

# 応用システム工学

## 第四回 回帰分析

平成23年05月06日

2011/05/06

1

## 母回帰係数の推定

- 標本回帰係数  $\hat{a}_1$  と母回帰係数  $a_1$  の関係

$$\hat{a}_1 = a_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})e_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

- 標本回帰係数  $\hat{a}_1$  の期待値  $E[\hat{a}_1]$

$$E[\hat{a}_1] = E[a_1] = a_1$$

- 標本回帰係数  $\hat{a}_1$  の分散  $V[\hat{a}_1]$

$$V[\hat{a}_1] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

2011/05/06

2

# 母回帰係数の推定

- 標本回帰係数  $\hat{a}_1$  は誤差  $e_i$  の一次関数として表される

$$\hat{a}_1 = a_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})e_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

→ 標本の誤差が正規分布なので標本回帰係数も正規分布する

$$\hat{a}_1 \sim N\left(a_1, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

– 標準正規分布化  $\frac{\hat{a}_1 - a_1}{\sigma^2} \sim N(0,1)$

- 期待値を引いて分散で割る

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} \rightarrow \text{ただし, } \sigma^2, a_1 \text{ は未知数}_3$$

2011/05/06

# 大数の法則

- 標本数 → 用いる母集団の数
- 標本サイズ → 母集団からとりだす標本の数
- 大数の法則 → 標本サイズを大きくすると、その期待値が母集団の期待値から離れた値となる確率は小さくなる
  - 標本平均の期待値  $E[\bar{X}_n]$  と分散  $V[\bar{X}_n]$  で評価
  - 標本  $X_1, X_2, \dots, X_n$  は母集団と同じ確率分布 (平均  $\mu$ , 分散  $\sigma^2$ ) となる確率変数
  - 標本  $X_1, X_2, \dots, X_n$  は独立

2011/05/06

# 確率変数の性質

- 確率変数 $X$ の関数 $g(X)$ の期待値 $E[g(X)]$ 
  - 確率密度関数: 確率変数 $X$ が値 $x$ をとる確率 $f(x)$ 
$$E[g(X)] = \sum_x g(x)f(x)$$
    - 確率変数の期待値 $E[X] \rightarrow g(x)=x$   
(原点の周りの1次のモーメント(積率))
    - 確率変数の分散 $V[X] \rightarrow g(x)=(x-\mu)^2$   
(期待値の周りの2次のモーメント)
  - 定数 $c$ に対する性質
    - 期待値  $E[cX] = \sum_x cxf(x) = c \sum_x xf(x) = cE[X]$
    - 分散  $V[cX] = \sum_x (cx - c\mu)^2 f(x) = c^2 \sum_x (x - \mu)^2 f(x) = c^2V[X]$

2011/05/06

5

# 同時確率分布

- 同時確率分布
  - 確率変数 $X$ の値が $x$ となり, 同時に確率変数 $Y$ の値が $y$ となる確率分布
  - $X, Y$ の同時確率密度関数  $f_{XY}(x, y)$
- 周辺確率分布
  - 同時確率分布から, ある確率変数の分布を抽出したもの
  - 確率変数 $X$ の周辺確率密度関数  $f_X(x)$

2011/05/06

6

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

## 複数の確率変数の性質

- 2つの確率変数の和に対する期待値

$$\begin{aligned}
 E[X + Y] &= \iint_{x,y} (x + y) f_{XY}(x, y) dx dy \\
 &= \iint_{x,y} x f_{XY}(x, y) dx dy + \iint_{x,y} y f_{XY}(x, y) dx dy \\
 &= \int_x x \int_y f_{XY}(x, y) dy dx + \int_x y \int_y f_{XY}(x, y) dx dy \\
 &= \int_x x f_X(x) dx + \int_x y f_Y(y) dy \\
 &= E[X] + E[Y]
 \end{aligned}$$

2011/05/06

7

## 複数の確率変数の性質

- 確率変数 $X, Y$ が独立である場合における、同時確率密度関数 $f_{XY}(x, y)$ と、周辺確率密度関数 $f_X(x), f_Y(y)$ の関係

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

- 独立な確率変数の積の期待値

$$\begin{aligned}
 E[XY] &= \iint_{x,y} xy f_{XY}(x, y) dx dy = \iint_{x,y} x f_X(x) y f_Y(y) dx dy \\
 &= \int_x x f_X(x) dx \int_y y f_Y(y) dy = E[X] E[Y]
 \end{aligned}$$

– この時共分散は

$$\begin{aligned}
 \text{Cov}[X, Y] &= E[(X - E[X])(Y - E[Y])] = E[XY - E[X]Y - XE[Y] + E[X]E[Y]] \\
 &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y]
 \end{aligned}$$

$$2011/05/06 \quad \stackrel{=}{=} E[XY] - E[X]E[Y] = 0$$

8

# 複数の確率変数の性質

- 分散と期待値の関係

$$\begin{aligned}V[X] &= E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2\end{aligned}$$

- 2つの確率変数の和に対する分散

$$\begin{aligned}V[X + Y] &= E[(X + Y)^2] - E[X + Y]^2 = E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2 \\ &= E[X^2] + 2E[XY] + E[Y^2] - (E[X]^2 + 2E[X]E[Y] + E[Y]^2) \\ &= (E[X^2] - E[X]^2) + (E[Y^2] - E[Y]^2) + 2(E[XY] - E[X]E[Y])\end{aligned}$$

- 2つの確率変数が独立の場合

$$\begin{aligned}V[X + Y] &= V[X] + V[Y] + 2(E[X]E[Y] - E[X]E[Y]) \\ &= V[X] + V[Y]\end{aligned}$$

2011/05/06

9

# 標本平均

- 標本平均の期待値

$$\begin{aligned}E[\bar{X}_n] &= E\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] = \frac{1}{n} \{E[X_1] + E[X_2] + \dots + E[X_n]\} \\ &= \frac{1}{n} \{\mu + \mu + \dots + \mu\} = \mu\end{aligned}$$

- 標本平均の分散

$$\begin{aligned}V[\bar{X}_n] &= V\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] = \frac{1}{n^2} V[X_1 + X_2 + \dots + X_n] \\ &= \frac{1}{n^2} \{V[X_1] + V[X_2] + \dots + V[X_n]\} = \frac{1}{n^2} \{\sigma^2 + \sigma^2 + \dots + \sigma^2\} \\ &= \frac{1}{n^2} n\sigma^2 = \frac{1}{n} \sigma^2\end{aligned}$$

2011/05/06

→標本の数nが小さいと分散が大きい  
標本の数を大きくすると期待値に近づく 10

# チェビシェフの不等式

- 集合  $I = \{x : |x - E[x]| \geq k\sqrt{V[x]}\}$       ただし  $k > 0$
- 確率変数  $X$  が  $x$  となる確率  $P(X=x)$  の確率頻度関数  $f(x)$   $P[|X - E[X]| \geq k\sqrt{V[X]}] = \sum_{x \in I} f(x)$   
– 確率  $P$  で分散を表す

$$\begin{aligned} V[X] &= \sum_{x \in I} (x - E[X])^2 f(x) \geq \sum_{x \in I} (k\sqrt{V[X]})^2 f(x) \\ &\geq k^2 V[X] \sum_{x \in I} f(x) = k^2 V[X] P[|X - E[X]| \geq k\sqrt{V[X]}] \end{aligned}$$

2011/05/06

11

# チェビシェフの不等式

- 確率  $P$  で分散を表す

$$V[X] \geq k^2 V[X] P[|X - E[X]| \geq k\sqrt{V[X]}]$$

$$\frac{1}{k^2} \geq P[|X - E[X]| \geq k\sqrt{V[X]}]$$

- 確率変数と期待値の差が分散の根より大きくなる確率は小さい ( $1/k^2$ )

2011/05/06

12