応用システム工学
第三回　確率統計の基礎

平成25年05月24日
多変数の確率分布
一般化線形モデル
相関分析
回帰分析

指数型分布族

- 確率分布が指数関数で表される

$$f(y;\theta) = s(y)\tau(\theta)e^{a(y)b(\theta)}$$

  - 既知の関数a,b,s,t
  - 二項分布　→　サイコロの目の出る確率

$$P(x) = {}_nC_x p^x (1-p)^{n-x}$$

  - ポアソン分布　→　一定の期間において，ごくまれに起こる事象の確率分布

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

  - 正規分布　→　自然界で起こる現象の多くがその分布に当てはまる(特に期待値に関する分布)
    - 二項分布，ポアソン分布の正規分布による近似もある

$$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

分散

- サンプルに対する算出
  - 相加平均 $$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
  - 分散 $$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$$
    分散は母集団の
    分散より常に小さい
  - 不偏分散 $$\sigma'^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$$
    バラツキがでるのは
    二個以上
  - 標準偏差 $$\sigma, \sigma'$$
    - (不偏)分散の平方根

期待値と分散

- 期待値の線形性

$$E[A+B]=E[A]+E[B]$$

- 期待値の線形性を用いた分散σ^2の表現

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

正規分布関数の分散1

- 正規分布の確率密度関数(確率変数X)

$$f_X\left(x;\mu,\sigma^2\right)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

- 分散

$$V\left[X\right]=\int_{-\infty}^{\infty}\left(x-\mu\right)^2 f_X\left(x\right)dx$$

$$=\int_{-\infty}^{\infty}\left(x^2-2x\mu+\mu^2\right)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx$$

$$=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\left(x^2-2x\mu+\mu^2\right)\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx$$

正規分布関数の分散2

$$V[X]=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}x^2\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx$$

$$-\frac{2\mu}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}x\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx$$

$$+\frac{\mu^2}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx$$

正規分布関数の分散3

- 第3項 正規分布関数の確率密度関数1参照

$$\frac{\mu^2}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx=\mu^2\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx$$

$$=\mu^2$$

- 第2項 正規分布関数の期待値2参照

$$\frac{-2\mu}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}y\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx=\frac{-2\mu}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}y\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx$$

$$=-2\mu^2$$

正規分布関数の分散4

- 第1項
  - 変数変換

$$z = \frac{x-\mu}{\sigma} \Rightarrow \frac{dz}{dx} = \frac{d}{dx}\frac{x-\mu}{\sigma} = \frac{1}{\sigma}$$

$$x = \sigma z + \mu \qquad x : -\infty \leftrightarrow \infty \Leftrightarrow z =: -\infty \leftrightarrow \infty$$

$$\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} x^2 \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}(z\sigma+\mu)^2 e^{-\frac{1}{2}z^2}\sigma dz$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\left(z^2\sigma^2 + 2z\sigma\mu + \mu^2\right)e^{-\frac{1}{2}z^2}\sigma dz$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(z^2\sigma^2 + 2z\sigma\mu + \mu^2\right)e^{-\frac{1}{2}z^2} dz$$

正規分布関数の分散5

$$\int_{-\infty}^{\infty} z e^{-\frac{1}{2}z^2} dz = 0 \qquad \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}z^2\right] dz = \sqrt{2\pi}$$

$$\int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2} dz = \int_{-\infty}^{\infty} zze^{-\frac{1}{2}z^2} dz = \left[z\left(-e^{-\frac{1}{2}z^2}\right)\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -e^{-\frac{1}{2}z^2} dz$$

$$= [0-0] + \sqrt{2\pi} = \sqrt{2\pi}$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(z^2\sigma^2 + 2z\sigma\mu + \mu^2\right) e^{-\frac{1}{2}z^2} dz = \frac{1}{\sqrt{2\pi}} \left(\sigma^2\sqrt{2\pi} + 2\sigma\mu \times 0 + \mu^2\sqrt{2\pi}\right)$$

$$= \sigma^2 + \mu^2$$

正規分布関数の分散6

$$V[X]=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} x^2 \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx$$

$$-\frac{2\mu}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} x \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx$$

$$+\frac{\mu^2}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx$$

$$=\left\{\sigma^2+\mu^2\right\}+\mu^2-2\mu^2=\sigma^2$$

用語

- 測定値または観測値（原因）
  - 説明変数
  - 予測変数
  - 独立変数
- 確率変数(上述の変数に応じて変動:結果)
  - 目的変数
  - 反応変数
  - 結果変数
  - 従属変数

多変量の確率分布1
2変数

- 同時確率密度関数
  - 二つの確率変数X,Y

$$x \le X \le x + dx$$

かつ

$$y \le Y \le y + dy$$

となる確率が

$$f_{XY}(x, y)dxdy$$

と書ける

  - XとYの同時確率密度関数

$$f_{XY}(x, y)$$

多変数の確率分布2
n変数

- 同時確率密度関数
  - n個の確率変数

$$X_1, X_2, \cdots, X_n$$

  - 各々 $x_i \le X_i \le x_i + dx_i \left(1 \le i \le n\right)$ となる確率が

$$f_{X_1 X_2 \cdots X_n}\left(x_1, x_2, \cdots, x_n\right) dx_1 dx_2 \cdots dx_n$$

  - $X_1, X_2, \cdots, X_n$ の同時確率密度関数

$$f_{X_1 X_2 \cdots X_n}\left(x_1, x_2, \cdots, x_n\right)$$

多変数の確率分布3
2変数

- 周辺密度関数
  - 二つの確率変数X,Y
  - Yを無視したXのみに関する確率密度関数を指す

    $f_X(x)$

  - 同時確率密度関数と周辺密度関数の関係 $f_{XY}(x,y)$ $f_X(x)$

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

多変数の確率分布4
n変数

- 周辺密度関数
  - n個の確率変数

$$X_1, X_2, \cdots, X_n$$

    - m個のみに注目した同時確率密度(m<n)

$$f_{X_1 X_2 \cdots X_m}\left(x_1, x_2, \cdots, x_m\right)$$

    - $X_1, X_2, \cdots, X_m$ についての周辺密度関数

$$f_{X_1 \cdots X_m}\left(x_1, \cdots, x_m\right) = \int_{-\infty}^{\infty} dx_{m+1} \cdots \int_{-\infty}^{\infty} dx_n \, f_{X_1 \cdots X_n}\left(x_1, \cdots, x_n\right)$$

確率変数の独立性1

- 独立性の定義
  - 二つの確率変数X,Y

$$P(X \in A\text{かつ}Y \in B) = P(X \in A)P(Y \in B)$$
$$\Leftrightarrow f_{XY}(x, y) = f_X(x) f_Y(y)$$

確率変数の独立性2

- 独立性の定義
  - n個の確率変数

$$X_1, X_2, \cdots, X_n$$

$$P\left(X_1 \in A_1 \text{かつ} X_2 \in A_2 \text{かつ} \cdots \text{かつ} X_n \in A_n\right)$$
$$= P\left(X_1 \in A_1\right) P\left(X_2 \in A_2\right) \cdots P\left(X_n \in A_n\right)$$
$$\Leftrightarrow f_{X_1 X_2 \cdots X_n}\left(x_1, x_2, \cdots, x_n\right) = f_{X_1}\left(x_1\right) f_{X_2}\left(x_2\right) \cdots f_{X_n}\left(x_n\right)$$

- 自由度
  - 独立に選べる確率変数の数

条件付き確率1

- 二つの確率変数X,Yが必ずしも独立でない

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

  – YのXに関する条件付き確率密度　$f(y|x)$
    - Xが値xを取る条件下でYが値yを取る確率密度
      – XとYが独立な場合　$f(y|x) = f_Y(y)$
- 同時確率密度$f_{XY}(x,y)$と条件付確率密度$f(y|x)$の関係

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

条件付き確率2

- 条件　$X_1 = x_1, X_2 = x_2, \cdots, X_m = x_m$
  に対する，$X_{m+1} = x_{m+1}, X_{m+2} = x_{m+2}, \cdots, X_n = x_n$
  となる確率密度関数

$$f\left(x_{m+1}, x_{m+2}, \cdots, x_n \mid x_1, x_2, \cdots, x_m\right)$$

$$\begin{aligned}
& f\left(x_1, x_2, \cdots, x_m, x_{m+1}, \cdots, x_n\right) \\
& = f\left(x_{m+1}, x_{m+2}, \cdots, x_n \mid x_1, x_2, \cdots, x_m\right) \\
& \times f\left(x_1, x_2, \cdots, x_m\right)
\end{aligned}$$

確率変数の和の分布

- 二つの独立な確率変数X,Y
- Z=X+Yの分布
  - 事象　$Z = z$

$$\Leftrightarrow X + Y = z$$

$$\Leftrightarrow Y = z - x \text{かつ} X = x$$

X,Yが独立の時，Y=z-xかつX=xとなる確率

$$P_X(x)P_Y(z-x)$$

  - 離散分布　$X \in M$
  - Z=zとなる確率

$$P(Z = z) = \sum_{x \in M} P_X(x)P_Y(z-x)$$

説明変数の種類

- 一般化線形モデル
  - 反応変数Yと説明変数$x_1, x_2, \cdots, x_m$の関係

$$g[E(Y)] = \underbrace{\beta_0 + \beta_1 x_1 + \cdots \beta_m x_m}_{\text{線形関数}}$$

- 量的な説明変数x
  - パラメータβは，説明変数の変化に伴う反応変数の変化を表す
- 質的な説明変数
  - 反応変数にパラメータが含まれるか否かを表す
  - ダミー変数，(0,1)の場合は指示変数

多変量解析 ⇒ 相関分析，回帰分析