Ch.6 統計的推定

§6.1 統計的推測

6.1.1 統計的推測の考え方

・ ~~母~~集団：分布 $f(x|\theta)$ に従う.

$X_1, \cdots, X_n \overset{i.i.d.}{\sim} f(x|\theta)$：サイズ n の標本.

$\downarrow \quad \downarrow$

$x_1, \cdots, x_n$：実現値.

・ θ：パラメータ

・統計的点推定：θ を $X_1 \sim X_n$ の値に基づいて予測すること.

・区間推定：θ の入る可能性の高い区間を求めること.

・統計的仮説検定：仮説の正しさを判断すること.

6.1.2 データの縮約

$X_1, \cdots, X_n \overset{i.i.d.}{\sim} f(x|\theta)$.

$\mathbb{X} = (X_1, \cdots, X_n)$, $\mathbb{x} = (x_1, \cdots, x_n)$ とおく.

Def. (十分統計量)

$\mathbb{X}$ を $T(\mathbb{X})$ に縮約しても、θ についての情報は失われない.

$T(\mathbb{X})$：θ に関して十分統計量

$\overset{def.}{\iff}$ $T(\mathbb{x}) = t$ をみたす $\mathbb{x}, t$ に対し、$T(\mathbb{X}) = t$ を与えたときの

$\mathbb{X} = \mathbb{x}$ の条件付き確率 $P(\mathbb{X} = \mathbb{x} \mid T(\mathbb{X}) = t)$ が θ に依らない. $\square$

Th. (Neymanの因子分解定理)

$T(X)$: θの十分統計量

$\Leftrightarrow$ $f_X(x|\theta) = h(x)g(T(x)|\theta)$ とかける. □

Def. (指数型分布族)

pdfの族が

$$f(x|\theta) = h(x)c(\theta)\exp(w(\theta)^T t(x))$$

の形になるとき, 指数型分布族 という □

ex) 正規分布, ガンマ分布, 2項分布, Poisson分布, 負の2項分布

は指数型分布族に入る.

§6.2 点推定量の導出方法

- パラメトリック・モデル $f(x|\theta)$　$\theta \in \mathbb{R}^k$.

θ を $X_1, \cdots, X_n \overset{i.i.d.}{\sim} f(x|\theta)$ から点推定.

θ の**推定量**を $\hat{\theta}(\mathbb{X})$ とかく.

6.2.1 モーメント法

- $X \sim f(x|\theta)$,

$\mathbb{E}[X] = \mu_1'(\theta), \mathbb{E}[X^2] = \mu_2'(\theta), \cdots, \mathbb{E}[X^k] = \mu_k'(\theta)$ とする.

- $X_1, \cdots, X_n \overset{i.i.d.}{\sim} f(x|\theta)$ について,

大数の法則より $\frac{1}{n}\sum_{i=1}^{n} X_i^r \xrightarrow{p} \mathbb{E}[X^r]$　$(r = 1, \cdots, k)$

$$\to \begin{cases} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu_1'(\theta) \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mu_2'(\theta) \\ \quad\vdots \\ \frac{1}{n}\sum_{i=1}^{n} X_i^k = \mu_k'(\theta) \end{cases}$$

を θ について解くことで $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_k)$ を得る.

これを**モーメント推定量**という.

ex $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$.

$\mathbb{E}[X_1] = \mu$, $\mathbb{E}[X_1^2] = \mathbb{V}[X_1] + \mathbb{E}[X_1]^2 = \sigma^2 + \mu^2$.

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} X_i = \hat{\mu} \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \hat{\sigma}^2 + \hat{\mu}^2 \end{cases}$$

モーメント推定量は.

$\hat{\mu} = \bar{X}$,

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \hat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S^2. \quad \square$$

6.2.2 最尤法

Def. (尤度函数)

$L(\theta|\mathbb{X}) := \prod_{i=1}^{n} f(X_i|\theta)$ を尤度函数 といい,

$\ell(\theta|\mathbb{X}) := \log L(\theta|\mathbb{X}) = \sum_{i=1}^{n} \log f(X_i|\theta)$ を

対数尤度函数 という. $\square$

Def. (最尤推定量)

$\hat{\theta}$: 最尤推定量

$\overset{\text{def.}}{\Longleftrightarrow} \quad \hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, L(\theta|\mathbb{X}) = \underset{\theta}{\operatorname{argmax}}\, \ell(\theta|\mathbb{X})$. $\square$

・MLEの候補：尤度方程式 $\frac{\partial}{\partial\theta}L(\theta|\mathbb{X})=0$ の解.

<u>Prop.</u>（MLEの不変性）

$\tau=g(\theta)$ の推定を考えるとき,

$\hat{\theta}$：θ の MLE $\Rightarrow$ $\hat{\tau}=g(\hat{\theta})$：$\tau$ の MLE. □

6.2.3 Bayes法.

・$f(x|\theta)$ で, θ：r.v. とみなして. pdf を仮定.

θ の事前分布：$\pi(\theta|\xi)$　（ξ：ハイパーパラメータ）

・モデル. $\begin{cases} \mathbb{X}|\theta \sim f(x|\theta) \\ \theta \sim \pi(\theta|\xi) \end{cases}$

・$\mathbb{X}=x$：given のとき. θ の事後分布は

$$\pi(\theta|x,\xi)=\frac{f(x|\theta)\pi(\theta|\xi)}{\int f(x|\theta)\pi(\theta|\xi)d\theta}$$ ← $\mathbb{X}$ の周辺分布

・$\mathbb{E}[\theta|\mathbb{X}]$：Bayes推定量

$\pi(\theta|x,\xi)$ のモード

$\hat{\theta}:=\underset{\theta}{\operatorname{argmax}}\,\pi(\theta|x,\xi)$：Bayes的最尤推定量.

・$\pi(\theta|\xi)$ と $\pi(\theta|x,\xi)$ が同じ分布族に入るような事前分布を

共役事前分布という.

§6.3 推定量の評価.

6.3.1. 不偏性

Def. (不偏推定量)

推定量 $\hat{\theta}$ が θ の不偏推定量

$\overset{\text{def.}}{\iff} \forall \theta, \quad \mathbb{E}_\theta[\hat{\theta}(\mathbb{X})] = \theta.$ $\square$

θ に依存している

Def. (バイアス)

$\hat{\theta}$: θ の推定量. $\hat{\theta}$ のバイアスは,

$$\mathrm{Bias}(\hat{\theta}) := \mathbb{E}_\theta[\hat{\theta}(\mathbb{X})] - \theta$$

で定められる. $\square$

Def. (平均2乗誤差)

$\mathrm{MSE}(\theta, \hat{\theta}) := \mathbb{E}[(\hat{\theta}(\mathbb{X}) - \theta)^2]$ を平均2乗誤差という. $\square$

Prop. (バイアス-バリアンス分解)

$$\mathrm{MSE}(\theta, \hat{\theta}) = \mathbb{V}[\hat{\theta}] + (\mathrm{Bias}(\theta))^2.$$ $\square$

Def. (BLUE)

$\hat{\theta} = \sum_{i=1}^{n} c_i X_i$ の形（線形）の推定量が，不偏かつ分散が最小となるとき，$\hat{\theta}$ を BLUE（最良線形不偏推定量）という． $\square$

6.3.2 Fisher情報量と Cramér-Rao 不等式．

・不偏推定量の分散はどこまで小さくできるか？

→ Cramér-Rao の下限まで．

Def. (Fisher情報量)

$X_1, \cdots, X_n \overset{i.i.d.}{\sim} f(x|\theta)$.

$$S_n(\theta, \mathbb{X}) := \frac{d}{d\theta} \ell(\mathbb{x}|\theta) = \frac{d}{d\theta} \sum_{i=1}^{n} \log f(x_i|\theta)$$

をスコア関数といい，その2乗の期待値

$$I_n(\theta) := \mathbb{E}[S_n(\theta, \mathbb{X})^2] = \mathbb{E}\left[\left(\frac{d}{d\theta}\ell(\mathbb{x}|\theta)\right)^2\right]$$

を Fisher情報量という． $\square$

Prop. 次の三つの条件を仮定する：

(C1) $f(x|\theta)$ のサポート $\{x \mid f(x|\theta)>0\}$ は θ に依らない.

(C2) $f(x|\theta)$ は θ に関して2階微分可能. さらに, $k=1,2$ として

$$\frac{d^k}{d\theta^k}\int f(x|\theta)dx = \int \frac{d^k}{d\theta^k} f(x|\theta)dx$$

(C3) $0 < I_1(\theta) < \infty$.

このとき, 次が成立：

(1) $\mathbb{E}[S_1(\theta, X_i)] = 0$.

(2) $I_n(\theta) = n I_1(\theta)$

n個のデータのFisher情報量は1個のデータのもののn倍

(3) $I_1(\theta) = -\mathbb{E}\left[\frac{d^2}{d\theta^2}\log f(X_i|\theta)\right]$. $\square$

Th. (Cramér-Raoの不等式)

上の(C1)~(C3)を仮定する. $\hat{\theta} = \hat{\theta}(X)$: θ の不偏推定量.

$\mathbb{V}[\hat{\theta}]$ が存在し

$$\frac{d}{d\theta}\int \hat{\theta}(x) f_n(x|\theta)dx = \int \hat{\theta}(x)\frac{d}{d\theta} f_n(x|\theta)dx$$

とする. このとき,

$$\forall\theta,\ \mathbb{V}_\theta[\hat{\theta}] \geq I_n(\theta)^{-1}.$$ $\square$

Cramér-Raoの下限

6.3.3 最尤推定量の一致性と漸近正規性，漸近有効性

Def. (一致性)

$X_1, X_2, \cdots \overset{\text{i.i.d.}}{\sim} f(x|\theta)$. $\hat{\theta}_n = \hat{\theta}_n(X_1, \cdots, X_n)$ とする.

$\hat{\theta}_n$: 一致性をもつ

$\overset{\text{def.}}{\Longleftrightarrow}$ $\hat{\theta}_n \xrightarrow{p} \theta$. $\square$

Prop. $\{a_n\}_{n=1}^{\infty}$: $a_n \nearrow \infty$.

$\exists U$: r.v. s.t. $a_n(\hat{\theta}_n - \theta) \xrightarrow{d} U$.

$\Rightarrow \hat{\theta}_n$: 一致性をもつ. $\square$

Def. (漸近分散，漸近有効)

・$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ となるとき，σ^2を漸近分散

という.

Cramér-Raoの下限

・$\hat{\theta}_n$の漸近分散が $\dfrac{1}{I_1(\theta)}$ のとき，i.e.,

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I_1(\theta)}\right)$$

のとき，$\hat{\theta}_n$は漸近有効であるという. $\square$