数理情報学９ (数理統計学)　**2. 統計的推定**

1　統計的推定問題

確率論と統計学では問題意識が以下のように異なっている．

確率：　ある確率法則のもとで，確率変数がどのような振舞いをするのかを考察する．

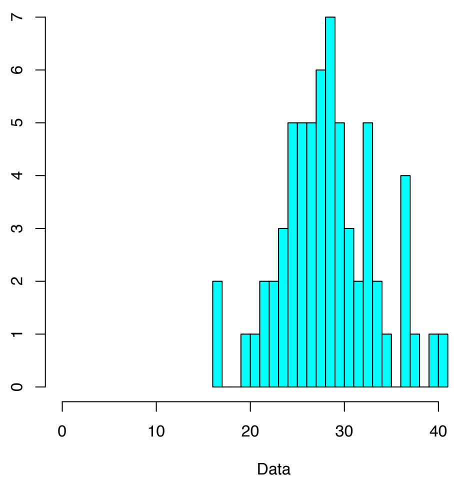
統計：　ある確率法則にしたがうと考えられる確率変数の実現値 (データ) を観測して，その元となる確率法則についての推論をおこなう．

とくに統計的推定の目的は，データの背後にある確率構造を特徴付ける量 (平均や分散など) を精度良く推測することである．

データ：ある確率分布にしたがう確率変数を観測した結果得られる値のこと．実現値とも言う．

例 **1.** 物理実験などの測定を考える．通常，物理的測定には測定誤差が生じる．このような誤差は実験をおこなうときの環境によって変動するものであり，完全にコントロールすることはできない．光の速度を測定したデータを図**??**に示す．

The classical data of Michaelson and Morley on the speed of light. The data consists of five experiments, each consisting of 20 consecutive 'runs'.
A. J. Weekes (1986) A Genstat Primer. London: Edward Arnold.



適当な物理的条件のもとで光の速度を測定した結果を，ヒストグラムで示している．横軸の単位は 1 万 km/sec．縦軸は度数．

このような観測データから，どのように光の速度を定めればよいだろうか？ 光速度は決まった値であるが，測定誤差があるために測定した結果 X_i は

$$\underset{(\text{測定結果})}{X_i} = \underset{(\text{光速度})}{c} + \underset{(\text{観測誤差，偶然変動})}{Z_i}, \qquad i = 1, \ldots, n \tag{1}$$

のような確率変数として記述されると考えるのが妥当であろう．　□

推定の問題を (1) のように定式化すると，次の事柄が気になってくる：

(1) 観測値 $X_1, \ldots, X_n$ から未知パラメータ (例 1 では光速度 c) を推定するためには，どのような推定量を用いればよいだろうか？
$\longrightarrow$ 不偏推定量や最尤推定量．

(2) いろいろな推定方法を使うとき，それらの推定精度をどのようにして比較すればいいだろうか？
$\longrightarrow$ 平均二乗誤差を比較

(3) これ以上の精度では推定することはできない，という限界はあるのだろうか？
$\longrightarrow$ クラメール・ラオの下限

2 不偏推定量

データ X が確率密度関数 $f(x;\theta)$, $\theta \in \mathbb{R}$ をもつ確率分布にしたがっているとする．推定量 $T(X)$ が

$$\text{任意の } \theta \text{ に対して } E_\theta[T(X)] = \theta$$

を満たすとき $T(X)$ はパラメータ θ の <u>不偏推定量</u> であると言う．ここで $E_\theta[\cdot]$ は $f(x;\theta)$ による期待値

$$E_\theta[T(X)] = \int T(x) f(x;\theta) dx$$

である．不偏推定量は「推定値は真のパラメータのまわりに分布している (偏りがない)」という意味で望ましい性質であると考えることができる．

期待値・分散の不偏推定量

確率変数 $X_1, X_2, \ldots, X_n$ が期待値 μ, 分散 σ^2 の分布にしたがうとする．

- 標本平均 $\displaystyle \frac{1}{n}\sum_{i=1}^{n} X_i$ は 期待値 μ の不偏推定量である．
- 標本分散 $\displaystyle \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ は分散 σ^2 の不偏推定量である．

標本分散はデータ数 n ではなく $n-1$ で割っていることに注意．

3 クラメール・ラオの下限

以下では，パラメータ θ の推定量 $T(X)$ の推定精度を平均二乗誤差

$$E_\theta[|T(X)-\theta|^2]$$

で測る[1]．もし推定量 $T(X)$ が θ の不偏推定量ならば，平均二乗誤差は

$$E_\theta[|T(X)-\theta|^2] = E_\theta[|T(X)-E_\theta[T(X)]|^2] = V_\theta[T(X)], \quad (f(x;\theta) \text{ のもとでの分散})$$

となり，$T(X)$ の分散に一致する．

確率変数 X が確率密度関数 $f(x;\theta)$, $\theta \in \boldsymbol{R}$ をもつ確率分布にしたがっているとき，パラメータ θ を不偏推定量 $T(X)$ で推定する．このとき $T(X)$ の平均二乗誤差の下限を与えるのがクラメール・ラオの定理である．

[1] 平均二乗誤差は推定量の精度を測る尺度のひとつとして非常によく用いられる．

定理：クラメール・ラオの下限

パラメータ θ の不偏推定量 $T(X)$ に対して，以下の不等式

$$E_\theta[|T(X)-\theta|^2] \geq \frac{1}{I(\theta)}$$

が成り立つ．ここで $I(\theta)$ は

$$I(\theta) = E_\theta\left[\left(\frac{\partial \log f(x;\theta)}{\partial\theta}\right)^2\right] = \int \left(\frac{\partial \log f(x;\theta)}{\partial\theta}\right)^2 f(x;\theta)dx$$

で定義され，フィッシャー情報量と呼ばれる

注意：必ずしも等号を達成する不偏推定量 $T(X)$ が存在するとは限らない．存在するかどうかは $f(x;\theta)$ の分布形に依存する．また存在するとしても，そのような T を簡単に構成できるとは限らない．

フィッシャー情報量 $I(\theta)$ は確率密度関数 $f(x;\theta)$ だけから計算できる量であり，推定量 T には依存していない．したがってクラメール・ラオの下限は，どのような不偏推定量 T に対してもその平均二乗誤差は $1/I(\theta)$ より小さくはならないことを主張している．つまり

(a) 不偏推定量には推定精度に限界がある

(b) 推定限界はフィッシャー情報量の逆数で与えられる (ただし達成可能とは限らない)

ことが示された．

クラメール・ラオの定理から不偏推定量 $T(X)$ の推定精度には限界があることが分かる．推定量の精度が推定限界にどのくらい近いかを測るために「推定量の効率」を用いる．

クラメール・ラオの定理より $I(\theta)^{-1} \leq V_\theta[T(X)]$ なので

$$\frac{1}{I(\theta)V_\theta[T(X)]} \leq 1$$

である．そこで不偏推定量 $T(X)$ の効率を

$$\frac{100}{I(\theta)V_\theta[T(X)]}\ (\%)$$

と定義する．ここで $V_\theta[T(X)]$ は $T(X)$ の $f(x;\theta)$ のもとでの分散とする．不偏推定量の効率は 0% から 100% の値をとることが分かる．効率が大きいほど推定精度の良い推定量ということになる．

4　標本分布 (正規分布から導かれる分布)

データが正規分布にしたがうとき，標本平均や標本分散がどのような分布にしたがうかについて解説する．

χ^2-分布 (カイ二乗分布)： $X_1, X_2, \ldots, X_n$ が独立に標準正規分布 $N(0,1)$ にしたがうとする．確率変数 Z_n を

$$Z_n = X_1^2 + X_2^2 + \cdots + X_n^2$$

と定める．Z_n がしたがう分布を自由度 n の χ^2分布 とよぶ．

例 **2** (自由度 1 の χ^2 分布の密度関数)．$X \sim N(0,1)$ のとき $Z_1 = X^2$ の密度関数を計算する．

$$\begin{aligned} P(Z_1 \leq z) &= P(-\sqrt{z} \leq X \leq \sqrt{z}) \\ &= 2\int_0^{\sqrt{z}} \phi(x;0,1)dx \\ &= 2\int_0^{z} \phi(\sqrt{y};0,1)\frac{1}{2\sqrt{y}}dy \quad (x=\sqrt{y} \text{ と変数変換}) \end{aligned}$$

したがって Z_1 の密度関数 $f(z)$ は

$$f(z) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}$$

となる (分布関数の微分は密度関数).

t-分布 (ティー分布)： 確率変数 $X, Y_1, \ldots, Y_n$ はすべて独立に標準正規分布 $N(0,1)$ にしたがうとする．このとき確率変数 T_n を

$$T_n = \sqrt{n} \cdot \frac{X}{\sqrt{Y_1^2 + \cdots + Y_n^2}} = \frac{X}{\sqrt{\frac{Y_1^2 + \cdots + Y_n^2}{n}}}$$

と定める．T_n がしがたう分布のことを $\underline{\text{自由度 } n \text{ の } t\text{-分布}}$ とよぶ.

T_n を χ^2 分布を用いて定義することもできる．すなわち X とは独立に自由度 n の χ^2 分布にしたがう確率変数を Z_n とすると，定義から T_n は

$$T_n = \sqrt{n} \cdot \frac{X}{\sqrt{Z_n}}$$

と書ける．自由度 n が無限大になる極限を考えると

$$T_n \xrightarrow{d} N(0,1), \ (n \to \infty)$$

が成立する (大数の法則，スラツキーの公式).

標本平均と標本分散の分布： 確率変数 $X_1, X_2, \ldots, X_n$ が独立に正規分布にしたがうとする：

$$X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2).$$

このとき

$$\text{標本平均} \quad \bar{X} = \frac{1}{n} \sum_{i=1} X_i$$

$$\text{標本分散} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

の分布は以下のようになる.

(1) $\dfrac{(n-1)S^2}{\sigma^2}$ は自由度 $n-1$ の χ^2 分布にしたがう

(2) $\bar{X}$ と S^2 は独立

(3) (1), (2) より $\sqrt{n}\ \dfrac{\bar{X} - \mu}{\sqrt{S^2}}$ は自由度 $n-1$ の t-分布にしたがう

(証明)
$X_1, \ldots, X_n$ を以下の直交行列を使って $Y_1, \ldots, Y_n$ に変換する：

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ * & \cdots & * \\ & & \\ * & \cdots & * \end{pmatrix} \begin{pmatrix} \frac{X_1 - \mu}{\sigma} \\ \vdots \\ \frac{X_n - \mu}{\sigma} \end{pmatrix}.$$

上の直交行列は第 1 行が $\left(\frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}}\right)$ であり，他の要素については，直交行列になるという制約を満たすかぎり，どのような値でもよいとする．このとき

$$\begin{aligned}\sqrt{n}\frac{\bar{X}-\mu}{\sigma} &= Y_1 \\ \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} &= \sum_{i=1}^{n} Y_i^2\end{aligned}$$

となる．正規分布の変数変換の公式から $(Y_1, \ldots, Y_n)$ は平均ベクトルが $(0, \ldots, 0)$ で分散共分散行列が I_n(ここで I_n は $n \times n$ の大きさの単位行列とする) の n 次元多変量正規分布にしたがう．なぜなら，上の直交行列を Q とおくと，

$$\left(\frac{X_1-\mu}{\sigma}, \ldots, \frac{X_n-\mu}{\sigma}\right) \sim N_n(0, I_n)$$

より

$$(Y_1, \ldots, Y_n) \sim N_n(0, QI_nQ^t) = N(0, I_n)$$

となる．分散共分散行列が対角行列であることから，$Y_1, \ldots, Y_n$ は独立になっている[2]．これより

$$\begin{aligned}Y_1^2+Y_2^2+\cdots Y_n^2 = \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} &= \frac{n(\bar{X}-\mu)^2}{\sigma^2}+\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{\sigma^2} \\ &= Y_1^2+\frac{(n-1)S^2}{\sigma^2}\end{aligned}$$

を得る．したがって

$$Y_2^2+\cdots+Y_n^2 = \frac{(n-1)S^2}{\sigma^2}$$

となるので，$(n-1)S^2/\sigma^2$ は自由度 $n-1$ の χ^2 分布にしたがい，かつ Y_1 とは独立である． □

演習問題

1. データ $X_1, X_2, \ldots, X_n$ は指数分布 $Ex[\lambda^{-1}]$ から独立に得られているとする．指数分布 $Ex[\lambda^{-1}]$ の確率密度関数は

$$f(x;\lambda) = \begin{cases}\dfrac{1}{\lambda}e^{-x/\lambda} & x \geq 0 \\ 0 & x < 0\end{cases}$$

と表せる．パラメータ λ の推定量として以下の 2 つ

$$T_1 = \frac{1}{n}\sum_{i=1}^{n}X_i, \qquad T_2 = n \cdot \min_{1\leq i\leq n} X_i$$

を考える．

(a) T_1 は不偏推定量であることを示せ．また T_1 の平均二乗誤差を計算せよ．

(b) T_2 は不偏推定量であることを示せ．また T_2 の平均二乗誤差を計算せよ．

2. データ $X_1, X_2, \ldots, X_n$ が区間 $(0, \theta)$ 上の一様分布 $U(0, \theta)$ から独立に得られているとする．一様分布の密度関数は

$$f(x;\theta) = \begin{cases}\dfrac{1}{\theta} & x \in (0, \theta) \\ 0 & x \notin (0, \theta)\end{cases}$$

[2] 多次元正規分布では無相関と独立は同値

パラメータ θ の推定量として

$$T_1 = \frac{2}{n}\sum_{i=1}^{n} X_i, \qquad T_2 = \frac{n+1}{n}\max\{X_1, \dots, X_n\}$$

の 2 つを考える．

(a) T_1, T_2 はともに θ の不偏推定量であることを示せ．

(b) T_1, T_2 の平均二乗誤差を求め，推定精度を比較せよ．

3. 成功の確率が p $(0 < p < 1)$ の独立な n 回のベルヌーイ試行を $X_1, X_2, \dots, X_n$ とする．つまり各 X_i は 0 か 1 の値をとりその確率は $P(X_i = 1) = p,\ P(X_i = 0) = 1 - p$ である．

   (a) $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ は p の不偏推定量であることを示せ．また $\bar{X}$ の分散を求めよ．

   (b) $n \geq 2$ とする． $\dfrac{n}{n-1}\bar{X}(1-\bar{X})$ はベルヌーイ試行の分散 $p(1-p)$ の不偏推定量であることを示せ．

4. $X_1, \dots, X_n$ は平均 μ で分散 1 の正規分布 $N(\mu, 1)$ に独立にしたがっているとする．このデータから μ を推定することを考える．ここでは

$$T(X_1, \dots, X_n) = c_1X_1 + c_2X_2 + \cdots + c_nX_n$$

という形の不偏推定量について考える．

   (a) T が μ の不偏推定量であるとき $c_1 + c_2 + \cdots + c_n = 1$ が成り立つことを示せ．

   (b) T が μ の不偏推定量であるときに，平均二乗誤差 $E[|T - \mu|^2]$ を最小にする $c_1, c_2, \dots, c_n$ は $c_1 = c_2 = \cdots = c_n = \frac{1}{n}$ であることを計算せよ．ヒント：シュワルツの不等式またはラグランジュの未定定数法．

以上の考察から $\sum_{i=1}^{n} c_iX_i$ と表される推定量のなかでは $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ が平均二乗誤差の意味で最適であることがわかる．

5. パラメータ $\theta \in R$ を含む確率密度 $f(x;\theta)$ のフィッシャー情報量を $I(\theta)$ とする．このとき $f(x;\theta)$ の分布にしたがう n 個の独立な確率変数 $X_1, \dots, X_n$ の密度関数 g は

$$g(x_1, \dots, x_n; \theta) = f(x_1;\theta) \times f(x_2;\theta) \times \cdots \times f(x_n;\theta)$$

となる．密度関数 $g(x_1, \dots, x_n; \theta)$ のフィッシャー情報量は $nI(\theta)$ になることを示せ．

6. データ $X_1, X_2, \dots, X_n$ は独立に同一のガンマ分布 $G_A(\alpha, \beta)$ にしたがうとする．ガンマ分布の確率密度関数は

$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right), \quad (x > 0)$$

と書ける．ここで $\Gamma(\alpha)$ はガンマ関数である．ガンマ分布の期待値と分散はそれぞれ $E[X_i] = \alpha\beta,\ V[X_i] = \alpha\beta^2$ である．α の値は既知のとき β を推定することを考える．

   (a) β の推定量として

$$\hat{\beta} = \frac{1}{\alpha n}\sum_{i=1}^{n} X_i$$

   を考える．$\hat{\beta}$ は β の不偏推定量であることを示せ．

(b) パラメータ β に関するフィッシャー情報量

$$I(\beta) = E_{\alpha,\beta}\left[\left(\frac{\partial \log f(x;\alpha,\beta)}{\partial \beta}\right)^2\right]$$

を計算せよ．

(c) $\hat{\beta}$ がクラメール・ラオの下限を達成しているかどうか調べよ．

7. $X_1, \ldots, X_n$ が独立に正規分布 $N(\mu, \mu)$ にしたがうとする．ここで $\mu > 0$ とする．パラメータ μ の推定量として $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ を考える．$\bar{X}$ がクラメール・ラオの下限を達成しているか調べよ．