

# 最適化理論

<http://www.cs.miyazaki-u.ac.jp/~date/lectures/optimization/>

伊達 章

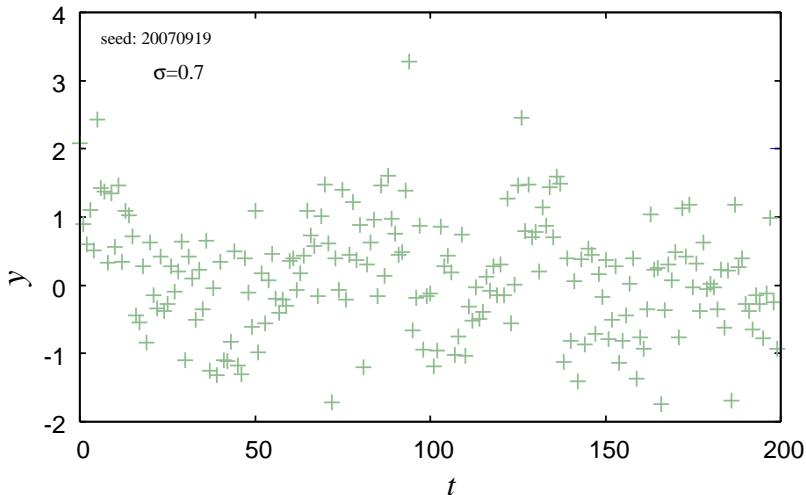
宮崎大学 工学部 情報システム工学科

2016年6月24日

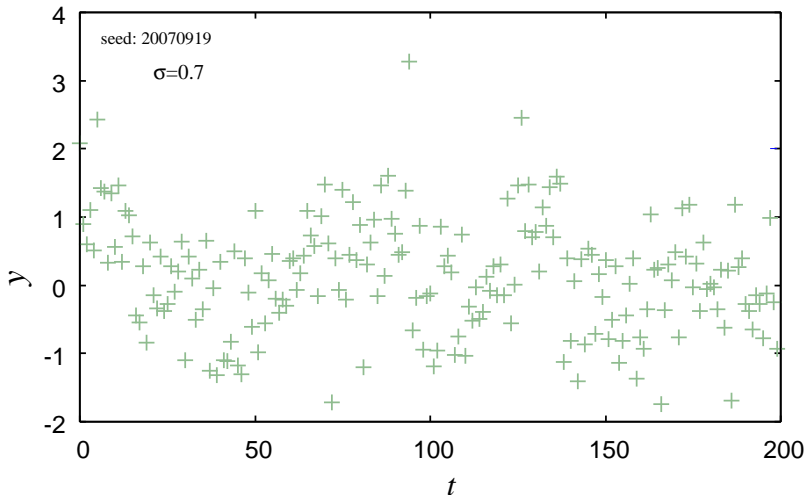
# 講義のスケジュール (案)

1. 講義の概要
2. 数学的準備：曲線と曲面
3. 数学的準備：1次形式と2次形式
4. 数学的準備：2次形式の標準形
5. 関数の極値：関数の勾配と等高線，関数の極値
6. 関数の極値：ラグランジュの未定乗数法
7. 関数の最適化：勾配法・ニュートン法
8. 関数の最適化：共役勾配法
9. 統計的最適化：正規分布，最尤推定
10. 動的計画法（その1）
11. 動的計画法（その2）
12. 最小二乗法：連立一次方程式，特異値分解と一般化逆行列
13. 最小二乗法（その2）
14. まとめ 7/15 金
15. 定期試験 7/22 金，解説

# 観測データ (時系列)

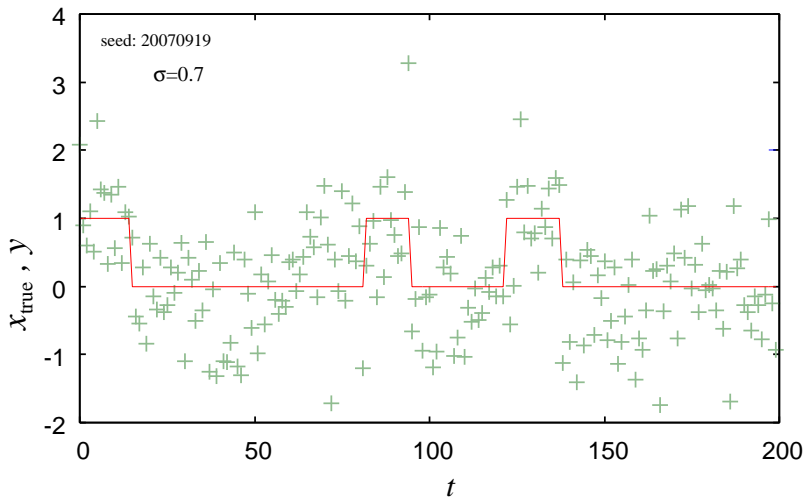


# 観測データ (時系列)



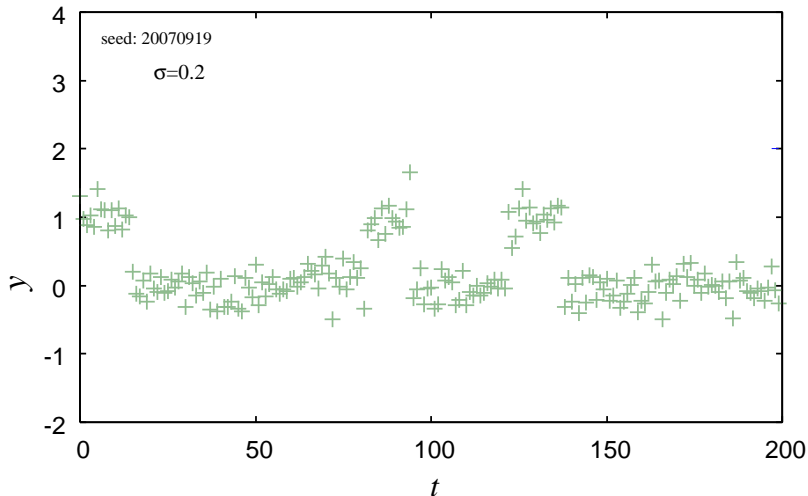
もとの信号は 0 か 1. 復元したい!

# 観測データ (時系列)



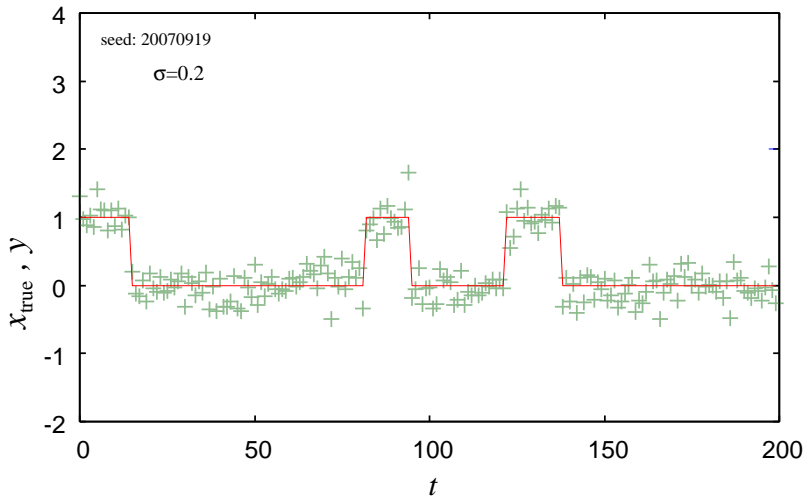
もとの信号は 0 か 1. 復元したい!

# 観測データ (時系列)



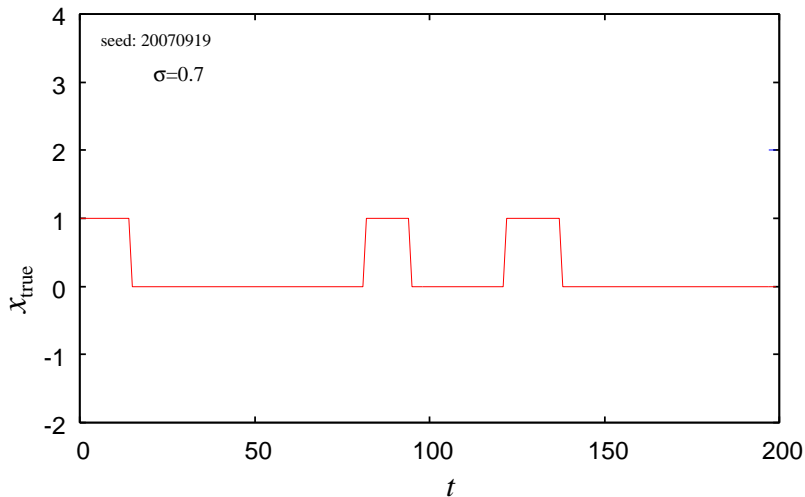
こちらのほうが復元しやすい

# 観測データ (時系列)



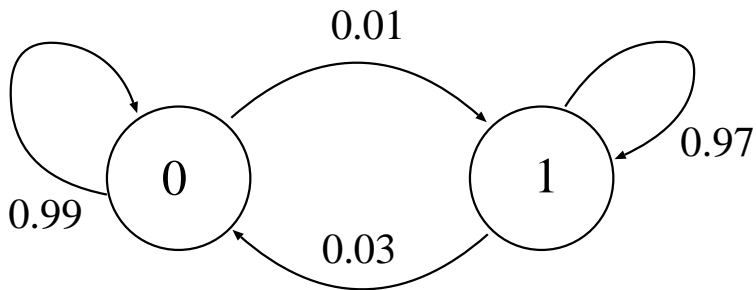
こちらのほうが復元しやすい

# もとの信号 $x_{\text{true}}$

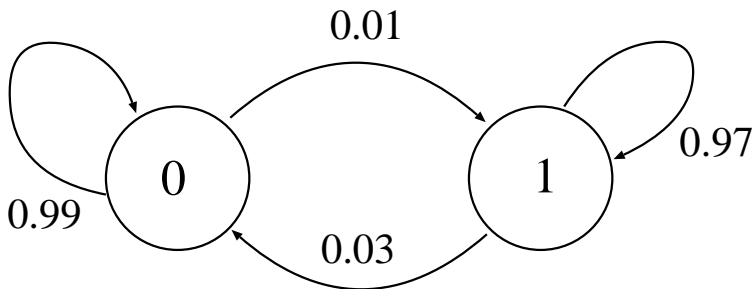


マルコフ的情報源

## マルコフ的情報源

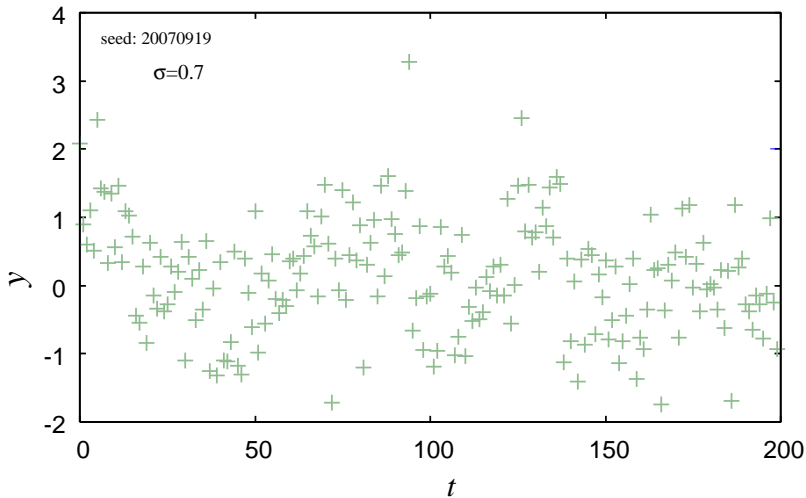


## マルコフ的情報源



0000000011111111110000000000

# 観測データ (時系列)



# 多段決定問題

多変数関数  $f(\boldsymbol{x})$  の最大化

$$J = f(x_1, x_2, \dots, x_n) \rightarrow \max$$

# 多段決定問題

多変数関数  $f(\mathbf{x})$  の最大化

$$J = f(x_1, x_2, \dots, x_n) \rightarrow \max$$

$n = 10, x_i \in \{0, 1\}$  の場合

	$\mathbf{x}$	$J$	
0	0000000000	$f(\mathbf{x}_0)$	
1	0000000001	$f(\mathbf{x}_1)$	
2	0000000010	$f(\mathbf{x}_2)$	
	$\vdots$		
$k$	0011101011	$f(\mathbf{x}_k)$	← 最大
	$\vdots$		
1023	1111111111	$f(\mathbf{x}_{2^n-1})$	

$$\max_i f(\mathbf{x}_i) = f(\mathbf{x}_k), \quad \operatorname{argmax}_i f(\mathbf{x}_i) = k$$

# 多段決定問題

多変数関数  $f(\mathbf{x})$  の最大化

$$J = f(x_1, x_2, \dots, x_n) \rightarrow \max$$

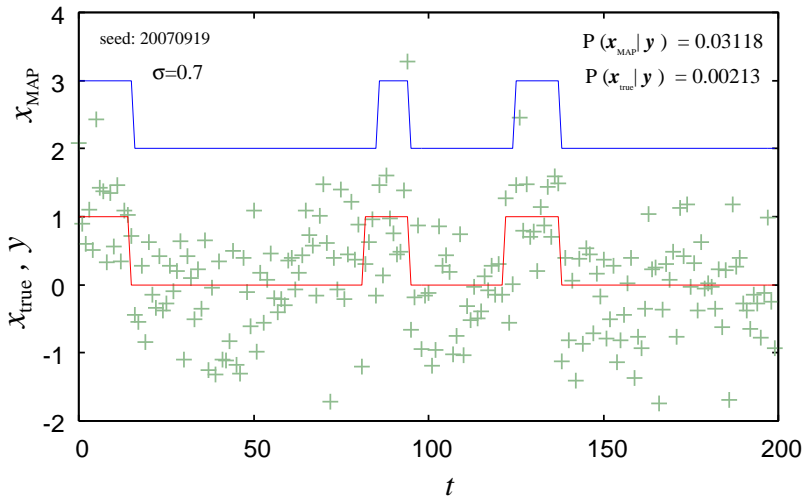
$n = 100, x_i \in \{0, 1\}$  の場合.  $2^{100} = (2^{10})^{10} \approx (10^3)^{10} = 10^{30}$

	$\mathbf{x}$	$J$	
0	0000000000	$f(\mathbf{x}_0)$	
1	0000000001	$f(\mathbf{x}_1)$	
2	0000000010	$f(\mathbf{x}_2)$	
3	0000000011	$f(\mathbf{x}_3)$	
	$\vdots$		
$k$	0011101011	$f(\mathbf{x}_k)$	← 最大
	$\vdots$		
$\approx 10^{30}$	111 $\dots$ 111	$f(\mathbf{x}_{2^n-1})$	

## 課題

- 例題：  $n = 200$ ,  $\mathbf{y} = (y_1, \dots, y_n)$
- 目的：  $\mathbf{x}^* = \operatorname{argmax}_x f(\mathbf{x}), x_i \in \{0, 1\}$  の計算
- まもとには計算できない.
- 動的計画法を使い, この問題を解決する.
- 仕組み理解し, コードを書き, 問題を解く.
- 口頭試問をクリア  
(1人2回まで挑戦可. 最大10分/回)  
⇒ 成績に10~20点を加算

# 事後確率最大にする値 $x_{\text{MAP}}$



# 多段決定問題

## 多変数関数の最大化

$$J = p(\mathbf{x}|\mathbf{y}) \rightarrow \max$$

$n = 200, x_i \in \{0, 1\}$  の場合.  $2^{200} = (2^{10})^{20} \approx (10^3)^{20} = 10^{60}$

	$\mathbf{x}$	$J$	
0	0000000000	$f(\mathbf{x}_0)$	
1	0000000001	$f(\mathbf{x}_1)$	
2	0000000010	$f(\mathbf{x}_2)$	
3	0000000011	$f(\mathbf{x}_3)$	
	$\vdots$		
$k$	0011101011	$f(\mathbf{x}_k)$	← 最大
	$\vdots$		
$\approx 10^{60}$	111 $\cdots$ 111	$f(\mathbf{x}_{2^n-1})$	

## 課題

- 例題：  $n = 200$ ,  $\mathbf{y} = (y_1, \dots, y_n)$
- 目的：  $\mathbf{x}^* = \operatorname{argmax}_x f(\mathbf{x}), x_i \in \{0, 1\}$  の計算
- まもとには計算できない.
- 動的計画法を使い, この問題を解決する.
- 仕組み理解し, コードを書き, 問題を解く.
- 口頭試問をクリア  
(1人2回まで挑戦可. 最大10分/回)  
⇒ 成績に10~20点を加算
- 締切：定期試験の一週間前 (7/14木) まで

## 基本知識（確率・統計の復習）

- 平均  $\mu$ , 分散  $\sigma^2$ , 標準偏差  $\sigma$
- 確率分布：一様分布, 正規分布
- 擬似乱数の生成
- 最尤推定
- 同時確率, 条件付き確率
- マルコフ的情報源
- ベイズの公式, 事前確率・事後確率
- 事後確率最大化
  
- 動的計画法（第8章）

## 確率，条件付き確率

	$B_1$ (風邪)	$B_2$ (風邪なし)	$p(A_i)$
$A_1$ (熱あり)	0.55	0.05	0.60
$A_2$ (熱なし)	0.10	0.30	0.40
$p(B_j)$	0.65	0.35	

例

同時確率  $p(A_1, B_1) = 0.55$

周辺確率  $p(A_1) = \sum_i p(A_1, B_i) = p(A_1) = 0.6$

条件付き確率

熱の有無を知る  $\Rightarrow$  風邪であるかどうか検討がつく：

$$p(B_1|A_1) = \frac{p(B_1)p(A_1|B_1)}{p(A_1)} = \frac{p(A_1, B_1)}{p(A_1)} = \frac{0.55}{0.6} \approx 0.92$$

# ベイズの公式

- ベイズの公式

熱があったとしよう。その時、風邪のあるなしの確率

$$p(B_1|A_1) = \frac{p(B_1)p(A_1|B_1)}{p(A_1)} = \frac{p(A_1, B_1)}{p(A_1)} = \frac{0.55}{0.6} \approx 0.92$$

$$p(B_2|A_1) = \frac{p(B_2)p(A_1|B_2)}{p(A_1)} = \frac{p(A_1, B_2)}{p(A_1)} = \frac{0.05}{0.6} \approx 0.08$$

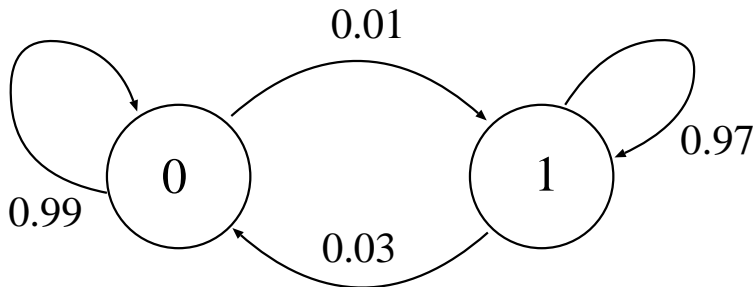
- 事後確率最大化（ベイズ推定）  $\operatorname{argmax}_i p(B_i|A_1) = 1$

風邪であることの方が確率が大  $\Rightarrow$  風邪であると推定

入力（観測値）： 熱のあるなし

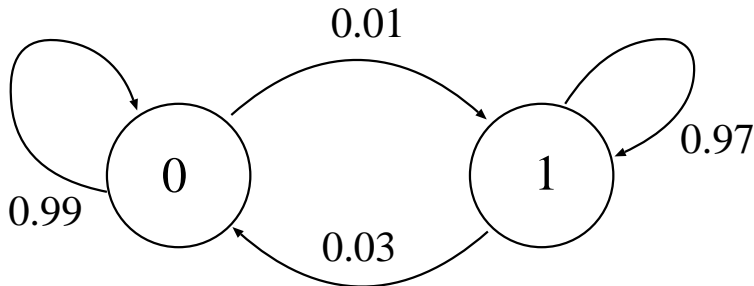
$\Rightarrow$  出力（推定値） 風邪かどうか

## マルコフ的情報源



0000000011111111110000000000

## マルコフ的情報源



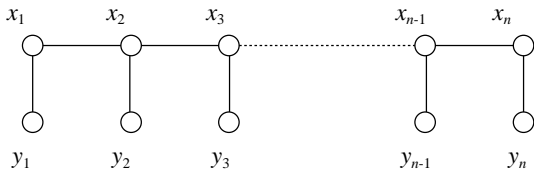
0000000011111111110000000000

$$p(\mathbf{x}) = 0.5 \times 0.99 \times 0.99 \times 0.99 \times 0.99 \dots$$

$$p(\mathbf{x}) = p(x_1) \times p(x_2|x_1) \times p(x_3|x_2) \times p(x_4|x_3) \dots$$

# 事後確率 $p(\mathbf{x}|\mathbf{y})$ の最大化

横棒は条件付確率，縦棒は尤度に対応



$$\mathbf{x}^* = \operatorname{argmax}_{x_1, x_2, \dots, x_n} p(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n)$$

$$\mathbf{x}^* = \operatorname{argmax}_{x_1, x_2, \dots, x_n} p(x_1) \prod_{i=2}^n p(x_i | x_{i-1}) \prod_{i=1}^n p(y_i | x_i)$$

## 使って良い知識

データ (観測値) :  $\mathbf{y} = (y_1, \dots, y_n)$

モデル :  $p(\mathbf{x}), p(\mathbf{y}|\mathbf{x})$

$$p(x_1) = \begin{cases} 0.5 & \text{if } x_1 = 0 \\ 0.5 & \text{if } x_1 = 1 \end{cases}$$

$$p(x_{i+1}|x_i) = \begin{cases} 0.99 & \text{if } x_i = 0, x_{i+1} = 0 \\ 0.01 & \text{if } x_i = 0, x_{i+1} = 1 \\ 0.97 & \text{if } x_i = 1, x_{i+1} = 1 \\ 0.03 & \text{if } x_i = 1, x_{i+1} = 0 \end{cases}$$

$$p(y_i|x_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(y_i - x_i)^2}{2\sigma^2} \right\}$$

以降, この課題では, 特に指定しない限り  $\sigma = 0.7$ .

# 事後確率最大化

$$\mathbf{x}^* = \operatorname{argmax}_x p(\mathbf{x}|\mathbf{y})$$

$p(\mathbf{x}|\mathbf{y})$  は知らない.  $p(\mathbf{x}), p(\mathbf{y}|\mathbf{x})$  は与えられている.

# 事後確率最大化

$$\mathbf{x}^* = \operatorname{argmax}_x p(\mathbf{x}|\mathbf{y})$$

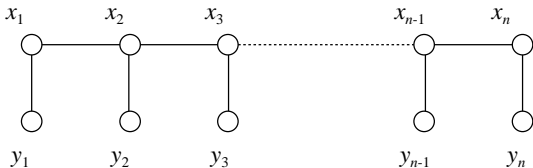
$p(\mathbf{x}|\mathbf{y})$  は知らない.  $p(\mathbf{x}), p(\mathbf{y}|\mathbf{x})$  は与えられている.

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}) &= \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})} \quad (\text{ベイズの公式}) \\ &= \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})} \end{aligned}$$

$\mathbf{y}$  は観測値  $\Rightarrow p(\mathbf{y})$  は定数

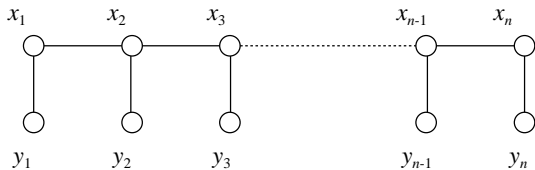
$$\mathbf{x}^* = \operatorname{argmax}_x p(\mathbf{x})p(\mathbf{y}|\mathbf{x})$$

# 事後確率最大化



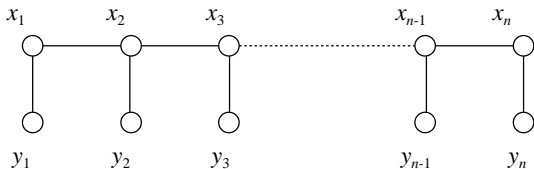
$$\begin{aligned} \mathbf{x}^* &= \operatorname{argmax}_{\mathbf{x}} p(\mathbf{x})p(\mathbf{y}|\mathbf{x}) \\ &= \operatorname{argmax}_{\mathbf{x}} p(x_1) \prod_{i=2}^n p(x_i|x_{i-1}) \prod_{i=1}^n p(y_i|x_i) \\ &= \operatorname{argmax}_{\mathbf{x}} \log p(x_1) + \sum_{i=2}^n \log p(x_i|x_{i-1}) + \sum_{i=1}^n \log p(y_i|x_i) \end{aligned}$$

# 事後確率最大化



$$\begin{aligned} J &= \log p(x_1) + \sum_{i=2}^n \log p(x_i | x_{i-1}) + \sum_{i=1}^n \log p(y_i | x_i) \\ &= \log p(x_1) + \log p(y_1 | x_1) + \sum_{i=1}^{n-1} \left\{ \log p(x_{i+1} | x_i) + \log p(y_{i+1} | x_{i+1}) \right\} \\ &\quad y_i \text{ は定数} \Rightarrow \text{教科書 p.216 (8.2) 式と同じ形} \\ &= f_1(x_1) + h_1(x_1, x_2) + h_2(x_2, x_3) + \cdots + h_{n-1}(x_{n-1}, x_n) \end{aligned}$$

# 事後確率最大化



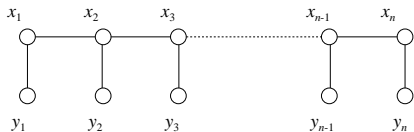
$$\begin{aligned} J &= \log p(x_1) + \sum_{i=2}^n \log p(x_i | x_{i-1}) + \sum_{i=1}^n \log p(y_i | x_i) \\ &= \log p(x_1) + \log p(y_1 | x_1) + \sum_{i=1}^{n-1} \left\{ \log p(x_{i+1} | x_i) + \log p(y_{i+1} | x_{i+1}) \right\} \\ &\quad y_i \text{ は定数} \Rightarrow \text{教科書 p.216 (8.2) 式と同じ形} \\ &= f_1(x_1) + h_1(x_1, x_2) + h_2(x_2, x_3) + \cdots + h_{n-1}(x_{n-1}, x_n) \end{aligned}$$

- $f_1(x_1) = \log p(x_1) + \log p(y_1 | x_1)$
- $h_i(x_i, x_{i+1}) = \log p(x_{i+1} | x_i) + \log p(y_{i+1} | x_{i+1})$

## 基本知識（確率・統計の復習）

- 平均  $\mu$ , 分散  $\sigma^2$ , 標準偏差  $\sigma$
- 確率分布：一様分布, 正規分布
- 擬似乱数の生成
- 最尤推定
- 同時確率, 条件付き確率
- マルコフ的情報源
- ベイズの公式, 事前確率・事後確率
- 事後確率最大化
  
- 動的計画法（第8章）

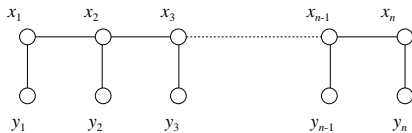
# 動的計画法



$$\begin{aligned} J &= \log p(x_1) + \log p(y_1|x_1) + \sum_{i=1}^{n-1} \left\{ \log p(x_{i+1}|x_i) + \log p(y_{i+1}|x_{i+1}) \right\} \\ &= f_1(x_1) + h_1(x_1, x_2) + h_2(x_2, x_3) + \dots + h_{n-1}(x_{n-1}, x_n) \end{aligned}$$

- $x_1$  に着目  $\Rightarrow f_1(x_1), h_1(x_1, x_2)$  にしか関係しない

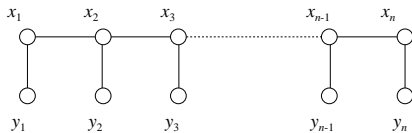
# 動的計画法



$$\begin{aligned} J &= \log p(x_1) + \log p(y_1|x_1) + \sum_{i=1}^{n-1} \left\{ \log p(x_{i+1}|x_i) + \log p(y_{i+1}|x_{i+1}) \right\} \\ &= f_1(x_1) + h_1(x_1, x_2) + h_2(x_2, x_3) + \dots + h_{n-1}(x_{n-1}, x_n) \end{aligned}$$

- $x_1$  に着目  $\Rightarrow f_1(x_1), h_1(x_1, x_2)$  にしか関係しない
- $f_1(x_1) + h_1(x_1, x_2)$  を最大にするよう  $x_1$  を選ぶ

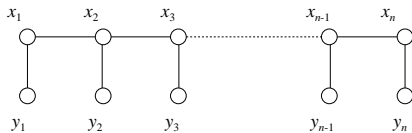
# 動的計画法



$$\begin{aligned} J &= \log p(x_1) + \log p(y_1|x_1) + \sum_{i=1}^{n-1} \left\{ \log p(x_{i+1}|x_i) + \log p(y_{i+1}|x_{i+1}) \right\} \\ &= f_1(x_1) + h_1(x_1, x_2) + h_2(x_2, x_3) + \cdots + h_{n-1}(x_{n-1}, x_n) \end{aligned}$$

- $x_1$  に着目  $\Rightarrow f_1(x_1), h_1(x_1, x_2)$  にしか関係しない
- $f_1(x_1) + h_1(x_1, x_2)$  を最大にするよう  $x_1$  を選ぶ
- $\uparrow x_2$  の値がわかっていないければ選べない

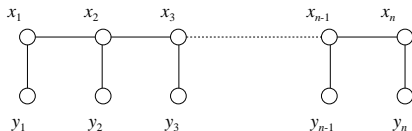
# 動的計画法



$$\begin{aligned} J &= \log p(x_1) + \log p(y_1|x_1) + \sum_{i=1}^{n-1} \left\{ \log p(x_{i+1}|x_i) + \log p(y_{i+1}|x_{i+1}) \right\} \\ &= f_1(x_1) + h_1(x_1, x_2) + h_2(x_2, x_3) + \cdots + h_{n-1}(x_{n-1}, x_n) \end{aligned}$$

- $x_1$  に着目  $\Rightarrow f_1(x_1), h_1(x_1, x_2)$  にしか関係しない
- $f_1(x_1) + h_1(x_1, x_2)$  を最大にするよう  $x_1$  を選ぶ
- $\uparrow x_2$  の値がわかっていないければ選べない
- そこで,  $x_2$  の可能なすべての値  $(0, 1)$  に対して, 以下を計算
- $f_2(x_2) = \max_{x_1} \{f_1(x_1) + h_1(x_1, x_2)\}$   
 $\hat{x}_1(x_2) = \operatorname{argmax}_{x_1} \{f_1(x_1) + h_1(x_1, x_2)\}$   
 $J = f_2(x_2) + h_2(x_2, x_3) + \cdots + h_{n-1}(x_{n-1}, x_n)$

# 動的計画法



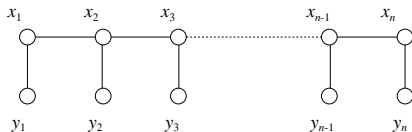
$$J = f_2(x_2) + h_2(x_2, x_3) + \cdots + h_{n-1}(x_{n-1}, x_n)$$

- 変数の数が1つ減った. これを続けていけばよい.
- $x_2$  に着目  $\Rightarrow f_2(x_2), h_2(x_2, x_3)$  にしか関係しない
- $f_2(x_2) + h_2(x_2, x_3)$  を最大にするよう  $x_2$  を選ぶ
- $\uparrow x_3$  の値がわかっていないければ選べない
- そこで,  $x_3$  の可能なすべての値  $(0, 1)$  に対して, 以下を計算

- $f_3(x_3) = \max_{x_2} \{f_2(x_2) + h_2(x_2, x_3)\}$

$$\hat{x}_2(x_3) = \operatorname{argmax}_{x_2} \{f_2(x_2) + h_2(x_2, x_3)\}$$

# 動的計画法



$$J = f_{n-1}(x_{n-1}) + h_{n-1}(x_{n-1}, x_n)$$

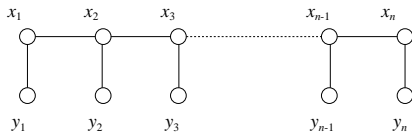
- $x_n$  の可能なすべての値  $(0, 1)$  に対して, 以下を計算

- $$f_n(x_n) = \max_{x_{n-1}} \{f_{n-1}(x_{n-1}) + h_{n-1}(x_{n-1}, x_n)\}$$

$$\hat{x}_{n-1}(x_n) = \operatorname{argmax}_{x_{n-1}} \{f_{n-1}(x_{n-1}) + h_{n-1}(x_{n-1}, x_n)\}$$

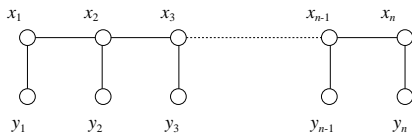
- $$x_n^* = \operatorname{argmax}_{x_n} f_n(x_n)$$

# 動的計画法



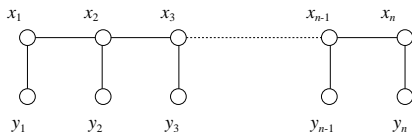
- $x_n^* = \operatorname{argmax}_{x_n} f_n(x_n)$

# 動的計画法



- $x_n^* = \operatorname{argmax}_{x_n} f_n(x_n)$
- $x_{n-1}^* = \hat{x}_{n-1}(x_n^*)$

# 動的計画法



- $x_n^* = \operatorname{argmax}_{x_n} f_n(x_n)$
- $x_{n-1}^* = \hat{x}_{n-1}(x_n^*)$
- $x_{n-2}^* = \hat{x}_{n-2}(x_{n-1}^*)$
- $\dots \rightarrow x_1^* = \hat{x}_1(x_2^*)$

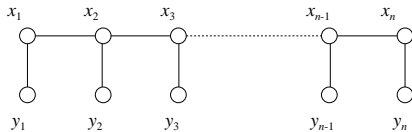
# データ構造

- $i = 1, \dots, 200, a = 0, 1$
- $x_i \cdots$  int x[i] 真の値
- $y_i \cdots$  double y[i] 観測データ
- $\hat{x}_i \cdots$  int xhat[i] 推定値
- $f_i(x_i) \cdots$  double f[i][a]  
 $i$  番目の変数の値が  $x_i = a$  のとき, そこまでに至る最適経路の尤もらしさ.
- $\hat{x}_i(x_{i+1}) \cdots$  int xhat[i][a]  
 $x_{i+1} = a$  のとき,  $x_i$  が取るべき値

# オプション課題について

- 例題 :  $n = 200, \mathbf{y} = (y_1, \dots, y_n)$
- 目的 :  $\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmax}} f(\mathbf{x}), x_i \in \{0, 1\}$  の計算
  - まもとには計算できない問題.
  - 動的計画法を使い解決.
  - 1. 仕組み理解し, 2. コードを書き, 問題を解く.
  - 3. 口頭試問をクリア  
(1人2回まで挑戦可. 最大10分/回)  
⇒ 成績に10~20点を加算
- 締切 : 定期試験の一週間前 (7/14木) まで
- 口頭試問 対応日 (A-333)
  - 7/4, 11 (月) 終日
  - 7/7, 14 (木) (14:50~16:20) を除く.
- 質問は私が不在でも A-334 の大学院生・4年生が対応

# 注意事項：口頭試問について



- 上記の絵を使い，説明をはじめる
  - 線を結ぶ両端の○に値が入ると，棒に値が付く
  - それを全部掛け算した値を最大にしたい
  - 下の○の値は与えられている.
  - 上の○には 0,1 が入る. 全部で  $2^{200}$  通り
  - どうする？
- 「分かった！」ことをアピールせよ
  - 動的計画法の仕組みを理解した
  - 自分でコードを書いて走らせた
- 最速3分以内に終了
- ノートパソコン持参が望ましい（実験結果を見せる）

# 情報工学特別講義について

- 4日間の集中講義：  
2016年8月29日（月）～9月1日（木）
  - 初回 10:30～. 場所は A-116.
  - 担当：松葉育雄先生（千葉大学）  
<http://www.eng.chiba-u.ac.jp/outProfile.tsv?no=1191>
- 内容：カオス入門  
カオスに対する入門的な解説  
<http://www.cs.miyazaki-u.ac.jp/date/lectures/2016sp/>
- 追加で履修登録可能！ → 伊達まで連絡下さい。

# 研究室のテーマ

- 脳は情報処理をする素晴らしい装置!

# 研究室のテーマ

- 脳は情報処理をする素晴らしい装置!
- コンピュータと脳の比較  
特に情報表現と計算の観点から

# 研究室のテーマ

- 脳は情報処理をする素晴らしい装置!
- コンピュータと脳の比較  
特に情報表現と計算の観点から
- 数理モデルによる脳の情報原理の理解

# 研究室のテーマ

- 脳は情報処理をする素晴らしい装置!
- コンピュータと脳の比較  
特に情報表現と計算の観点から
- 数理モデルによる脳の情報原理の理解
  - 学習・記憶の神経回路モデル
  - 自己組織化のモデル
  - 思考・意志決定のモデル

# 研究室紹介について

- 脳の情報処理, パターン情報処理  
学習・記憶, 自己組織化の数理モデル
- 7/12火曜日は4年生が対応 (私は席を外す?)
- A-333 に話をしにきてください  
(いつでもよい)

# 講義のスケジュール (案)

1. 講義の概要
2. 数学的準備：曲線と曲面
3. 数学的準備：1次形式と2次形式
4. 数学的準備：2次形式の標準形
5. 関数の極値：関数の勾配と等高線，関数の極値
6. 関数の極値：ラグランジュの未定乗数法
7. 関数の最適化：勾配法・ニュートン法
8. 関数の最適化：共役勾配法
9. 統計的最適化：正規分布，最尤推定
10. 動的計画法（その1）
11. 動的計画法（その2）
12. 最小二乗法：連立一次方程式，特異値分解と一般化逆行列
13. 最小二乗法（その2）
14. まとめ 7/15 金
15. 定期試験 7/22 金，解説

終