データ生成観測メカニズムの確率モデルからの機械学習法

CDS 早稲田大学 データ科学センター

概要と注意点

- 意思決定写像[松嶋ら，2022]によるデータ科学の体系化を応用した研究事例
- 授業や教科書に載るかは未定（おそらく載らない）
- 以降に登場する用語はまだ定着していないものを含む

目次

- 目的：予測問題
- 設定：関数木とモデルツリー
- 評価基準：ベイズリスク関数
- ベイズ基準のもと最適な予測とその効率的計算
  - メタツリーを用いたモデルツリーの総和計算
  - 特徴量割当ベクトルに関する総和計算
- 実験結果
- まとめ

目的：予測問題

- 予測問題とは

説明変数または特徴量　　　　目的変数

性別	クーポンの配布	購買額
男	あり	1000
女	なし	2000
⋮	⋮	⋮
男	あり	1500
女	なし	?

本研究では
説明変数はカテゴリカル

連続量→回帰

目的：予測問題

- 予測問題とは

説明変数または特徴量　　目的変数

性別	クーポンの配布	購買
男	あり	あり
女	なし	あり
⋮	⋮	⋮
男	あり	なし
女	なし	？

本研究では
説明変数はカテゴリカル

カテゴリカル→分類

目的：予測問題

- 予測問題とは

説明変数または特徴量 : $\boldsymbol{x}$ 　　目的変数 : y

	x_1	x_2	y
$\{\boldsymbol{x}_i, y_i\}_{i=1}^n =: \mathcal{D}$	x_{11}	x_{12}	y_1
	x_{21}	x_{22}	y_2
	$\vdots$	$\vdots$	$\vdots$
	x_{n1}	x_{n2}	y_n
$\boldsymbol{x}_{n+1}$	$x_{n+1,1}$	$x_{n+1,2}$	$y_{n+1} = ?$

過去の説明変数と目的変数の組$\mathcal{D} = \{\boldsymbol{x}_i, y_i\}_{i=1}^n$から
新規説明変数$\boldsymbol{x}_{n+1}$に対応する未知目的変数y_{n+1}を予測する

意思決定写像[松嶋ら，2022]としての整理

目的： $\boldsymbol{x}_{n+1}$に対応した$\underset{\sim}{y}_{n+1}$を予測したい.

設定：

評価基準：

データ集合 $\mathcal{X}^{n+1} \times \mathcal{Y}^n$

$\{\boldsymbol{x}_i, \underset{\sim}{y}_i\}_{i=1}^{n}, \boldsymbol{x}_{n+1}$

→ **意思決定写像** →

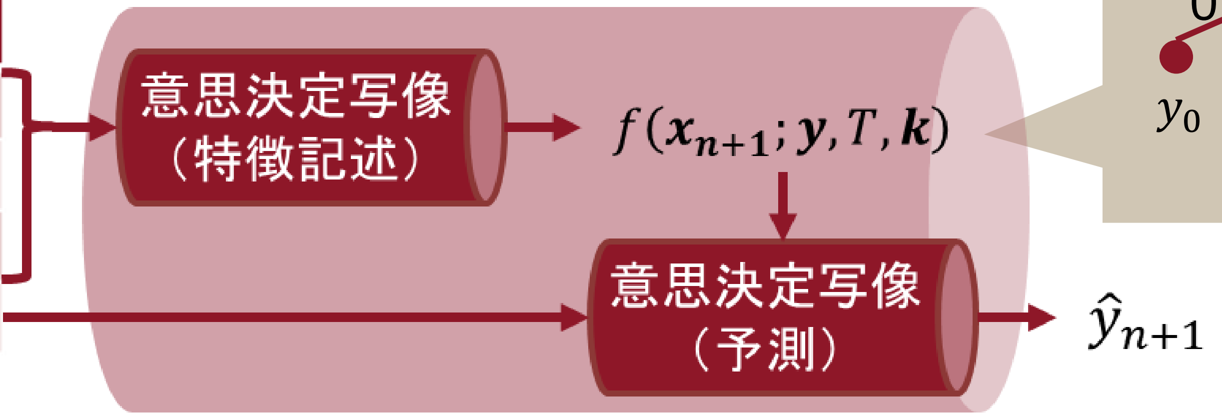
決定集合 $\mathcal{Y}$

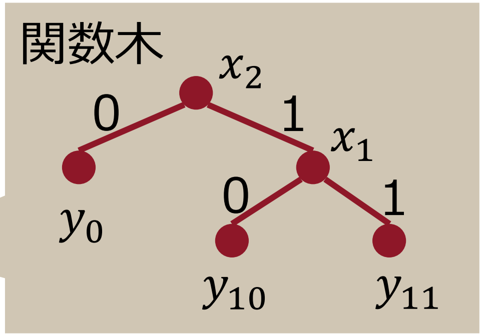
$\hat{\underset{\sim}{y}}_{n+1}$

関数木とモデルツリー

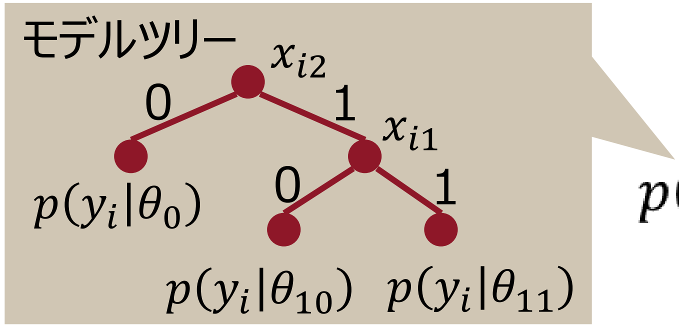
- 関数木による予測（例，CART[Breiman et al, 1984]）

x_1	x_2	y
0	1	0
1	1	1
⋮	⋮	⋮
0	1	?





- モデルツリーのもとでの予測（例，[須子ら，2003]）

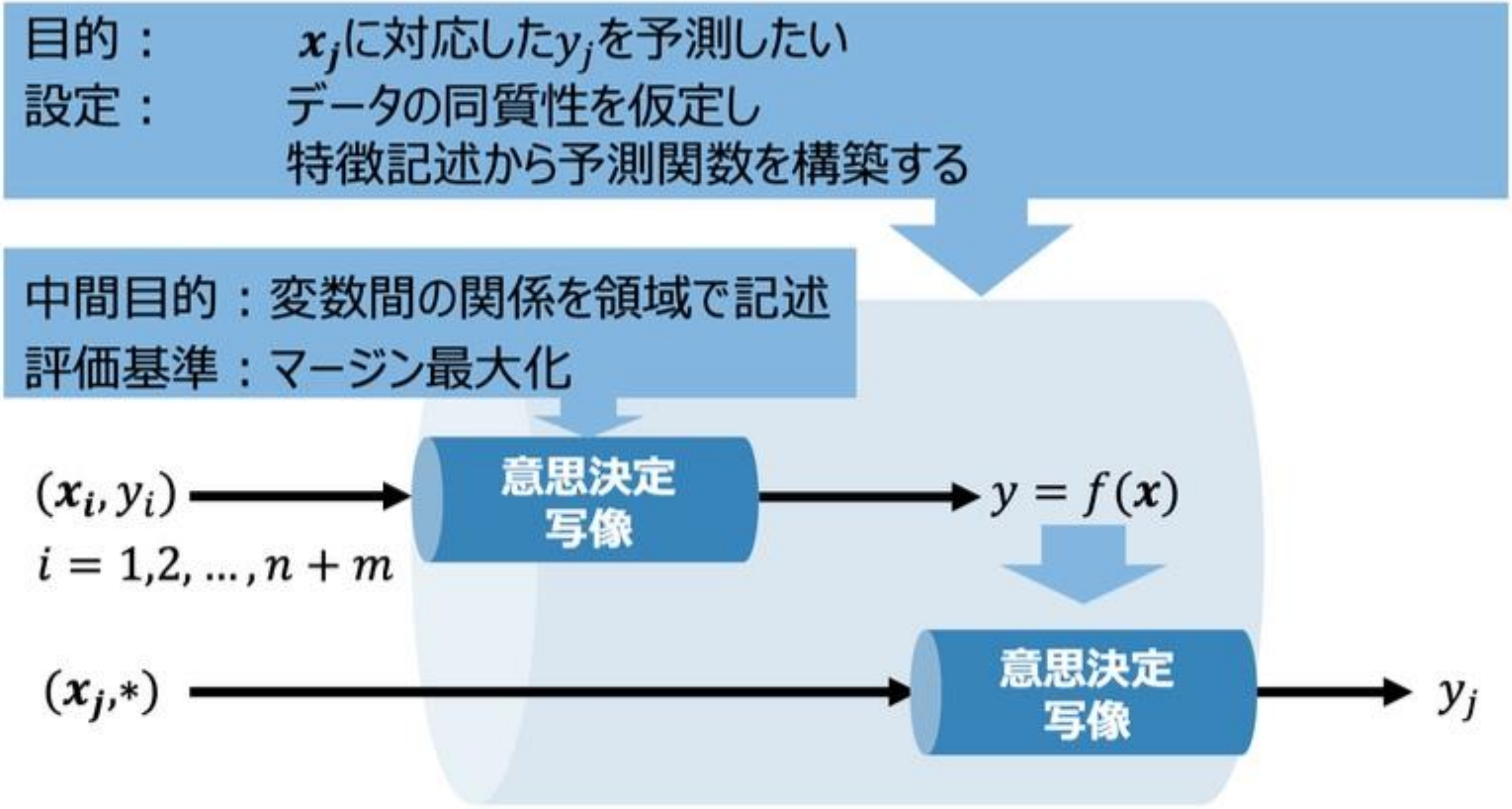


$p(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}, T, \boldsymbol{k})$

x_1	x_2	y
0	1	0
1	1	1
⋮	⋮	⋮
0	1	?

予測としての分類問題の例

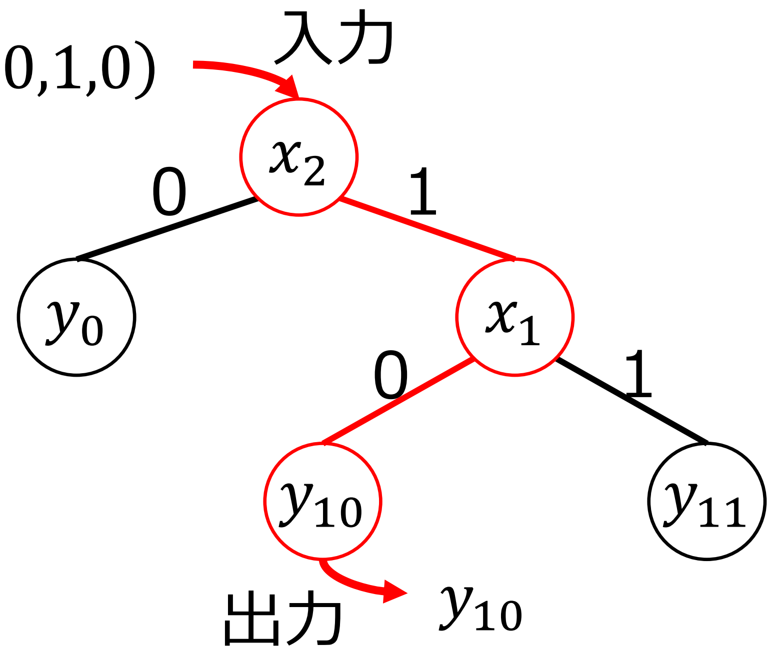
[データの同質性（新語）] 学習データと予測の対象となる新規データは同じメカニズムで発生する



関数木とモデルツリー

- 関数木（例，CART[Breiman et al, 1984]）

$(x_1, x_2, x_3) = (0,1,0)$



この関数を特徴づけるのは．．．

$\boldsymbol{k} \in \mathcal{K}$：各内部ノードに割り当てられた説明変数の添字列（特徴量割当ベクトル）

$T \in \mathcal{T}$：木の形

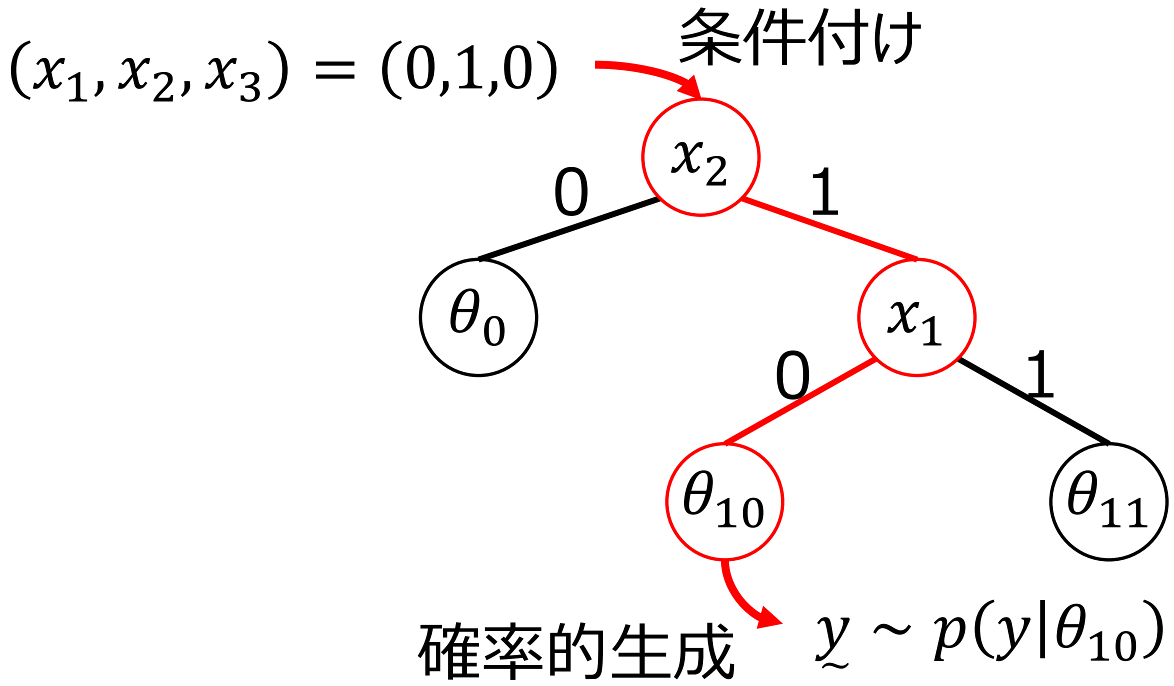
$\boldsymbol{y} \in \mathcal{Y}$：葉ノードに割り当てられた予測値

これらを学習データDの特徴を記述するようあらかじめ調整

- 代表的な性能改善手法（例，Random Forest[Breiman, 2001]）
  - 複数の関数木の構築・利用：$\frac{1}{L}\sum_{l=1}^{L} f(\boldsymbol{x}_{n+1}; \boldsymbol{\theta}_l, T_l, \boldsymbol{k}_l)$

関数木とモデルツリー

- モデルツリー



この分布を特徴づけるのは．．．

$\boldsymbol{k} \in \mathcal{K}$：各内部ノードに割り当てられた説明変数の添字列（特徴量割当ベクトル）

$T \in \mathcal{T}$：木の形

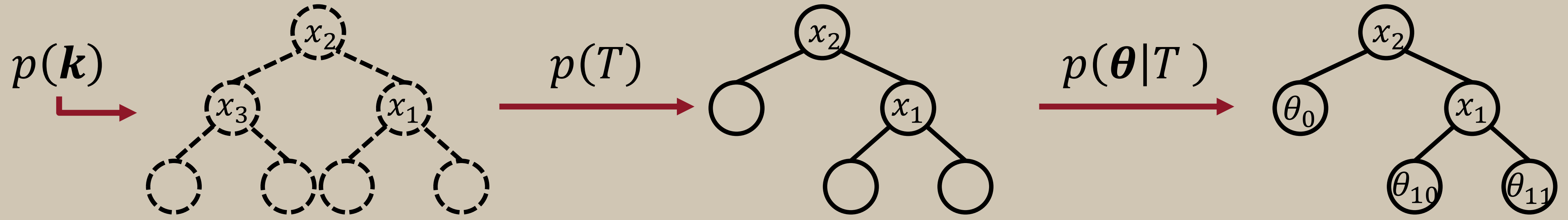
$\boldsymbol{\theta} \in \boldsymbol{\Theta}$：葉ノードに割り当てられた生成分布パラメータ

- 確率モデルの仮定を利用することで，予測の統計的性質の議論が可能に.
- 今回は特にベイズ基準のもとでの良さを議論→事前分布$p(\boldsymbol{\theta}, T, \boldsymbol{k})$を仮定.

関数木とモデルツリー

●モデルツリーの事前分布

事前分布 $p(\boldsymbol{\theta}, T, \boldsymbol{k})$



$\boldsymbol{k} \in \mathcal{K}$ ：特徴量割当ベクトル，$T \in \mathcal{T}$ ：木の形，$\boldsymbol{\theta} \in \boldsymbol{\Theta}$ ：生成分布パラメータ

$p(y_i | \boldsymbol{x}_i, \boldsymbol{\theta}, T, \boldsymbol{k})$

$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{y}$
0	1	0
1	1	1
⋮	⋮	⋮
0	1	?

$\hat{y}_{n+1}$

以降，δで表す

意思決定写像[松嶋ら，2022]としての整理

目的： $\boldsymbol{x}_{n+1}$に対応した$\underset{\sim}{y}_{n+1}$を予測したい．
設定：i.i.d.に$\underset{\sim}{y}_i \sim p(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}, T, \boldsymbol{k})$（モデルツリーによるデータ生成観測メカニズム）
$\underset{\sim}{\boldsymbol{\theta}} \sim p(\boldsymbol{\theta}|T)$（yの共役事前分布）
$\underset{\sim}{T} \sim p(T)$（[Nakahara et al.,2022]の分布）
$\underset{\sim}{\boldsymbol{k}} \sim p(\boldsymbol{k})$（一様分布）
評価基準：

データ集合 $\mathcal{X}^{n+1} \times \mathcal{Y}^n$ → **意思決定写像** → 決定集合 $\mathcal{Y}$

$\{\boldsymbol{x}_i, \underset{\sim}{y}_i\}_{i=1}^n, \boldsymbol{x}_{n+1}$ → $\hat{\underset{\sim}{y}}_{n+1}$

ベイズ基準のもと最適な予測

- 予測の評価基準（ベイズ危険関数$BR(\delta)$）
  - 損失関数

0-1誤差，二乗誤差など

$$L\left((\boldsymbol{\theta}, T, \boldsymbol{k}), \delta(\mathcal{D}, x_{n+1})\right) = \sum_{y_{n+1}} p(y_{n+1} | \boldsymbol{x}_{n+1}, \boldsymbol{\theta}, T, \boldsymbol{k}) \ell\left(y_{n+1}, \delta(\mathcal{D}, x_{n+1})\right)$$

  - ベイズ危険関数

$$BR(\delta) = \sum_{\boldsymbol{\theta}, T, \boldsymbol{k}} p(\boldsymbol{\theta}, T, \boldsymbol{k}) \sum_{\mathcal{D}, x_{n+1}} p(\mathcal{D}, x_{n+1} | \boldsymbol{\theta}, T, \boldsymbol{k}) \, L\left((\boldsymbol{\theta}, T, \boldsymbol{k}), \delta(\mathcal{D}, x_{n+1})\right)$$

→ベイズ危険関数を最小化する予測を行う

予測の良し悪しを統計的に評価できる

意思決定写像[松嶋ら，2022]としての整理

目的： $\boldsymbol{x}_{n+1}$に対応した$\underset{\sim}{y}_{n+1}$を予測したい.

設定：i.i.d.に$\underset{\sim}{y}_i \sim p(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}, T, \boldsymbol{k})$（モデルツリーによるデータ生成観測メカニズム）

$\underset{\sim}{\boldsymbol{\theta}} \sim p(\tilde{\boldsymbol{\theta}}|T)$（$y$の共役事前分布）

$\underset{\sim}{T} \sim p(T)$（[Nakahara et al.,2022]の分布）

$\underset{\sim}{\boldsymbol{k}} \sim p(\boldsymbol{k})$（一様分布）

評価基準：ベイズ危険関数$BR(\delta)$の最小化

データ集合 $\mathcal{X}^{n+1} \times \mathcal{Y}^n$

$\{\boldsymbol{x}_i, \underset{\sim}{y}_i\}_{i=1}^n, \boldsymbol{x}_{n+1}$

→ **意思決定写像** →

決定集合 $\mathcal{Y}$

$\underset{\sim}{\hat{y}}_{n+1}$

ベイズ基準のもと最適な予測

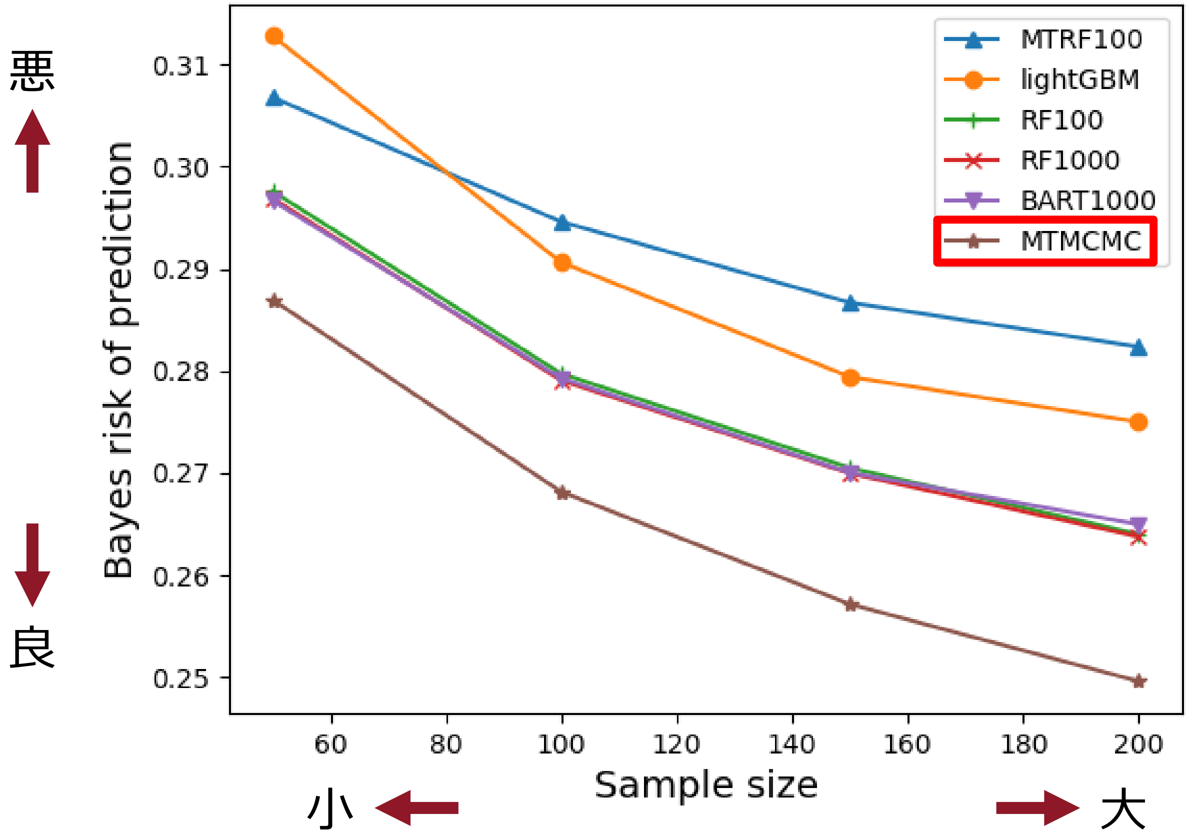
- ベイズ基準のもと最適な予測δ^*
  - 例，0-1損失関数に基づくベイズリスク関数を評価基準とする場合

$$
\begin{aligned}
&\delta^*(\mathcal{D}, \boldsymbol{x}_{n+1}) \\
&= \arg\max_{y_{n+1}} p(y_{n+1} | \boldsymbol{x}_{n+1}, \mathcal{D}) \\
&= \arg\max_{y_{n+1}} \sum_{\boldsymbol{k} \in \mathcal{K}} \sum_{T \in \mathcal{T}} \int \underbrace{p(y_{n+1} | \boldsymbol{x}_{n+1}, \boldsymbol{\theta}, T, \boldsymbol{k})}_{\text{モデルツリー}} \underbrace{p(\boldsymbol{\theta}, T, \boldsymbol{k} | \mathcal{D})}_{\text{事後分布}} \mathrm{d}\boldsymbol{\theta}
\end{aligned}
$$

  - あらゆるモデルツリーの事後分布による重みつき和計算が理論上最適
  - ただし，計算量が膨大（$\boldsymbol{k}, T$の和の範囲が問題サイズに対し指数的に増加）

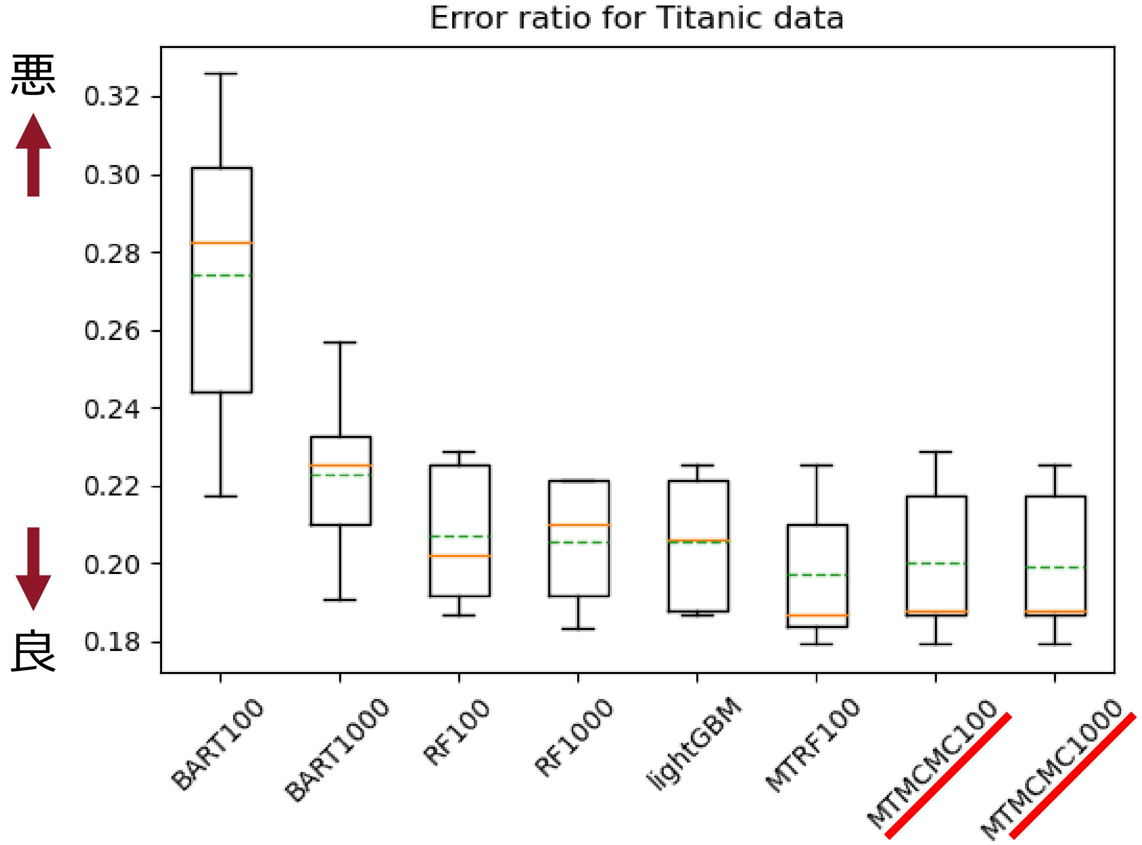
人工データ実験結果

- 人工データで平均予測誤り確率（ベイズリスクの近似値）を確認
  - $\mathcal{X} = \mathcal{Y} = \{0,1\}$
  - $K = 20, D_{\max} = 10$
  - $\boldsymbol{k}, T$：100回生成
  - $\boldsymbol{\theta}, \boldsymbol{x}, y$：100回生成
  - バーンイン期間：500
  - MCMCサンプルサイズ：1000
  - 比較手法
    - Random Forest[Breiman, 2001]
    - LightGBM[Ke et al, 2017]
    - BART[Chipman et al., 2010]
    - メタツリーランダムフォレスト [Dobashi et al., 2010]



実データ実験結果

- Titanicデータでの予測実験
  - 説明変数
    - タイタニック号の乗客の属性
  - 目的変数
    - 生存 / 死亡
  - 実験方法
    - 連続変数は5段階に量子化
      →全変数をOne-hot表現に
    - 5-foldクロスバリデーション
  - 実験結果
    - 5回分の平均誤り率の箱ひげ図



ベイズ基準のもと最適な予測の効率的計算

- ベイズ基準のもと最適な予測を効率的に計算するには？

$$
\begin{aligned}
&\delta^*(\mathcal{D}, x_{n+1}) \\
&= \arg\max_{y_{n+1}} p(y_{n+1} | \boldsymbol{x}_{n+1}, \mathcal{D}) \\
&= \arg\max_{y_{n+1}} \sum_{\boldsymbol{k} \in \mathcal{K}} \sum_{T \in \mathcal{T}} \int p(y_{n+1} | \boldsymbol{x}_{n+1}, \boldsymbol{\theta}, T, \boldsymbol{k})\, p(\boldsymbol{\theta}, T, \boldsymbol{k} | \mathcal{D}) \mathrm{d}\boldsymbol{\theta} \\
&= \arg\max_{y_{n+1}} \sum_{\boldsymbol{k} \in \mathcal{K}} p(\boldsymbol{k} | \mathcal{D}) \sum_{T \in \mathcal{T}} p(T | \mathcal{D}, \boldsymbol{k}) \underbrace{\int p(y_{n+1} | \boldsymbol{x}_{n+1}, \boldsymbol{\theta}, T, \boldsymbol{k}) p(\boldsymbol{\theta} | \mathcal{D}, T, \boldsymbol{k}) \mathrm{d}\boldsymbol{\theta}}
\end{aligned}
$$

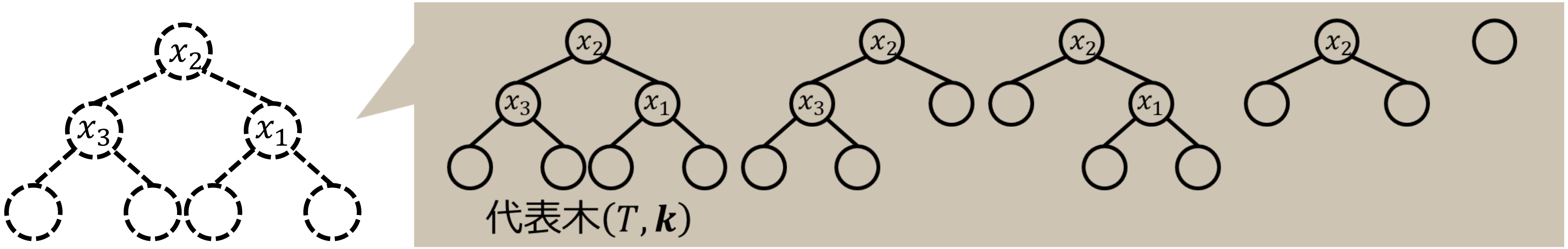
未解決

メタツリーによる
効率的厳密計算
[須子ら，2003]
[Dobashi et al., 2021]

$\boldsymbol{\theta}$に共役事前分布を仮定
→解析的に解ける
$p(y_{n+1} | x_{n+1}, \mathcal{D}, T, \boldsymbol{k})$

ベイズ基準のもと最適な予測の効率的計算

- メタツリー$M_{T,\boldsymbol{k}}$
  - 特徴量割当ベクトルが$\boldsymbol{k}$であるようなモデルツリーの集合



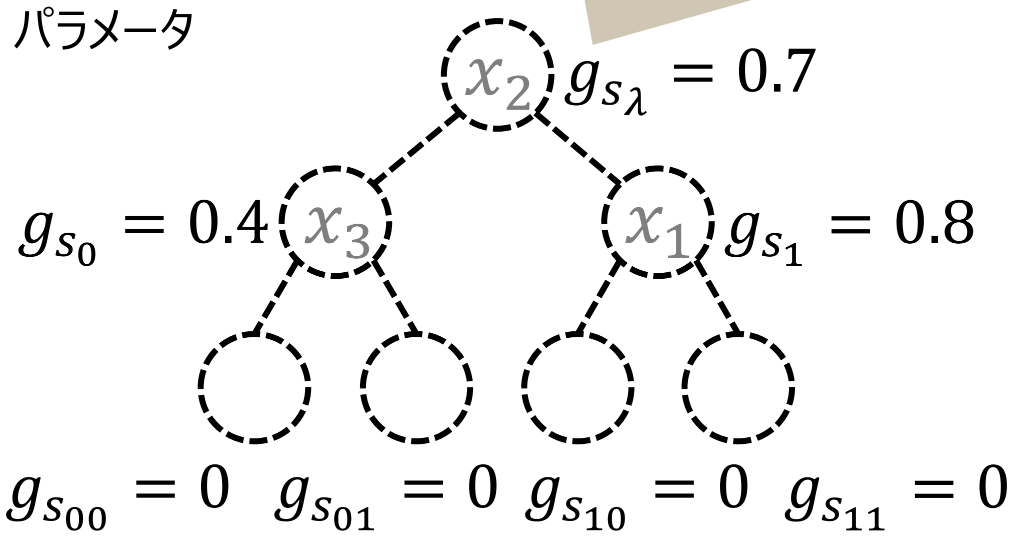
定義：$M_{T,\boldsymbol{k}} \coloneqq \{(T',\boldsymbol{k}') \in \mathcal{T} \times \mathcal{K} \,|\, \boldsymbol{k}' = \boldsymbol{k}, T'$は$T$の部分木$\}$

性質：$M_{T,\boldsymbol{k}}$に含まれる全てのモデルツリーに関する期待値を厳密かつ効率的に計算するアルゴリズムが存在[須子ら, 2003] [Dobashi et al., 2021].
→代表木の形を$\mathcal{T}$の最大木$T_{\max}$にとれば，$\mathcal{T}$上の期待値が計算可能.

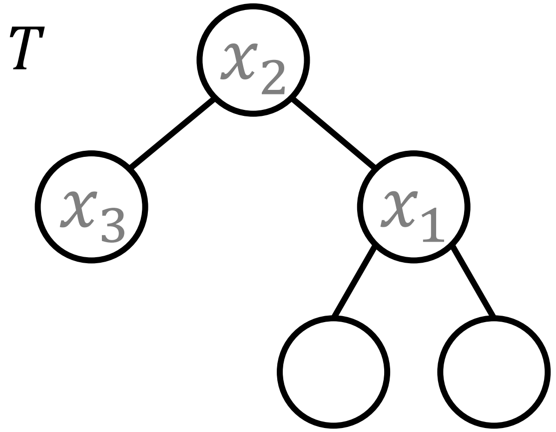
メタツリーを用いたモデルツリーの総和計算

- メタツリー上のモデルツリーの事前分布[Nakahara et al.,2022]
  - 具体例

枝分かれ確率g_s
（メタツリー$M_{T,\boldsymbol{k}}$の各ノード$s \in \mathcal{S}$に付与）



$M_{T,\boldsymbol{k}}$の部分木
として生成



$$
\begin{aligned}
&p(T) \\
&= g_{s_\lambda}(1 - g_{s_0})g_{s_1}(1 - g_{s_{10}})(1 - g_{s_{11}}) \\
&= 0.7 \times (1 - 0.4) \times 0.8 \times (1 - 0) \times (1 - 0) = 0.336
\end{aligned}
$$

メタツリーを用いたモデルツリーの総和計算

- メタツリー上のモデルツリーの事前分布[Nakahara et al.,2022]

- Definition 1
  - $(g_s)_{s\in\mathcal{S}} \in [0,1]^{|\mathcal{S}|}$：枝分かれ確率（葉ノードに対しては$g_s = 0$）
  - このとき，以下で分布を定義

$$p(T) \coloneqq \prod_{s\in\mathcal{I}_T} g_s \prod_{s'\in\mathcal{L}_T} (1-g_{s'})$$

- この分布が確率分布の定義を満たすこと，すなわち

$$\sum_T p(T) = \sum_T \prod_{s\in\mathcal{I}_T} g_s \prod_{s'\in\mathcal{L}_T} (1-g_{s'}) = 1$$

が木の再帰的構造を利用した数学的帰納法で証明可能

メタツリーを用いたモデルツリーの総和計算

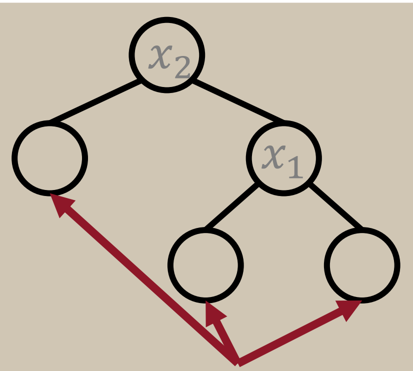
- メタツリー上のモデルツリーの事後分布
  - 計算したかったもの：$\sum_{T\in\mathcal{T}} p(T|\mathcal{D},\boldsymbol{k})\, p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T,\boldsymbol{k})$
  - 事後分布

$$
\begin{aligned}
&p(T|\mathcal{D},\boldsymbol{k})\\
&=\frac{1}{C}\prod_{s\in\mathcal{I}_T} g_s \prod_{s'\in\mathcal{L}_T}(1-g_{s'})\prod_{i=1}^{n}\prod_{s\in\mathcal{L}_T} p(y_i|\boldsymbol{x}_i,s,\boldsymbol{k})^{I\{s_{T,\boldsymbol{k}}(\boldsymbol{x})=s\}}\\
&=\frac{1}{C}\prod_{s\in\mathcal{I}_T} g_s \prod_{s'\in\mathcal{L}_T}\left((1-g_{s'})\prod_{i=1}^{n} p(y_i|\boldsymbol{x}_i,s,\boldsymbol{k})^{I\{s_{T,\boldsymbol{k}}(\boldsymbol{x})=s\}}\right)
\end{aligned}
$$

パラメータに吸収（後述）

$$
=\prod_{s\in\mathcal{I}_T} g_{s|\mathcal{D},\boldsymbol{k}} \prod_{s'\in\mathcal{L}_T}\left(1-g_{s'|\mathcal{D},\boldsymbol{k}}\right)
$$

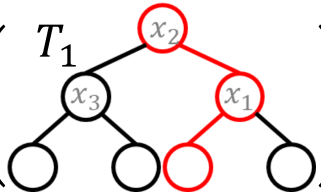
木の共役事前分布！

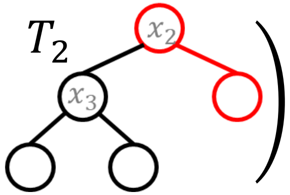


$I\{s_{T,\boldsymbol{k}}(\boldsymbol{x})=s\}$により
葉ノードから
一つが選ばれる

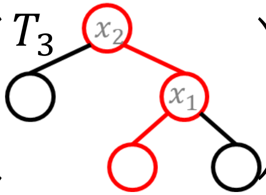
メタツリーを用いたモデルツリーの総和計算

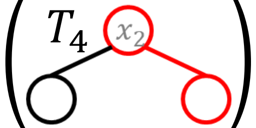
- メタツリー上のモデルツリーに関する期待値
  - 計算したかったもの：$\sum_{T\in\mathcal{T}} p(T|\mathcal{D},\boldsymbol{k})\, p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T,\boldsymbol{k})$
  - 例，最大深さ2の2分木，$(x_{n+1,1},x_{n+1,2},x_{n+1,3})=(0,1,0)$

$p\Bigg($ 

$\Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T_1,\boldsymbol{k}) + p\Bigg($ 

$\Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T_2,\boldsymbol{k})$

$+p\Bigg($ 

$\Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T_3,\boldsymbol{k}) + p\Bigg($ 

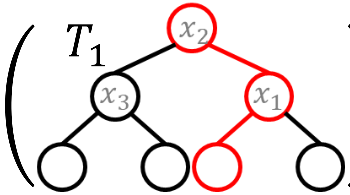
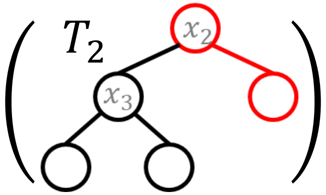
$\Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T_4,\boldsymbol{k})$

$+p\Bigg($ 

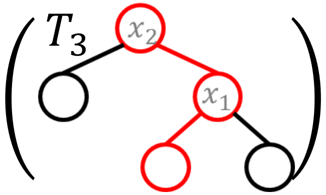
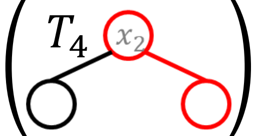
$\Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T_5,\boldsymbol{k})$

メタツリーを用いたモデルツリーの総和計算

- メタツリー上のモデルツリーに関する期待値
  - 計算したかったもの：$\sum_{T\in\mathcal{T}} p(T|\mathcal{D},\boldsymbol{k})\, p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T,\boldsymbol{k})$
  - 例，最大深さ2の2分木，$(x_{n+1,1},x_{n+1,2},x_{n+1,3})=(0,1,0)$

$p\Bigg($  $\Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T_1,\boldsymbol{k}) + p\Bigg($  $\Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T_2,\boldsymbol{k})$

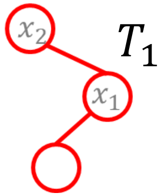
等しい　　　　等しい

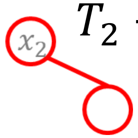
$+p\Bigg($  $\Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T_3,\boldsymbol{k}) + p\Bigg($  $\Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T_4,\boldsymbol{k})$

$+p\Bigg($  $\Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T_5,\boldsymbol{k})$

メタツリーを用いたモデルツリーの総和計算

- メタツリー上のモデルツリーに関する期待値
  - 計算したかったもの：$\sum_{T\in\mathcal{T}} p(T|\mathcal{D},\boldsymbol{k})\, p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T,\boldsymbol{k})$
  - 例，最大深さ2の2分木，$(x_{n+1,1}, x_{n+1,2}, x_{n+1,3}) = (0,1,0)$

$p\Bigg($ 

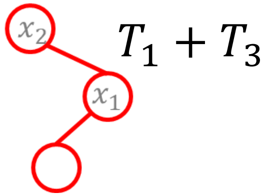
$T_1 + T_3 \Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_{10},\boldsymbol{k}) + p\Bigg($ 

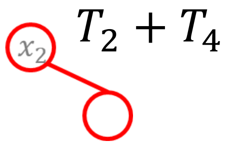
$T_2 + T_4 \Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_0,\boldsymbol{k})$

$+p\Bigg(T_5$  $\Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_\lambda,\boldsymbol{k})$

メタツリーを用いたモデルツリーの総和計算

- メタツリー上のモデルツリーに関する期待値
  - 計算したかったもの：$\sum_{T\in\mathcal{T}} p(T|\mathcal{D},\boldsymbol{k})\, p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T,\boldsymbol{k})$
  - 例，最大深さ2の2分木，$(x_{n+1,1}, x_{n+1,2}, x_{n+1,3}) = (0,1,0)$

$p\Bigg($ 

$\Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_{10},\boldsymbol{k}) + p\Bigg($ 

$\Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_0,\boldsymbol{k})$

$+p\Bigg($ 

$\Bigg) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_\lambda,\boldsymbol{k})$

$$
\begin{aligned}
= &\, g_{s_\lambda|\mathcal{D},\boldsymbol{k}} g_{s_1|\mathcal{D},\boldsymbol{k}} \left(1 - g_{s_{10}|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_{10},\boldsymbol{k}) \\
&+ g_{s_\lambda|\mathcal{D},\boldsymbol{k}} \left(1 - g_{s_1|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_0,\boldsymbol{k}) \\
&+ \left(1 - g_{s_\lambda|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_\lambda,\boldsymbol{k})
\end{aligned}
$$

メタツリーを用いたモデルツリーの総和計算

- メタツリー上のモデルツリーに関する期待値
  - 計算したかったもの：$\sum_{T\in\mathcal{T}} p(T|\mathcal{D},\boldsymbol{k})\, p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T,\boldsymbol{k})$
  - 例，最大深さ2の2分木，$(x_{n+1,1},x_{n+1,2},x_{n+1,3})=(0,1,0)$

$$
\begin{aligned}
&= g_{s_\lambda|\mathcal{D},\boldsymbol{k}} g_{s_1|\mathcal{D},\boldsymbol{k}} \left(1-g_{s_{10}|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_{10},\boldsymbol{k}) \\
&\quad + g_{s_\lambda|\mathcal{D},\boldsymbol{k}} \left(1-g_{s_1|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_0,\boldsymbol{k}) \\
&\quad + \left(1-g_{s_\lambda|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_\lambda,\boldsymbol{k})
\end{aligned}
$$

メタツリーを用いたモデルツリーの総和計算

- メタツリー上のモデルツリーに関する期待値
  - 計算したかったもの：$\sum_{T \in \mathcal{T}} p(T|\mathcal{D}, \boldsymbol{k})\, p(y_{n+1}|\boldsymbol{x}_{n+1}, \mathcal{D}, T, \boldsymbol{k})$
  - 例，最大深さ2の2分木，$(x_{n+1,1}, x_{n+1,2}, x_{n+1,3}) = (0,1,0)$

$$
\begin{aligned}
&= g_{s_\lambda|\mathcal{D},\boldsymbol{k}} g_{s_1|\mathcal{D},\boldsymbol{k}} \left(1 - g_{s_{10}|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1}, \mathcal{D}, s_{10}, \boldsymbol{k}) \\
&\quad + g_{s_\lambda|\mathcal{D},\boldsymbol{k}} \left(1 - g_{s_1|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1}, \mathcal{D}, s_0, \boldsymbol{k}) \\
&\quad + \left(1 - g_{s_\lambda|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1}, \mathcal{D}, s_\lambda, \boldsymbol{k})
\end{aligned}
$$

$$
\begin{aligned}
&= \left(1 - g_{s_\lambda|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1}, \mathcal{D}, s_\lambda, \boldsymbol{k}) \\
&\quad + g_{s_\lambda|\mathcal{D},\boldsymbol{k}} \Big(\left(1 - g_{s_1|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1}, \mathcal{D}, s_0, \boldsymbol{k}) \\
&\qquad\qquad + g_{s_1|\mathcal{D},\boldsymbol{k}} \left(1 - g_{s_{10}|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1}, \mathcal{D}, s_{10}, \boldsymbol{k}) \Big)
\end{aligned}
$$

メタツリーを用いたモデルツリーの総和計算

- メタツリー上のモデルツリーに関する期待値
  - 計算したかったもの：$\sum_{T\in\mathcal{T}} p(T|\mathcal{D},\boldsymbol{k})\, p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},T,\boldsymbol{k})$
  - 例，最大深さ2の2分木，$(x_{n+1,1},x_{n+1,2},x_{n+1,3})=(0,1,0)$

$$
\begin{aligned}
&= g_{s_\lambda|\mathcal{D},\boldsymbol{k}}\, g_{s_1|\mathcal{D},\boldsymbol{k}} \left(1-g_{s_{10}|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_{10},\boldsymbol{k}) \\
&\quad + g_{s_\lambda|\mathcal{D},\boldsymbol{k}} \left(1-g_{s_1|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_0,\boldsymbol{k}) \\
&\quad + \left(1-g_{s_\lambda|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_\lambda,\boldsymbol{k})
\end{aligned}
$$

$$
\begin{aligned}
&= \left(1-g_{s_\lambda|\mathcal{D},\boldsymbol{k}}\right) \boxed{p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_\lambda,\boldsymbol{k})} \\
&\quad + g_{s_\lambda|\mathcal{D},\boldsymbol{k}} \boxed{\left(\left(1-g_{s_1|\mathcal{D},\boldsymbol{k}}\right) \boxed{p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_0,\boldsymbol{k})} + g_{s_1|\mathcal{D},\boldsymbol{k}} \boxed{\left(1-g_{s_{10}|\mathcal{D},\boldsymbol{k}}\right) p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_{10},\boldsymbol{k})} \right)}
\end{aligned}
$$

メタツリーを用いたモデルツリーの総和計算

- メタツリー上のモデルツリーに関する期待値

再帰関数によって計算可能

$$
\begin{aligned}
&q(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s,\boldsymbol{k})\\
&= \left(1-g_{s|\mathcal{D},\boldsymbol{k}}\right)\boxed{p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s,\boldsymbol{k})} + g_{s|\mathcal{D},\boldsymbol{k}}\boxed{q(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_{\mathrm{child}},\boldsymbol{k})}
\end{aligned}
$$

$$
\begin{aligned}
&= \left(1-g_{s_\lambda|\mathcal{D},\boldsymbol{k}}\right)\boxed{p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_\lambda,\boldsymbol{k})}\\
&\quad +g_{s_\lambda|\mathcal{D},\boldsymbol{k}}\boxed{\left(\left(1-g_{s_1|\mathcal{D},\boldsymbol{k}}\right)\boxed{p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_0,\boldsymbol{k})} + g_{s_1|\mathcal{D},\boldsymbol{k}}\boxed{\left(1-g_{s_{10}|\mathcal{D},\boldsymbol{k}}\right)p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_{10},\boldsymbol{k})}\right)}
\end{aligned}
$$

メタツリーを用いたモデルツリーの総和計算

- メタツリー上のモデルツリーに関する期待値

再帰関数によって計算可能

sが枝分かれしない
場合の予測分布

sが枝分かれする
場合の予測分布

$$q(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s,\boldsymbol{k})$$
$$=\left(1-g_{s|\mathcal{D},\boldsymbol{k}}\right)\boxed{p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s,\boldsymbol{k})}+g_{s|\mathcal{D},\boldsymbol{k}}\boxed{q(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_{\mathrm{child}},\boldsymbol{k})}$$

$$=\left(1-g_{s_\lambda|\mathcal{D},\boldsymbol{k}}\right)\boxed{p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_\lambda,\boldsymbol{k})}$$
$$+g_{s_\lambda|\mathcal{D},\boldsymbol{k}}\boxed{\left(\left(1-g_{s_1|\mathcal{D},\boldsymbol{k}}\right)\boxed{p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_0,\boldsymbol{k})}+g_{s_1|\mathcal{D},\boldsymbol{k}}\boxed{\left(1-g_{s_{10}|\mathcal{D},\boldsymbol{k}}\right)p(y_{n+1}|\boldsymbol{x}_{n+1},\mathcal{D},s_{10},\boldsymbol{k})}\right)}$$

メタツリーを用いたモデルツリーの総和計算

- 事後分布の正規化

$$\frac{1}{C}\prod_{s\in\mathcal{I}_T} g_s \prod_{s'\in\mathcal{L}_T}\left((1-g_{s'})\prod_{i=1}^{n} p(y_i|\boldsymbol{x}_i,s,\boldsymbol{k})^{I\{s_{T,\boldsymbol{k}}(\boldsymbol{x})=s\}}\right)=\prod_{s\in\mathcal{I}_T} g_{s|\mathcal{D},\boldsymbol{k}}\prod_{s'\in\mathcal{L}_T}\left(1-g_{s'|\mathcal{D},\boldsymbol{k}}\right)$$

パラメータに吸収

- $g_{s|\mathcal{D},\boldsymbol{k}}$は以下の式に従って逐次的に更新可能

$$g_{s|\boldsymbol{x}^{i+1},y^{i+1},\boldsymbol{k}} = \frac{g_{s|\boldsymbol{x}^i,y^i,\boldsymbol{k}}\,q\left(y_{i+1}\middle|\boldsymbol{x}_{i+1},\boldsymbol{x}^i,y^i,s_{\mathrm{child}},\boldsymbol{k}\right)}{\left(1-g_{s|\boldsymbol{x}^i,y^i,\boldsymbol{k}}\right)p\left(y_{i+1}\middle|\boldsymbol{x}_{i+1},\boldsymbol{x}^i,y^i,s,\boldsymbol{k}\right)+g_{s|\boldsymbol{x}^i,y^i,\boldsymbol{k}}\,q\left(y_{i+1}\middle|\boldsymbol{x}_{i+1},\boldsymbol{x}^i,y^i,s_{\mathrm{child}},\boldsymbol{k}\right)}$$

メタツリーを用いたモデルツリーの総和計算

- 事後分布の正規化

$$\frac{1}{C}\prod_{s\in\mathcal{I}_T} g_s \prod_{s'\in\mathcal{L}_T}\left((1-g_{s'})\prod_{i=1}^{n} p(y_i|\boldsymbol{x}_i,s,\boldsymbol{k})^{I\{s_{T,\boldsymbol{k}}(\boldsymbol{x})=s\}}\right) = \prod_{s\in\mathcal{I}_T} g_{s|\mathcal{D},\boldsymbol{k}} \prod_{s'\in\mathcal{L}_T}(1-g_{s'|\mathcal{D},\boldsymbol{k}})$$

パラメータに吸収

- $g_{s|\mathcal{D},\boldsymbol{k}}$は以下の式に従って逐次的に更新可能

sが枝分かれする事前確率

sが枝分かれする場合の周辺尤度

$$g_{s|\boldsymbol{x}^{i+1},y^{i+1},\boldsymbol{k}} = \frac{g_{s|\boldsymbol{x}^i,y^i,\boldsymbol{k}}\, q(y_{i+1}|\boldsymbol{x}_{i+1},\boldsymbol{x}^i,y^i,s_{\mathrm{child}},\boldsymbol{k})}{\underbrace{(1-g_{s|\boldsymbol{x}^i,y^i,\boldsymbol{k}})p(y_{i+1}|\boldsymbol{x}_{i+1},\boldsymbol{x}^i,y^i,s,\boldsymbol{k}) + g_{s|\boldsymbol{x}^i,y^i,\boldsymbol{k}}\, q(y_{i+1}|\boldsymbol{x}_{i+1},\boldsymbol{x}^i,y^i,s_{\mathrm{child}},\boldsymbol{k})}}$$

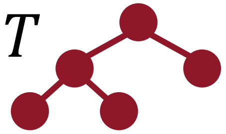
sが枝分かれしない場合の周辺尤度

sについても周辺化した周辺尤度

➡ ベイズの定理そのもの！

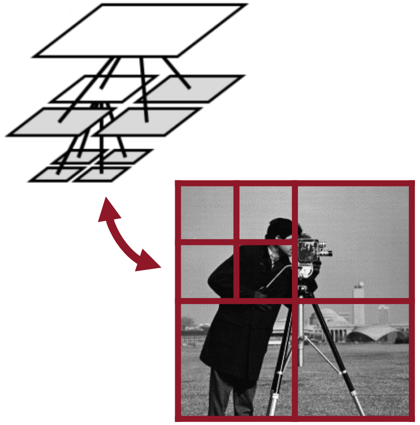
木の事前分布の応用

- 木構造上の確率モデル [Nakahara et al., 2022][Nakahara et al., 2022]
  - データの背後の木構造に関する事前分布とベイズ推論アルゴリズム
    （期待値，最頻値，事後分布，予測分布，⋯）



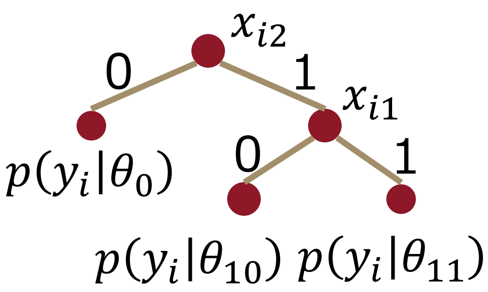
$$T \sim p(T) = \prod_{s\in\mathcal{L}_T} g_s \prod_{s'\in\mathcal{I}_T} (1-g_s)$$

四分木領域分割



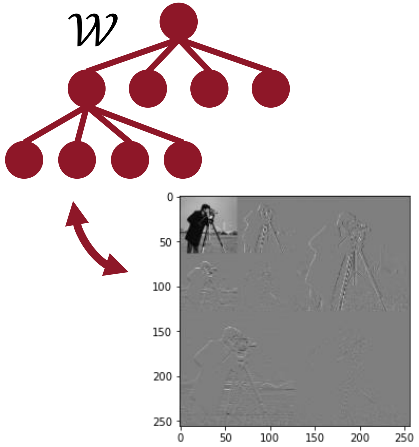
[Nakahara et al., 2021]
[Nakahara et al., 2022]

決定木



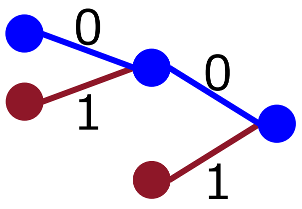
[Dobashi et al., 2021]
[中原ら，2022]

ウェーブレットパケット木



[岡ら, 2021]

文脈木



$01\underline{\textcolor{blue}{00}1}01\cdots$

[Nakahara et al., 2022]

ベイズ基準のもと最適な予測の効率的計算

- ベイズ基準のもと最適な予測を効率的に計算するには？

$$\delta^*(\mathcal{D}, x_{n+1})$$
$$= \arg\max_{y_{n+1}} p(y_{n+1} | \boldsymbol{x}_{n+1}, \mathcal{D})$$
$$= \arg\max_{y_{n+1}} \sum_{\boldsymbol{k}\in\mathcal{K}} \sum_{T\in\mathcal{T}} \int p(y_{n+1} | \boldsymbol{x}_{n+1}, \boldsymbol{\theta}, T, \boldsymbol{k})\, p(\boldsymbol{\theta}, T, \boldsymbol{k} | \mathcal{D}) \mathrm{d}\boldsymbol{\theta}$$
$$= \arg\max_{y_{n+1}} \sum_{\boldsymbol{k}\in\mathcal{K}} p(\boldsymbol{k}|\mathcal{D}) \sum_{T\in\mathcal{T}} p(T|\mathcal{D}, \boldsymbol{k}) \underbrace{\int p(y_{n+1} | \boldsymbol{x}_{n+1}, \boldsymbol{\theta}, T, \boldsymbol{k}) p(\boldsymbol{\theta} | \mathcal{D}, T, \boldsymbol{k}) \mathrm{d}\boldsymbol{\theta}}$$

未解決

メタツリーによる
効率的厳密計算
[須子ら，2003]
[Dobashi et al., 2021]

$\boldsymbol{\theta}$に共役事前分布を仮定
→解析的に解ける
$$p(y_{n+1} | x_{n+1}, \mathcal{D}, T, \boldsymbol{k})$$

特徴量割当ベクトルに関する総和計算

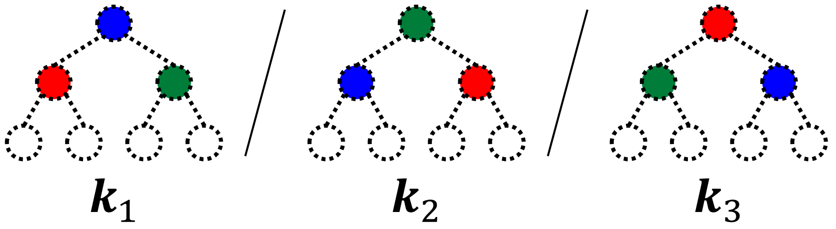
- 特徴量割当ベクトル（メタツリー）上の総和計算量削減アイデア

$$\sum_{\boldsymbol{k}\in\mathcal{K}} \Rightarrow \sum_{\boldsymbol{k}\in\{\boldsymbol{k}_1,\boldsymbol{k}_2,\dots,\boldsymbol{k}_B\}\subset\mathcal{K}}$$

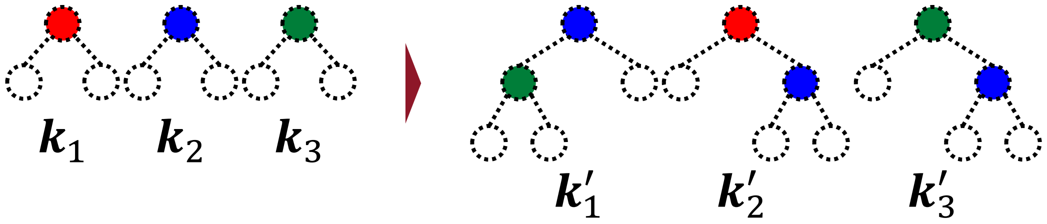
$\mathcal{K}$の部分集合$\{\boldsymbol{k}_1,\boldsymbol{k}_2,\dots,\boldsymbol{k}_B\}$上での総和に置き換える

- 部分集合構築法

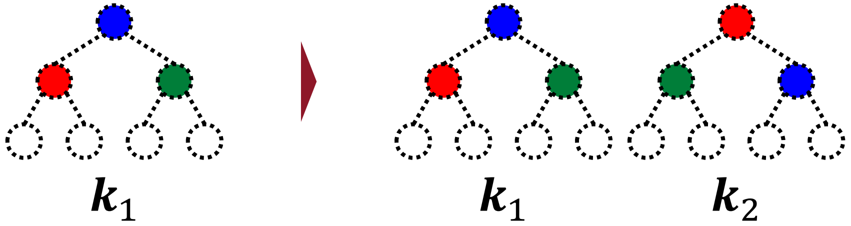
並列的（独立）[Dobashi et al., 2021]



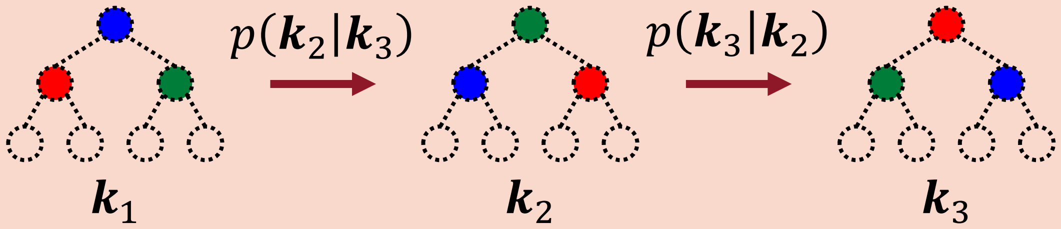
並列的（同時）[田島ら，2022]



直列的 [于ら，2022][馬庭ら，2022]



サンプリング的[中原ら，2022]



特徴量割当ベクトルに関する総和計算

- 特徴量割当ベクトル（メタツリー）上の総和計算量削減アイデア

$$\sum_{\boldsymbol{k}\in\mathcal{K}} \quad \Rightarrow \quad \sum_{\boldsymbol{k}\in\{\boldsymbol{k}_1,\boldsymbol{k}_2,\dots,\boldsymbol{k}_B\}\subset\mathcal{K}}$$

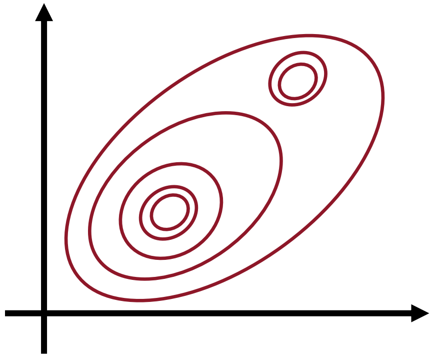
$\mathcal{K}$の部分集合$\{\boldsymbol{k}_1,\boldsymbol{k}_2,\dots,\boldsymbol{k}_B\}$上での総和に置き換える

- 部分集合構築法

考え方\構築対象		メタツリー集合	関数木集合
並列的	独立	メタツリーランダムフォレスト[Dobashi et al., 2021]	Random Forest[Breiman, 2001]
	同時	メタツリー集合並列構成法[田島ら，2022]	
直列的		メタツリーブースティング（回帰）[于ら，2022] メタツリーブースティング（分類）[馬庭ら,2022]	XGBoost[Chen et al., 2016], LigntGBM[Ke et al, 2017]
サンプリング的		**メタツリーマルコフ連鎖モンテカルロ法**	

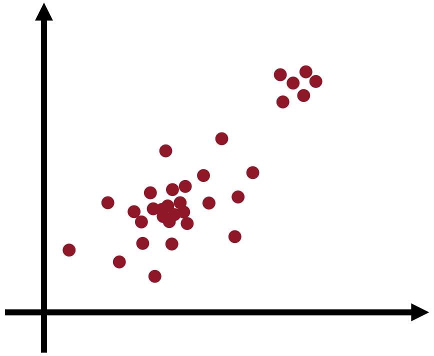
マルコフ連鎖モンテカルロ法の概要

●サンプリング法



複雑な分布$\frac{1}{c}\tilde{p}(z)$

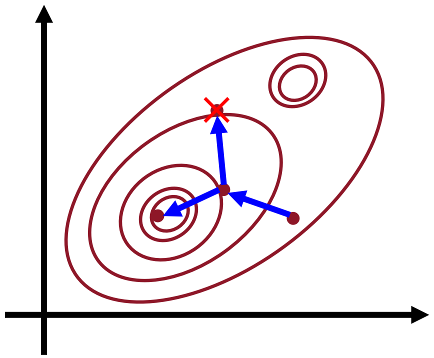
近似



別手段によるサンプル$\left\{z^{(t)}\right\}_{t=1}^{t_{\mathrm{end}}}$

経験分布に基づき
期待値，最頻値
なども算出可能

●MCMC法（特にメトロポリスヘイスティングス法）



どれほど離れた点を
提案するかが問題

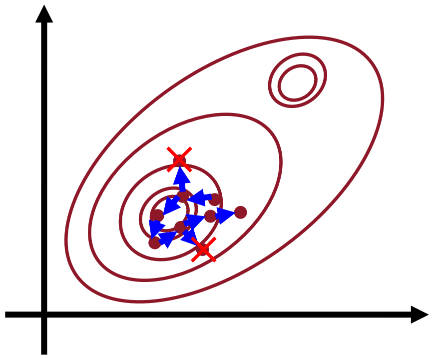
- 提案分布$q\left(z^{(t+1)}\middle|z^{(t)}\right)$により近くの点を提案
- 採択（移動）/ 棄却

反復

詳細つり合い条件
を満たすように

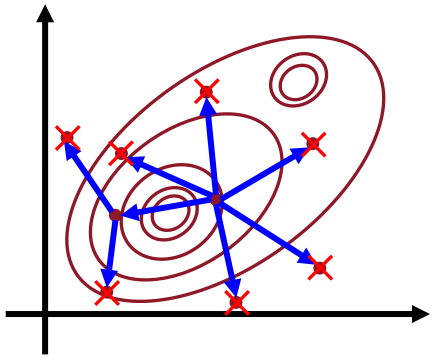
マルコフ連鎖モンテカルロ法の概要

- 提案分布の設計における課題



提案分布の分散：小

分布を覆うのに時間がかかる



提案分布の分散：大

棄却率が上がる

メタツリーマルコフ連鎖モンテカルロ法

- MCMC法をメタツリーの積分に適用

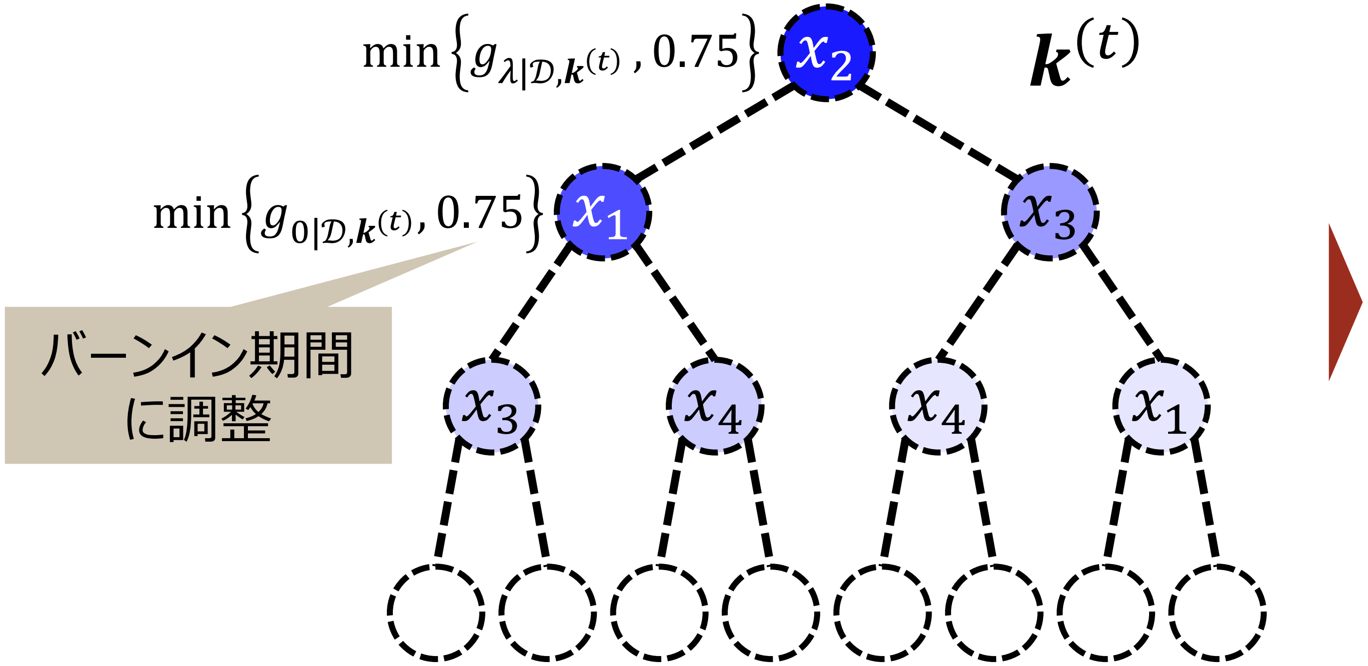
$$\delta^*(\mathcal{D}, x_{n+1})$$
$$= \arg\max_{y_{n+1}} \sum_{\boldsymbol{k}\in\mathcal{K}} p(\boldsymbol{k}|\mathcal{D}) \sum_{T\in\mathcal{T}} p(T|\mathcal{D},\boldsymbol{k}) \int p(y_{n+1}|\boldsymbol{x}_{n+1},\boldsymbol{\theta},T,\boldsymbol{k})p(\boldsymbol{\theta}|\mathcal{D},T,\boldsymbol{k})\mathrm{d}\boldsymbol{\theta}$$
$$= \arg\max_{y_{n+1}} \frac{1}{t_{\mathrm{end}}} \sum_{t=1}^{t_{\mathrm{end}}} \sum_{T\in\mathcal{T}} p\left(T\middle|\mathcal{D},\boldsymbol{k}^{(t)}\right) \int p\left(y_{n+1}\middle|\boldsymbol{x}_{n+1},\boldsymbol{\theta},T,\boldsymbol{k}^{(t)}\right)p\left(\boldsymbol{\theta}\middle|\mathcal{D},T,\boldsymbol{k}^{(t)}\right)\mathrm{d}\boldsymbol{\theta}$$

  - MCMCサンプル$\{\boldsymbol{k}^{(1)},\boldsymbol{k}^{(2)},\dots,\boldsymbol{k}^{(t_{\mathrm{end}})}\}$上の標本平均で近似

- 提案分布$q\left(\boldsymbol{k}^{(t+1)}\middle|\boldsymbol{k}^{(t)}\right)$をどのように設計するか？
  - 特徴量割当ベクトルはカテゴリカルなので分散の調整という考え方は使えない.
  →木の事後分布$p\left(T\middle|\mathcal{D},\boldsymbol{k}^{(t)}\right)$を利用し，固定する変数の数を調整！

メタツリーマルコフ連鎖モンテカルロ法

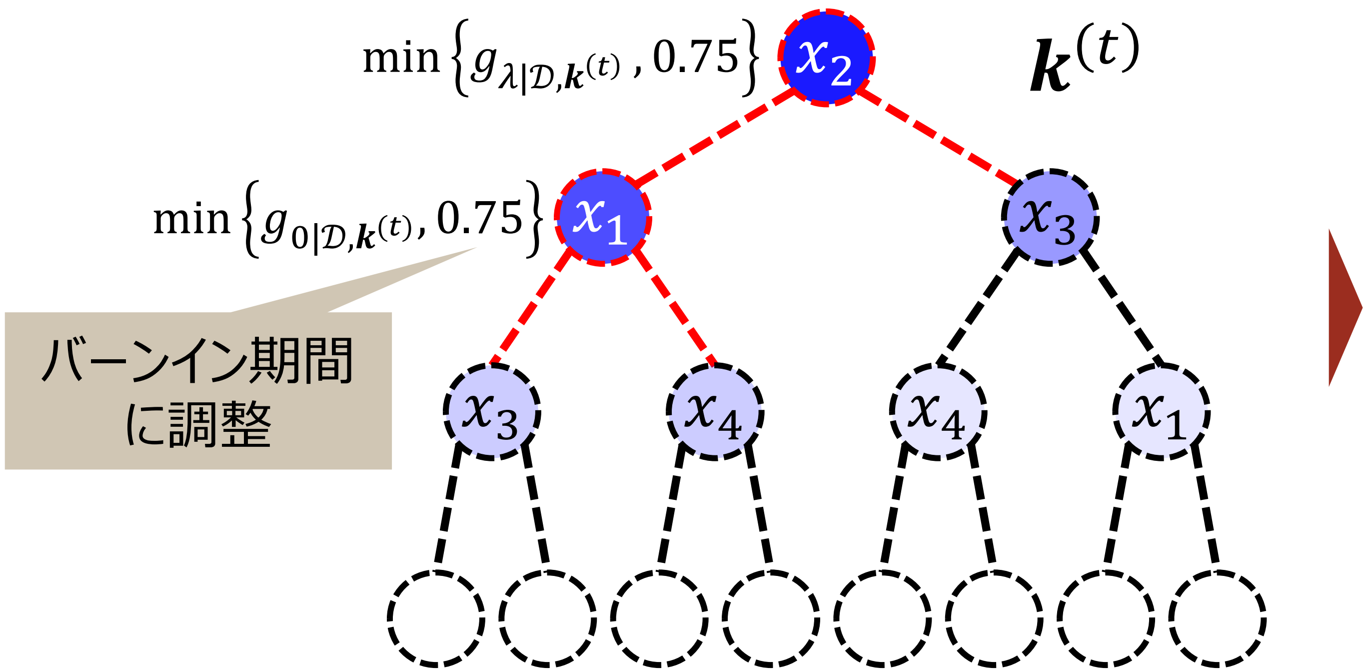
- 木の形Tに関する事後分布$p\left(T \middle| \mathcal{D}, \boldsymbol{k}^{(t)}\right)$を計算可能
  - 各ノードで分岐が行われる事後確率$g_{s|\mathcal{D},\boldsymbol{k}^{(t)}}$を計算可能
  - $g_{s|\mathcal{D},\boldsymbol{k}^{(t)}}$が小さいほど頻繁に$k_s$を更新



※各ノードの含まれやすさを色で表現

メタツリーマルコフ連鎖モンテカルロ法

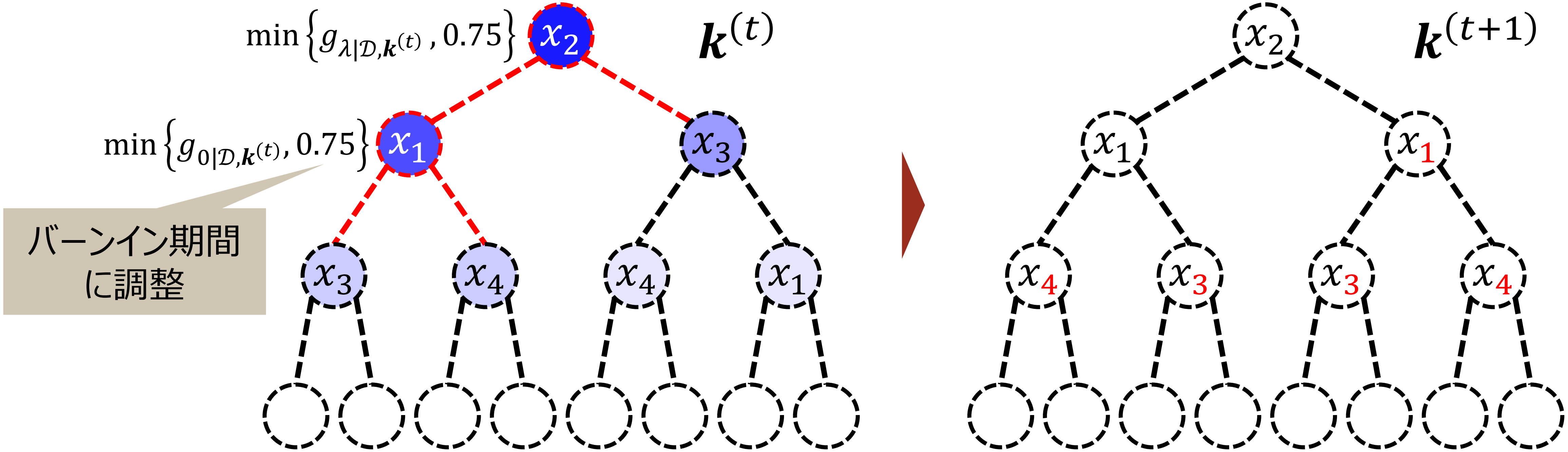
- 木の形Tに関する事後分布$p\left(T \middle| \mathcal{D}, \boldsymbol{k}^{(t)}\right)$を計算可能
  - 各ノードで分岐が行われる事後確率$g_{s|\mathcal{D},\boldsymbol{k}^{(t)}}$を計算可能
  - $g_{s|\mathcal{D},\boldsymbol{k}^{(t)}}$が小さいほど頻繁に$k_s$を更新



※各ノードの含まれやすさを色で表現

メタツリーマルコフ連鎖モンテカルロ法

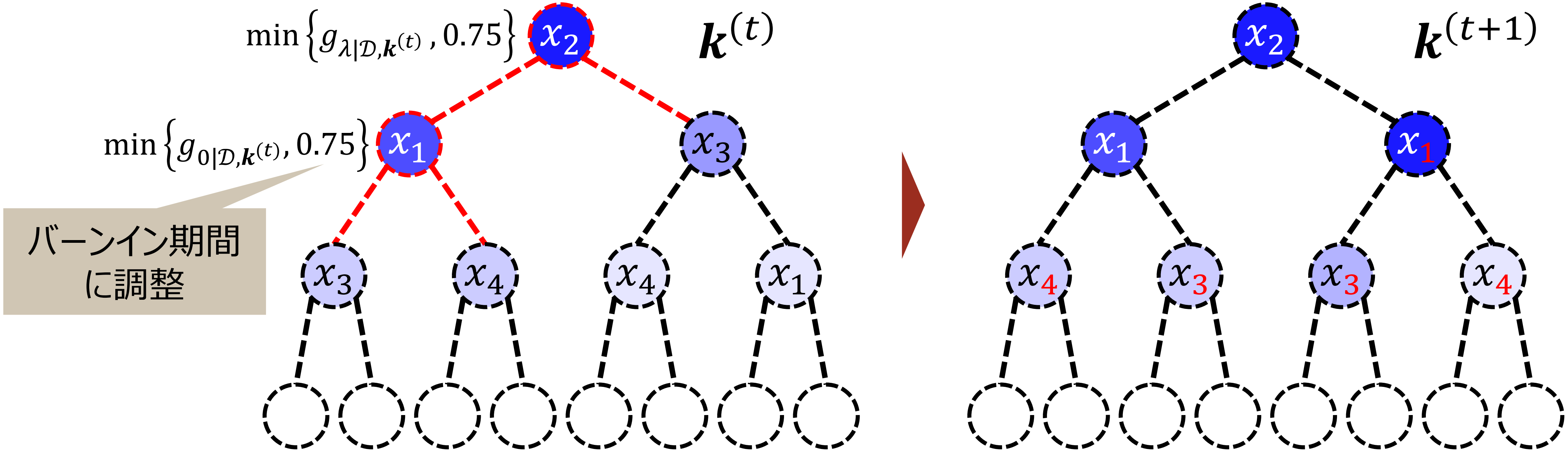
- 木の形Tに関する事後分布$p\left(T \middle| \mathcal{D}, \boldsymbol{k}^{(t)}\right)$を計算可能
  - 各ノードで分岐が行われる事後確率$g_{s|\mathcal{D},\boldsymbol{k}^{(t)}}$を計算可能
  - $g_{s|\mathcal{D},\boldsymbol{k}^{(t)}}$が小さいほど頻繁に$k_s$を更新



※各ノードの含まれやすさを色で表現

メタツリーマルコフ連鎖モンテカルロ法

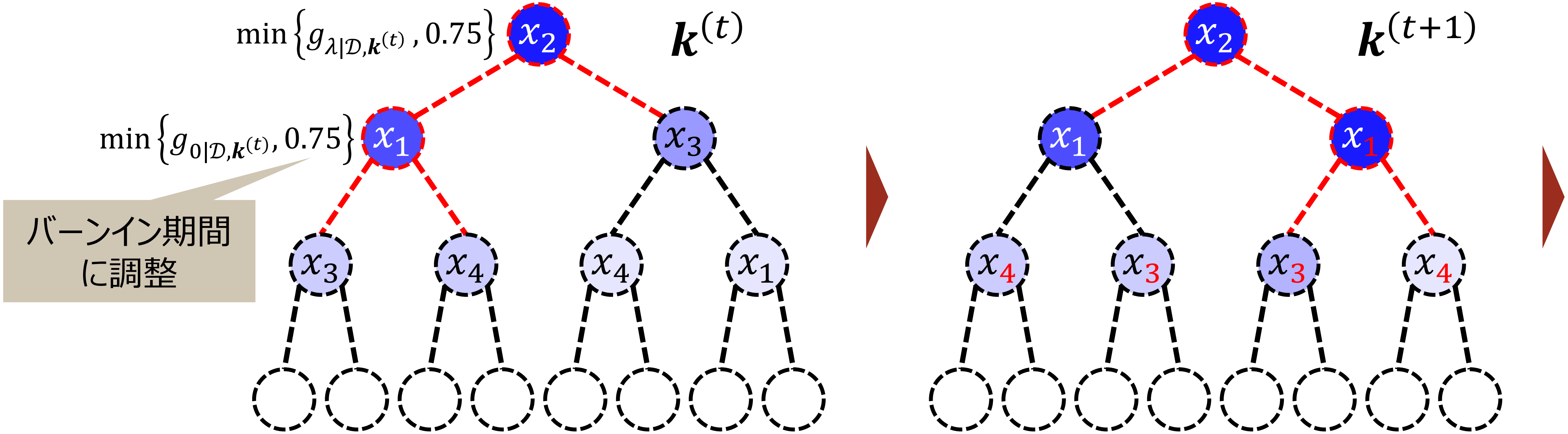
- 木の形Tに関する事後分布$p(T|\mathcal{D},\boldsymbol{k}^{(t)})$を計算可能
  - 各ノードで分岐が行われる事後確率$g_{s|\mathcal{D},\boldsymbol{k}^{(t)}}$を計算可能
  - $g_{s|\mathcal{D},\boldsymbol{k}^{(t)}}$が小さいほど頻繁に$k_s$を更新



※各ノードの含まれやすさを色で表現

メタツリーマルコフ連鎖モンテカルロ法

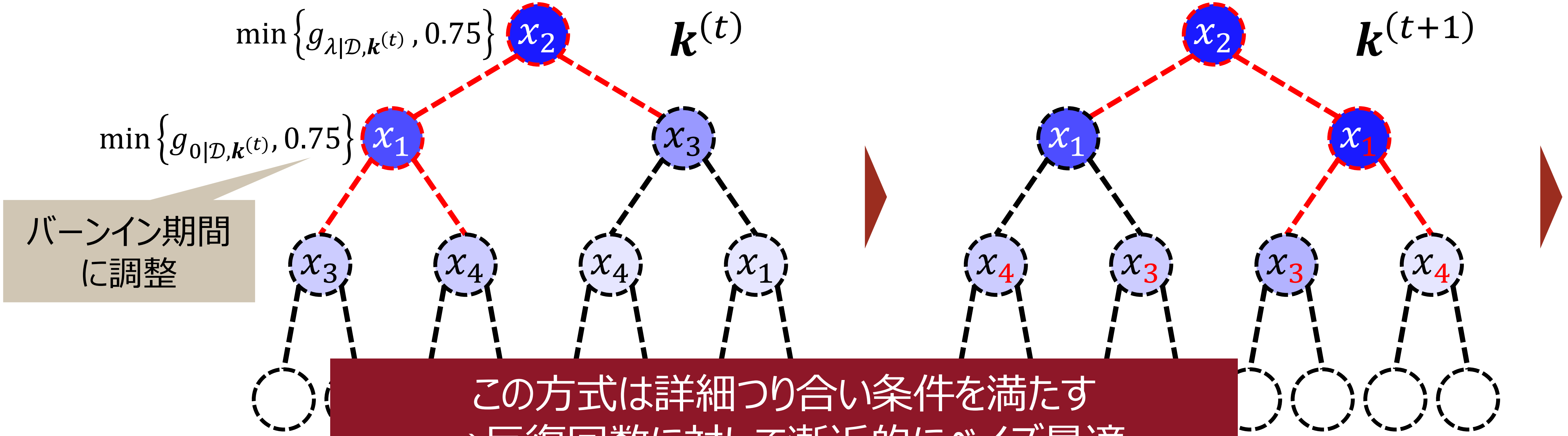
- 木の形Tに関する事後分布$p(T|\mathcal{D}, \boldsymbol{k}^{(t)})$を計算可能
  - 各ノードで分岐が行われる事後確率$g_{s|\mathcal{D},\boldsymbol{k}^{(t)}}$を計算可能
  - $g_{s|\mathcal{D},\boldsymbol{k}^{(t)}}$が小さいほど頻繁に$k_s$を更新



※各ノードの含まれやすさを色で表現

メタツリーマルコフ連鎖モンテカルロ法

- 木の形Tに関する事後分布$p\left(T \middle| \mathcal{D}, \boldsymbol{k}^{(t)}\right)$を計算可能
  - 各ノードで分岐が行われる事後確率$g_{s|\mathcal{D},\boldsymbol{k}^{(t)}}$を計算可能
  - $g_{s|\mathcal{D},\boldsymbol{k}^{(t)}}$が小さいほど頻繁に$k_s$を更新



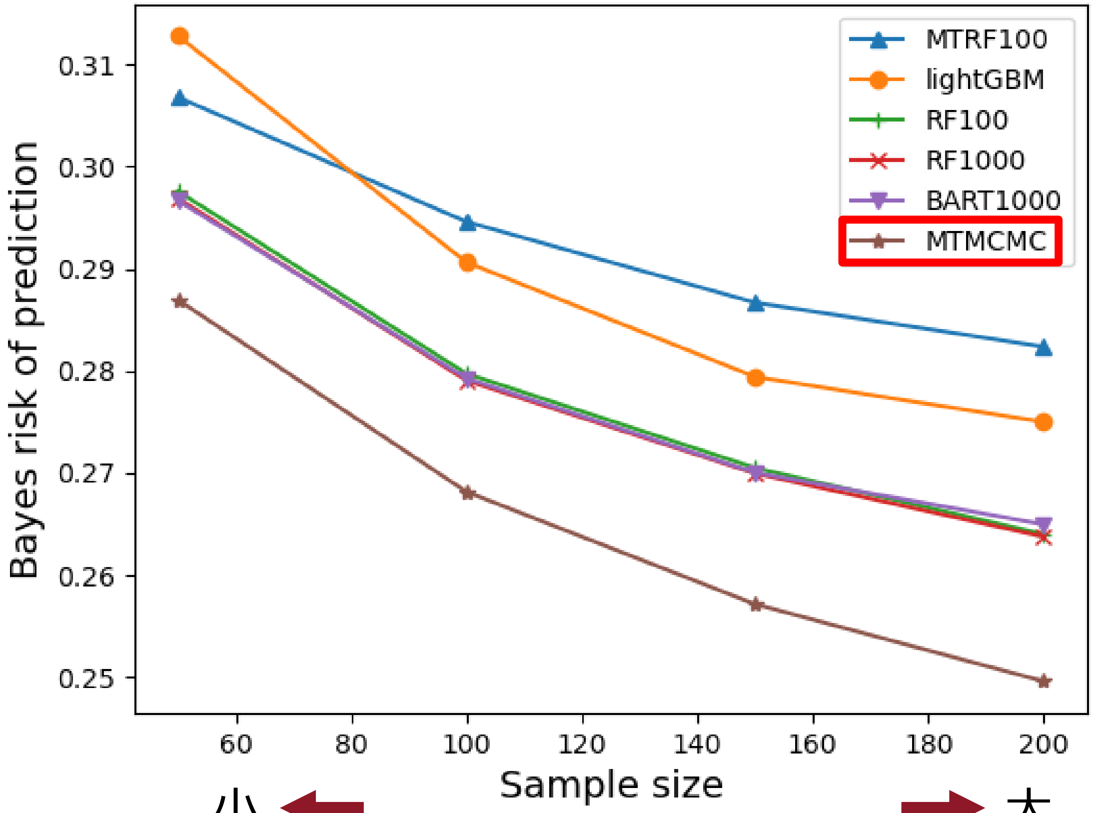
この方式は詳細つり合い条件を満たす
→反復回数に対して漸近的にベイズ最適

※各

実験結果

- 人工データで平均予測誤り確率（ベイズリスクの近似値）を確認
  - $\mathcal{X} = \mathcal{Y} = \{0,1\}$
  - $K = 20, D_{\max} = 10$
  - $\boldsymbol{k}, T$：100回生成
  - $\boldsymbol{\theta}, \boldsymbol{x}, y$：100回生成
  - バーンイン期間：500
  - MCMCサンプルサイズ：1000
  - 比較手法
    - Random Forest[Breiman, 2001]
    - LightGBM[Ke et al, 2017]
    - BART[Chipman et al., 2010]
    - メタツリーランダムフォレスト [Dobashi et al., 2010]

悪

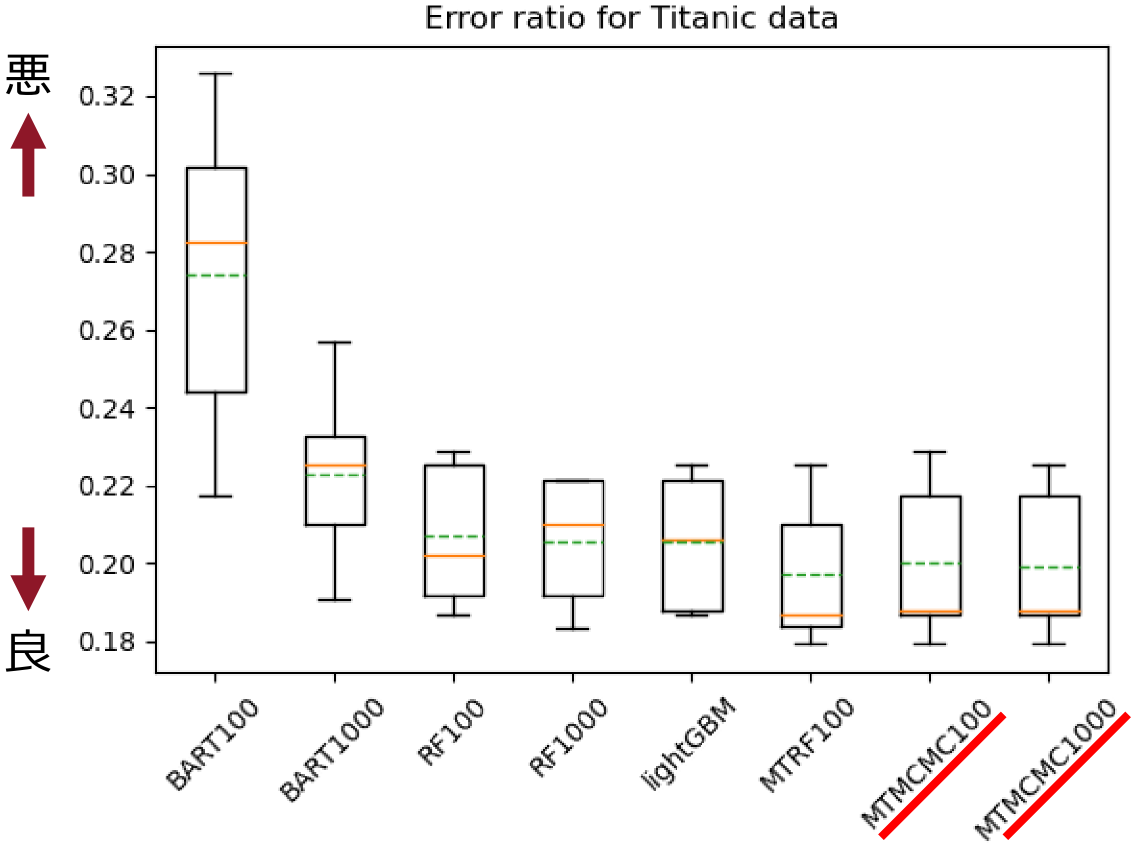


良

小 大

実データ実験

- Titanicデータでの予測実験
  - 説明変数
    - タイタニック号の乗客の属性
  - 目的変数
    - 生存 / 死亡
  - 実験方法
    - 連続変数は5段階に量子化
      →全変数をOne-hot表現に
    - 5-foldクロスバリデーション
  - 実験結果
    - 5回分の平均誤り率の箱ひげ図

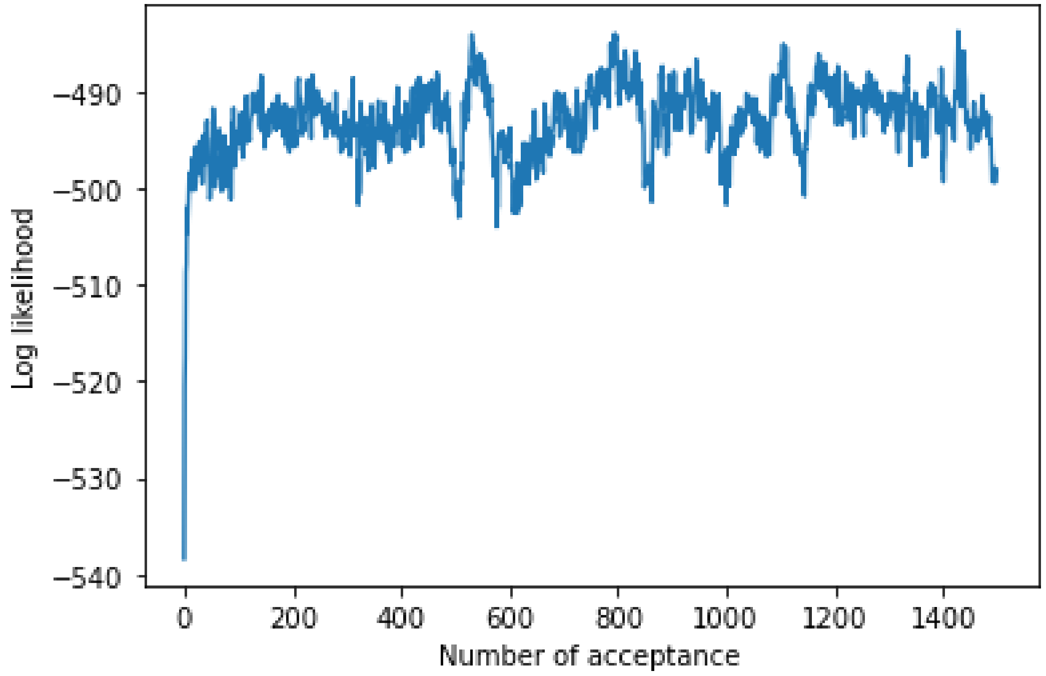


まとめ

- 従来の決定木は特徴記述や予測の関数を表現
  →我々は確率的データ生成観測メカニズムとしての決定木を提案
- ベイズ最適な予測の理論式を導出
- ベイズ最適な予測の計算困難性を以下のアイデアで回避
  - 木の形に関する総和計算：メタツリー
  - 特徴量割当ベクトルに関する総和計算：
    MCMC法（その他の探索的手法も研究中）
- 人工データ上でRandom ForestやLightGBMを凌駕．実データ上でもそれに匹敵する性能を示している.

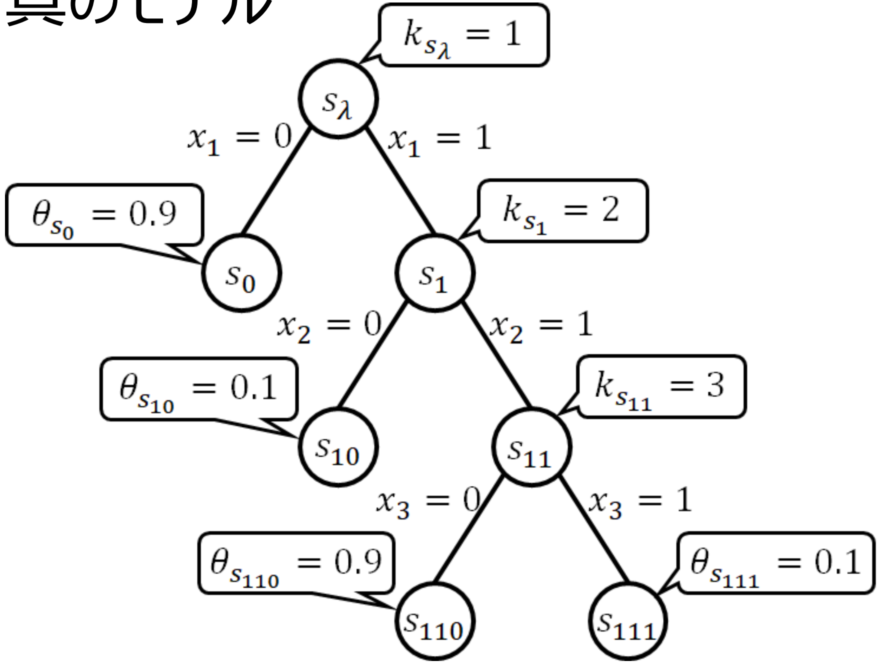
付録：実データ実験

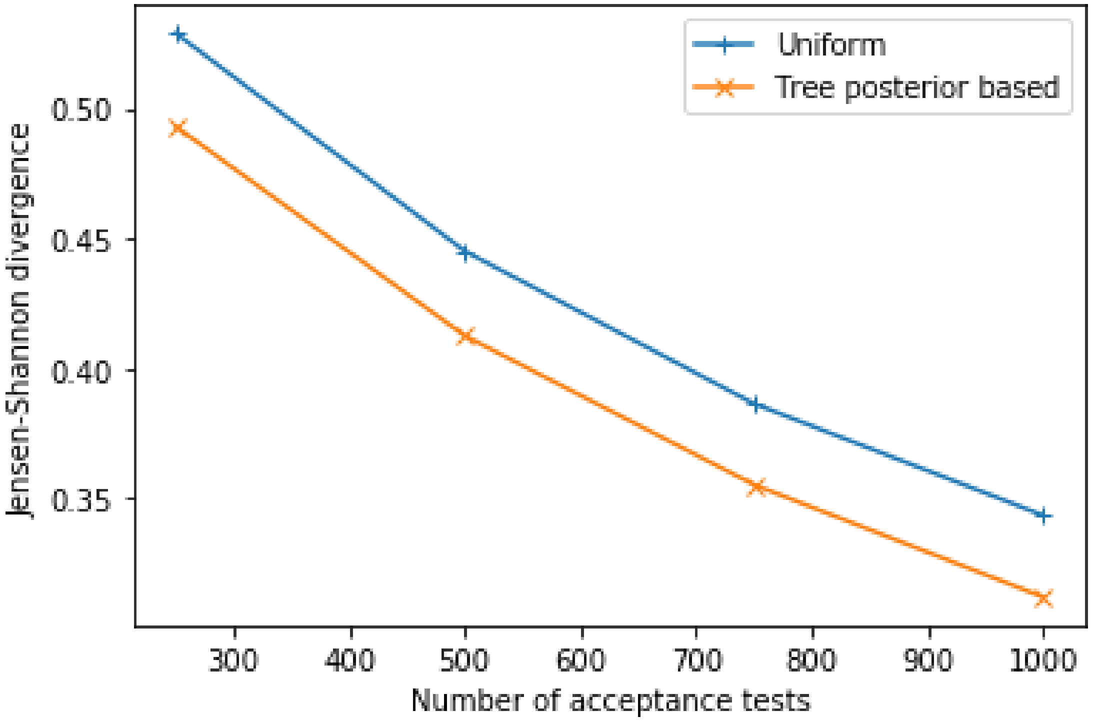
- Titanicデータにおける MCMC採択回数に対する尤度の挙動



付録：真の事後分布への収束確認

- 真の事後分布とのJensen-Shannonダイバージェンス
  - 実験条件
    - $\mathcal{X} = \mathcal{Y} = \{0,1\}$
    - $D_{\max} = 3, K = 5$
      →モデル総数78,125
    - 真のモデル





参考文献

- 松嶋敏泰，早稲田大学データ科学教育チーム，データ科学入門I データに基づく意思決定の基礎，サイエンス社，2022.
- L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, Classification and regression trees. CRC press, 1984.
- 須子統太，野村亮，松嶋敏泰，平澤茂一，決定木モデルにおける予測アルゴリズムについて，電子情報通信学会技術研究報告，vol. 103, np. 246, pp.93-98, 2003.
- Leo Breiman. Random forests. Machine learning, 45(1):5–32, 2001.
- Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30:3146–3154, 2017.
- Hugh A. Chipman, Edward I. George, and Robert E. McCulloch. BART: Bayesian additive regression trees. The Annals of Applied Statistics, 4(1):266 – 298, 2010.
- Dobashi, N.; Saito, S.; Nakahara, Y.; Matsushima, T. Meta-Tree Random Forest: Probabilistic Data-Generative Model and Bayes Optimal Prediction. Entropy 2021, 23, 768. https://doi.org/10.3390/e23060768

参考文献

- Nakahara, Y.; Saito, S.; Kamatsuka, A.; Matsushima, T. Probability Distribution on Full Rooted Trees. Entropy 2022, 24, 328. https://doi.org/10.3390/e24030328
- Y. Nakahara, S. Saito, A. Kamatsuka and T. Matsushima, "Probability Distribution on Rooted Trees," 2022 IEEE International Symposium on Information Theory (ISIT), 2022, pp. 174-179, doi: 10.1109/ISIT50566.2022.9834481.
- Nakahara, Y.; Matsushima, T. A Stochastic Model for Block Segmentation of Images Based on the Quadtree and the Bayes Code for It. *Entropy* **2021**, *23*, 991. https://doi.org/10.3390/e23080991
- Nakahara, Y.; Matsushima, T. Stochastic Model of Block Segmentation Based on Improper Quadtree and Optimal Code under the Bayes Criterion. Entropy 2022, 24, 1152. https://doi.org/10.3390/e24081152
- 中原悠太，齋藤翔太，一條尚希，風間皐希，松嶋敏泰，決定木モデルにおけるメタツリーに対するマルコフ連鎖モンテカルロ法，第45回情報理論とその応用シンポジウム（SITA2022），pp.47-52，2022.
- 岡 凌平，中原悠太，松嶋敏泰，二次元離散ウェーブレットパケット変換の基底が未知の場合のベイズ基準のもと最適なノイズ除去アルゴリズム，第44回情報理論とその応用シンポジウム（SITA2021），pp.47-52，2022.

参考文献

- 田島慶斗，一條尚希，島田航志，松嶋敏泰，決定木モデル上のベイズ最適な予測を近似するためのメタツリー集合並列構成アルゴリズムに関する一考察，第45回情報理論とその応用シンポジウム (SITA2022)，pp.59-64，2022.
- 于 文斌，風間皐希，中原悠太，一條尚希，齋藤翔太，松嶋敏泰，決定木モデルに対するベイズ最適な予測のメタツリーブースティング法による近似，信学技報, vol. 121, no. 327, IT2021-67, pp. 219-224, 2022年1月.
- 馬庭良太，一條尚希，島田航志，松嶋敏泰，決定木モデル上の分類におけるメタツリー集合のブースティング的構築についての一考察，第45回情報理論とその応用シンポジウム (SITA2022)，pp.53-58，2022.
- Chen, T.; and Guestrin, C. 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’16, 785–794.