

点推定の基礎 (最尤法から AIC まで)

erutuf

2015 年 12 月 30 日

1 準備

まず、確率論や統計学の基礎事項について準備を行う。

1.1 確率論

この記事では測度論に基いて数学的に確率の概念を導入する。

定義 1.1. Ω を集合、 \mathcal{B} を Ω の部分集合族とする。 \mathcal{B} が完全加法族であるとは以下の条件をみたすときに言う：

1. $\emptyset \in \mathcal{B}$
2. $A_1, A_2 \in \mathcal{B}$ のとき $A_1 \cap A_2 \in \mathcal{B}$
3. $A \in \mathcal{B}$ のとき $A^c (= \Omega \setminus A) \in \mathcal{B}$
4. $A_1, A_2, \dots \in \mathcal{B}$ のとき $\bigcup_{k=1}^{\infty} A_k \in \mathcal{B}$.

また、集合とその完全加法族の組 (Ω, \mathcal{B}) を可測空間と呼ぶ。

定義 1.2. (Ω, \mathcal{B}) を可測空間とする。以下の条件をみたす \mathcal{B} 上の関数 $P : \mathcal{B} \rightarrow [0, 1]$ を確率と呼ぶ：

1. $P(\Omega) = 1$
2. $B_1, B_2, \dots \in \mathcal{B}$ が $B_i \cap B_j = \emptyset$ ($i \neq j$) をみたすとき $P(\bigcup_{k=1}^{\infty} B_k) = \sum_{k=1}^{\infty} P(B_k)$.

可測空間と確率の組 (Ω, \mathcal{B}, P) を確率空間と呼ぶ。

可測空間の基本的な例として、 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ がある。ここで、 $\mathcal{B}(S)$ は 位相空間 S の任意の開集合を含む最小の完全加法族であり、Borel 集合族と呼ばれる。

定義 1.3. (Ω, \mathcal{B}) を可測空間とする。関数 $X : \Omega \rightarrow \mathbb{R}$ が (Ω, \mathcal{B}) 上の可測関数であるとは、任意の $a \in \mathbb{R}$ に対して $X^{-1}(\{\omega \in \Omega \mid X(\omega) \leq a\}) \in \mathcal{B}$ が成り立つときに言う。また、確率空間上の可測関数のことを確率変数と呼ぶ。

慣習的に、集合 $\{\omega \in \Omega \mid X(\omega) \leq a\}$ の確率 $P(\{\omega \in \Omega \mid X(\omega) \leq a\})$ を $P(X \leq a)$ と略記する。 $P(X < a)$ 、 $P(X \geq a)$ 、 $P(X > a)$ などについても同様である。また、可測関数 f と確率変数 X の合成 $f \circ X$ のこ

とをしばしば $f(X)$ と書く。

定義 1.4. 確率変数 $X : \Omega \rightarrow \mathbb{R}$ に対し、その累積分布関数 (culminative distribution function, cdf) $F_X : \mathbb{R} \rightarrow \mathbb{R}$ を以下で定義する：

$$F_X(a) = P(X \leq a)$$

また、cdf F_X が関数 $p_X : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ を用いて

$$F_X(a) = \int_{-\infty}^a p_X(x) dx$$

と書けるとき p_X を確率密度関数 (probability density function, pdf) と呼ぶ。

確率変数 X に対しその期待値 $E[X]$ を

$$E[X] = \int_{-\infty}^{\infty} x p_X(x) dx$$

で定める。また、可測関数 $f : \mathbb{R} \rightarrow \mathbb{R}$ に対して $f(X)$ の期待値を

$$E[f(X)] = \int_{-\infty}^{\infty} f(x) p_X(x) dx$$

で定める。

(TODO)

1.2 統計量

可測関数 $T : \mathbb{R}^n \rightarrow \mathbb{R}$ 、無作為標本 $X = (X_1, \dots, X_n)$ に対し、 $T(X)$ を統計量と呼ぶ。統計量の例として、標本平均

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

がある。

確率ベクトル $\mathbf{X} = (X_1, \dots, X_n)$ の jpdf を $p(x, \theta) = f^\theta(x)$ とする。統計量 $T(\mathbf{X})$ が θ に関して情報損失を起こさないという概念を以下のように定式化する：

定義 1.5. 統計量 $T(\mathbf{X})$ 、実現値 t に対し \mathbf{X} の cpdf $f^\theta(x | t)$ が θ に依存しないとき T は θ に対する十分統計量であると言う。

2 情報量、最尤推定

3 バイアス、一様最小分散

4 赤池情報量基準