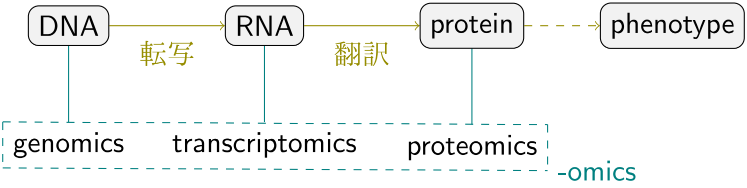
テンソル同時分解の拡張による オミクスデータの統合

阿部興[1] ・島村徹平[2]

2023 年 6 月 3 日

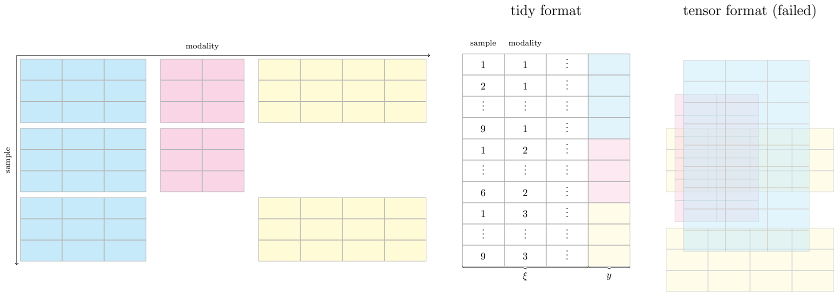
[1]東京医科歯科大学難治疾患研究所
[2]名古屋大学医学系研究科・東京医科歯科大学難治疾患研究所

動機：分析対象



- オミクス（-omics）データを統合して分析したい
  - 積極的理由：データを補い合い普遍的な特徴を抽出
  - 消極的理由：対応のあるサンプルなので非独立

動機：分析手法



- semi-paired なデータが多い
- モダリティごとに分布が変わる

切断, Rectified, 離散化

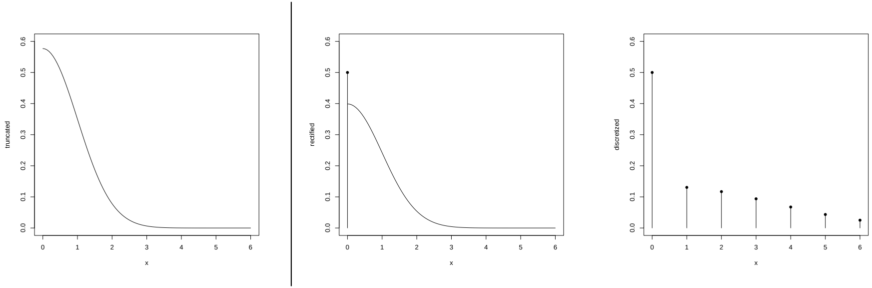


Fig: left: 潜在変数の事前分布. right: 観測されるデータのモデル. 一点 0 で確率を持つ（eg. 重さ, 計数）

data: $Y = (y_{ijk})$

$y = (y_n) = \mathrm{vec}(Y)$ as given .

Let the m-th subscript of y_{ijk} be ξ_m. Canonical decomposition and parallel factor analysis (also known as CP decomposition) seeks matrices $v^{(m)}$ such that

$$y_{ijk} \approx \sum_l v_{il}^{(1)} v_{jl}^{(2)} v_{kl}^{(3)}.$$

This equation can be represented as follows:

$$y_n \approx \sum_l \prod_{d=1}^{D} v_{dr}^{x_{nd}} \tag{1}$$

where

$$V = (v_{dl}) = \begin{pmatrix} v^{(1)} \\ v^{(2)} \\ v^{(3)} \end{pmatrix}$$

モデル 0

式 1 $y_n \approx \sum_l \prod_{d=1}^{D} v_{dl}^{x_{nd}}$; はややあいまい
次のように書き直す:

$$y_n \sim \mathcal{N}\left(y_n \mid \sum_{l=1}^{L} \prod_{d=1}^{D} v_{dl}^{x_{nd}}, \lambda^{-1}\right) \tag{2}$$

$$v_{dl} \sim \mathcal{N}(v_{dl}|0, \tau^{-1}) \tag{3}$$

$$\lambda \sim \mathcal{G}(\lambda|a, b)$$

このモデルを修正・拡張する：

- 非負制約（解釈性）
- 分布を変える（マルチオミクス）

事前分布：非負制約

中間変数