確率統計基礎講義A

若木宏文

wakaki@hiroshima-u.ac.jp
http://home.hiroshima-u.ac.jp/ wakaki/lecture/index.shtml

2019.4.12

目 次

1．1次元正規分布とその拡張

1.1. 1次元正規分布

定義

確率密度関数

$$f(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}\ (\mu\in\mathbb{R},\sigma>0)$$

によって定まる連続型分布を，1 次元正規分布といい，$N(\mu,\sigma^2)$ と表す．

特に，$N(0,1)$ を標準正規分布と呼び，その確率密度関数を $\phi(x)$, 分布関数を $\Phi(x)$ と書く．

何故正規分布？

多くのデータの分布に当てはまりが良い.
（特に生物のデータ）

中心極限定理

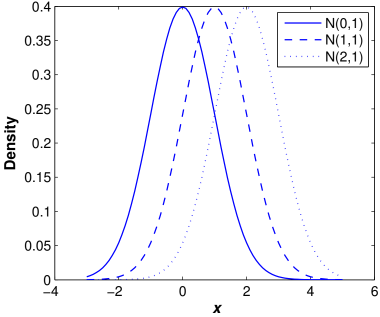
$X_1, X_2, \ldots \overset{i.i.d.}{\sim} \{\mathsf{E}[X_i] = \mu, \mathsf{Var}[X_i] = \sigma^2\}$

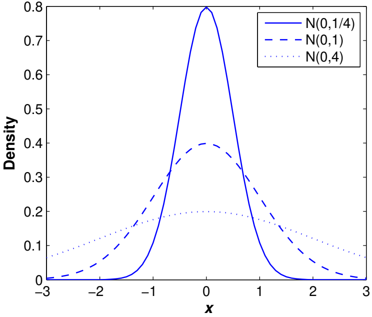
$$\frac{\sqrt{n}}{\sigma}\left\{\frac{1}{n}(X_1 + \cdots + X_n) - \mu\right\} \xrightarrow{d} N(0,1) \ (n \to \infty)$$

注. $X_1, X_2, \ldots$ が同じ分布でなくても，適当な条件下で漸近正規性が成り立つ. （リンドバーグ・フェラーの中心極限定理）

分布の特性

1. 正規分布は μ を中心に左右対称な分布であり，μ は中心位置，σ^2 は裾の広がりを表している．





分布の特性(続き)

2. $X \sim N(\mu, \sigma^2) \;\Rightarrow\; Z = \dfrac{X-\eta}{\psi} \sim N\left(\dfrac{\mu-\eta}{\psi}, \dfrac{\sigma^2}{\psi^2}\right)$

   $X \sim N(\mu, \sigma^2) \;\Rightarrow\; Z = \dfrac{X-\mu}{\sigma^2} \sim N(0, 1)$

   $Z \sim N(0, 1) \;\Rightarrow\; X = \sigma Z + \mu \sim N(\mu, \sigma^2)$

3. $X \sim N(\mu, \sigma^2) \;\Rightarrow\; \mathsf{E}[X] = \mu, \mathsf{Var}[X] = \sigma^2$

$$\alpha_r = \begin{cases} 0 & (r \text{ が奇数}) \\ 1 \times 3 \times \cdots \times (r-1)\sigma^r & (r \text{ が偶数}) \end{cases}$$

4. $X \sim N(\mu, \sigma^2)$ のとき，X の特性関数は

$$\mathcal{C}(t) = \exp\left(it\mu - \frac{1}{2}t^2\sigma^2\right)$$

分布の特性(続き)

5. 再生性：$X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$ で，かつ，X_1, X_2 が独立ならば $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
6. 乱数発生法（ボックス＝ミュラー法）$U_1, U_2 \overset{i.i.d.}{\sim} U(0,1)$ とし，

$$Z_1 = \sqrt{-2 \log U_1} \cos(2\pi U_2)$$
$$Z_2 = \sqrt{-2 \log U_1} \sin(2\pi U_2)$$

   とすると $Z_1, Z_2 \overset{i.i.d.}{\sim} N(0,1)$

正規分布の発見

ド・モアブル＝ラプラスの定理
最初に正規分布を提示したのは（恐らく）ド・モアブル
$X \sim B(n,p)$ (2 項分布) ならば

$$\frac{X-np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1) \ (n \to \infty)$$

ガウス分布，誤差分布
測定誤差の分布として，ガウスが導出した.
同じ対象を繰り返し測定して得られる観測値 $x_1, \ldots, x_n$ の統計モデルとして
$X_1, \ldots, X_n \overset{i.i.d.}{\sim} f(x-\mu)$ （確率密度関数）とする.

正規分布の発見（続き）

μ の尤度関数は

$$L(\mu; x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i - \mu)$$

任意の $x_1, \ldots, x_n$ に対して，$L(\mu; x_1, \ldots, x_n)$ が $\mu = (x_1 + \ldots + x_n)/n$ で最大となるような，滑らかな関数 f を導出すると

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}$$

が得られる．