2014年後期
応用確率統計

⑧　統計的推測

河野 行雄
kawano@pe.titech.ac.jp
2014年12月11日

確率・統計とは？

- ランダムな（確率的な）ものには情報がある
- ランダムな観測値から情報を統計的に取り出す

■ 確率論（基礎的）

- 確率的現象を理論的に扱う
- 確率空間を設定しその上で理論体系を構築

■ 統計学（応用的）

- 現実に起こる現象を実験的に処理する
- 確率的構造を用いてその背後にある真実を推測する

統計的推測

■ 観測データ

例: $X_i = \theta + \varepsilon_i \quad (i = 1, \cdots, n)$

確率変数　未知母数　誤差・偶然変動

■ 統計的推測

観測値 $X_1, \cdots, X_n$ から未知母数 θ を推定すること

■ 仮定

1. $\varepsilon_1, \cdots, \varepsilon_n$ は互いに独立
2. $\varepsilon_1, \cdots, \varepsilon_n$ は同じ分布

i.i.d.
(independent identically distributed)

3. 誤差の平均 $E(\varepsilon_i) = 0$
4. 誤差の分散 $V(\varepsilon_i) < \infty$

推定量と推定値

■ 推定量 （estimator)

確率変数 $X_1, \cdots, X_n$ から未知母数 θ を推定する方式

$\hat{\theta}(X_1, \cdots, X_n)$ ← 関数、確率変数

■ 推定値 （estimate)

実現値 $x_1, \cdots, x_n$ を推定量 $\hat{\theta}$ に代入して得られる値

$\hat{\theta}(x_1, \cdots, x_n)$ ← 関数の値、実現値

平均の推定量

■ 標本平均

$$\hat{\theta} = \frac{X_1 + \cdots + X_n}{n}$$

■ 中央値

大きさの順に並べ替え

$$\{\tilde{X}_1, \cdots, \tilde{X}_n\} = \mathrm{sort}\{X_1, \cdots, X_n\}$$

$$\theta = \begin{cases} X_{(n+1)/2} & n : \mathrm{odd} \\ \left(X_{n/2} + X_{n/2+1}\right)/2 & n : \mathrm{even} \end{cases}$$

■ 加重平均

$$\sum_{i=1}^{n} c_n = 1$$ ← データの重み

$$\hat{\theta} = c_1 X_1 + c_2 X_2 + \cdots + c_n X_n$$

■ トリム平均

大きい m 個と小さい m 個を削除

$$\hat{\theta} = \frac{1}{n-2m} \sum_{i=m+1}^{n-m} X_i$$

■ 幾何平均

$$\hat{\theta} = \left(\prod_{i=1}^{n} X_i \right)^{1/n}$$

X_i は正の数

■ 一般化平均

$$\hat{\theta} = \left(\frac{1}{n} \sum_{i=1}^{n} X_i^m \right)^{1/m}$$

$m=1$ で標本平均、$m \to 0$ で幾何平均

不偏性

■ 推定量 $\hat{\theta}$ が不偏

観測値に基づく推定値の平均が真の母数に一致すること

$$E\left(\hat{\theta}\,|\,\theta\right)=E^{X_1,\cdots,X_n}\left(\hat{\theta}\left(X_1,\cdots,X_n\right)|\,\theta\right)=\theta$$

（※ E^X : X に対する期待値）

■ 推定量の不偏性

- 標本平均、加重平均: 誤差の平均が 0 であれば不偏
- 中央値、トリム平均: 分布が対称であれば不偏
- 幾何平均: 不偏ではない

標本平均(相加平均)

■ 推定量

$$\hat{\theta} = \frac{X_1 + \cdots + X_n}{n} = \left(\frac{1}{n} \sum_{i=1}^{n} X_i^m \right)^{1/m} \Bigg|_{m=1}$$

■ 不偏性

$$E\left(\hat{\theta}\right) = \frac{E\left(X_1\right) + \cdots + E\left(X_n\right)}{n} = \theta$$

標本平均(相加平均)は不偏

幾何平均(相乗平均)

■ 推定量

$$\hat{\theta} = \left(\prod_{i=1}^{n} X_i \right)^{1/n} = \exp\left(\frac{1}{n} \log \left(\prod_{i=1}^{n} X_i \right) \right) = \exp\left(\frac{1}{n} \sum_{i=1}^{n} \log X_i \right)$$

$$= \left(\frac{1}{n} \sum_{i=1}^{n} X_i^m \right)^{1/m} \Bigg|_{m \to 0}$$

イエンセンの不等式

上に凸な関数 $f(x)$ に対して

$$\sum_{i=1}^{n} p(x_i) f(x_i) \le f\left(\sum_{i=1}^{n} p(x_i) x_i \right)$$

■ 標本平均との関係

$$\log\left(\prod_{i=1}^{n} X_i \right)^{1/n} = \frac{1}{n} \sum_{i=1}^{n} \log X_i \le \log\left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)$$

$$\therefore \quad \left(\prod_{i=1}^{n} X_i \right)^{1/n} \le \frac{1}{n} \sum_{i=1}^{n} X_i$$ ⇨ 幾何平均(相乗平均)は不偏ではない

中央値

■ 推定量

大きさの順に並べ替え

$$\{\tilde{X}_1,\cdots,\tilde{X}_n\}=\mathrm{sort}\{X_1,\cdots,X_n\}$$

$$\theta=\begin{cases}X_{(n+1)/2} & n:\mathrm{odd}\\ \left(X_{n/2}+X_{n/2+1}\right)/2 & n:\mathrm{even}\end{cases}$$

■ 不偏性

確率分布が対称なとき

$$\int_{-\infty}^{M}p(x)dx=\int_{M}^{\infty}p(x)dx=1/2$$

$$E(\hat{\theta})=M=E(X)=\theta \quad \Rightarrow \text{確率分布が対称なとき中央値は不偏}$$

回帰直線

■ 観測データ（2つの物理量の関係）

従属な確率変数　　独立変数

例：　$Y_i = ax_i + b + \varepsilon_i \quad (i = 1, \cdots, n)$

未知母数（回帰係数）　　誤差・偶然変動

■ 最小二乗法（最小分散）

$$(\hat{a}, \hat{b}) = \arg\min_{a,b} \sum_{i=1}^{n} \varepsilon_i^2 = \arg\min_{a,b} \sum_{i=1}^{n} (y_i - ax_i - b)^2$$

$$\Rightarrow \quad \hat{a} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad \hat{b} = \bar{y} - \hat{a}\bar{x}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

例題：実験データの回帰

■ 実験データ

x	20	25	30	35	40	45	50
y	9.137	8.913	8.665	8.528	8.242	8.203	7.972

■ 最小二乗法（最小分散）

$$(\hat{a},\hat{b}) = \arg\min_{a,b} \sum_{i=1}^{n} \varepsilon_i^2 = \arg\min_{a,b} \sum_{i=1}^{n} \left(y_i - ax_i - b\right)^2$$

$$\Rightarrow \quad \hat{a} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \hat{b} = \bar{y} - \hat{a}\bar{x}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

実験データの回帰

■ 実験データ

x	20	25	30	35	40	45	50
y	9.137	8.913	8.665	8.528	8.242	8.203	7.972

■ 回帰直線

$$\bar{x} = \frac{1}{7}\sum_{i=1}^{7} x_i = 35 \qquad \bar{y} = \frac{1}{7}\sum_{i=1}^{7} y_i = 8.523$$

$$\hat{a} = \frac{\sum_{i=1}^{7}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{7}(x_i - \bar{x})^2} = -0.03812 \qquad \hat{b} = \bar{y} - \hat{a}\bar{x} = 9.858$$

$$y = -0.03812x + 9.859$$

まとめ

■ 推定量と推定値

推定量: $\hat{\theta}(X_1, \cdots, X_n)$　　推定値: $\hat{\theta}(x_1, \cdots, x_n)$

■ 不偏性

$$E(\hat{\theta} \mid \theta) = E^{X_1, \cdots, X_n}\left(\hat{\theta}(X_1, \cdots, X_n) \mid \theta\right) = \theta$$

■ 標本平均と幾何平均

$$\hat{\theta} = \frac{X_1 + \cdots + X_n}{n} \quad \Rightarrow \quad \text{不偏}$$

$$\hat{\theta} = \left(\prod_{i=1}^{n} X_i \right)^{1/n} \quad \Rightarrow \quad \text{不偏ではない}$$

一般化平均と幾何平均

$$\lim_{m\to 0}\left(\frac{1}{n}\sum_{i=1}^{n}X_i^m\right)^{1/m} = \lim_{m\to 0}\exp\frac{\log\sum_{i=1}^{n}X_i^m}{nm}$$

$$= \exp\frac{\sum_{i=1}^{n}\log X_i}{n} = \left(\prod_{i=1}^{n}X_i\right)^{1/n}$$

$$\because \quad \lim_{m\to 0}\frac{\log\sum_{i=1}^{n}X_i^m}{nm} = \lim_{m\to 0}\frac{1}{1/n\sum_{i=1}^{n}X_i^m}\left(1/n\sum_{i=1}^{n}X_i^m\right)'$$

ロピタルの法則

$$= \lim_{m\to 0}\frac{1}{n}\sum_{i=1}^{n}\log(X_i)X_i^m = \frac{\sum_{i=1}^{n}\log X_i}{n}$$