \left(\frac{1}{\sigma_y^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\left(\frac{\sum(X-\mu_x)^2}{\sigma_x^2} + \frac{\sum(Y-\mu_y)^2}{\sigma_y^2}\right)} \tag{4.20}$$

[13]ただし、これを母集団とするのが適当かは分からない。

[14] $T(\bar{X}, \bar{Y})$ の値は

$$T(\bar{X}, \bar{Y}) = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}}$$

かもしれない。

である。これは p.60 の (3.14) 式から x についての式と y についての式をかけ合わせることで得られる。次に平均値 μ_x、μ_y および分散 σ_1^2、σ_2^2 の最尤推定値を求める。まず、p.60 の例のように L のログをとると、

$$\log L = -\frac{m+n}{2}\log(2\pi) - \frac{m}{2}\log\sigma_x^2 - \frac{n}{2}\log\sigma_y^2 - \frac{1}{2}\left(\frac{\sum(X-\mu_x)^2}{\sigma_x^2} + \frac{\sum(Y-\mu_y)^2}{\sigma_y^2}\right)$$

となる。μ_x、μ_y、σ_x^2、σ_y^2 について微分すると、

$$\hat{\mu_x} = \bar{X}, \quad \hat{\mu_y} = \bar{Y}, \quad \frac{\partial}{\partial\sigma^2}\log L = 0$$

を解くことで以下のように得られる。

$$\hat{\mu_x} = \bar{X}, \qquad \hat{\mu_y} = \bar{Y}$$

$$\hat{\sigma_x}^2 = \frac{1}{m}\sum^{m}(X-\bar{X})^2, \quad \hat{\sigma_y}^2 = \frac{1}{n}\sum^{n}(Y-\bar{Y})^2$$

従って

$$L(\hat{\Phi}) = \left(\frac{1}{2\pi}\right)^{\frac{m+n}{2}} \left(\frac{m}{\sum(X-\bar{X})^2}\right)^{\frac{m}{2}} \left(\frac{n}{\sum(Y-\bar{Y})^2}\right)^{\frac{n}{2}} \cdot e^{-\frac{m+n}{2}}$$

である。次に仮説 $H_0 : \mu_x = \mu_y = \mu$ が真のとき、平均値、分散の最尤推定値は (4.20) 式において $\mu_x = \mu_y = \mu$ とし、

$$\frac{\partial}{\partial\mu}\log L = 0, \quad \frac{\partial}{\partial\sigma_x^2}\log L = 0, \quad \frac{\partial}{\partial\sigma_y^2}\log L = 0$$

を解いて次のようになる。

$$\hat{\mu} = \frac{m\bar{X}\sigma_y^2 + n\bar{Y}\sigma_x^2}{m\sigma_y^2 + n\sigma_x^2}.$$

ここで $\hat{\sigma_x^2}$、$\hat{\sigma_y^2}$ がうまく求められず挫折。

4.3.2　2 つの母分散 σ_x^2, σ_y^2 は未知だが、$\sigma_1^2 = \sigma_2^2$ と仮定できる場合

このとき、大きさ m と n の二つの標本
$\{X_1, X_2, \cdots, X_m\}, \{Y_1, Y_2, \cdots, Y_n\}$ による母平均の差 $\mu_1 - \mu_2$ の検定統計量

$$T(\bar{X}, \bar{Y}) = \frac{\bar{X} - \bar{Y}}{\sqrt{(\frac{1}{m} + \frac{1}{n}) \cdot \frac{\sum(X-\bar{X})^2 + \sum(Y-\bar{Y})^2}{m+n-2}}}$$

で、$T(\bar{X}, \bar{Y})$ の分布は自由度 $m+n-2$ の t 分布に従う。
検定統計量が上記のようになることは p.76 の尤度比検定の例 3 を参照のこと。

応用例

この検定が当てはまるのは、例えば、二種の製品の品質の優劣を双方からの抜取標本の検査成績によって判定する場合、あるいは新生児の体重が今と 10 年前とで変化しているか判定する場合などである。

例　新生児の体重の変化 1978 年と 1988 年の新生児の体重のデータが以下の表のように与えられた。10 年間で新生児の体重に変化があるかを検定したい。二つの母集団の平均を μ_x、μ_y とする。仮説は $H_0 : \mu_x = \mu_y$ とする。対立仮説は、新生児の体重は年々減少してきているという情報より、$\mu_x > \mu_y$ とする。

新生児の体重 (g)										
1978 年	3840	3540	3920	2920	3820	3910	3300	2770	3000	3900
1988 年	3470	2550	2920	2530	3280	2840	2520	3350	3610	3430

データより、標本平均は $\bar{X} = 3492$ g、$\bar{Y} = 3050$ g である。このとき検定統計量は

$$\begin{aligned} T(\bar{X}, \bar{Y}) &= \frac{\bar{X} - \bar{Y}}{\sqrt{(\frac{1}{m} + \frac{1}{n}) \cdot \frac{\sum (X-\bar{X})^2 + \sum (Y-\bar{Y})^2}{m+n-2}}} \\ &= \frac{3492 - 3050}{\sqrt{(\frac{1}{10} + \frac{1}{10}) \cdot \frac{1882760 + 1639600}{10+10-2}}} = \frac{442}{\sqrt{\frac{3522360}{5 \times 18}}} = \frac{442}{197.8} \approx 2.235 \end{aligned}$$

となる。ここで、自由度 18 の t 分布における有意水準 0.05 の棄却域は $t_{18}(0.05) = 1.734$ 以上である。検定統計量 $T(\bar{X}, \bar{Y})$ は棄却域に含まれるので、有意水準 0.05 で仮説は棄却される。従って、10 年間で新生児の体重は減少していると考えてよさそうである。

私が修論ではじめに行ったのはこちらの検定。まず、二つの集団を「循環の強い年 (6 年分)」と、「解析期間全体 (52 年分)」とした[15]。この 2 つの集団の平均 x_1, x_2 ににについての仮説 $x_1 = x_2$ が棄却されれば、シグナルは有意と言えるだろうと考えた。ただ、$\sigma_x^2 = \sigma_y^2$ はグラフからも明らかに成り立たないので、不適切であるということになり、「母分散が未知の場合の母平均の検定」でやり直した。下のウェルチの検定を用いれば問題ないのかも知れない。結論として、どの検定をどのように用いればよいのか、まだよく分からない。

[15] 比べるとしたら「循環の強い年」と「循環の強い年を除いた (52-6) 年」とすべきかもしれない。

4.3.3　母分散 σ_x^2, σ_y^2 が共に未知の場合

このとき、大きさ m と n の二つの標本
$\{X_1, X_2, \cdots, X_n\}, \{Y_1, Y_2, \cdots, Y_n\}$ による母平均の差 $\mu_1 - \mu_2$ の検定統計量 $T(\bar{X}, \bar{Y}, s_1^2, s_2^2)$ は、

$$T(\bar{X}, \bar{Y}, s_x^2, s_y^2) = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

ただし

$$s_1^2 = \frac{\sum (X_i - \bar{X})^2}{m-1}, \quad s_2^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1}$$

で、$T(\bar{X}, \bar{Y}, s_1^2, s_2^2)$ の分布は自由度 k の t 分布に従う。この自由度 k は

$$k = \frac{(\frac{s_1^2}{m} + \frac{s_2^2}{n})^2}{\frac{s_1^4}{m^2(m-1)} + \frac{s_2^4}{n^2(n-1)}}$$

で近似する。k が整数値でないときは、その最も近い整数値を k と定める[16]。これをウェルチの検定という。

4.4　無相関の検定

大きさ n の標本の相関係数が r のとき、無相関検定の検定統計量 $T(r, 0)$ は

$$T(r, 0) = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

で、この分布は自由度 $n-2$ の t 分布に従う。[17]このとき仮説および対立仮説は次のようになる。

$$H_0 \quad : \quad \text{無相関である},$$
$$H_1 \quad : \quad \text{相関がある}.$$

このとき $T(r, 0) \in R$ ならば[18]、有意水準 α で仮説 H_0 を棄却する。注意すべきは、この検定は両側検定であることである。$\pm t_{n-2}(\frac{\alpha}{2})$ の値を外れると棄却域になる。

[16] 証明は省略。
[17] 証明は省略。
[18] R は棄却域である。

第5章　未解決な問題

1. 4.1.3 節の検出力関数の意味がよく分からない。
2. 4.1.3 節 p.71 のネイマン・ピアソンの定理について。その意味と使い方がよく分からない。1 の検出力関数の問題と関連している。
3. 標本分散 s^2 の考え方。テキストによって n で割るものと $n-1$ で割るものとに分かれる。近年のテキストには後者が多いように思われる。ただし、検定統計量はどちらの立場でも、同じ量となっている。
4. 3 に関連して、自由度のとり方もいろいろ考えると難しい。SST など (東西風も?)、前年と相関が高い物理量に関しては、(データの数-1) を自由度とするのは問題、という話もあるらしい。
5. 4.3.1 節、「2 つの母分散 σ_x^2, σ_y^2 が既知の場合」の検定統計量の証明が途中で挫折している。

Appendix A. 正規分布の重要性とその応用

正規分布は経験的な分布にもしばしば見られる。例えば物理現象、天文現象における測定値の分布、生体測定値の分布、大量生産における品質の変異など、その若干例である。しかし正規分布の重要さはむしろ理論的にみたときの確率模型としての一般性に基づいている。正規分布は多くの確率分布模型の極限形と考えることができる。二項分布、ポアソン分布、超幾何分布などがそうである。また標本分布の理論模型であるカイ二乗分布や t 分布も正規分布に対して同様の関係にある。従って正規分布は確率模型として統計理論で最も基本的なものということができる。

また、確率変数が互いに独立な多数の確率変数の一次結合であるとき、その分布は個々の要素変数の分布法則に関係なく正規分布で近似されることが証明される。中心極限定理がそれである。標本平均値の確率分布が正規分布に従うのもこの定理に基づいている。平均値だけでなく、標本理論では誤差の分布法則に関して正規分布は重要な役割を持っている。このため、いろいろな統計問題の処理にはしばしばこの正規分布を援用しなければならないのである。

Appendix B. ガンマ関数とその性質

ガンマ関数は次の式で定義される。ここで n は任意の正の実数である。

$$\Gamma(n) = \int_0^\infty e^{-x} x^{n-1} dx.$$

次にガンマ関数のいくつかの性質について述べる。

1. $\Gamma(n) = (n-1)\Gamma(n-1)$
   <証明> $\Gamma(n)$ を部分積分法によって積分すれば

$$\begin{aligned} \int_0^\infty e^{-x} x^{n-1} dx &= \left[-e^{-x} x^{n-1}\right]_0^\infty + (n-1)\int_0^\infty e^{-x} x^{n-2} dx \\ &= (n-1)\int_0^\infty e^{-x} x^{n-2} dx. \end{aligned} \tag{1}$$

   ガンマ関数記号を用いれば

$$\Gamma(n) = (n-1)\Gamma(n-1).$$

   (証明終わり)

2. $n > 0$ が整数のとき $\Gamma(n) = (n-1)!$
   <証明> 1. を繰り返して置き換えれば

$$\Gamma(n) = (n-1)(n-2)\cdots 2 \cdot 1 \cdot \Gamma(1)$$

   となる。そして

$$\Gamma(1) = \int_0^\infty e^{-x} dx = \left[-e^{-x}\right]_0^\infty = 1$$

   であるから

$$\Gamma(n) = (n-1)!.$$

   ($n = 1$ のときも $0! = 1$ で成り立つ) (証明終わり)

3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
   <証明> $x = y^2$ とおけば

$$\Gamma(n) = \int_0^\infty e^{-x} x^{n-1} dx = 2\int_0^\infty y^{2n-1} e^{-y^2} dy. \tag{2}$$

   故に、

$$\Gamma(\frac{1}{2}) = 2\int_0^\infty e^{-y^2} dy,$$

$$\left[\Gamma(\frac{1}{2})\right]^2 = 4\int_0^\infty \int_0^\infty e^{-(y^2+z^2)}dydz.$$

極座標変換して $(0 < r < \infty, 0 < \theta < \frac{\pi}{2})$

$$\begin{aligned}\left[\Gamma(\frac{1}{2})\right]^2 &= 4\int_0^{\frac{1}{2}\pi} \int_0^\infty e^{-r^2} r dr d\theta \\ &= 4\left[\theta\right]_0^{\frac{1}{2}\pi} \cdot \left[-\frac{1}{2}e^{-r^2}\right]_0^\infty \\ &= 4 \cdot \frac{1}{2}\pi \cdot \frac{1}{2} = \pi.\end{aligned} \tag{3}$$

故に

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}.$$

(証明終わり)

4. n のいくつかの値に対して $\Gamma(n)$ を計算すると次のとおりである[1]。

n	$\frac{1}{2}$	1	2	3	4	5
$\Gamma(n)$	$\sqrt{\pi}$	1	1	2	6	24

[1] $n = 0$ のとき $\Gamma(n) = \infty$ であるが、この証明は難しい

Appendix C. ベータ関数とその性質

ガンマ関数と密接な関係を持つベータ関数が次の式で定義される。

$$B(m,n) = \int_0^1 x^{m-1}(1-x)^{n-1}dx \tag{4}$$

ただし、$m > 0, n > 0$ である。ベータ関数は次のような性質を持っている。

1. ベータ関数は変数 m, n に対して対称である。即ちこの二つの変数を交換しても関数の値は変わらない。(4) 式を $x = 1 - z$ で置換すると

   $$B(m,n) = \int_0^1 x^{m-1}(1-x)^{n-1}dx = \int_0^1 z^{n-1}(1-z)^{m-1}dz = B(n,m)$$

   となる。
2. ベータ関数は次のようにガンマ関数で示すことができる。

   $$B(m,n) = (\Gamma(m)\Gamma(n))/\Gamma(m+n) \tag{5}$$

   < 証明 >
   まず、ガンマ関数は次の形で記述できる。

   $$\Gamma(n) = \int_0^\infty e^{-x}x^{n-1}dx = 2\int_0^\infty y^{2n-1}e^{-y^2}dy \tag{6}$$

   $\Gamma(m)$,$\Gamma(n)$ を (6) の形で書けば

   $$\Gamma(m)\Gamma(n) = 4\int_0^\infty \int_0^\infty x^{2m-1}y^{2n-1}e{-}(x^2+y^2)dxdy. \tag{7}$$

   極座標に変換すれば[2]

   $$= 4\int_0^{\pi/2} \int_0^\infty r^{2(m+n-1)}e^{-r^2} r\sin^{2n-1}\theta\cos^{2m-1}\theta drd\theta$$

[2]極座標変換は以下のようになる。まず xy 座標系の x,y は $r\theta$ 座標系では $x = r\cos\theta$, $y = r\sin\theta$ であるから、

$$dx = \cos\theta dr - r\sin d\theta, \quad dy = -\sin\theta dr + r\cos\theta d\theta.$$

ここで xy 座標系におけるベクトル $(dx, 0)$,$(0, dy)$ 間の面積を考えると、

$$|(dx, 0) \times (0, dy)| = dxdy.$$

一方、$r\theta$ 系におけるベクトル $(\cos\theta dr, -r\sin d\theta)$, $(-\sin\theta dr, r\cos\theta d\theta)$ 間の面積は

$$|(\cos\theta dr, -r\sin d\theta) \times (-\sin\theta dr, r\cos\theta d\theta) = rdrd\theta.$$

以上より、$dxdy = rdrd\theta$ となる。また、θ に関する積分区間が $0 \sim \pi/2$ となるのは、はじめの積分区間が x, y ともに $0 \sim \infty$、つまり $x > 0$, $y > 0$ だったからである。

$$= \left[2\int_0^{\pi/2} \sin^{2n-1}\theta \cos^{2m-1}\theta d\theta \right] \left[2\int_0^{\infty} r^{2(m+n)-1} e^{-r^2} dr \right]$$

この式の第一項を $\sin^2\theta = z$ で変換すると

$$\int_0^1 (1-z)^{m-1} z^{n-1} dz = B(m,n)$$

となり、また第二項は (6) 式により $\Gamma(m+n)$ である。よって

$$\Gamma(m)\Gamma(n) = B(m,n)\Gamma(m+n)$$

となる。　　(証明終わり)

Appendix D. 用語

推測統計学　推計学とも。観察された集団のみを扱う記述統計学に対し、観察された集団を母集団から抽出された標本とみなし、これから確率法則に基づいて母集団の諸性質を推論する方法、またその方法の研究をいう。W. S. ゴセットの小標本の理論を端緒に、R. A. フィッシャーにより確立された。標本から母集団の母数を推定する推定論 (一つの推定値を求める点推定と、母数がある確率のもとで存在する区間を推定する区間推定 (信頼度がある) と、母集団に関するある仮説の成否を判定する検定論 (仮説検定) がある。応用面として標本調査、品質管理、実験計画法等があり、世論調査、製品検査、医薬の効用、教育・心理関係など、自然・社会科学にわたり多方面に利用される。

仮説検定　単に検定とも。母集団について立てた仮説 (帰無仮説)H_0 を捨てるか捨てないかを、母集団から抽出した標本 (測定値) によって定めること。H_0 が成立すると仮定して母集団から任意に n 個の標本の組を抽出したときの分布を調べ、それが $(100-\alpha)\%$の確率で存在する範囲を決め、実際に測定した n 個の値の組がその範囲の外 (棄却域) に落ちれば H_0 を捨てて、その範囲内にあれば H_0 を捨てない。α を危険率または有意水準といい、普通 5%または 1%にとる。この場合、仮説 H_0 が正しいのに測定値によって H_0 を捨てる誤りを第 1 種の誤りといい、その確率が危険率である。H_0 が正しくないのに H_0 を捨てない誤りを第 2 種の誤りという。

参考文献

『統計数理入門』 森田 優三 著 日本評論社 1968 年

『統計解析のはなし』 石村 貞夫 著 東京図書 1989 年

『すぐわかる統計解析』 石村 貞夫 著 東京図書 1993 年

『統計学講義 第 2 版』 林 周二 著 丸善 1973 年

『岩波数学辞典 第 3 版』 日本数学会 編 岩波書店 1993 年

『統計数学入門』 本間 鶴千代 著 森北出版 1970 年

『高校数学解法事典』 寺田文行、本部均 編 旺文社 1993 年

『高等学校 確率・統計』 高橋 睦男 他 著 数研出版株式会社 1989 年

『自然科学者のための数学概論 [増訂版]』 寺沢 寛一 著 岩波書店 2001 年 21 版

『マイペディア』 平凡社