正規分布 (1) 付録

坂田 綾香

二変量正規分布の行列表現

- 二変量をx_1, x_2として，ベクトル$\boldsymbol{x}$で表す
- x_1, x_2の平均値をμ_1, μ_2として，ベクトル$\boldsymbol{\mu}$で表す
- 共分散行列$\boldsymbol{\Sigma}$を導入する

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_x^2 & \rho \\ \rho & \sigma_y^2 \end{pmatrix}$$

- 二変量正規分布は次のように表される

$$p(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}\right)$$

  - Σ^{-1}はΣの逆行列
  - $|\Sigma|$はΣの行列式
  - Tは転置

n変量正規分布の行列表現

- n変量を$x_1, x_2, \dots, x_n$として，ベクトル$\boldsymbol{x}$で表す
- $x_1, \dots, x_n$の平均値を$\mu_1, \dots, \mu_n$として，ベクトル$\boldsymbol{\mu}$で表す
- 共分散行列$\boldsymbol{\Sigma}$を導入する

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix},\ \Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & \rho_{n-1,n} \\ \rho_{1n} & \cdots & \rho_{n-1,n} & \sigma_n^2 \end{pmatrix}$$

- n変量正規分布は次のように表される

$$p(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}\right)$$

複数の峰を持つ分布の例

成分数2の混合正規分布(normal mixture)
※ 有限混合分布(finite mixture)の一種

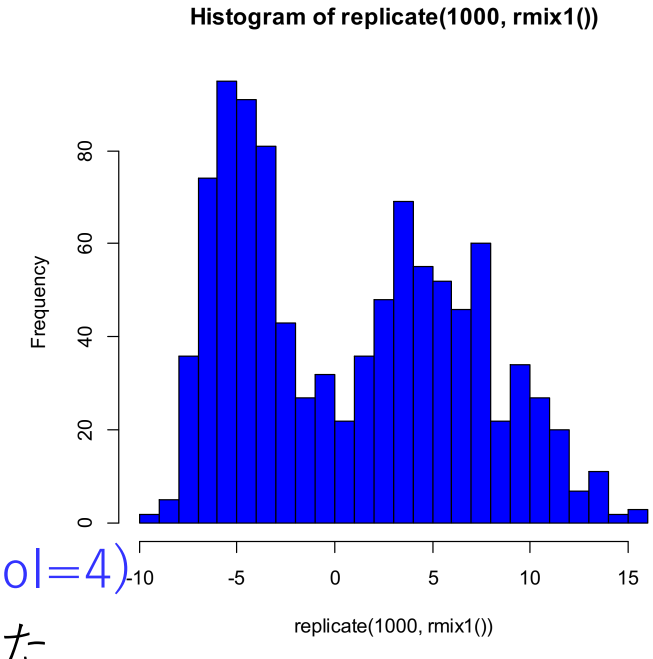
確率qで $\mathrm{N}(\mu_1, \sigma_1^2)$からのサンプルを出力
確率$1-q$で$\mathrm{N}(\mu_2, \sigma_2^2)$からのサンプルを出力

$$p(x; \mu_1, \mu_{2,} \sigma_1^2, \sigma_2^2) = \frac{q}{\sqrt{2\pi\sigma_1^2}} \mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} + \frac{1-q}{\sqrt{2\pi\sigma_2^2}} \mathrm{e}^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$$

扱い方が難しい点もあるが「分類」の基本モデル
として機械学習や統計学で重要

Rで混合正規分布

```
set.seed(1985); q=0.4
rmix1=function(){
 id=sample(c(0,1),prob=c(q,1-q),size=1)
 if(id==0){
  rnd=rnorm(n=1,mean=-5,sd=1.5)
 }else{
  rnd=rnorm(n=1,mean=+5,sd=4.0)
 }
 return(rnd)
 }
hist(replicate(1000,rmix1()),breaks=20,col=4)
```



#一般の有限混合分布に使える方法を示した

リンデベルグ条件

- 同分布ではないときにCLTが成立する条件

(1)$E[X_k] = 0,\ k = 1,2,\dots,n$

(2)$V[X_k] = \sigma_k^2 < \infty,\ k = 1,2,\dots,n$

(3)$B = \sqrt{\sum_{k=1}^{n} \sigma_k^2}$として，任意の$M(>0)$に対して

$$\lim_{n\to\infty} \frac{1}{B^2} \sum_{k=1}^{n} E[X_k^2, \text{ただし} |X_k| > MB] = 0$$

が成り立つならばS/Bの分布は標準正規分布に収束する

証明は確率論 (西尾 真喜子)などを参照

(1), (2)は満たすが(3)は満たさない例

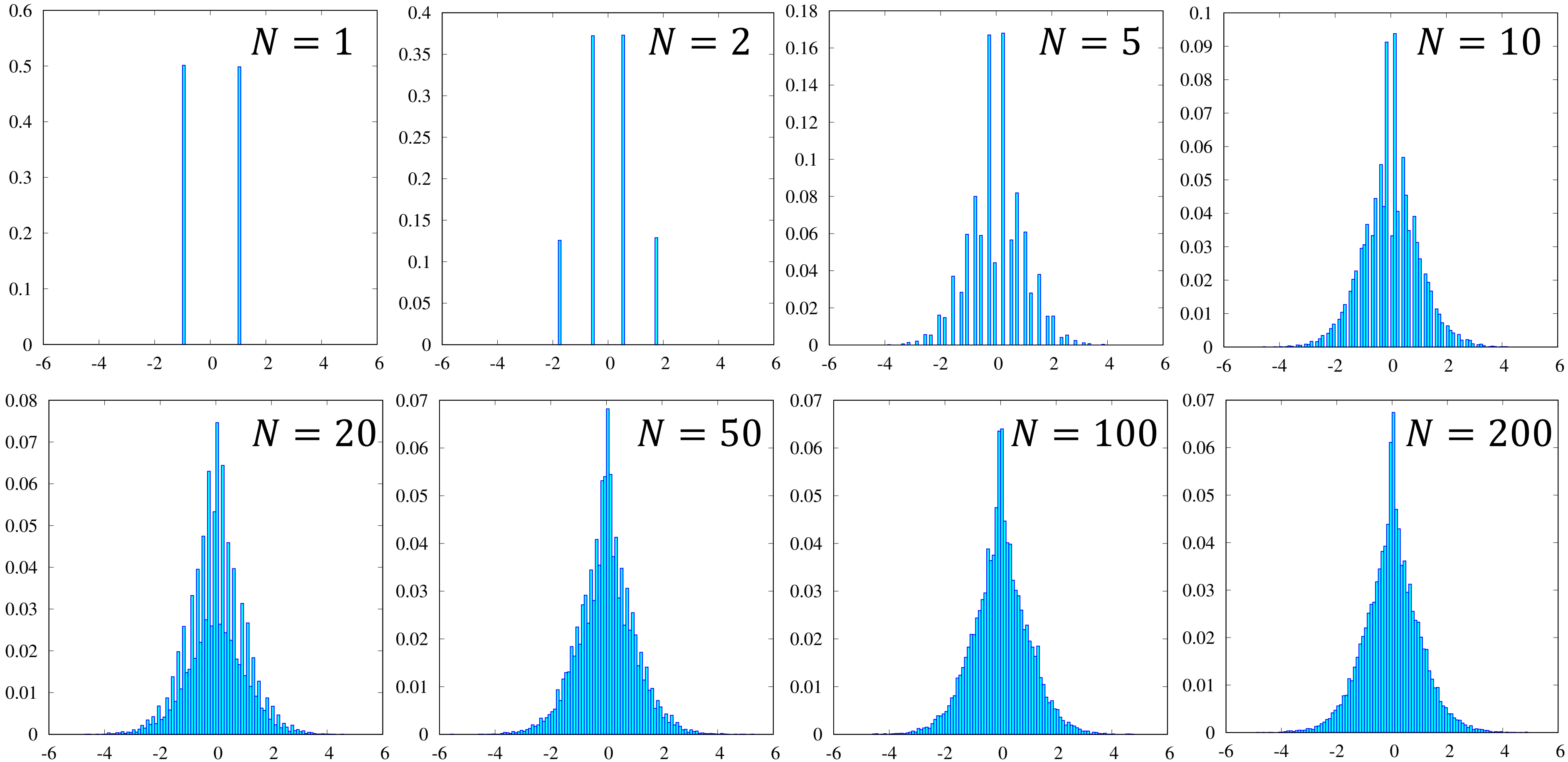
- k番目の確率変数X_kは$\{-k, 0, k\}$いずれかの値をとる

$$X_k = \begin{cases} k & \text{確率 } \dfrac{1}{2k} \\ -k & \text{確率 } \dfrac{1}{2k} \\ 0 & \text{確率 } 1 - \dfrac{1}{k} \end{cases}$$

- 平均0，分散$k(< \infty)$なので条件(1), (2)は満たす
- $B^2 = \sum_{k=1}^{n} k$，$\sum_{k=1}^{n} E[X_k^2$, ただし$|X_k| > MB] = k$ なので(3)は成立しない

リンデベルグ条件つづき

$\frac{S}{\sqrt{B}}$の分布のNによる変化



標準正規分布に収束しない

大学共同利用機関法人 情報・システム研究機構
統計数理研究所