

応用システム工学

第三回 確率統計の基礎

平成26年04月25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

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

- 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

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

期待値

- 確率変数の平均みたいなもの
- 期待値 $E[X]$

– 離散分布の確率変数 X

- X の取り得る値の集合 M

$$E[X] = \sum_{x \in M} xP(X = x)$$

– 連続分布の確率変数 X

$$E[X] = \int_{-\infty}^{\infty} xf_X(x)dx$$

正規分布関数の期待値1

- 正規分布の確率密度関数(確率変数 X)

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

- 期待値

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

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

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

正規分布関数の期待値2

- 変数変換

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

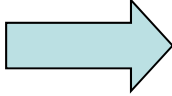
$$x = \sigma z + \mu \quad x: -\infty \leftrightarrow \infty \Leftrightarrow z: -\infty \leftrightarrow \infty$$

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

正規分布関数の期待値3

- 右辺第一項 $\frac{d}{dz} e^{-\frac{1}{2}z^2} = -\frac{1}{2} 2ze^{-\frac{1}{2}z^2} = -ze^{-\frac{1}{2}z^2}$

$$\int_{-\infty}^{\infty} ze^{-\frac{1}{2}z^2} dz = \left[e^{-\frac{1}{2}z^2} \right]_{-\infty}^{\infty} = e^{-\infty} - e^{-\infty} = 0$$

- 右辺第二項 $\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz = \sqrt{2\pi}$  正規分布関数の
確率密度関数3
参照

- 全体
$$E[X] = \frac{1}{\sqrt{2\pi}} \left[\sigma \int_{-\infty}^{\infty} ze^{-\frac{1}{2}z^2} dz + \mu \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz \right]$$
$$= \frac{1}{\sqrt{2\pi}} \left[\sigma 0 + \mu \sqrt{2\pi} \right]$$
$$= \mu$$

分散

- 確率変数 X のばらつきの程度を表す
- 分散 $V[X]$ の定義

$$V[X] = E[(X - E[X])^2]$$

分散

- サンプルに対する算出

- 相加平均
$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- 分散
$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

分散は母集団の
分散より常に小さい

- 不偏分散
$$\sigma'^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

バラツキができるのは
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

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

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)

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

- 分散

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

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

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

正規分布関数の分散2

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

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

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

- 第2項 正規分布関数の期待値2参照

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

正規分布関数の分散4

- 第1項

– 変数変換

$$z = \frac{x - \mu}{\sigma} \quad \Rightarrow \quad \frac{dz}{dx} = \frac{d}{dx} \frac{x - \mu}{\sigma} = \frac{1}{\sigma}$$
$$x = \sigma z + \mu \quad x: -\infty \leftrightarrow \infty \Leftrightarrow z: -\infty \leftrightarrow \infty$$

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

正規分布関数の分散5

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

$$\begin{aligned} \int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2} dz &= \int_{-\infty}^{\infty} z z e^{-\frac{1}{2}z^2} dz = \left[z \left(-e^{-\frac{1}{2}z^2} \right) \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -e^{-\frac{1}{2}z^2} dz \\ &= [0 - 0] + \sqrt{2\pi} = \sqrt{2\pi} \end{aligned}$$

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

正規分布関数の分散6

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

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

$$x \leq X \leq x + dx$$

$$y \leq Y \leq y + dy \quad \begin{array}{l} \text{かつ} \\ \text{となる確率が} \end{array}$$

$$f_{XY}(x, y) dx dy \quad \text{と書ける}$$

- 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

$$X_1, X_2, \dots, X_n$$

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

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

- X_1, X_2, \dots, X_n の同時確率密度関数

$$f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)$$

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 $f_{XY}(x, y)$
と周辺密度関数 $f_X(x)$
の関係

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

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

$$X_1, X_2, \dots, X_n$$

- m個のみに注目した同時確率密度($m < n$)

$$f_{X_1 X_2 \dots X_m}(x_1, x_2, \dots, x_m)$$

- X_1, X_2, \dots, X_m についての周辺密度関数

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

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

$$X_1, X_2, \dots, X_n$$

$$P(X_1 \in A_1 \text{かつ} X_2 \in A_2 \text{かつ} \dots \text{かつ} X_n \in A_n)$$

$$= P(X_1 \in A_1)P(X_2 \in A_2) \cdots P(X_n \in A_n)$$

$$\Leftrightarrow f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

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

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

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

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 $Z = z$

X, Y が独立の時, $Y=z-x$ かつ $X=x$ となる確率

$$\Leftrightarrow X + Y = z$$

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

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

– 離散分布 $X \in M$

– $Z=z$ となる確率

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