

# ベイズ統計モデリングによるバイオインフォマティクス研究事例の紹介

～HP 公開版～

細田至温

早稲田大学理工学術院  
招聘研究員

September 20, 2025

# 目次

- 1 Umibato: estimation of time-varying microbial interaction using continuous-time regression hidden Markov model
- 2 BayesianSSA: a Bayesian statistical model based on structural sensitivity analysis for predicting responses to enzyme perturbations in metabolic networks
- 3 本日のまとめ

Umibato

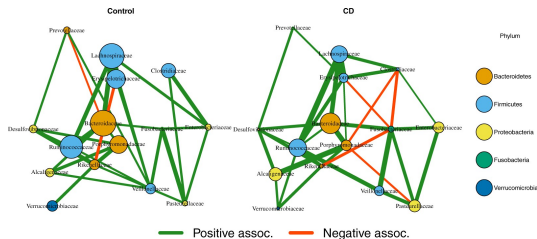
OXFORD

## 4 / 75

- 1 Umibato: estimation of time-varying microbial interaction using continuous-time regression hidden Markov model
  - 背景
  - 手法
  - 結果
  - まとめ
- 2 BayesianSSA: a Bayesian statistical model based on structural sensitivity analysis for predicting responses to enzyme perturbations in metabolic networks
- 3 本日のまとめ

## 微生物相互作用は重要な研究対象

例えば、微生物相互作用が疾患を持つ場合と持たない場合で異なることが示唆されている (McGregor, Labbe, and Greenwood, 2020)



微生物相互作用を微生物叢データから推定する手法:

- 相関ベースの手法
- 一般化ロトカ・ヴォルテラ方程式 (gLVE) ベースの手法





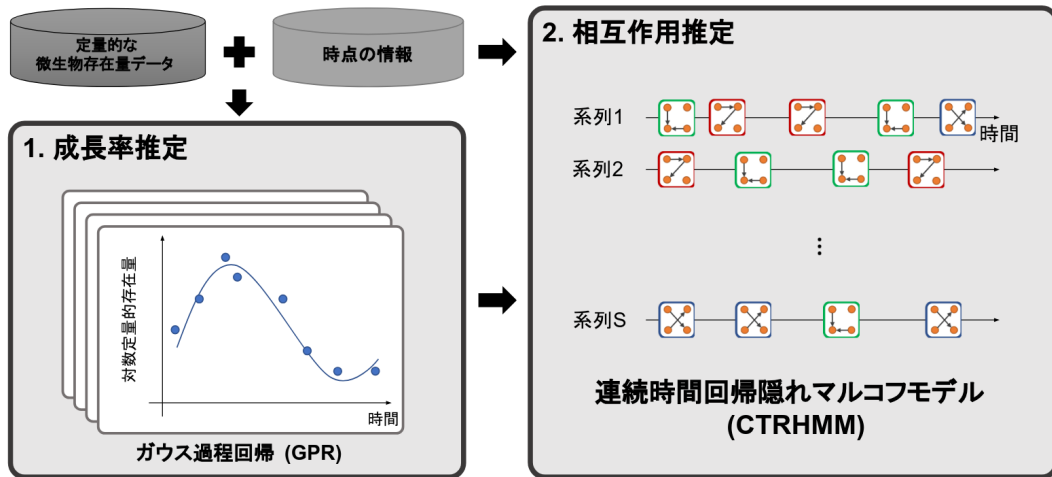




# Outline

- 1 Umibato: estimation of time-varying microbial interaction using continuous-time regression hidden Markov model
  - 背景
  - 手法
  - 結果
  - まとめ
- 2 BayesianSSA: a Bayesian statistical model based on structural sensitivity analysis for predicting responses to enzyme perturbations in metabolic networks
- 3 本日のまとめ

# Umibato 概観











- $z(t)$  を連続時間マルコフ連鎖でモデル化する
  - 現在の値のみに依存して未来の値が決まる
  - 短時間の経過であればあるほど状態が遷移しにくい
- 離散的に観測された  $x(t)$  から  $z(t)$  の値の系列を推定する
- 実用上は離散時間マルコフ連鎖と比較して以下の利点がある
  - 1 不均一なサンプリング間隔を自然に扱える
    - 離散時間マルコフ連鎖では欠損値を駆使しなければならない
  - 2 任意の時間が経過したときの遷移確率を計算できる



$$\mathbf{P}(\Delta t) = \exp(\mathbf{Q}\Delta t), \quad (5)$$

- ただし,  $\exp(\cdot)$  は行列指数関数,  $Q$  は遷移速度行列

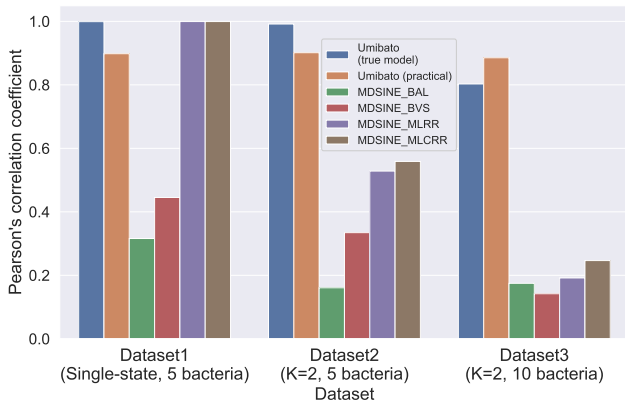
直感的には観測点間の確率過程を周辺化するような操作になる



---

- 1 Umibato: estimation of time-varying microbial interaction using continuous-time regression hidden Markov model
  - 背景
  - 手法
  - 結果
  - まとめ
- 2 BayesianSSA: a Bayesian statistical model based on structural sensitivity analysis for predicting responses to enzyme perturbations in metabolic networks
- 3 本日のまとめ

人工データ実験では Umibato の性能が従来の手法より良いことが示された



## 結果

## マウスの腸内細菌叢データを使って実データ実験を行った

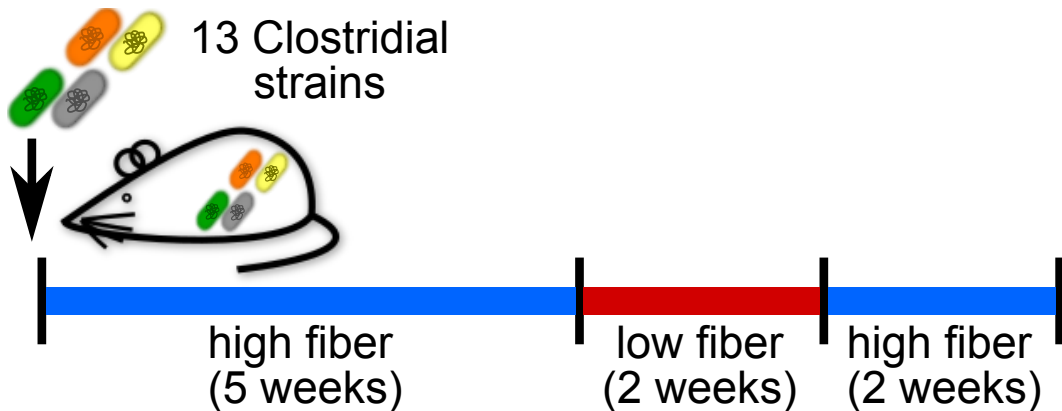
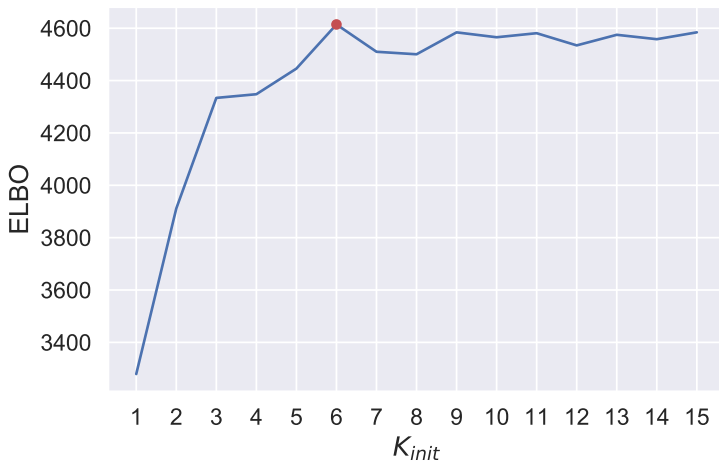
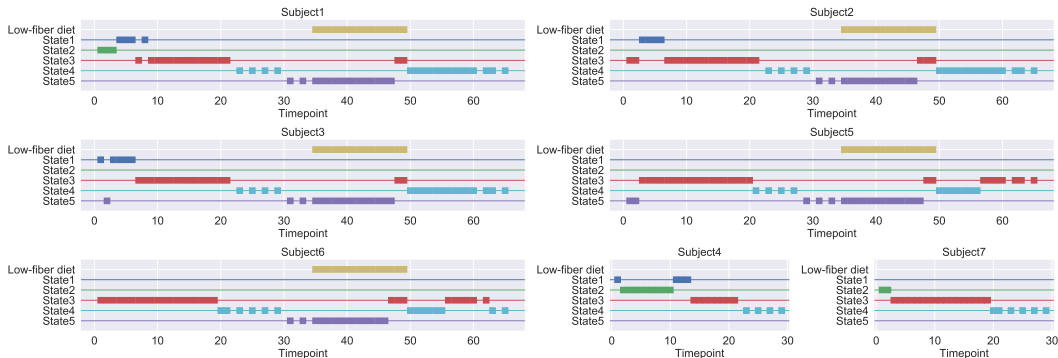


Figure: Bucci et al., 2016より引用

ELBO が最大を示すのは初期状態数 6 のときの試行であった



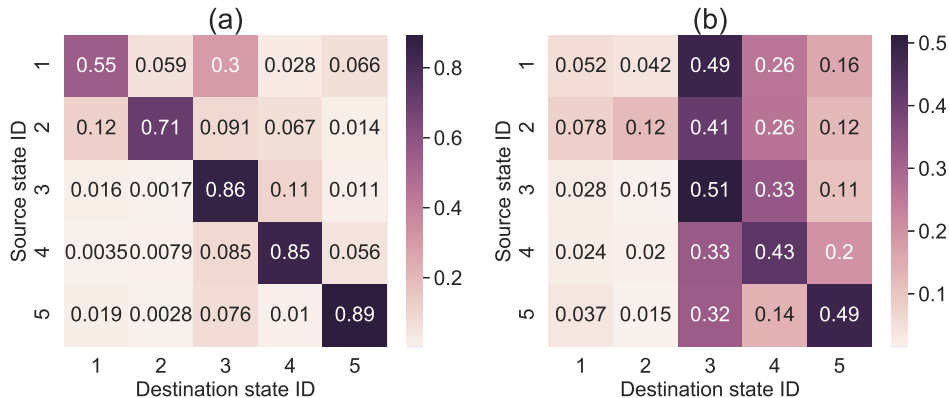
食事の変化に伴って相互作用が変化していることが示唆された



## 結果

## 状態 3, 4, 5 が主な状態であることが分かる

左が一日経過した場合の遷移確率行列，右が一週間経過した場合の遷移確率行列を示す．



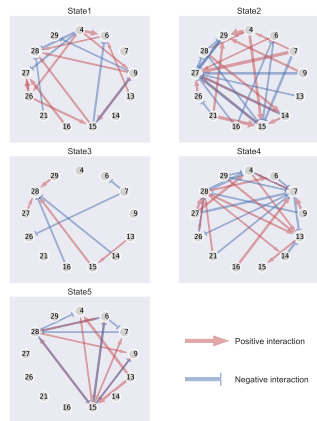


## 結果

# 微生物相互作用ネットワークは各状態で大きく異なる

右図は各状態に対応する相互作用ネットワーク

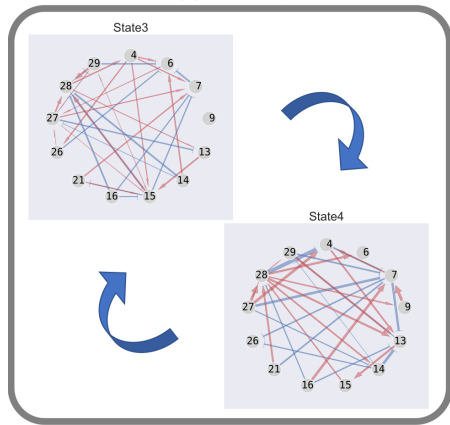
- 各円が微生物を示す
- 例えば寄生関係が確認できる
  - 寄生関係とは、ここでは A が B の増加に寄与し、B が A の減少に寄与するような A と B の関係を指す
- 特に状態 5 で多く見られる



## 結果

## ここまでの結果から示唆される微生物相互作用の動態の概念図

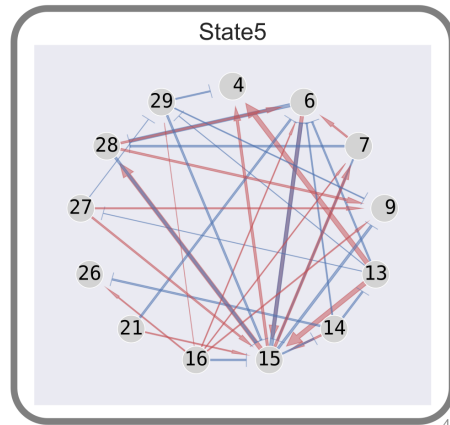
## 高繊維の食事



食事の変化

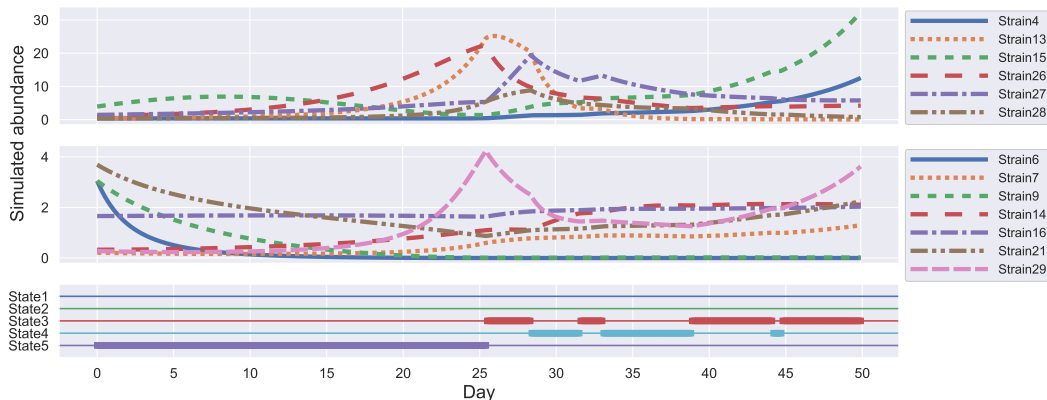


## 低繊維の食事



45

推定されたパラメタによるシミュレーションから、長期間の低繊維の食事は復元不可能な細菌叢多様性の減少を招くことが示唆



- 1 Umibato: estimation of time-varying microbial interaction using continuous-time regression hidden Markov model
  - 背景
  - 手法
  - 結果
  - まとめ
- 2 BayesianSSA: a Bayesian statistical model based on structural sensitivity analysis for predicting responses to enzyme perturbations in metabolic networks
- 3 本日のまとめ

# まとめ

Umibato というアルゴリズムを提案し時間変化する微生物相互作用を推定した

- 離散的に時間変化する変数を導入して時間変化する相互作用を表現した
- 食事の変化に伴って相互作用ネットワークが変化していた
- 同じ高繊維の食事でも二つの相互作用状態で構成されていることが示唆された

# BayesianSSA

## Open Access



BayesianSSA: a Bayesian statistical model based on structural sensitivity analysis for predicting responses to enzyme perturbations in metabolic networks

Shion Hosoda<sup>1\*</sup>, Hisashi Iwata<sup>1</sup>, Takuya Miura<sup>1</sup>, Maiko Tanabe<sup>1</sup>, Takashi Okada<sup>2</sup>, Atsushi Mochizuki<sup>2</sup> and Miwa Sato<sup>1</sup>





# 生体内では複数の化学反応が代謝ネットワークを構成して働いている

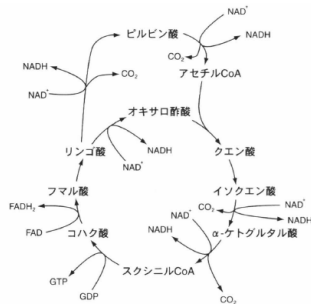


Figure: <https://photosyn.jp/pwiki/index.php?TCA%E5%9B%9E%E8%B7%AF>

# 代謝ネットワークにおける応答予測は重要な問題

反応を改変したときの代謝の流れ（フラックス）の変化を知りたいことがある。  
例えばバイオものづくりでは、どの遺伝子を過剰発現させれば生産量が増えるのかが分かれば効率を上げることができる。



Figure: Voigt, 2020より引用

実験すれば結果が分かるがコストがかかる → 予測できると嬉しい

- 代謝ネットワークのシミュレーションにより、化学反応を促進させたときの物質の生産量の変化を予測できる
- しかし、それには代謝ネットワークに含まれる全ての化学反応の反応速度を同定する必要がある
- 反応速度を同定することは、関数形及び関数のパラメタを同定することであり、それを化学反応の数だけ考慮すると膨大な空間を探索することになる

## 簡単に適用できる方法として SSA がある

一方で、近年開発された構造感度解析 (SSA: Structural sensitivity analysis) (Mochizuki and Fiedler, 2015) では、代謝ネットワークの構造情報のみから定性的な応答予測を行える

シミュレーションベースの方法と異なるのは、

- 量の情報は分らない
- 反応速度の関数系やパラメタが不要

という点で、適用範囲が広い



本研究では実験結果を用いて SSA に環境情報を統合することを目指す

SSA の予測に実験結果から抽出した環境情報を統合するベイズ統計モデル, BayesianSSA を開発した

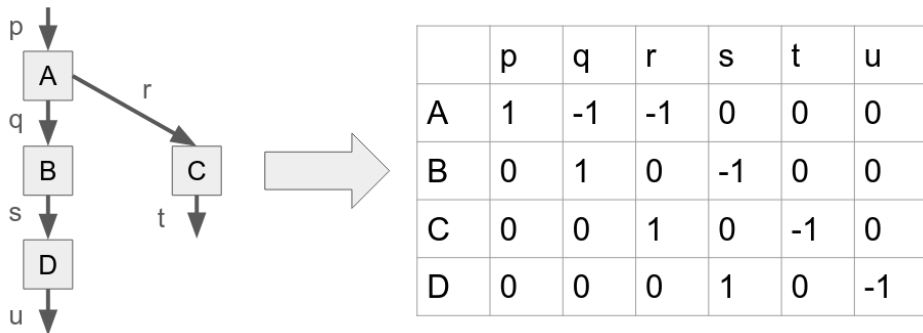
このモデルは以下のような特徴を持つ

- 摂動実験（遺伝子の過剰発現などにより反応を促進して物質の生産量を測定する）による実験結果をデータとして用いる
- モデル内に SSA のアルゴリズムを関数として含む
- 事後分布により計算される正值性確信度により確率的な予測が可能
- 予測分布により実験結果の予測も可能



# 化学量論行列

代謝ネットワークは化学量論行列  $\nu$  と呼ばれる行列で表現される  
 $\nu_{m,j}$  は反応  $j$  により起こる代謝物  $m$  の増減を示す（図では 1 または -1 のみだが化学反応式によっては他の整数もありうる）





# 代謝ネットワークの ODE

代謝ネットワークにおける代謝物濃度のシステムは以下で表される:

$$\frac{dx_m}{dt} = \sum_{j=1}^J \nu_{m,j} F_j(k_j, \mathbf{x}), \quad (6)$$

$x_m$ : 代謝物  $m$  の濃度

$\nu_{m,j}$ : 化学量論行列の  $(m, j)$  要素

$F_j(\cdot)$ : 反応速度関数 (反応速度定数  $k_j$  とシステム内全ての代謝物の濃度  $\mathbf{x}$  に依存)









## 構造感度解析 (SSA : Structural sensitivity analysis) 3/8

右辺第一項をテイラー展開して,

$$\delta F_j = \frac{\partial F_j}{\partial k_j}(k_j, \bar{\mathbf{x}}) \delta k_{j^*} \delta_{j^*, j} + \sum_{m=1}^M \frac{\partial F_j}{\partial x_m}(k_j, \bar{\mathbf{x}}) \delta x_m. \quad (8)$$

ここで, 定常状態において  $\sum_j \nu_{m,j} F_j = 0$  がどの  $m$  についても成立するので,

$$\delta F_j = \sum_{k=1}^K \delta \mu_k c_{k,j}, \quad (9)$$

が成立する. これを(8)に代入して,

$$\sum_{m=1}^M \frac{\partial F_j}{\partial x_m}(k_j, \bar{\mathbf{x}}) \delta x_m - \sum_{k=1}^K \delta \mu_k c_{k,j} = -\frac{\partial F_j}{\partial k_j}(k_j, \bar{\mathbf{x}}) \delta k_{j^*} \delta_{j^*, j}. \quad (10)$$

## 構造感度解析 (SSA : Structural sensitivity analysis) 4/8

簡単のため,  $r_{j,m} := \frac{\partial F_j}{\partial x_m}(k_j, \bar{\mathbf{x}})$  とおくと,

$$\sum_{m=1}^M r_{j,m} \delta x_m - \sum_{k=1}^K \delta c_{k,j} \mu_k = -\frac{\partial F_j}{\partial k_j}(k_j, \bar{\mathbf{x}}) \delta k_{j^*} \delta_{j^*,j}. \quad (11)$$

また, 定義から, 保存量  $\lambda_l$  は

$$\lambda_l := \sum_{m=1}^M d_{l,m} x_m, \quad (12)$$

であり,  $\lambda_l$  の初期値にのみ依存し, どの時間においても成り立つことから,

$$\sum_{m=1}^M d_{l,m} \delta x_m = 0. \quad (13)$$

## 構造感度解析 (SSA : Structural sensitivity analysis) 5/8

$\mathbf{A}$  行列に  $\delta \mathbf{x}$  と  $\delta \boldsymbol{\mu}$  を縦に並べたベクトルを右からかけると,

$$\mathbf{A}(\mathbf{r}) \begin{pmatrix} \delta \mathbf{x} \\ \delta \boldsymbol{\mu} \end{pmatrix} = \left( \begin{array}{c|c} \mathbf{R}(\mathbf{r}) & -\mathbf{C} \\ \hline -\mathbf{D}^T & \mathbf{0}_{L \times K} \end{array} \right) \begin{pmatrix} \delta \mathbf{x} \\ \delta \boldsymbol{\mu} \end{pmatrix} \quad (14)$$

$$= \mathbf{R}(\mathbf{r})\delta \mathbf{x} - \mathbf{C}\delta \boldsymbol{\mu} - \mathbf{D}^T\delta \mathbf{x} + \mathbf{0}, \quad (15)$$

(11)および (13)から,

$$\mathbf{A}(\mathbf{r}) \begin{pmatrix} \delta \mathbf{x} \\ \delta \boldsymbol{\mu} \end{pmatrix} = -\mathbf{e}_{j^*}. \quad (16)$$

ここで,  $\mathbf{e}_{j^*} \in \mathbb{R}^{J+L}$  は,  $j^*$  番目の成分が  $\delta F_{j^*}^{(k)} = \frac{\partial F_{j^*}}{\partial k_{j^*}}(k_{j^*}, \bar{\mathbf{x}})\delta k_{j^*}$  で, それ以外が 0 であるベクトルである.



## 構造感度解析（SSA : Structural sensitivity analysis） 6/8

両辺を  $\delta F_{j^*}^{(k)}$  で割ると,

$$\mathbf{A}(\mathbf{r}) \left( \frac{\delta \mathbf{x} / \delta F_{j^*}^{(k)}}{\delta \boldsymbol{\mu} / \delta F_{j^*}^{(k)}} \right) = -\mathbf{h}_{j^*}, \quad (17)$$

と表現できる. ここで,  $\mathbf{h}_{j^*}$  は  $j^*$  番目の要素が 1 でそれ以外の要素が 0 のベクトルである.



## 構造感度解析（SSA : Structural sensitivity analysis） 8/8

以上より，摂動する反応または保存量  $j^*$  すべてについて考えて， $\mathbf{A}(\mathbf{r})$  が正則とすると，

$$\left( \frac{\delta \tilde{\mathbf{X}}}{\delta \tilde{\mathbf{M}}} \right) = -\mathbf{A}(\mathbf{r})^{-1} \mathbf{I} = \mathbf{S}(\mathbf{r}). \quad (21)$$

$F_j(\cdot)$  が  $k_j$  について単調増加であることから  $\delta F_{j^*}^{(k)} > 0$  なので，

$$\text{sign} \left( \left( \frac{\delta \mathbf{X}}{\delta \mathbf{M}} \right) \right) = \text{sign}(\mathbf{S}(\mathbf{r})). \quad (22)$$

$\delta k_{j^*}$  及び  $\delta \lambda_{j^*}$  は微小であるが， $\mathbf{r}$  に符号が依存しない要素については，微小でない場合においても微小な変化の積分と考えると成立することが分かる。  $\square$



# SSA による定性的予測

計算機上は SSA を用いてどのように定性的予測を得るのか？

$A(r)$  及び  $S(r)$  は  $r$  に依存するが、その符号が  $r$  に依存せず一定であれば、それが定性的予測になる。この条件を満たすか判定する方法としては以下の2つ。

- 代数的計算： $r$  を変数としたまま  $q_{m,j}(r)$  を計算し、 $r$  の符号から  $q_{m,j}(r)$  ( $Q(r)$  の  $(m,j)$  成分) の符号を決定する
- 数值的計算： $r$  の乱数を取ることで、 $q_{m,j}(r)$  を計算し、 $r$  の符号から  $q_{m,j}(r)$  の符号を決定する

いずれの場合でも、 $r_{j,m} := \frac{\partial F_j}{\partial x_m}(k_j, \bar{x})$  の符号は典型的には正（代謝物濃度の反応への依存関係は、多くが化学反応と基質の関係によるものであるため）なので、この情報から  $q_{m,j}(r)$  の符号を求める。

一方で、アロステリック制御のようなケースでは  $r_{j,m}$  の符号は負として扱う。

# SSA による定性的予測

具体例を示す．  $r_1 > 0, r_2 > 0$  として，

- 1  $S_{m,j} = r_1 + r_2 \Rightarrow q_{m,j}(\mathbf{r}) = 1$
- 2  $S_{m,j} = -r_1 - r_2 \Rightarrow q_{m,j}(\mathbf{r}) = -1$
- 3  $S_{m,j} = r_1 - r_2 \Rightarrow q_{m,j}(\mathbf{r})$  は不定

BayesianSSA は，3 のケースに対して  $\mathbf{r}$  を確率変数とみなすことで対処する手法である．

- これは，（主に  $\{k_j\}_j$  を介して依存する）環境条件の確率を考えて特定の条件にフォーカスすることに該当する

# SSA による定性的予測

摂動する対象である反応・保存量と、観測する対象である代謝物濃度・反応フラックスをそれぞれ総称するため、以下の語句を定義する。

## 摂動対象

代謝ネットワークにおける化学反応及び保存量を**摂動対象**と呼ぶ。

## 観測対象

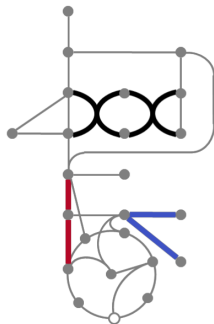
代謝ネットワークにおける代謝物濃度及び反応フラックスを**観測対象**と呼ぶ。

- 1 Umibato: estimation of time-varying microbial interaction using continuous-time regression hidden Markov model
- 2 BayesianSSA: a Bayesian statistical model based on structural sensitivity analysis for predicting responses to enzyme perturbations in metabolic networks
  - 背景
  - 理論的背景
  - 手法
  - 結果
  - まとめ
- 3 本日のまとめ



# BayesianSSA 概観

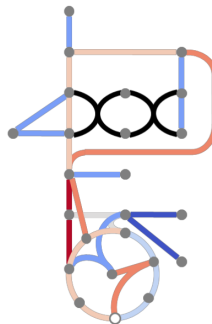
### General consequences by original SSA



- Target chemical
- Positive
- Negative
- Zero
- Indefinite

The predictive response of the target chemical flux

### Specific consequences



**BayesianSSA makes  
indefinite predictions  
interpretable stochastically**











## 使用するデータ

- Toya, Shiraki, and Shimizu, 2015による大腸菌中央代謝系のネットワークを対象とした.
- 生物学的実験により, 摂動対象に対応する, 大腸菌の特定の遺伝子を過剰発現させたデータを使用.
- 観測対象はコハク酸排出フラックス

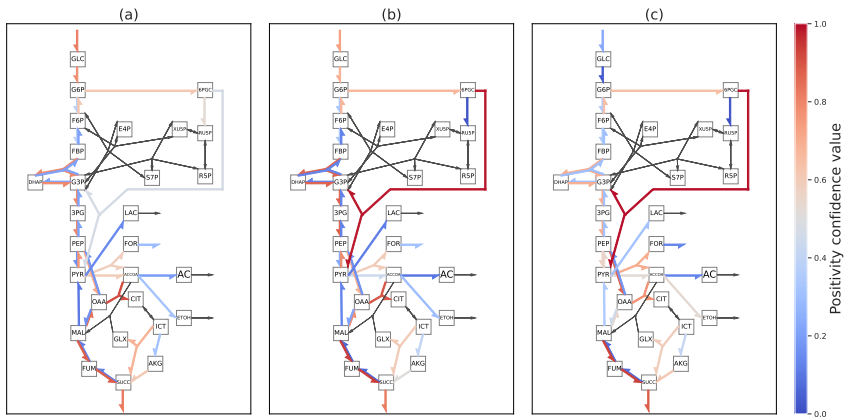
- 1 Umibato: estimation of time-varying microbial interaction using continuous-time regression hidden Markov model
- 2 BayesianSSA: a Bayesian statistical model based on structural sensitivity analysis for predicting responses to enzyme perturbations in metabolic networks
  - 背景
  - 理論的背景
  - 手法
  - **結果**
  - まとめ
- 3 本日のまとめ



## 結果

## 疑似データによる推定結果の変化

文献に基づいてトイデータを与えてみると、正值性確信度が変化するのが見て取れる.



## 結果

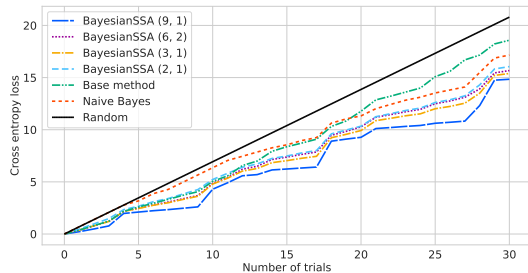
## 予測精度では BayesianSSA が最も良かった

データへのフィットをしない

BayesianSSA として base method, データへのフィットのみをする統計モデルとして naive Bayes を比較対象とした. 以下の結論を得た:

- データ使う必要は? → base method より良いので必要ある
- SSA 使う必要は? → naive Bayes より良いので必要ある

→ BayesianSSA によりデータと SSA を統合した意味があった

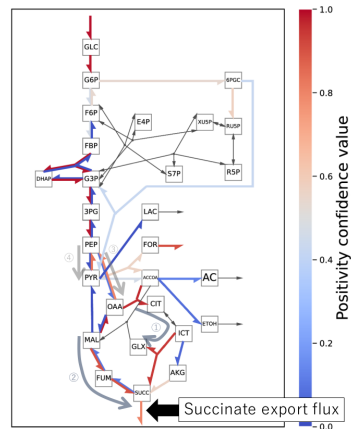


## 結果

## 最終的に得られた事後分布は一部解釈可能だった

最終的に得られた事後分布で、正值性確信度を計算、  
解釈性を検証した

文献	反応	一致
Vemuri, Eiteman, and Altman, 2002	① Glyoxylate pathway	✓
van Heerden and Nicol, 2013		
Vemuri, Eiteman, and Altman, 2002	② Reductive pathway	✓
van Heerden and Nicol, 2013		
Millard et al., 1996	③ PPC	×
Chatterjee et al., 2001; Liang et al., 2015	④ PTS	×

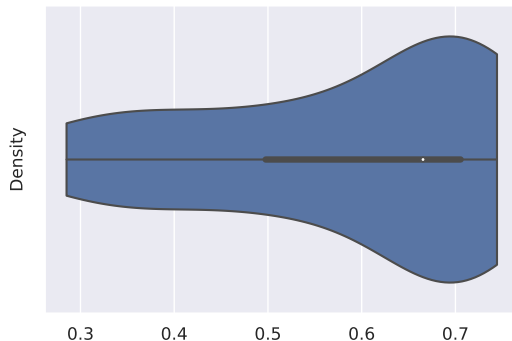


## 結果

## Out-of-sample に対する予測はランダムより良かった

Naive Bayes との比較から推測すると, BayesianSSA は新規の  $(m, j)$  に対しての予測にその他の結果を利用できている

そこで, フィットさせたデータにない  $(m, j)$  の実験結果に対する予測精度を可視化した





## まとめ

- 代謝ネットワークにおける理論である SSA にデータを統合する統計モデル BayesianSSA を開発した
- 実データによる精度評価ではデータと理論をそれぞれ片方のみ用いる場合より性能が良く，手法の有効性が示された
- 最終的に得られた事後分布の解釈は過去の文献と部分的に一致，部分的に不一致であった。

## 本日のまとめ

# ベイズ統計モデルによる柔軟なモデリング

本ご紹介した2つの研究はどちらもベイズ統計モデルの開発を主軸にしています。




- Umibato : 時間変化する確率変数を導入
- BayesianSSA : 特定の環境での関数の値を示す確率変数を導入

このような手法の開発に興味があれば、ぜひお声がけください！






ご清聴ありがとうございました.




# Reference I

-  Bucci, Vanni et al. (June 2016). “MDSINE: Microbial Dynamical Systems INference Engine for Microbiome Time-Series Analyses”. In: **Genome Biology** 17.1, p. 121. ISSN: 1474-760X. DOI: 10.1186/s13059-016-0980-6.
-  Chatterjee, Ranjini et al. (Jan. 2001). “Mutation of the ptsG Gene Results in Increased Production of Succinate in Fermentation of Glucose by Escherichia Coli”. In: **Applied and Environmental Microbiology** 67.1, pp. 148–154. DOI: 10.1128/AEM.67.1.148-154.2001. (Visited on 11/17/2023).
-  Embree, Mallory et al. (Dec. 2015). “Networks of Energetic and Metabolic Interactions Define Dynamics in Microbial Communities”. In: **Proceedings of the National Academy of Sciences of the United States of America** 112.50, pp. 15450–15455. ISSN: 1091-6490. DOI: 10.1073/pnas.1506034112.




## Reference II

-  Liang, Quanfeng et al. (Aug. 2015). “Comparison of Individual Component Deletions in a Glucose-Specific Phosphotransferase System Revealed Their Different Applications”. In: **Scientific Reports** 5.1, p. 13200. ISSN: 2045-2322. DOI: 10.1038/srep13200. (Visited on 06/23/2023).
-  McGregor, Kevin, Aurélie Labbe, and Celia M T Greenwood (Mar. 2020). “MDiNE: A Model to Estimate Differential Co-Occurrence Networks in Microbiome Studies”. In: **Bioinformatics** 36.6, pp. 1840–1847. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btz824.
-  Millard, C S et al. (May 1996). “Enhanced Production of Succinic Acid by Overexpression of Phosphoenolpyruvate Carboxylase in Escherichia Coli”. In: **Applied and Environmental Microbiology** 62.5, pp. 1808–1810. DOI: 10.1128/aem.62.5.1808-1810.1996. (Visited on 06/23/2023).

## Reference III

-  Mochizuki, Atsushi and Bernold Fiedler (Feb. 2015). “Sensitivity of Chemical Reaction Networks: A Structural Approach. 1. Examples and the Carbon Metabolic Network”. In: **Journal of Theoretical Biology** 367, pp. 189–202. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2014.10.025. (Visited on 04/21/2023).
-  Okada, Takashi and Atsushi Mochizuki (Aug. 2017). “Sensitivity and Network Topology in Chemical Reaction Systems”. In: **Physical Review E** 96.2, p. 022322. DOI: 10.1103/PhysRevE.96.022322. (Visited on 04/21/2023).
-  Toya, Yoshihiro, Takanori Shiraki, and Hiroshi Shimizu (2015). “SSDesign: Computational Metabolic Pathway Design Based on Flux Variability Using Elementary Flux Modes”. In: **Biotechnology and Bioengineering** 112.4, pp. 759–768. ISSN: 1097-0290. DOI: 10.1002/bit.25498. (Visited on 04/21/2023).

## Reference IV

-  van Heerden, Carel D. and Willie Nicol (Sept. 2013). “Continuous and Batch Cultures of Escherichia Coli KJ134 for Succinic Acid Fermentation: Metabolic Flux Distributions and Production Characteristics”. In: **Microbial Cell Factories** 12.1, p. 80. ISSN: 1475-2859. DOI: 10.1186/1475-2859-12-80. (Visited on 11/16/2023).
-  Vemuri, G. N., M. A. Eiteman, and E. Altman (Apr. 2002). “Effects of Growth Mode and Pyruvate Carboxylase on Succinic Acid Production by Metabolically Engineered Strains of Escherichia Coli”. In: **Applied and Environmental Microbiology** 68.4, p. 1715. DOI: 10.1128/AEM.68.4.1715-1727.2002. (Visited on 11/21/2023).
-  Voigt, Christopher A. (Dec. 2020). “Synthetic Biology 2020–2030: Six Commercially-Available Products That Are Changing Our World”. In: **Nature Communications** 11.1, p. 6379. ISSN: 2041-1723. DOI: 10.1038/s41467-020-20122-2. (Visited on 06/14/2023).