応用統計学 2017 第7回 最尤推定

2017年11月22日 (水)
清 智也 sei@mist.i.u-tokyo.ac.jp
http://ur0.pw/yTzt

今回は，多くの統計モデルに対して汎用的に用いられる最尤推定量を扱う。

- 尤度関数，最尤推定，尤度方程式，指数型分布族，十分統計量，凸関数[1]。

演習問題

問題 7-1. 以下の各統計モデルに対する最尤推定量を求めよ。すなわち，与えられたデータ $x_1, \ldots, x_n$ に対して尤度関数 $L(\theta) = \prod_{t=1}^n f(x_t; \theta)$ を最大化する θ を求めよ。また，その不偏性を調べよ。

(i) ポアソン分布 $f(x;\theta) = (\theta^x/x!)e^{-\theta}$, $x \in \mathbb{Z}_{\geq 0} = \{0, 1, \cdots\}$.

(ii) 正規分布 $f(x;\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2}e^{-(x-\mu)^2/(2\sigma^2)}$, $x \in \mathbb{R}$, $\theta = (\mu, \sigma^2)$.

(iii) 指数分布 $f(x;\theta) = \theta e^{-\theta x}$, $x > 0$.

問題 7-2. 以下の各統計モデルは指数型分布族であることを示せ。

(i) Bernoulli 分布（サイズ 1 の二項分布）$f(x;p) = p^x(1-p)^{1-x}$, $x \in \{0, 1\}$.

(ii) 負の二項分布 $f(x;p) = \binom{r+x-1}{x}p^r(1-p)^x$, $x \in \mathbb{Z}_{\geq 0}$. ただし r は既知とする。

(iii) カテゴリー数 k，サイズ 1 の多項分布 $f(\boldsymbol{x};\boldsymbol{p}) = \prod_{i=1}^k p_i^{x_i}$, $x_i \in \{0, 1\}$, $\sum_{i=1}^k x_k = 1$.

問題 7-3 (順序統計量)**.** $X_1, \ldots, X_n$ は独立同一分布に従い，それぞれの確率密度関数は $f(x;\theta) = \theta e^{-\theta x}$ $(x \geq 0)$ であるとする。また，$X_1, \ldots, X_n$ を小さい順に並べ替えて得られる確率変数列を $X_{(1)}, \ldots, X_{(n)}$ とおく[2]。このようにして得られる $X_{(k)}$ $(1 \leq k \leq n)$ は一般に順序統計量 (order statistic) と呼ばれる。以下の問いに答えよ。

(i) $X_{(1)}, \ldots, X_{(n)}$ の同時密度関数を求めよ。

(ii) $1 < k < n$ とする。$X_{(1)}, \ldots, X_{(k)}$ のみが観測された場合の θ の最尤推定量を求めよ。

---

[1] likelihood function, maximum likelihood estimator (MLE), likelihood equation, exponential family, sufficient statistic, convex function.

[2] 正確には，確率空間 $(\Omega, \mathcal{F}, P)$ で定義された確率変数列 $X_1, \ldots, X_n$ に対し，根元事象 $\omega \in \Omega$ ごとに $X_1(\omega), \ldots, X_n(\omega)$ を小さい順に並べ替えたものを $X_{(1)}(\omega), \ldots, X_{(n)}(\omega)$ とおく。これによって確率変数列 $X_{(1)}, \ldots, X_{(n)}$ が定義される。

問題 7-4 (やや難しい). 初回の宿題で実施したアンケートの回答時刻のデータを，生存時間解析と呼ばれる考え方でモデリングしてみよう。受講者は全部で J 人おり，J は既知とする（$J = 122$)。各受講者の回答時刻は独立同一分布に従うと仮定する。また，ある特定の受講者一人について，時刻 t まで回答しなかったという条件のもとで，時刻 t から $t + \mathrm{d}t$ の間に回答する確率を $\lambda(t)\mathrm{d}t$ とおく。$\lambda(t)$ はハザード関数 (hazard function) と呼ばれる。回答時刻の累積分布関数を $F(t)$，密度関数を $f(t) = F'(t)$ とおくとき，

$$\lambda(t) = \frac{f(t)}{1 - F(t)}$$

の関係がある（なぜか？）。また $1 - F(t)$ は生存関数 (survival function) と呼ばれる。

さて，アンケートの受付開始時刻を 0，受付打ち切り時刻を c で固定し，実際の回答者数を n，回答のあった時刻を $t_1 < \cdots < t_n$ とおく。ハザード関数を

$$\lambda(t) = \frac{\phi}{\eta}\left(\frac{t}{\eta}\right)^{\phi-1}, \quad \eta > 0, \quad \phi > 0, \tag{$*$}$$

と仮定したときの尤度関数を求めよ。また，授業で集めたアンケートの回答時刻に対し，(η, ϕ) の最尤推定値を計算機で求めよ。なお，式 $(*)$ のハザード関数に対応する確率分布を Weibull 分布と呼ぶ。

宿題 7

問題 7-5. 指数型分布族

$$f(x; \boldsymbol{\theta}) = a(x)e^{\boldsymbol{\theta}'\boldsymbol{s}(x) - \psi(\boldsymbol{\theta})}, \quad x \in \mathbb{R}, \quad \boldsymbol{\theta} \in \mathbb{R}^d,$$

において，$\psi(\boldsymbol{\theta})$ は $\boldsymbol{\theta}$ の凸関数であることを示せ。また，狭義凸となるための条件を調べよ。