統計的推定

目次

本スライドの内容

このスライドは，次の書籍の第 6 章「統計的推定」の内容に基づく．

- 『ガイダンス 確率統計：基礎から学び本質の理解へ』，発行：サイエンス社，ISBN：978-4-7819-1526-5.

書籍に関する最新の情報は，以下の URL から入手することができます．

`https://www.saiensu.co.jp`

この URL は，サイエンス社が運営しているホームページです．

概要

このスライドでは，統計的推定で必要となる統計量の性質を紹介する．また，これらの統計量と，標本調査で得られた標本データを利用して，未知パラメータである母数を推測する「統計的推定の考え方」を解説する．

はじめに (1)

調査の対象とする集合から得られる特性値 (数値) の集まりは母集団とよばれ，その値の分布は母集団分布とよばれる．母集団の大きさ (データの個数) が小さい場合は，母集団分布そのものや，母集団分布の代表値 (平均や分散など) を直接調べることができ，この調査方法は全数調査とよばれる．全数調査では，記述統計の考え方を用いて母集団のデータの特徴を明らかにする．これに対して，「母集団の大きさ」が大きく，母集団すべてを調べることが困難なとき，母集団から無作為に標本を抽出し，この標本を調べることにより元の母集団の特徴を推測する調査方法がある．この調査方法は標本調査とよばれ，標本調査に基づいて行われる (統計的推定や統計的仮説検定などの) 統計分析手法は推測統計とよばれる．

はじめに (2)

n 個の標本を取る操作は，確率空間 (Ω, P) 上の確率変数 $X_1, X_2, \cdots, X_n$ で表すことができ，これら n 個の確率変数は標本変量とよばれ，n は標本の大きさ，または標本サイズとよばれる．なお，X_k は k 番目に標本を取る操作を表し，X_k の実現値 (観測値) は小文字 x_k で表す．この標本変量の n 個の実現値 $x_1, x_2, \cdots, x_n$ は (大きさ n の) 標本データとよばれる．

以降で紹介する推測統計では，母集団の大きさは標本の大きさ n より十分大きく，かつ標本は無作為に抽出する (つまり，母集団の各要素を等しい確率で抽出する) ことを想定する．そのため，以降では，標本変量 $X_1, X_2, \cdots, X_n$ は独立であり (独立性)，かつ各 X_k の分布は母集団分布と同じであると仮定する (同分布性).

本書では，独立性と同分布性の 2 つの性質をみたす標本変量を，この母集団からの (大きさ n の) 無作為標本とよぶ.

はじめに (3)

母集団分布を特徴付ける定数 (またはベクトル) は**母数**とよばれ，一般には母数を θ，母集団分布を D_θ という記号で表す．以降では，多くの場合，母集団分布には有限な平均，分散，標準偏差が存在すると仮定している．このとき，これらの母数をそれぞれ**母平均**，**母分散**，**母標準偏差**とよび，それぞれ記号 μ, σ^2, σ で表す．一般に，母集団分布が D_θ のとき，その母集団を D_θ 母集団とよぶ．なお，母集団分布が正規分布 $N(\mu, \sigma^2)$ のとき，その母集団を正規母集団 $N(\mu, \sigma^2)$ ともよぶ．同様に，母集団分布が指数分布 $\mathrm{Exp}(\lambda)$ のとき，その母集団を指数母集団 $\mathrm{Exp}(\lambda)$ ともよぶ．他にも，母集団分布がベルヌーイ分布 $Be(p)$ のとき，母数 p を**母比率**とよび，その母集団を二項母集団 $Be(p)$ ともよぶ．

定義 6.1.1 (統計量と標本分布), 注意 6.1.1

定義 6.1.1 (統計量と標本分布)

標本変量 $X_1, X_2, \cdots, X_n$ の関数 $T_n = T(X_1, X_2, \cdots, X_n)$ を統計量といい，統計量 T_n が従う分布を標本分布とよぶ．$X_1, X_2, \cdots, X_n$ にそれぞれの実現値 $x_1, x_2, \cdots, x_n$ を代入した統計量の実現値は，小文字で $t_n = T(x_1, x_2, \cdots, x_n)$ と表す．特に，母数 θ を推定する目的で使われる統計量 T_n を θ の推定量とよび，その実現値 t_n を θ の推定値とよぶ．

注意 6.1.1　定義 6.1.1 において，各 X_k は (Ω, P) 上の確率変数であるため，統計量 T_n も (Ω, P) 上の確率変数であり，

$$T_n(\omega) = T(X_1(\omega), X_2(\omega), \cdots, X_n(\omega)) \quad (\omega \in \Omega)$$

と定義される．

基本的な統計量とその平均 (分散) (1)

ある母集団からの大きさ n の無作為標本 $X_1, X_2, \cdots, X_n$ を考え，母平均 $\mu = E(X_k)$ が存在し，母分散 $\sigma^2 = V(X_k)$ が有限とする．次の 3 つの基本的な統計量の標本分布に関する結果を紹介する．

$$\text{標本平均：}\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \tag{6.1}$$

$$\widehat{s}_n^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)^2 \tag{6.2}$$

$$\text{不偏標本分散：}U_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2 \qquad (n \geq 2) \tag{6.3}$$

まず，例 3.2.7 より，$E(\overline{X}_n) = \mu$ かつ $V(\overline{X}_n) = \sigma^2/n$ である．

基本的な統計量とその平均 (分散) (2)

次に，系 3.2.1 より，次式が成り立つ．

$$E(\hat{s}_n^2) = \frac{1}{n}\sum_{k=1}^{n} E((X_k - \mu)^2) = \frac{1}{n}\sum_{k=1}^{n}\sigma^2 = \sigma^2. \qquad (6.4)$$

次に，$E(U_n^2)$ を計算する．まず，系 3.2.1 より，式変形

$$E\left(\sum_{k=1}^{n}(X_k - \overline{X}_n)^2\right) = \sum_{k=1}^{n} E((X_k - \mu + \mu - \overline{X}_n)^2)$$

$$= \sum_{k=1}^{n} E((X_k - \mu)^2) + \sum_{k=1}^{n} E((\overline{X}_n - \mu)^2) - 2\sum_{k=1}^{n} E((X_k - \mu)(\overline{X}_n - \mu))$$

$$= n\sigma^2 + nV(\overline{X}_n) - 2\sum_{k=1}^{n} E((X_k - \mu)(\overline{X}_n - \mu)) \qquad (6.5)$$

が成り立つ．

基本的な統計量とその平均 (分散) (3)

ここで，$X_1, X_2, \cdots, X_n$ は独立であるため，$\mathrm{Cov}(X_k, X_j) = 0\ (k \neq j)$ である．したがって，次の計算結果が得られる．

$$
\begin{aligned}
E((X_k - \mu)(\overline{X}_n - \mu)) &= \frac{1}{n}\sum_{j=1}^{n} E((X_k - \mu)(X_j - \mu)) \\
&= \frac{1}{n}V(X_k) + \frac{1}{n}\sum_{\substack{1\le j\le n \\ j\neq k}} \mathrm{Cov}(X_k, X_j) = \frac{\sigma^2}{n}. \qquad (6.6)
\end{aligned}
$$

(6.5), (6.6) および $V(\overline{X}_n) = \sigma^2/n$ より，次が成り立つ．

$$
E\left(\sum_{k=1}^{n}(X_k - \overline{X}_n)^2\right) = n\sigma^2 + nV(\overline{X}_n) - 2n\frac{\sigma^2}{n} = (n-1)\sigma^2,
$$

$$
E(U_n^2) = \frac{1}{n-1}E\left(\sum_{k=1}^{n}(X_k - \overline{X}_n)^2\right) = \sigma^2. \qquad (6.7)
$$

基本的な統計量とその平均 (分散) (4)

定義	平均 (分散)	名称
$\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$	$E(\overline{X}_n) = \mu \quad (V(\overline{X}_n) = \sigma^2/n)$	標本平均
$\widehat{s}_n^2 = \frac{1}{n}\sum_{k=1}^n (X_k - \mu)^2$	$E(\widehat{s}_n^2) = \sigma^2$	
$U_n^2 = \frac{1}{n-1}\sum_{k=1}^n (X_k - \overline{X}_n)^2$	$E(U_n^2) = \sigma^2$	不偏標本分散

注意 6.1.2

(6.3) の U_n^2 の分布の理論的性質 (定理 6.1.2) を理解するためには，(6.2) の $\widehat{s}_n^2$ の分布の理論的性質 (定理 6.1.1) を先に理解しておくことが望ましい．そのため，$\widehat{s}_n^2$ を基本的な統計量の 1 つとして取り上げて解説する．しかし，μ の値が未知の場合は，$X_1, X_2, \cdots, X_n$ の値が定まっても $\widehat{s}_n^2$ の数値を具体的に計算できない．そのため，U_n^2 と比較すると，$\widehat{s}_n^2$ は区間推定や統計的仮説検定での応用の機会が限られる．

定理 6.1.1

次に紹介する定理 6.1.1 と定理 6.1.2 は，区間推定や統計的仮説検定で必要となる重要な定理である．定理 6.1.2 は結果を理解できれば十分である．

定理 6.1.1

$N(\mu, \sigma^2)$ 母集団からの大きさ n の無作為標本を $X_1, X_2, \cdots, X_n$ とする．このとき，(6.1), (6.2) で定めた $\overline{X}_n$ と $\widehat{s}_n^2$ に対して次が成り立つ．

$$(1) \quad \frac{n\widehat{s}_n^2}{\sigma^2} = \sum_{k=1}^{n} \left(\frac{X_k - \mu}{\sigma} \right)^2 \sim \chi^2(n) \qquad (6.8)$$

$$(2) \quad \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \sim N(0, 1) \qquad (6.9)$$

定理 6.1.1

[証明]　まず，例 3.2.7 の (3.21) より，(6.9) が成り立つ．次に，各 k に対して $Z_k = (X_k - \mu)/\sigma$ とおく．注意 2.1.6 より，各 Z_k は $N(0,1)$ に従い，補題 3.1.1 より，$Z_1, Z_2, \cdots, Z_n$ は独立である．したがって，定義 3.2.8 より，(6.8) が成り立つ．

定理 6.1.2

定理 6.1.1 の (6.9) の分母の σ を $\sqrt{U_n^2}$ に置き換えると，次の定理 6.1.2 の (6.11) となる．定理 6.1.2 は結果を理解できれば十分である．

定理 6.1.2

$N(\mu, \sigma^2)$ 母集団からの大きさ n の無作為標本を $X_1, X_2, \cdots, X_n$ とする ($n \geq 2$)．このとき，(6.1), (6.3) で定めた $\overline{X}_n$ と U_n^2 は独立であり，次が成り立つ．

(1) $$\frac{(n-1)U_n^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2 \sim \chi^2(n-1) \quad (6.10)$$

(2) $$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sqrt{U_n^2}} \sim t(n-1) \quad (6.11)$$

点推定：はじめに (1)

母数 θ の推定量 $T_n = T(X_1, X_2, \cdots, X_n)$ を決めて，「標本変量 $X_1, X_2, \cdots, X_n$ の実現値 $x_1, x_2, \cdots, x_n$ から求められる 1 つの推定値 $t_n = T(x_1, x_2, \cdots, x_n)$ が θ である」と推測するのが点推定の考え方．

推定量のばらつきの大きさを測定するための指標 (ものさし) として，分散 $V(T_n)$ や標準偏差

$$\mathrm{se}(T_n) := \sqrt{V(T_n)} = \sqrt{E((T_n - E(T_n))^2)}$$

などが用いられ，$\mathrm{se}(T_n)$ は推定量 T_n の標準誤差ともよばれる．母数 θ が実数値のとき，θ と推定量 T_n との「近さ」を測るために

$$\mathrm{MSE}(T_n, \theta) = E((T_n - \theta)^2)$$

を用いることにし，この $\mathrm{MSE}(T_n, \theta)$ を「θ の推定量 T_n に対する平均二乗誤差」とよぶ．$\mathrm{MSE}(T_n, \theta)$ が小さいほど「T_n は θ の良い推定量である」と考えられる．

点推定：はじめに (2)

母数 θ の真の値が未知のとき，一般には $\mathrm{MSE}(T_n, \theta)$ の値を計算できない．$\mathrm{MSE}(T_n, \theta)$ と分散 $V(T_n)$ の間には，大小関係

$$\begin{aligned} \mathrm{MSE}(T_n, \theta) &= E((T_n - E(T_n) + E(T_n) - \theta)^2) \\ &= E((T_n - E(T_n))^2) + 2(E(T_n) - \theta)E(T_n - E(T_n)) + (E(T_n) - \theta)^2 \\ &= V(T_n) + (E(T_n) - \theta)^2 \geq V(T_n) \qquad (6.26) \end{aligned}$$

が成り立つ．なお，T_n が $E(T_n) = \theta$ をみたす場合は $\mathrm{MSE}(T_n, \theta) = V(T_n)$ が成り立つため，この場合は $V(T_n)$ の計算を通じて $\mathrm{MSE}(T_n, \theta)$ の値を計算できる．このように $E(T_n) = \theta$ をみたす T_n は「θ の不偏推定量」とよばれる (定義 6.2.1).

点推定では，良い性質を持った推定量を採用することが重要である．推定量の望ましい性質として不偏性，一致性，最尤性などがあり，以下ではこれらの性質を順に紹介する．

定義 6.2.1 (不偏性)

定義 6.2.1 (不偏性)

母数 θ は実数値とする．標本変量 $X_1, X_2, \cdots, X_n$ に対し，統計量 $T_n = T(X_1, X_2, \cdots, X_n)$ が θ の不偏推定量であるとは，$E(T_n) = \theta$ が成立することである．また，$T_n = T(X_1, X_2, \cdots, X_n)$ と $S_n = S(X_1, X_2, \cdots, X_n)$ がともに θ の不偏推定量で，標準誤差の大小関係 $\mathrm{se}(S_n) \leq \mathrm{se}(T_n)$，つまり

$$\sqrt{E((S_n - \theta)^2)} \leq \sqrt{E((T_n - \theta)^2)}$$

をみたすとき，S_n は T_n よりも有効であるという．

例 6.2.1

$X_1, X_2, \cdots, X_n$ は，ある母集団からの大きさ n の無作為標本とし，母平均 $\mu = E(X_k)$ が存在し，母分散 $\sigma^2 = V(X_k)$ は有限とする．このとき，標本平均 $\overline{X}_n$ は $E(\overline{X}_n) = \mu$ をみたすため，μ の不偏推定量である．さらに，次も成り立つ．

$$E(\overline{X}_k) = \mu, \quad V(\overline{X}_k) = \frac{\sigma^2}{k} \quad (1 \leq k \leq n).$$

したがって，$1 \leq j < k \leq n$ に対し，$\overline{X}_j$ と $\overline{X}_k$ はともに μ の不偏推定量であり，$\overline{X}_k$ は $\overline{X}_j$ より有効である．

例 6.2.1

次に，定数 α_k $(1 \leq k \leq n)$ が $\sum_{k=1}^{n} \alpha_k = 1$ をみたすとき，統計量

$$T_n = \sum_{k=1}^{n} \alpha_k X_k$$

は線形推定量とよばれる．このとき，系 3.2.1 より，

$$E(T_n) = \sum_{k=1}^{n} \alpha_k E(X_k) = \sum_{k=1}^{n} \alpha_k \mu = \mu$$

であるため，T_n は μ の不偏推定量である．一方で，$X_1, X_2, \cdots, X_n$ の独立性と注意 3.2.3 より，次式が成り立つ．

$$V(T_n) = \sum_{k=1}^{n} \alpha_k^2 V(X_k) = \sigma^2 \sum_{k=1}^{n} \alpha_k^2. \qquad (6.27)$$

例 6.2.1

(6.27) と $\sum_{k=1}^{n} \alpha_k = 1$ より，次の不等式が成り立つ．

$$0 \leq \sigma^2 \sum_{k=1}^{n} \left(\alpha_k - \frac{1}{n} \right)^2 = \sigma^2 \left(\sum_{k=1}^{n} \alpha_k^2 - \frac{2}{n} \sum_{k=1}^{n} \alpha_k + \sum_{k=1}^{n} \frac{1}{n^2} \right)$$

$$= \sigma^2 \sum_{k=1}^{n} \alpha_k^2 - \frac{\sigma^2}{n} = V(T_n) - V(\overline{X}_n). \qquad (6.28)$$

不等式 (6.28) より，線形推定量の中で分散が最小となるのは標本平均 $\overline{X}_n$ である．このことが，母平均 μ の推定量として標本平均 $\overline{X}_n$ が広く用いられる根拠となる．

例題 6.2.1

例題 6.2.1

X_1, X_2 を $N(\mu, \sigma^2)$ 母集団からの大きさ 2 の無作為標本とする．このとき，次の統計量が σ^2 の不偏推定量となるように定数 c_1，c_2 を求めよ．

$$(1) \quad c_1(X_1 + X_2 - 2\mu)^2 \qquad (2) \quad c_2(X_1 - X_2)^2$$

[解答]　まず，系 3.2.4 より，$X_1 + X_2 \sim N(2\mu, 2\sigma^2)$ がわかる．よって，$E((X_1 + X_2 - 2\mu)^2) = 2\sigma^2$ がわかり，$c_1 = 1/2$ である．次に，系 3.2.5 より，$X_1 - X_2 \sim N(0, 2\sigma^2)$ がわかる．よって，$E((X_1 - X_2)^2) = 2\sigma^2$ がわかり，$c_2 = 1/2$ である．

定義 6.2.2 (一致性)

標本の大きさ n ごとに，母数 θ の推定量 $T_n = T(X_1, X_2, \cdots, X_n)$ が与えられることが多く，このとき，$T_1, T_2, \cdots, T_k, \cdots$ は推定量の (無限) 系列である．次に紹介する一致性は，「標本サイズ n ごとに与えられた推定量の (無限) 系列に対する性質」であり，標本サイズ n を大きく取れば，「推定量 T_n が母数 θ に近い値を取る確率」が 1 に近づくことを意味する．

定義 6.2.2 (一致性)

母数 θ は実数値とする．大きさ n の標本変量 $X_1, X_2, \cdots, X_n$ に対し定義される統計量 $T_n = T(X_1, X_2, \cdots, X_n)$ が θ の一致推定量であるとは，任意の $\varepsilon > 0$ に対して

$$\lim_{n\to\infty} P(|T_n - \theta| \leq \varepsilon) = 1$$

が成り立つことをいう．

例 6.2.2

$X_1, X_2, \cdots, X_n$ は，ある母集団からの大きさ n の無作為標本とし，母平均 $\mu = E(X_k)$ が存在し，母分散 $\sigma^2 = V(X_k)$ は有限とする．
定数 α_k $(1 \le k \le n)$ は $\sum_{k=1}^n \alpha_k = 1$ をみたすとし，線形推定量 $T_n = \sum_{k=1}^n \alpha_k X_k$ について考察する．例 6.2.1 より，

$$E(T_n) = \mu, \quad V(T_n) = \sigma^2 \sum_{k=1}^n \alpha_k^2.$$

よって，チェビシェフの不等式より，任意の $\varepsilon > 0$ に対して，不等式

$$P(|T_n - \mu| > \varepsilon) \le \frac{1}{\varepsilon^2} V(T_n) = \frac{\sigma^2}{\varepsilon^2} \sum_{k=1}^n \alpha_k^2$$

が成り立つ．たとえば，$\alpha_k = 1/n$ $(1 \le k \le n)$ の場合は

$$\sum_{k=1}^n \alpha_k^2 = n \left(\frac{1}{n} \right)^2 = \frac{1}{n} \to 0 \quad (n \to \infty)$$

であるため，$\lim_{n \to \infty} P(|\overline{X}_n - \mu| > \varepsilon) = 0$ がわかる．したがって，$\overline{X}_n$ は μ の一致推定量である．

「最尤法」導入の背景

ここまでは，「良い推定量」がみたすべき性質として，不偏性や一致性について解説した．推定量の一致性とは，標本サイズを大きくするにつれ (期待値ではなく) 推定量自身が母数に近づくという性質をいう．そのため，一致性は「良い推定量」が当然みたすべき性質と言える．一方で，ある不偏推定量に「期待値 0 の任意の確率変数」を加えた新たな推定量は無限に存在するが，これらはすべて不偏推定量である．そのため，すべての不偏推定量が「良い推定量」とは限らない．また，後程説明する「最も良い推定量の 1 つと考えられている最尤推定量」は，一般に不偏性をみたすとは限らない．このように，不偏性は「良い推定量」がみたすべき性質ではあるものの，必須の要件ではない．さらに，不偏推定量が存在しない場合や，存在してもその中で最も有効な不偏推定量を求めることが困難な場合もある．そこで以下では，別の推定法として，最も良い統計的推定法の 1 つと考えられている最尤推定法について説明する．

定義 6.2.3 (最尤法)

$X_1, X_2, \cdots, X_n$ は θ を母数とする母集団分布 D_θ からの大きさ n の無作為標本とし，$X_1, X_2, \cdots, X_n$ の実現値 $x_1, x_2, \cdots, x_n$ が与えられているとする．このとき，尤度関数とよばれる母数 θ の関数 $L(\theta)$ を以下で定義する．まず，母集団分布 D_θ が密度関数 $f_\theta(x)$ から定まる分布の場合は，$L(\theta)$ を次式で定義する．

$$L(\theta) = \prod_{k=1}^{n} f_\theta(x_k). \qquad (6.29)$$

次に，母集団分布 D_θ が離散分布の場合は，$L(\theta)$ を次式で定義する．

$$L(\theta) = \prod_{k=1}^{n} P(X_k = x_k). \qquad (6.30)$$

定義 6.2.3 (最尤法)

このとき，関係式

$$L(\widehat{\theta}_n) = \max_{\theta} L(\theta) \qquad (6.31)$$

をみたし，尤度関数を最大にする $\widehat{\theta}_n$ を θ の推定値とする方法を最尤法とよび，この $\widehat{\theta}_n$ を θ の最尤推定値とよぶ.

$\widehat{\theta}_n$ は実現値 $x_1, x_2, \cdots, x_n$ の関数であるから，n 変数関数 $\widehat{\theta}(x_1, x_2, \cdots, x_n)$ を用いて

$$\widehat{\theta}_n = \widehat{\theta}(x_1, x_2, \cdots, x_n)$$

と表せる．この最尤推定値 $\widehat{\theta}(x_1, x_2, \cdots, x_n)$ の実現値 $x_1, x_2, \cdots, x_n$ を確率変数 $X_1, X_2, \cdots, X_n$ に置き換えた確率変数

$$\overline{\theta}_n = \widehat{\theta}(X_1, X_2, \cdots, X_n)$$

を θ の最尤推定量とよぶ.

注意 6.2.2

最尤法の考え方と「最尤推定量が良い推定量である」ことを説明する．定義 6.2.3 において，標本変量 $X_1, X_2, \cdots, X_n$ の独立性より，(6.29) は n 変量確率変数 $(X_1, X_2, \cdots, X_n)$ の $(x_1, x_2, \cdots, x_n)$ における同時密度関数の値であり，(6.30) は確率 $P(X_1 = x_1, \cdots, X_n = x_n)$ の値である．そのため最尤法とは，標本データ $x_1, x_2, \cdots, x_n$ が得られたときに，「この標本データにおける同時密度関数の値 (6.29)」，または「この標本データが得られる確率 (6.30)」を最大にするように母数 θ を推定する方法である．確率的には起こりにくい標本データが得られることもあるため，この方針で得られた最尤推定量が上手く機能するかは自明ではない．しかし，最尤推定量は漸近正規性とよばれる性質を持ち，この漸近正規性が「最尤推定量が良い推定量である」ことを保証する．

注意 6.2.3

定義 6.2.3 において，(6.31) の最大値を求めることは，対数尤度関数 $l(\theta) = \log L(\theta)$ の最大値を求めることと同値である．最尤推定値を計算するときは $L(\theta)$ と $l(\theta)$ のどちらを用いてもよいが，対数尤度関数 $l(\theta)$ を用いた方が，最尤推定値を求めるための計算が簡単であることが多い．

例 6.2.3

$X_1, X_2, \cdots, X_n$ を二項母集団 $Be(p)$ からの大きさ n の無作為標本とし，母比率 p の最尤推定量を計算する．まず，$x_1, x_2, \cdots, x_n$ を $X_1, X_2, \cdots, X_n$ の実現値とすると，これらの値は 0 または 1 である．以下では，$x = \sum_{k=1}^{n} x_k$ とおく．このとき，尤度関数 $L(p)$ と対数尤度関数 $l(p) = \log L(p)$ は

$$L(p) = \prod_{k=1}^{n} P(X_k = x_k) = \prod_{k=1}^{n} p^{x_k}(1-p)^{1-x_k} = p^x(1-p)^{n-x},$$
$$l(p) = \log L(p) = x \log p + (n-x)\log(1-p)$$

と計算できる．さらに，対数尤度関数の微分は

$$\frac{d}{dp} l(p) = \frac{x}{p} - \frac{n-x}{1-p} = \frac{x-np}{p(1-p)} \qquad (0 < p < 1)$$

と計算できる．

例 6.2.3

したがって，$L(p)$ の最大値は $\widehat{p}_n = x/n$ で達成され，

$$\widehat{p}_n = \frac{1}{n}(x_1 + x_2 + \cdots + x_n), \quad \overline{p}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \overline{X}_n$$

である．よって，母比率 p の最尤推定量は標本平均 $\overline{X}_n$ である．

したがって，たとえば表が出る確率が p である硬貨を 100 回投げて 70 回表が出たとすると，$\widehat{p}_{100} = 70/100 = 0.7$ が最尤推定値である．

もちろん $p = 0.1$ の場合でも 70 回表が出る可能性はあるが，その確率は $L(0.1) = (0.1)^{70}(0.9)^{30}$ であり，この確率は $p = 0.7$ の場合に 70 回表が出る確率 $L(0.7) = (0.7)^{70}(0.3)^{30}$ よりはるかに小さい．

そのため，母比率 p を $\widehat{p}_{100} = 0.1$ と推測するより $\widehat{p}_{100} = 0.7$ と推測するほうが尤もらしい，と考えるのが最尤法の考え方である．

例題 6.2.3

例題 6.2.3

$X_1, X_2, \cdots, X_n$ を $N(\mu, \sigma^2)$ 母集団からの大きさ n の無作為標本とするとき，母数 $\theta = (\mu, \sigma^2)$ の最尤推定量を求めよ.

[解答]　$x_1, x_2, \cdots, x_n$ を $X_1, X_2, \cdots, X_n$ の実現値とすると，これらの値は実数であり，尤度関数 $L(\mu, \sigma^2)$ と対数尤度関数 $l(\mu, \sigma^2) = \log L(\mu, \sigma^2)$ は

$$L(\mu, \sigma^2) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_k - \mu)^2}{2\sigma^2}\right\},$$

$$l(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k - \mu)^2$$

と計算できる.

例題 6.2.3

[解答 (続き)]　$l(\mu, \sigma^2)$ を μ と σ^2 の 2 変数関数と考えて偏微分すると，

$$\frac{\partial}{\partial \mu} l(\mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{k=1}^{n} (x_k - \mu),$$

$$\frac{\partial}{\partial \sigma^2} l(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{k=1}^{n} (x_k - \mu)^2$$

が成り立つ．条件 $\frac{\partial}{\partial \mu} l(\mu, \sigma^2) = \frac{\partial}{\partial \sigma^2} l(\mu, \sigma^2) = 0$ をみたす μ と σ^2 の値のとき，$l(\mu, \sigma^2)$ は最大となる．したがって，母数 $\theta = (\mu, \sigma^2)$ の最尤推定値 $\widehat{\theta}_n = (\widehat{\mu}_n, \widehat{(\sigma^2)}_n)$ は

$$\widehat{\mu}_n = \frac{1}{n} \sum_{k=1}^{n} x_k (= \overline{x}_n), \quad \widehat{(\sigma^2)}_n = \frac{1}{n} \sum_{k=1}^{n} (x_k - \overline{x}_n)^2$$

である．よって，$\theta = (\mu, \sigma^2)$ の最尤推定量 $\overline{\theta}_n = (\overline{\mu}_n, \overline{(\sigma^2)}_n)$ は

$$\overline{\mu}_n = \frac{1}{n} \sum_{k=1}^{n} X_k = \overline{X}_n, \quad \overline{(\sigma^2)}_n = \frac{1}{n} \sum_{k=1}^{n} (X_k - \overline{X}_n)^2.$$

「区間推定」導入の背景

母数 θ は実数値とする．大きさ n の標本調査で得られた標本データ $x_1, x_2, \cdots, x_n$ に対し，「この標本データを用いた推定値 $\widehat{\theta}_n = \widehat{\theta}(x_1, x_2, \cdots, x_n)$ が母数 θ である」と的確に推測するのが点推定の考え方であった．しかし，$\widehat{\theta}_n$ は統計的に θ に近い値を取ると考えられるものの，$\widehat{\theta}_n$ と θ が一致することはほとんど起こり得ない．そこで，「(確率的に評価した) 一定の幅を持つ区間を作り，母数 θ はその区間の中にある」と幅を持たせて推測するのが区間推定の考え方である．

信頼度，信頼区間

D_θ 母集団からの大きさ n の標本変量を $X_1, X_2, \cdots, X_n$ と表す．また，定数 α は $0 < \alpha < 1$ をみたすとし，この α は小さい値であることを想定する．ここで，2 つの統計量

$$S_n = S(X_1, X_2, \cdots, X_n), \quad T_n = T(X_1, X_2, \cdots, X_n)$$

が大小関係 $S_n \leq T_n$ と，関係式

$$P(S(X_1, X_2, \cdots, X_n) \leq \theta \leq T(X_1, X_2, \cdots, X_n)) = 1 - \alpha \quad (6.32)$$

をみたすとき，S_n, T_n を両端とする (無作為な) 区間

$$[S(X_1, X_2, \cdots, X_n), T(X_1, X_2, \cdots, X_n)] \qquad (6.33)$$

を信頼度 $1 - \alpha$ の θ の信頼区間，または θ の $100(1 - \alpha)\%$ 信頼区間とよぶ．

信頼度，信頼区間

このとき，$X_1, X_2, \cdots, X_n$ の実現値 $x_1, x_2, \cdots, x_n$ が得られるごとに，2 つの統計量の実現値 $S(x_1, x_2, \cdots, x_n)$, $T(x_1, x_2, \cdots, x_n)$ と，(6.33) に対応した 1 つの信頼区間

$$[S(x_1, x_2, \cdots, x_n), T(x_1, x_2, \cdots, x_n)] \qquad (6.34)$$

が確定する．たとえば，大きさ n の観測を 1000 回実施した場合，信頼度 95% ($\alpha = 0.05$) の信頼区間 (6.34) は，1000 回のうち 950 回程度は母数 θ を含むことが期待される．

区間推定では，信頼係数 $1 - \alpha$ の値を大きくする (α の値を小さくする) と，「信頼区間の幅

$$T(X_1, X_2, \cdots, X_n) - S(X_1, X_2, \cdots, X_n)$$

が広くなり鋭い推定ができなくなる」という相反の関係がある．

分位点

区間推定の考え方に基づき，いくつかの具体的な母集団分布の場合に，母数の信頼区間の構成方法を紹介する．そのための準備として，分布の分位点の記号を定義する．定数 α は $0 < \alpha < 1$ をみたし，X は連続型確率変数とする．このとき，関係式

$$P(X \geq u) = \alpha$$

をみたす実数 u の値を，この確率変数 X が従う分布の「上側 α 分位点」とよぶ．また，上側 $1-\alpha$ 分位点を「下側 α 分位点」とよぶ．これらの分位点は，X の分布ごとに次表の記号で表す．

分布	確率変数 X	上側 α 分位点	下側 α 分位点
標準正規分布	$X \sim N(0,1)$	z_α	$z_{1-\alpha} = -z_\alpha$
t-分布	$X \sim t(n)$	$t_\alpha^{(n)}$	$t_{1-\alpha}^{(n)} = -t_\alpha^{(n)}$
カイ二乗分布	$X \sim \chi^2(n)$	$c_\alpha^{(n)}$	$c_{1-\alpha}^{(n)}$

図 6.1　標準正規分布や t-分布の密度関数の概形.

標準正規分布や t-分布の密度関数は，原点に関して左右対称である．そのため，$0 < \alpha < 1/2$ をみたす α に対して，分位点に関して次が成立.

$$z_{1-\alpha} = -z_\alpha, \quad t^{(n)}_{1-\alpha} = -t^{(n)}_\alpha. \qquad (6.35)$$

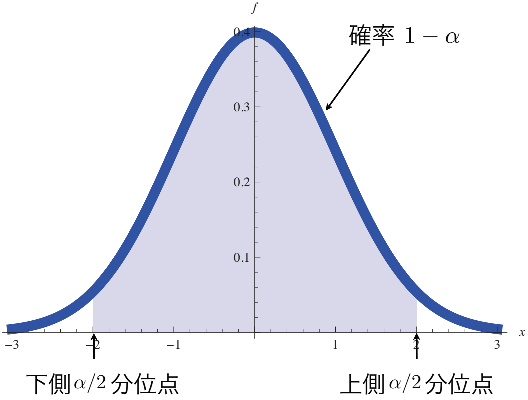
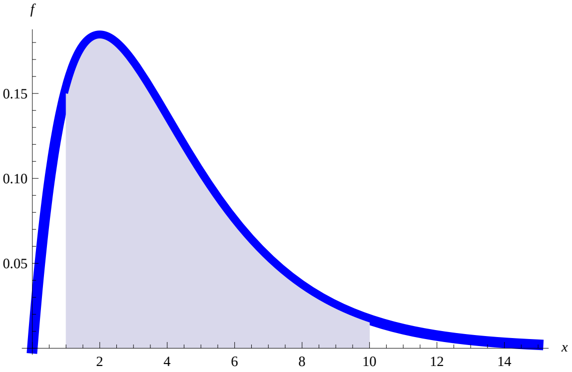


図 6.2 カイ二乗分布の密度関数の概形.

カイ二乗分布の密度関数は，原点に関して左右対称ではないため，分位点 $c_{\alpha}^{(n)}$ については (6.35) に対応する関係式は成立しない.



系 6.3.1

正規母集団 $N(\mu, \sigma^2)$ における母平均 μ や母分散 σ^2 の区間推定を行う場合に，次の系 6.3.1 が有用である．

系 6.3.1

$X_1, X_2, \cdots, X_n$ を $N(\mu, \sigma^2)$ 母集団からの大きさ n の無作為標本とする．このとき，3 つの統計量 $\overline{X}_n$, $\widehat{s}_n^2$, U_n^2 と，$0 < \alpha < 1$ に対して次が成り立つ．

$$1 - \alpha = P\left(\overline{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right), \tag{6.36}$$

$$1 - \alpha = P\left(\overline{X}_n - t_{\alpha/2}^{(n-1)}\frac{\sqrt{U_n^2}}{\sqrt{n}} \leq \mu \leq \overline{X}_n + t_{\alpha/2}^{(n-1)}\frac{\sqrt{U_n^2}}{\sqrt{n}}\right), \tag{6.37}$$

$$1 - \alpha = P\left(\frac{n\widehat{s}_n^2}{c_{\alpha/2}^{(n)}} \leq \sigma^2 \leq \frac{n\widehat{s}_n^2}{c_{1-\alpha/2}^{(n)}}\right), \tag{6.38}$$

$$1 - \alpha = P\left(\frac{(n-1)U_n^2}{c_{\alpha/2}^{(n-1)}} \leq \sigma^2 \leq \frac{(n-1)U_n^2}{c_{1-\alpha/2}^{(n-1)}}\right). \tag{6.39}$$

系 6.3.1

[証明]　定理 6.1.1 より，統計量の分布に関して次が成り立つ．

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \sim N(0,1), \quad \frac{n\widehat{s}_n^2}{\sigma^2} \sim \chi^2(n).$$

これらの標本分布の上側と下側の $\alpha/2$ 分位点を用いた 2 つの確率

$$P\left(-z_{\alpha/2} \leq \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \leq z_{\alpha/2}\right), \quad P\left(c_{1-\alpha/2}^{(n)} \leq \frac{n\widehat{s}_n^2}{\sigma^2} \leq c_{\alpha/2}^{(n)}\right)$$

の値はいずれも $1-\alpha$ である．一方で，事象の同値な表現として

$$\left\{-z_{\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \leq z_{\frac{\alpha}{2}}\right\} = \left\{\overline{X}_n - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right\},$$

$$\left\{c_{1-\alpha/2}^{(n)} \leq \frac{n\widehat{s}_n^2}{\sigma^2} \leq c_{\alpha/2}^{(n)}\right\} = \left\{\frac{n\widehat{s}_n^2}{c_{\alpha/2}^{(n)}} \leq \sigma^2 \leq \frac{n\widehat{s}_n^2}{c_{1-\alpha/2}^{(n)}}\right\}$$

が成り立つため，(6.36) と (6.38) が得られる．同様にして，(6.37) と (6.39) も得られる．

例 6.3.1

$N(\mu, \sigma^2)$ 母集団からの大きさ 20 の無作為標本を $X_1, X_2, \cdots, X_{20}$ とする．

- 標本平均 $\overline{X}_{20}$ の実現値が 100 で $\sigma^2 = 10^2$ のとき，(6.36) と $z_{0.05} = 1.645$ より，母平均 μ の 90% 信頼区間は

$$\left[100 - \frac{10}{\sqrt{20}} z_{0.05},\ 100 + \frac{10}{\sqrt{20}} z_{0.05}\right] = [96.32,\ 103.68].$$

- 標本平均 $\overline{X}_{20}$ の実現値が 100 で $U_{10}^2 = 10^2$ のとき，(6.37) と $t_{0.05}^{(19)} = 1.729$ より，母平均 μ の 90% 信頼区間は

$$\left[100 - \frac{10}{\sqrt{20}} t_{0.05}^{(19)},\ 100 + \frac{10}{\sqrt{20}} t_{0.05}^{(19)}\right] = [96.13,\ 103.87].$$

例 6.3.1

$N(\mu, \sigma^2)$ 母集団からの大きさ 20 の無作為標本を $X_1, X_2, \cdots, X_{20}$ とする．

- $\widehat{s}_{20}^2 = \frac{1}{20}\sum_{k=1}^{20}(X_k - \mu)^2$ の実現値が 100 のとき，(6.38) と $c_{0.95}^{(20)} = 10.85$ と $c_{0.05}^{(20)} = 31.41$ より，母分散 σ^2 の 90% 信頼区間は

$$\left[\frac{20 \times 100}{c_{0.05}^{(20)}}, \ \frac{20 \times 100}{c_{0.95}^{(20)}}\right] = [63.67, \ 184.33].$$

- 不偏標本分散 U_{20}^2 の実現値が 100 のとき，(6.39) と $c_{0.95}^{(19)} = 10.12$ と $c_{0.05}^{(19)} = 30.14$ より，σ^2 の 90% 信頼区間は

$$\left[\frac{19 \times 100}{c_{0.05}^{(19)}}, \ \frac{19 \times 100}{c_{0.95}^{(19)}}\right] = [63.04, \ 187.75].$$