

IBIS2008

2008年10月29日～31日

# 密度比推定の手法と応用

杉山将 (東工大)



# 重点サンプリング

- 確率密度  $p'(x)$  に関する期待値を, **別の確率密度**  $p(x)$  を用いて計算する統計手法
- $p'(x)$  から標本を取るのが難しいとき, 代わりに簡単な分布  $p(x)$  から標本を取る.

$$\begin{aligned}\mathbb{E}_{\mathbf{x} \sim p'(\mathbf{x})} [A(\mathbf{x})] &= \int A(\mathbf{x}) p'(\mathbf{x}) d\mathbf{x} \\ &= \int A(\mathbf{x}) \frac{p'(\mathbf{x})}{p(\mathbf{x})} p(\mathbf{x}) d\mathbf{x} \\ &= \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x})} \left[ A(\mathbf{x}) \frac{p'(\mathbf{x})}{p(\mathbf{x})} \right]\end{aligned}$$

密度比



# 本発表の構成

3

## 1. 密度比は、重点サンプリングだけでなく、様々な統計的データ解析に役立つ！

- パターン認識, 非定常環境下での適応学習, ドメイン適応, マルチタスク学習, 異常値検出, 時系列の変化点検知, 特徴選択, 次元削減, 独立成分分析, 条件付き確率推定, 二標本検定など

## 2. 密度比推定の手法

## 3. 実験

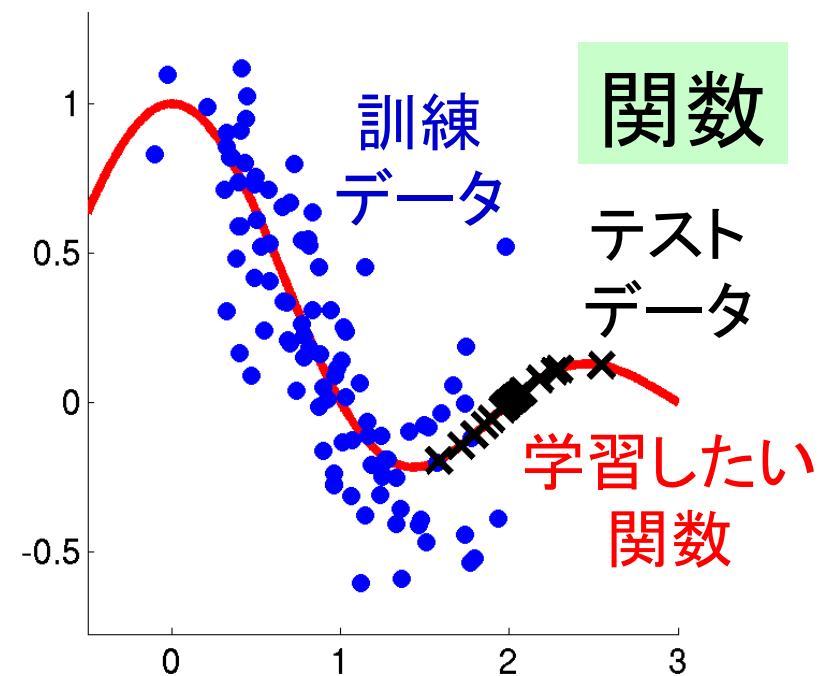
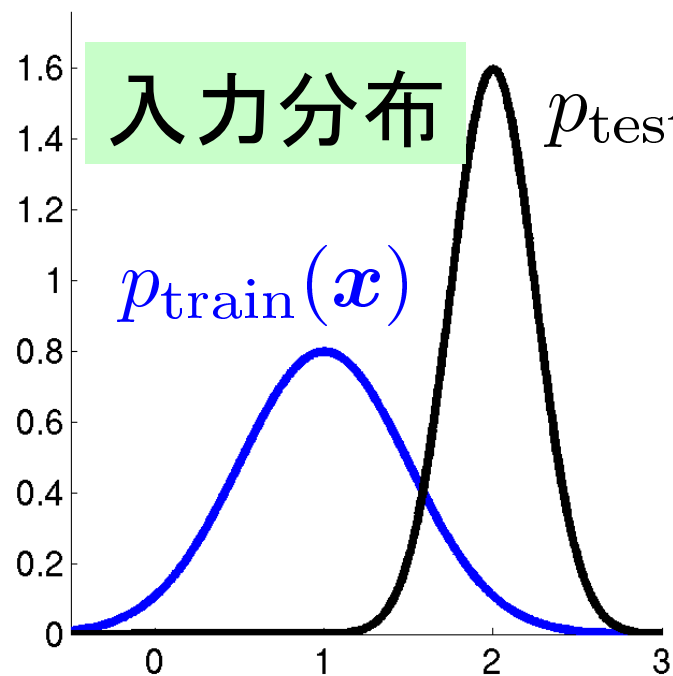
$$\frac{p'(\boldsymbol{x})}{p(\boldsymbol{x})}$$

# 例：非定常環境下での適応学習 4

■ **共変量シフト**：訓練時とテスト時で入力分布が変化するが，入出力関数は不変

- ロボット制御則の獲得
- アンケートデータの学習
- 脳波，音声，言語データの学習

Shimodaira (JSPI2000)



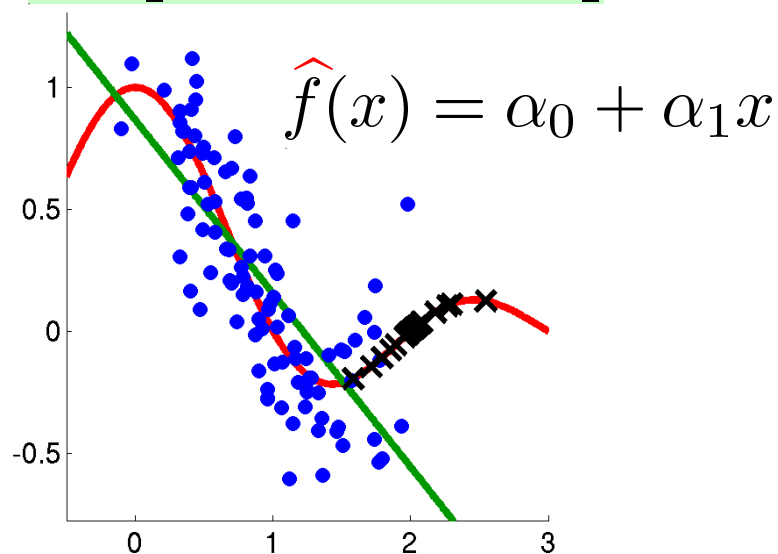
# 共変量シフト下での学習

5

Shimodaira (JSPI2000), Sugiyama & Müller (Stat&Deci2007),  
Sugiyama, Krauledat & Müller (JMLR2007)

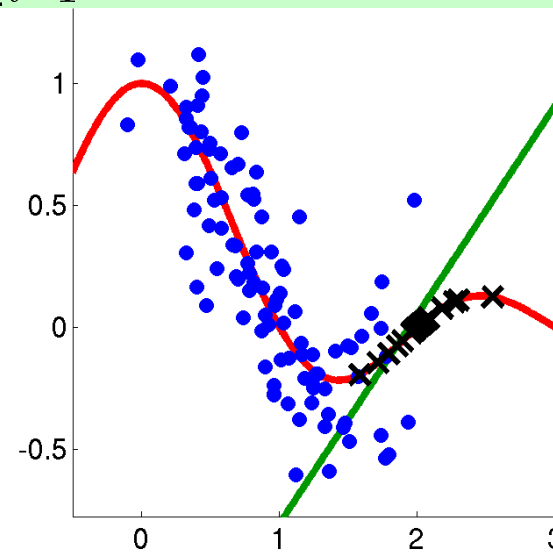
■ 通常の最小二乗法は  
一貫性がない

$$\min_{\alpha} \left[ \sum_{i=1}^n (\hat{f}(x_i) - y_i)^2 \right]$$



■ 密度比重み付きの最小  
二乗法は一貫性がある

$$\min_{\alpha} \left[ \sum_{i=1}^n \frac{p_{\text{test}}(x_i)}{p_{\text{train}}(x_i)} (\hat{f}(x_i) - y_i)^2 \right]$$



訓練誤差(尤度)に基づく学習法全てに適応可能!

# ドメイン適応とマルチタスク学習

6

- **ドメイン適応**: あるドメインでの学習結果を別のドメインでの学習に活用する

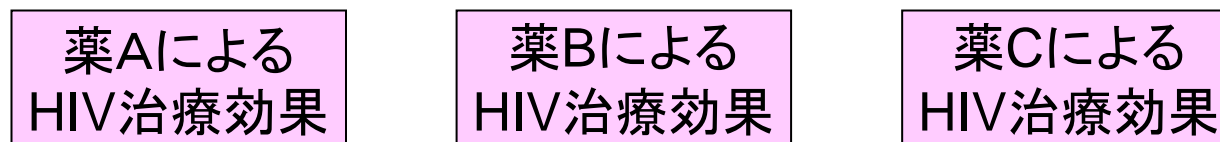
Storkey & Sugiyama (NIPS2006)

Tsuboi, Kashima, Hido, Bickel & Sugiyama (SDM2008)



- **マルチタスク学習**: 複数の似た問題を同時に解くことにより, それぞれを解くよりも良い解を得る

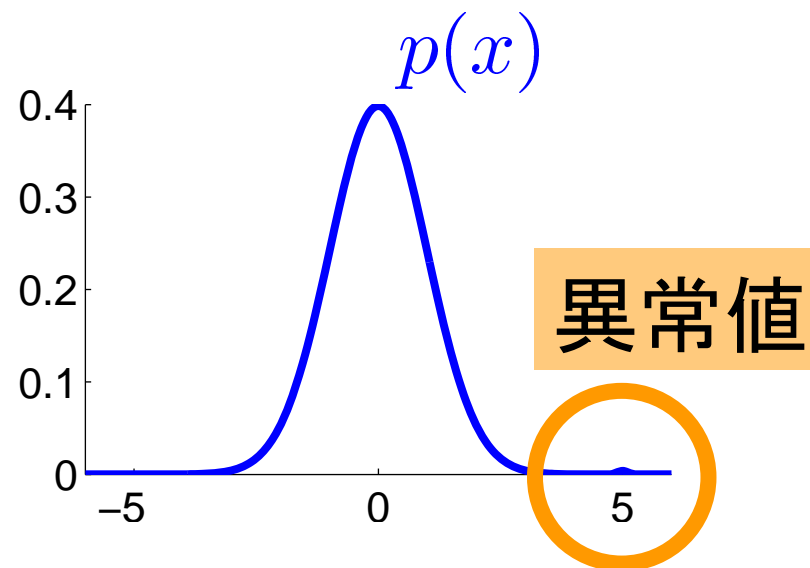
Bickel, Bogojeska, Lengauer & Scheffer (ICML2008)



- 同様な**密度比重み付け**によりうまく対応できる.

# 例：異常値検出

- **異常値検出**：標本に含まれる異常値を見つける
  - 生産ラインの不良品検出
  - ネットワークシステムへの不正侵入検出
  - 文書集合からの新規トピック発見
- **確率密度の低い**標本を異常値とみなす

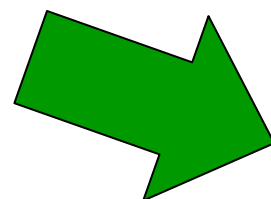
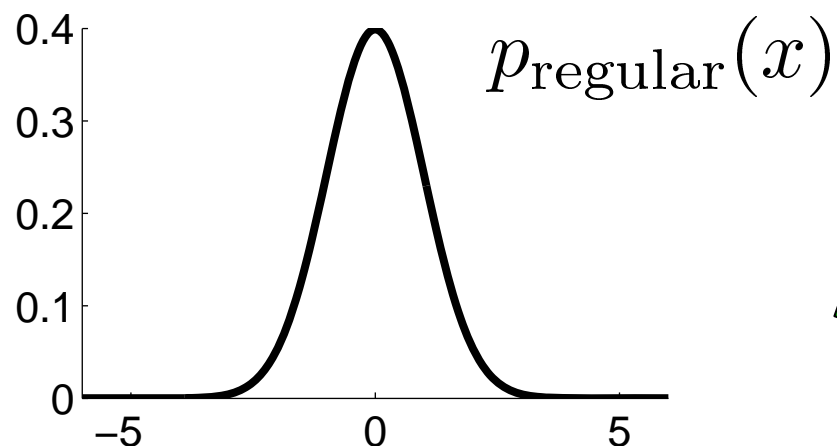


# 例：異常値検出（続き）

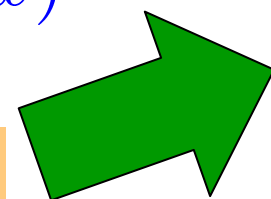
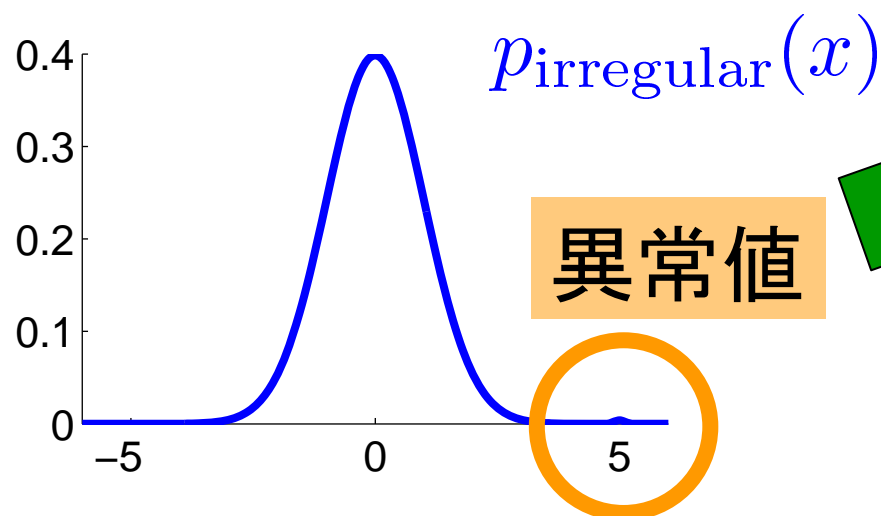
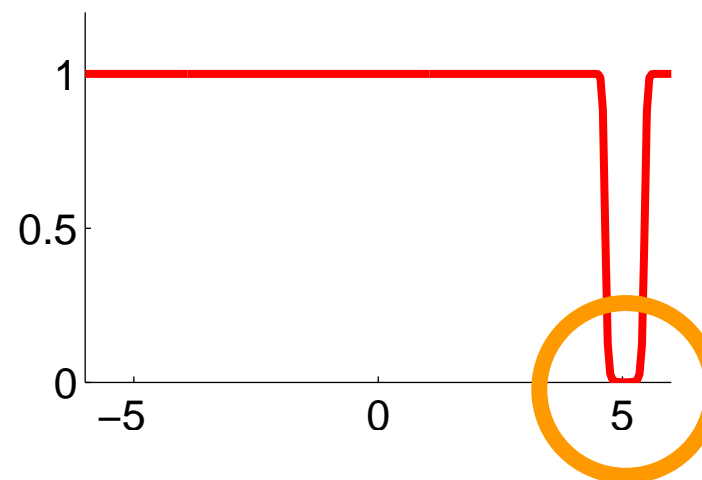
8

Hido, Tsuboi, Kashima & Sugiyama (ICDM2008)

■ 異常値は密度比が小さい



$$\frac{p_{\text{regular}}(x)}{p_{\text{irregular}}(x)}$$



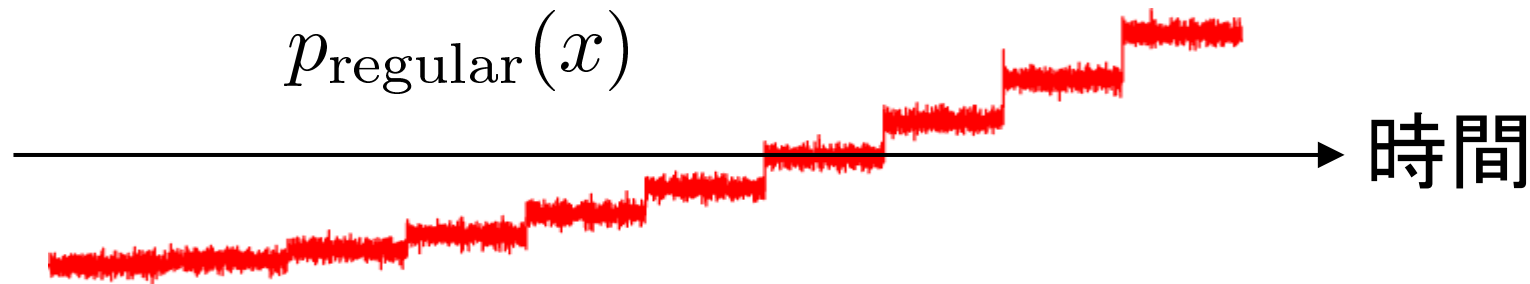


# 時系列の変化点検知

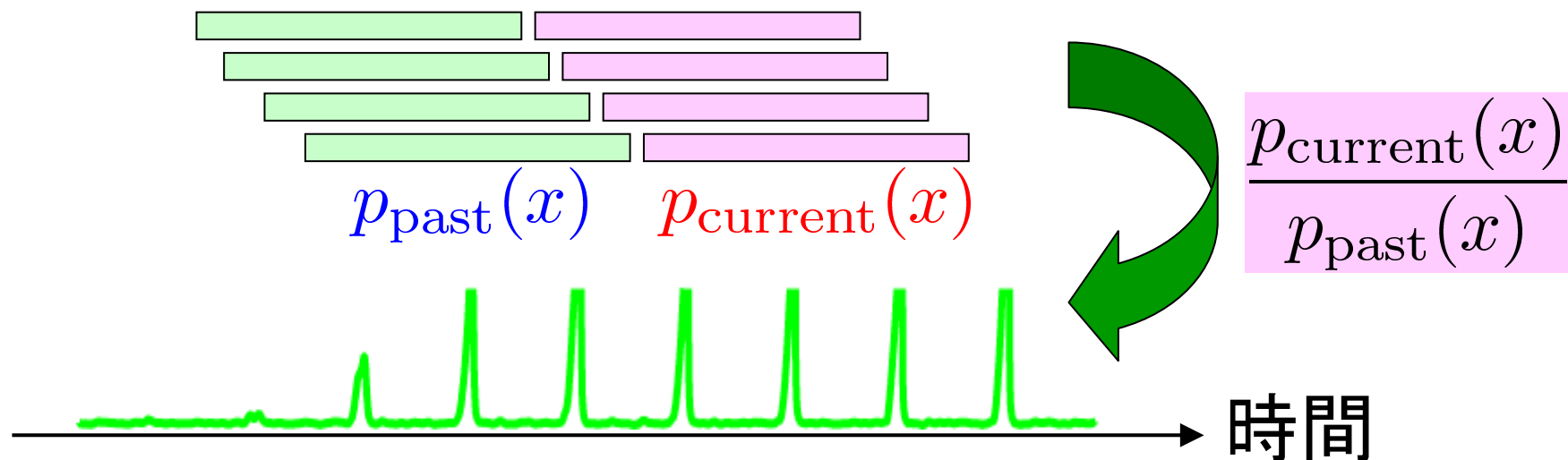
9

Kawahara & Sugiyama (IBIS2008)

- 時系列の傾向が変化する点を検知する

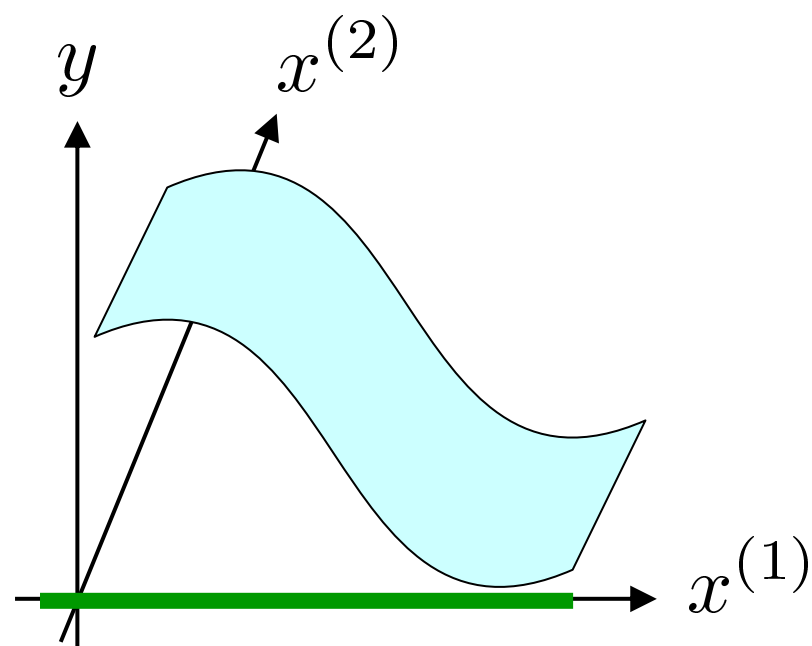
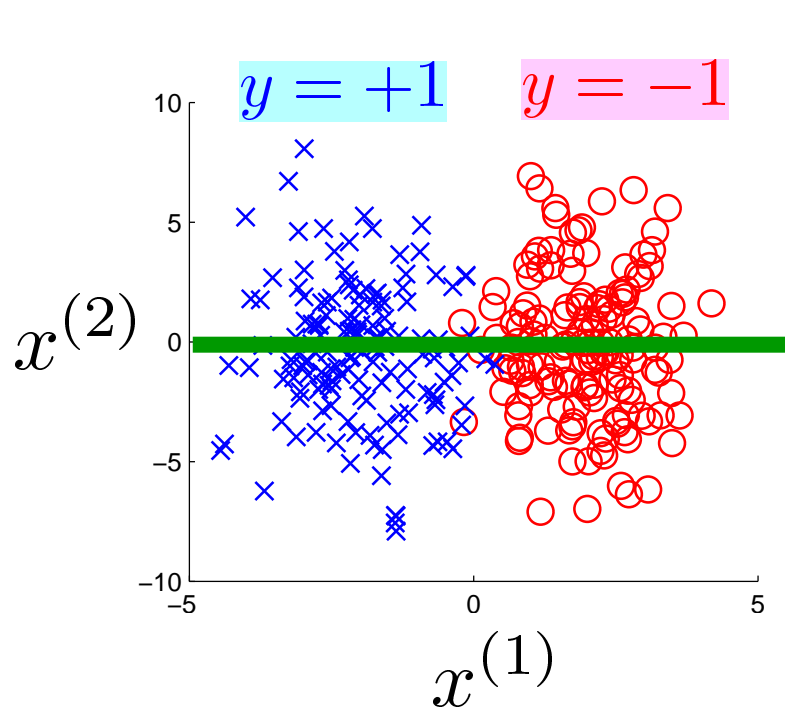


- 密度比を逐次的に計算:



# 例：特徴選択

- **特徴選択**：入出力関係  $y = f(x^{(1)}, \dots, x^{(d)})$  で、出力  $y$  を最もよく説明する入力変数  $x^{(k)}$  を選ぶ
  - 遺伝子選択, 脳の活動部位の特定, 創薬など

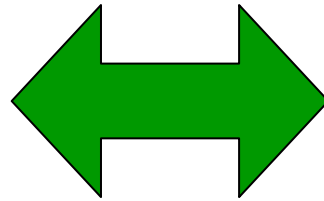


# 相互情報量に基づく独立性判定<sup>11</sup>

Suzuki, Sugiyama, Sese & Kanamori (FSDM2008)

■ 相互情報量:  $\int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy$

相互情報量  
がゼロ



$x$  と  $y$  は  
独立

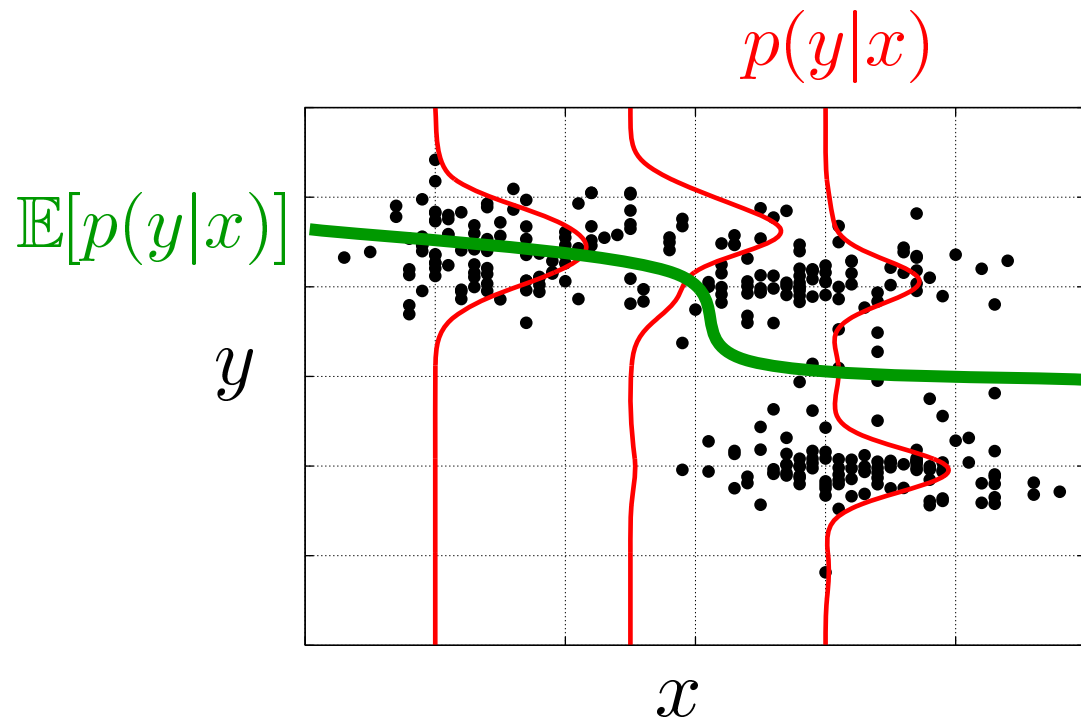
- 相互情報量は**密度比**  $\frac{p(x, y)}{p(x)p(y)}$  を用いて計算できる.
- 特徴選択, 独立成分分析などに適用可

# 例：条件付き確率推定

12

- **回帰分析**：条件付き期待値  $\mathbb{E}[p(y|x)]$  の推定法
- 条件付き分布が複雑なときは，回帰分析では情報不足

- 多峰性
- 非対称性
- 不等分散性



- **条件付き確率**  $p(y|x)$  そのものを推定したい！

# 例：条件付き確率推定（続き）

13

Sugiyama, Takeuchi, Suzuki, Kanamori & Hachiya (準備中)

- 条件付き確率は，**密度比**を用いて計算できる.

$$p(\mathbf{y}|\mathbf{x}) = \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})}$$



# 本発表の構成

14

1. 密度比は, 重点サンプリングだけでなく, 様々な統計的データ解析に役立つ!
2. 密度比推定の手法
  - Kullback-Leibler Importance Estimation Procedure (KLIEP)
  - Least-Squares Importance Fitting (LSIF)
  - Unconstrained LSIF (uLSIF)
3. 実験

# 密度比の推定

$$w(\boldsymbol{x}) = \frac{p'(\boldsymbol{x})}{p(\boldsymbol{x})}$$

- 密度比が分かれば, 色々な事ができる.
- 実用上は, 密度比を**標本から推定**する必要がある.

$$\{\boldsymbol{x}_i\}_{i=1}^n \stackrel{i.i.d.}{\sim} p(\boldsymbol{x})$$

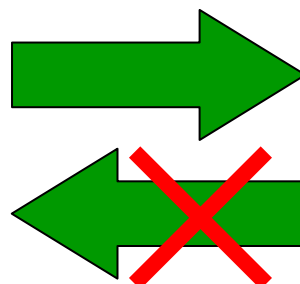
$$\{\boldsymbol{x}'_i\}_{i=1}^{n'} \stackrel{i.i.d.}{\sim} p'(\boldsymbol{x})$$

- **ナイーブな方法**: それぞれの密度を推定し, 比を取る

$$\hat{w}(\boldsymbol{x}) = \frac{\hat{p}'(\boldsymbol{x})}{\hat{p}(\boldsymbol{x})}$$

ある問題を解くとき、その問題よりも  
難しい問題を途中段階で解いてはならない

$p(x), p'(x)$   
が分かる



$w(x) = \frac{p'(x)}{p(x)}$   
が分かる

- 密度を求めるよりも、密度比を求めるほうが易しい！
- 密度推定を経由せず、密度比を推定することにする。





# 本発表の構成

17

1. 密度比は, 重点サンプリングだけでなく, 様々な統計的データ解析に役立つ!
2. 密度比推定の手法
  - Kullback-Leibler Importance Estimation Procedure (KLIEP)
  - Least-Squares Importance Fitting (LSIF)
  - Unconstrained LSIF (uLSIF)
3. 実験

# Kullback-Leibler Importance Estimation Procedure (KLIEP)

18

Sugiyama, Nakajima, Kashima, von Bünau & Kawanabe (NIPS2007)

## ■ 線形密度比モデル:

$$\begin{aligned}\hat{w}(\boldsymbol{x}) &= \sum_{\ell=1}^b \alpha_{\ell} \phi_{\ell}(\boldsymbol{x}) \\ &= \boldsymbol{\alpha}^{\top} \boldsymbol{\phi}(\boldsymbol{x})\end{aligned}$$

$$\alpha_{\ell} \geq 0$$

$$\phi_{\ell}(\boldsymbol{x}) \geq 0$$

(例: ガウスカーネル)

- $p'(\boldsymbol{x})$  から  $\hat{p}'(\boldsymbol{x}) = \hat{w}(\boldsymbol{x})p(\boldsymbol{x})$  のKLダイバージェンスを最小にするようにパラメータを学習

$$\min_{\boldsymbol{\alpha}} \text{KL}[p'(\boldsymbol{x}) || \hat{p}'(\boldsymbol{x})]$$

## ■ KL情報量の分解:

$$\begin{aligned} \text{KL}[p'(\mathbf{x})||\hat{p}'(\mathbf{x})] &= \int p'(\mathbf{x}) \log \frac{p'(\mathbf{x})}{\hat{w}(\mathbf{x})p(\mathbf{x})} d\mathbf{x} \\ &= \underbrace{\int p'(\mathbf{x}) \log \frac{p'(\mathbf{x})}{p(\mathbf{x})} d\mathbf{x}}_{\text{定数}} - \int p'(\mathbf{x}) \log \hat{w}(\mathbf{x}) d\mathbf{x} \end{aligned}$$

$$\min_{\alpha} \text{KL}[p'(\mathbf{x})||\hat{p}'(\mathbf{x})] \longleftrightarrow \max_{\alpha} \int p'(\mathbf{x}) \log \hat{w}(\mathbf{x}) d\mathbf{x}$$

■  $\hat{p}'(\mathbf{x}) = \hat{w}(\mathbf{x})p(\mathbf{x})$  は確率密度:  $\int \hat{w}(\mathbf{x})p(\mathbf{x})d\mathbf{x} = 1$

拘束条件

# KLIEP: 解法 (続き)

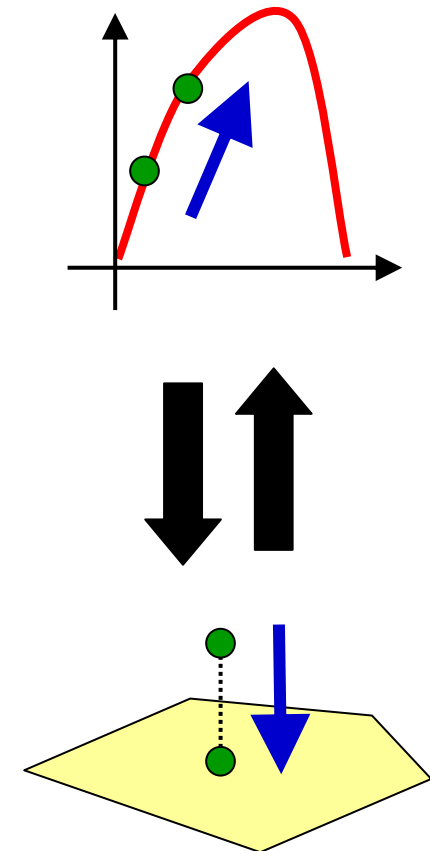
20

- 期待値を標本平均で近似すれば,

$$\max_{\alpha} \sum_{i=1}^{n'} \log(\alpha^{\top} \phi(\mathbf{x}'_i))$$

$$\text{subject to } \sum_{i=1}^n \alpha^{\top} \phi(\mathbf{x}_i) = n \text{ and } \alpha \geq 0$$

- これは**凸最適化**問題なので,
  - 勾配上昇
  - 拘束充足を繰り返せば**大域的最適解**に収束
- 最適解は**スパース**!



# KLIEP: 収束性

21

Sugiyama, Suzuki, Nakajima, Kashima, von Bünau & Kawanabe (AISM2008)  
Nguyen, Wainwright & Jordan (NIPS2007)

## ■ パラメトリックの場合:

$$\hat{w}(\boldsymbol{x}) = \sum_{\ell=1}^b \alpha_{\ell} \phi_{\ell}(\boldsymbol{x})$$

- 標本から学習したパラメータは、**最適な値に**  $1/\sqrt{\bar{n}}$  のオーダーで収束

$$\bar{n} = \min(n, n')$$

(少ない方の標本数)

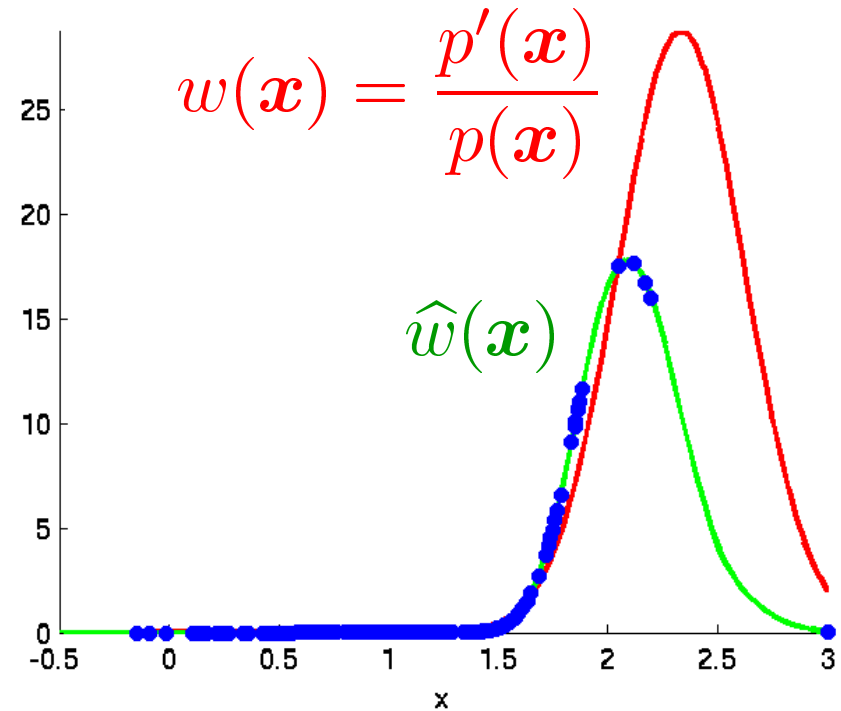
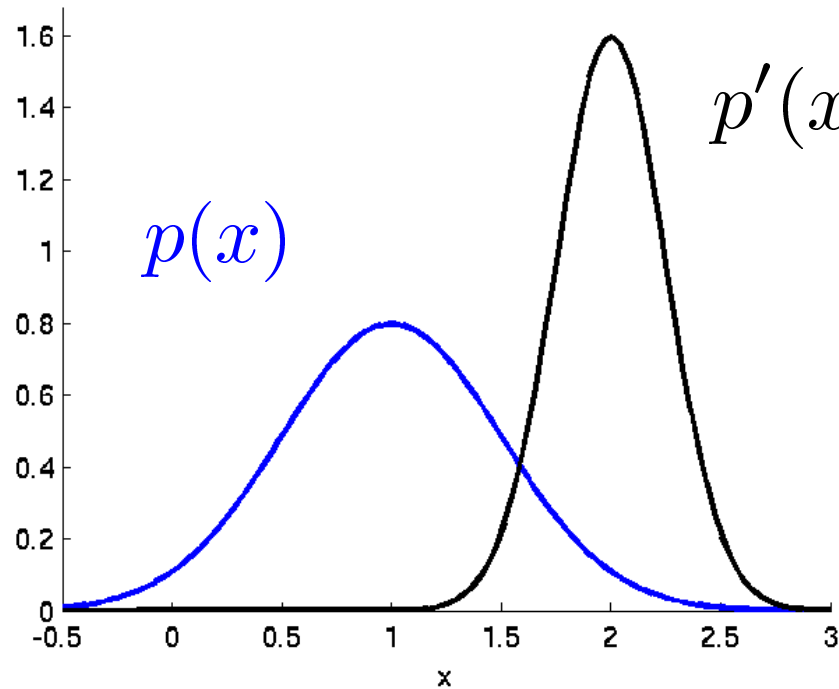
## ■ ノンパラメトリックの場合:

$$\hat{w}(\boldsymbol{x}) = \sum_{\ell=1}^{n'} \alpha_{\ell} K(\boldsymbol{x}, \boldsymbol{x}_{\ell})$$

- 標本から学習した関数は、**真の関数に**  $1/\sqrt{\bar{n}}$  より少し遅いオーダーで収束 (関数集合の**ブラケットティングナンバー**に依存)

# KLIEP: 実行例

22



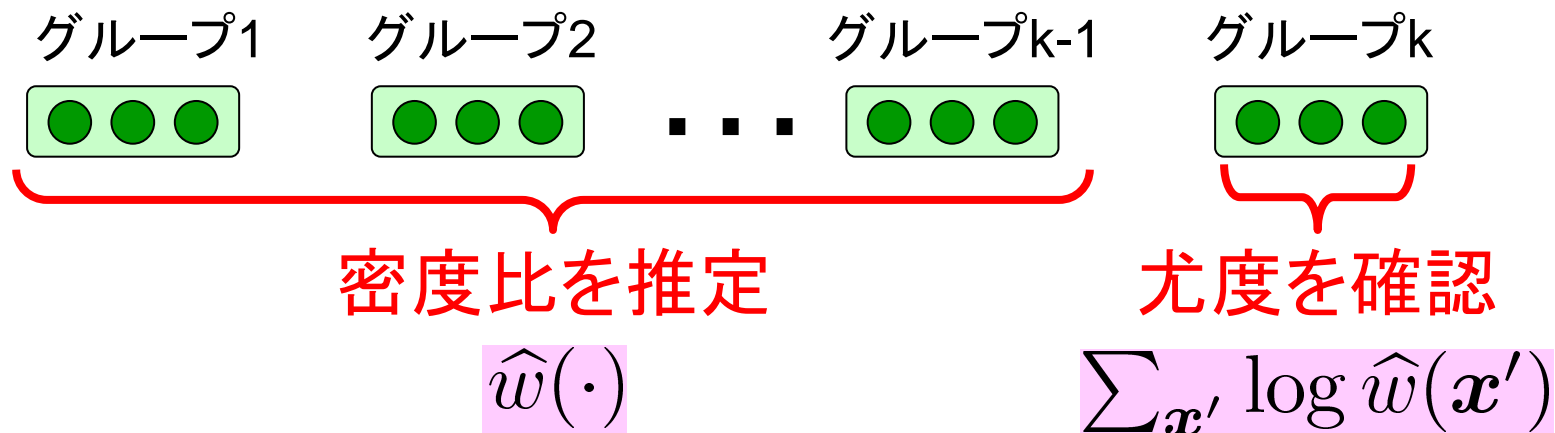
$$\hat{w}(\mathbf{x}) = \sum_{\ell=1}^{n'} \alpha_{\ell} K(\mathbf{x}, \mathbf{x}'_{\ell})$$

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right)$$

■ ガウシアンのはどうやって決めればよいか？

■ 交差確認法:

- テスト標本を分割

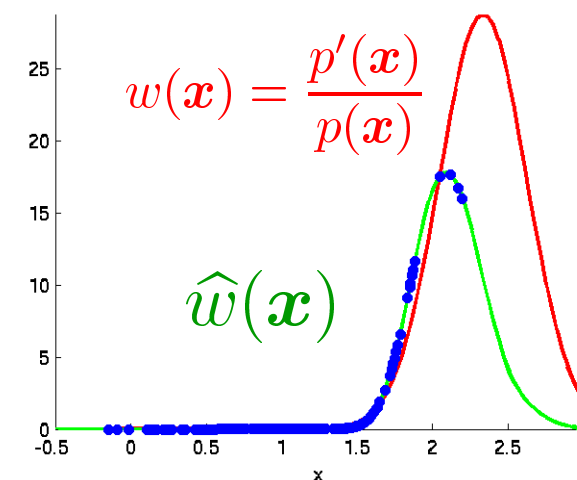
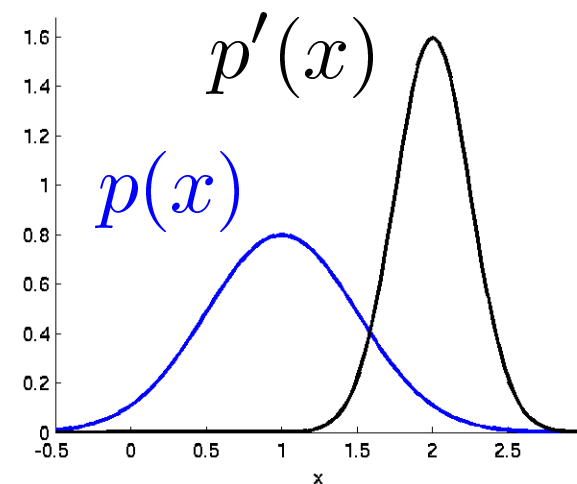
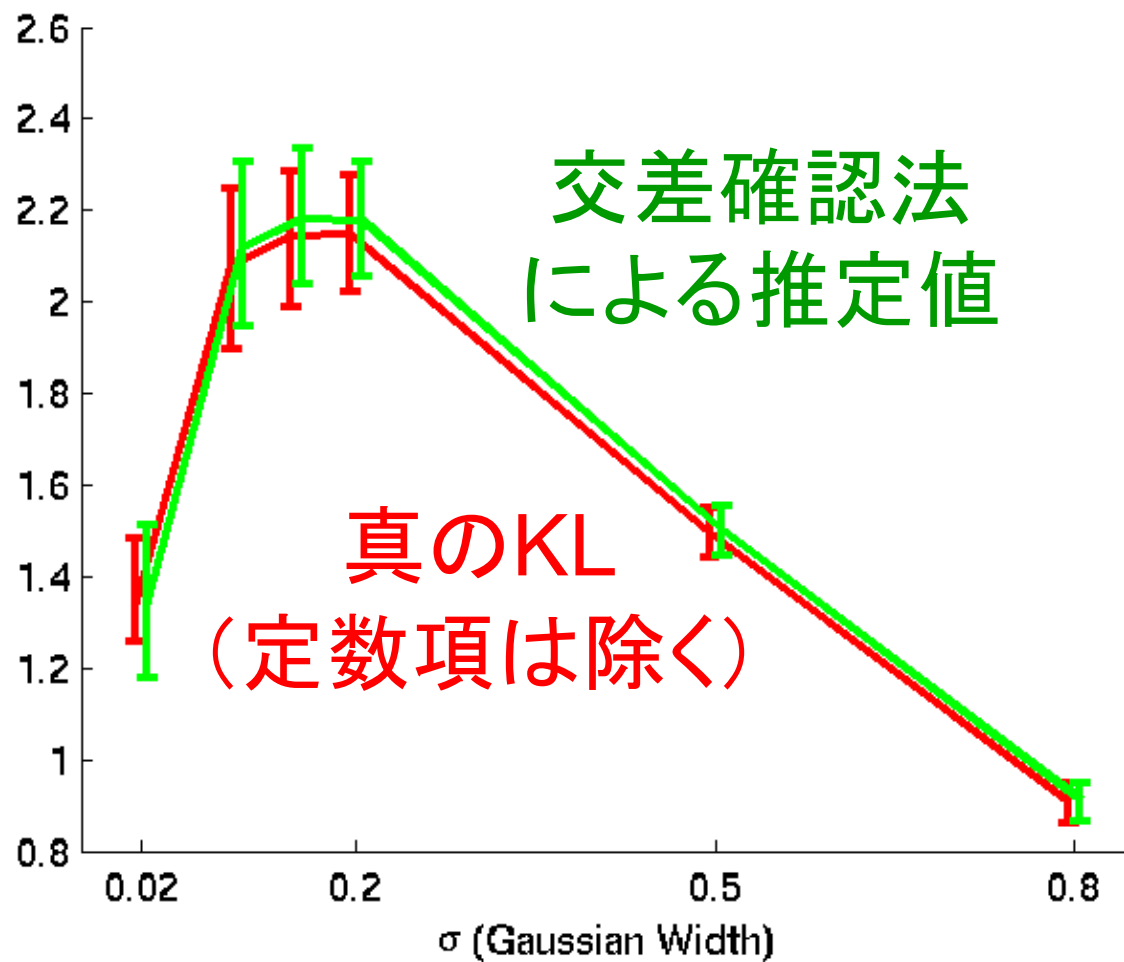


- 全ての組み合わせに対して繰り返し, 平均を出力

■ これにより, KLの**不偏推定量**が得られる

# KLIEP: 尤度交差確認法の例

24



■ 非常に精度が良い！



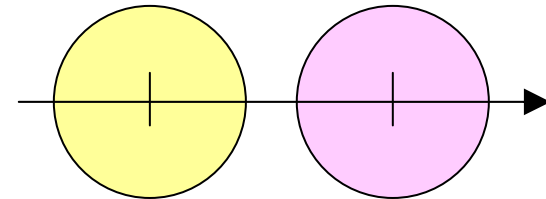
# 計算機実験の設定

25

■ 実験設定:  $d$ 次元データ, 分散1の正規分布で平均は

● 訓練:  $(0, 0, 0, \dots, 0)$

● テスト:  $(1, 0, 0, \dots, 0)$



■ カーネル密度推定(KDE):

● 訓練密度とテスト密度をそれぞれ推定

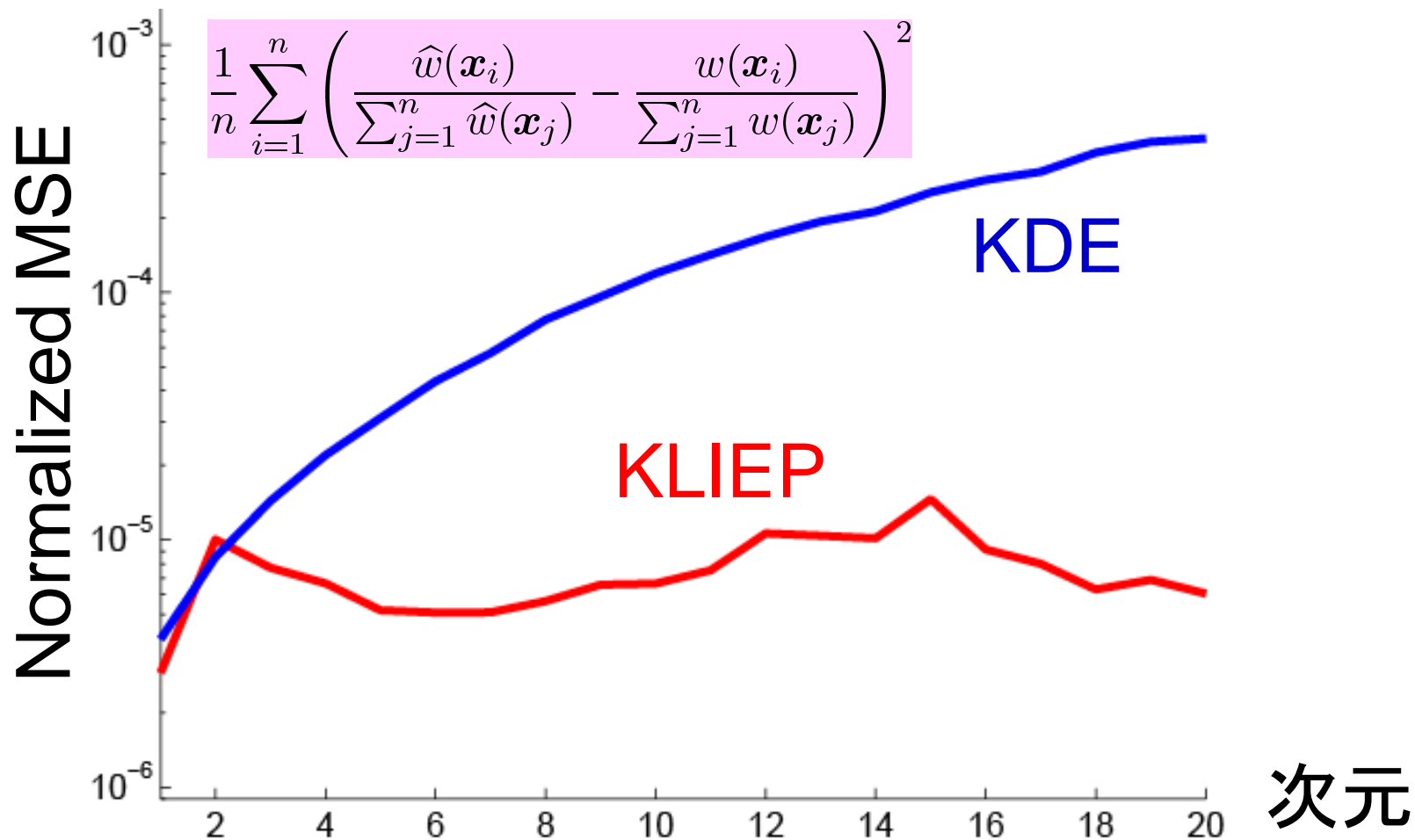
● ガウシアンカーネルの幅は尤度交差確認法で決定

■ KLIEP:

● ガウシアンカーネルの幅は交差確認法で決定

# 次元が増えたときの精度比較

26



- KDE: 次元が増えると誤差は単調に増加
- KLIEP: 高次元でも比較的うまくいく

- 比較的高次元でも精度がよい (Vapnikの原理)
- スパースな大域的最適解が求まる
- 交差確認法により**モデル選択**ができる
- **分母と分子のドメイン**が異なってもよい (相互情報量や条件付き確率の推定で重要)
- KLロスは、相互情報量 (KLダイバージェンス) 推定と相性がよい
- **対数線形密度比モデル**を用いれば、分子のデータが大量の場合でも効率よく計算できる

Tsuboi, Kashima, Hido, Bickel & Sugiyama (SDM2008)



# 本発表の構成

28

1. 密度比は, 重点サンプリングだけでなく, 様々な統計的データ解析に役立つ!
2. 密度比推定の手法
  - Kullback-Leibler Importance Estimation Procedure (KLIEP)
  - Least-Squares Importance Fitting (LSIF)
  - Unconstrained LSIF (uLSIF)
3. 実験

# Least-Squares Importance Fitting (LSIF)

Kanamori, Hido & Sugiyama (NIPS2008)

## ■ 線形密度比モデル:

$$\begin{aligned}\hat{w}(\mathbf{x}) &= \sum_{\ell=1}^b \alpha_{\ell} \phi_{\ell}(\mathbf{x}) \\ &= \boldsymbol{\alpha}^{\top} \boldsymbol{\phi}(\mathbf{x})\end{aligned}$$

$$\alpha_{\ell} \geq 0$$

$$\phi_{\ell}(\mathbf{x}) \geq 0$$

(例: ガウスカーネル)

## ■ 二乗損失:

$$J_0(\boldsymbol{\alpha}) = \frac{1}{2} \int \left( \hat{w}(\mathbf{x}) - w(\mathbf{x}) \right)^2 p(\mathbf{x}) d\mathbf{x}$$

$$w(\mathbf{x}) = \frac{p'(\mathbf{x})}{p(\mathbf{x})}$$

## ■ 二乗損失の分解:

$$\begin{aligned} J_0(\boldsymbol{\alpha}) &= \frac{1}{2} \int \left( \hat{w}(\boldsymbol{x}) - w(\boldsymbol{x}) \right)^2 p(\boldsymbol{x}) d\boldsymbol{x} \\ &= \frac{1}{2} \int \left( \hat{w}(\boldsymbol{x}) \right)^2 p(\boldsymbol{x}) d\boldsymbol{x} - \int \hat{w}(\boldsymbol{x}) p'(\boldsymbol{x}) d\boldsymbol{x} \\ &\quad + \underbrace{\frac{1}{2} \int \left( w(\boldsymbol{x}) \right)^2 p(\boldsymbol{x}) d\boldsymbol{x}}_{\text{定数}} \end{aligned}$$

## ■ 拘束条件: $\boldsymbol{\alpha} \geq \mathbf{0}$

$$w(\boldsymbol{x}) = \frac{p'(\boldsymbol{x})}{p(\boldsymbol{x})}$$

# LSIF: 解法 (続き)

31

- 期待値を標本平均で近似し, 正則化項を加えれば

$$\min_{\alpha} \left[ \frac{1}{2} \alpha^{\top} \widehat{H} \alpha - \widehat{h}^{\top} \alpha + \lambda \alpha^{\top} \mathbf{1} \right] \quad \text{subject to } \alpha \geq 0$$

$$\widehat{H} = \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{x}_i)^{\top} \phi(\mathbf{x}_i) \quad \widehat{h} = \frac{1}{n'} \sum_{j=1}^{n'} \phi(\mathbf{x}'_j)$$

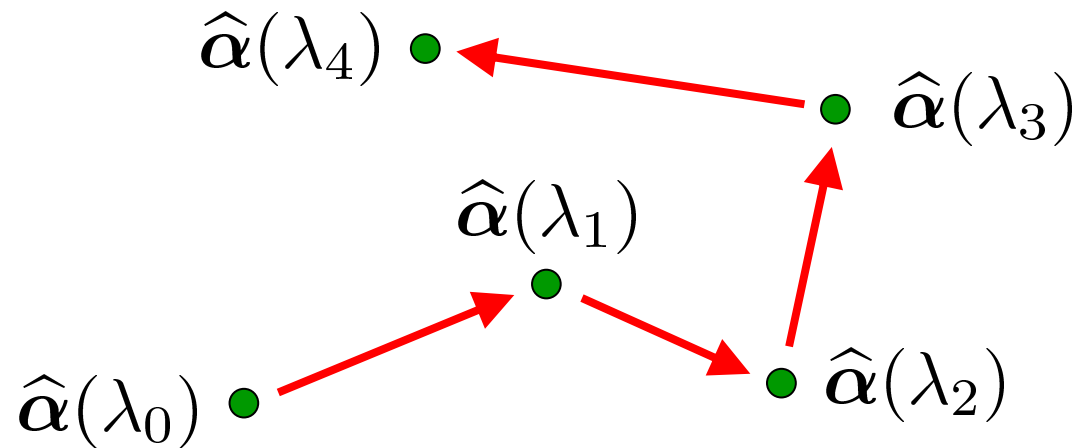
- これは凸二次計画問題なので, 標準的な二次計画ソルバーで大域的最適解が求まる
- 最適解はスパース!

# LSIF: 正則化パス追跡

32

$$\min_{\alpha} \left[ \frac{1}{2} \alpha^{\top} \widehat{H} \alpha - \widehat{h}^{\top} \alpha + \lambda \alpha^{\top} \mathbf{1} \right] \quad \text{subject to } \alpha \geq 0$$

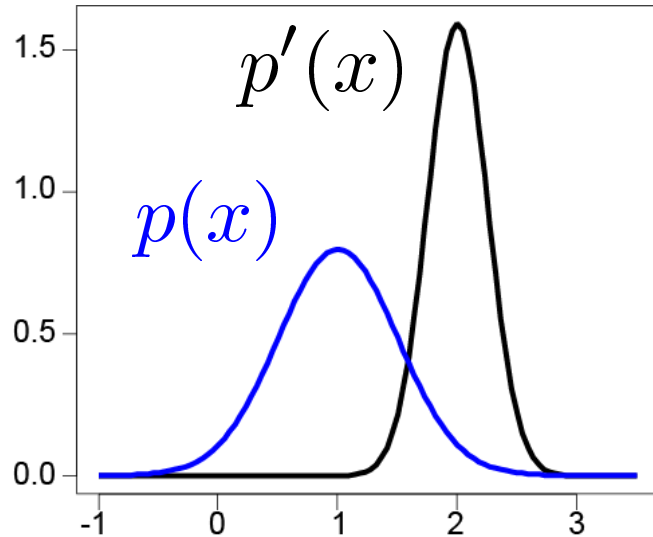
- 解は正則化パラメータ  $\lambda$  に関して**区分線形**



$$\lambda_0 > \lambda_1 > \dots > 0$$

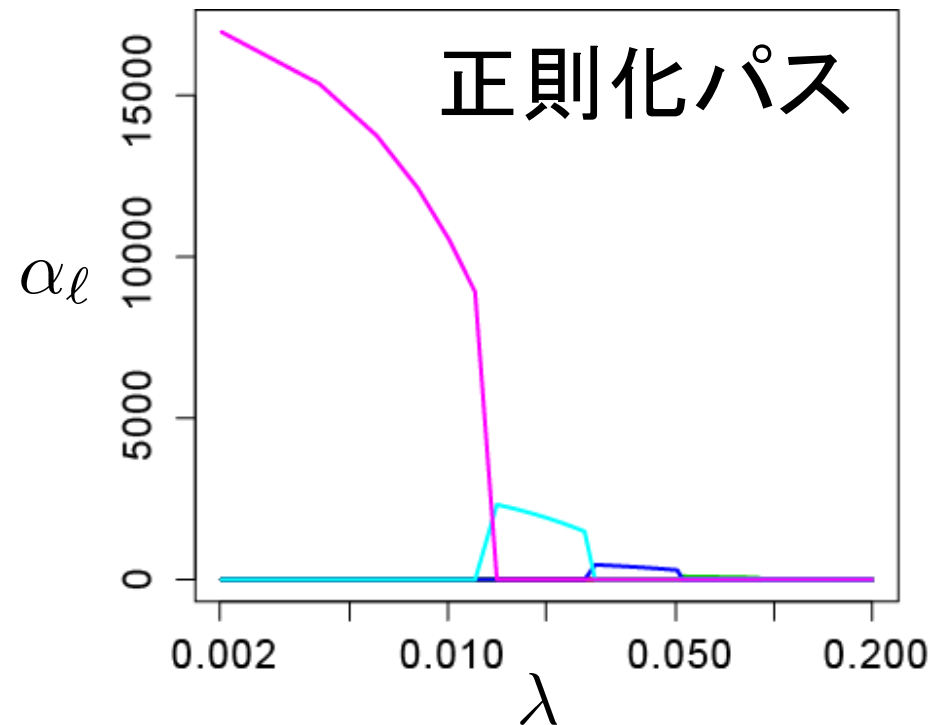
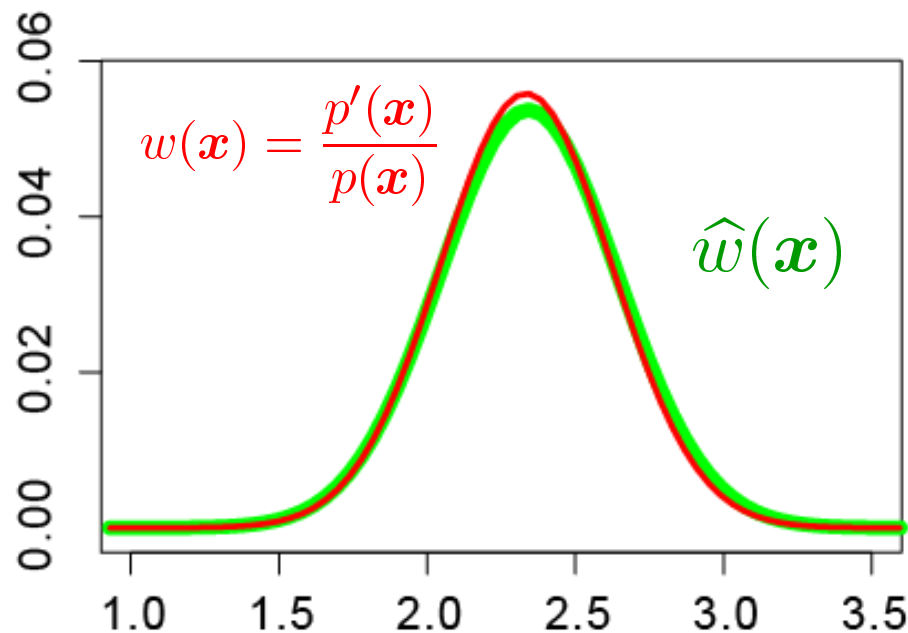
- **全ての**  $\lambda$  に対する解が効率よく計算できる！





- 正則化パラメータとカーネル幅は交差確認法で決定

$$\hat{w}(x) = \sum_{\ell=1}^{n'} \alpha_{\ell} \exp\left(-\frac{\|x - x'_{\ell}\|^2}{2\sigma^2}\right)$$



# LSIF: 特徴

34

- 二乗ロスは条件付き確率推定には都合が良い(裾はあまり気にしない)
- **正則化パス**が計算できるため計算効率が良い
- しかし, 正則化パス計算は数値的にやや不安定



# 本発表の構成

35

1. 密度比は, 重点サンプリングだけでなく, 様々な統計的データ解析に役立つ!
2. 密度比推定の手法
  - Kullback-Leibler Importance Estimation Procedure (KLIEP)
  - Least-Squares Importance Fitting (LSIF)
  - Unconstrained LSIF (uLSIF)
3. 実験

# Unconstrained LSIF (uLSIF) 36

Kanamori, Hido & Sugiyama (NIPS2008)

$$\min_{\alpha} \left[ \frac{1}{2} \alpha^\top \widehat{H} \alpha - \widehat{h}^\top \alpha + \lambda \alpha^\top \mathbf{1} \right] \quad \text{subject to } \alpha \geq \mathbf{0}$$

## ■ LSIFを少し改変:

- 非負拘束を無視する
- 正則化項を2次にする

$$\tilde{\alpha} = \operatorname{argmin}_{\alpha} \left[ \frac{1}{2} \alpha^\top \widehat{H} \alpha - \widehat{h}^\top \alpha + \frac{\lambda}{2} \alpha^\top \alpha \right]$$

$$\widehat{H} = \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_i)$$

$$\widehat{h} = \frac{1}{n'} \sum_{j=1}^{n'} \phi(\mathbf{x}'_j)$$

# uLSIF: 解法

$$\tilde{\alpha} = \underset{\alpha}{\operatorname{argmin}} \left[ \frac{1}{2} \alpha^\top \widehat{H} \alpha - \widehat{h}^\top \alpha + \frac{\lambda}{2} \alpha^\top \alpha \right]$$

- 解  $\tilde{\alpha}$  は**解析的**に計算できる！

$$\tilde{\alpha} = (\widehat{H} + \lambda I)^{-1} \widehat{h}$$

- 無視した非負拘束は, **事後補正**する:

$$\widehat{\alpha} = \max(\mathbf{0}, \tilde{\alpha})$$

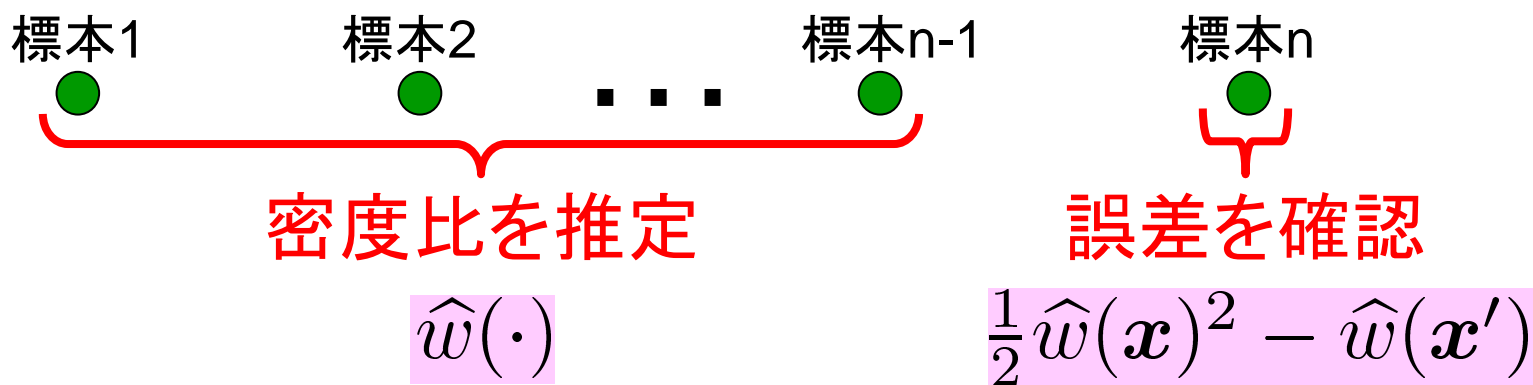
- LSIFとuLSIFの解の差:

- 正則化パラメータ  $\lambda$  が大きいとき, 差が小さい
- ガウス幅  $\sigma$  が大きいとき, 差が小さい

# uLSIF: 一つ抜き交差確認法

38

## ■ 一つ抜き交差確認法:



- 通常はn回繰り返して計算する必要があるが、uLSIFでは**解析的に一発**で計算できる  
(Sherman-Woodbury-Morrison公式を使う)
- モデル選択を含めた計算時間が**飛躍的に短縮**される！

# 密度比推定法のまとめ

39

手法	密度推定	分子と分母の ドメイン	モデル 選択	計算速度
KDE	あり	違ってても良い	あり	超速い
KMM	なし	同じ	なし	やや遅い
LogReg	なし	同じ	あり	やや遅い
KLIEP	なし	違ってても良い	あり	やや遅い
LSIF	なし	違ってても良い	あり	速い
uLSIF	なし	違ってても良い	あり	とても速い

- Kernel mean matching (KMM)

Huang, Smola, Gretton, Borgwardt & Schölkopf (NIPS2006)

- Logistic regression method (LogReg)

Qin (Biometrika1998), Cheng & Chu (Bernoulli2004)

Bickel, Brückner & Scheffer (ICML2007)



# 本発表の構成

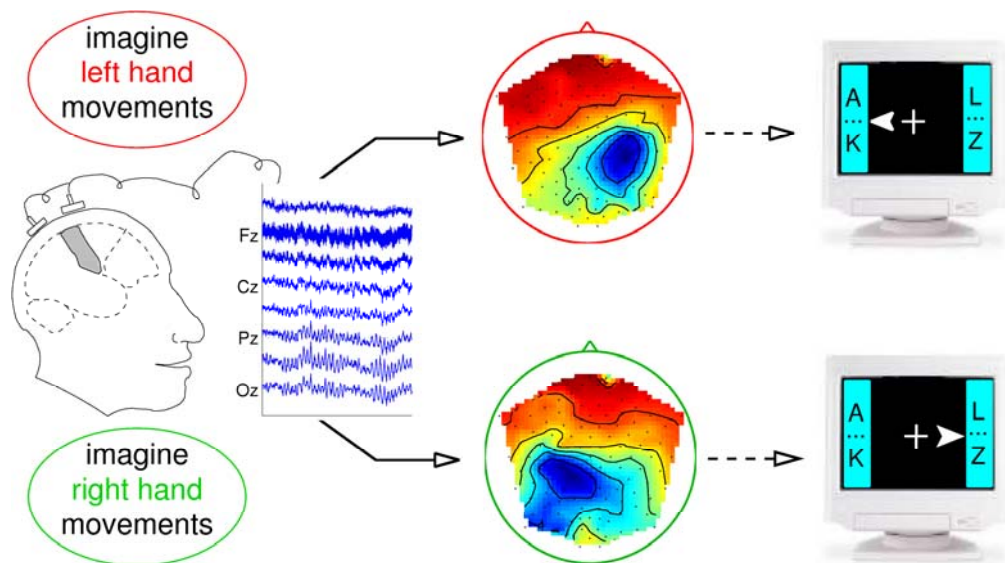
1. 密度比は, 重点サンプリングだけでなく, 様々な統計的データ解析に役立つ!
2. 密度比推定の手法
3. 実験
  - 共変量シフト適応
  - 外れ値検出
  - 特徴選択
  - 条件付き確率推定



# 共変量シフト適応： ブレイン・コンピュータインターフェース

Sugiyama, Krauledat & Müller (JMLR2007)

- 脳波(EEG)でコンピューターを操作する
- 入力は脳波, 出力は左右の方向
- 分類規則をデータから学習する



## 訓練風景

- 画面に現れる指示に従って、**左手／右手を動かす**イメージをする。



動画:ドイツFraunhofer研究所提供

# テスト風景

- 左右の手を動かすイメージをすることにより、脳波でパッドを操作する



動画:ドイツFraunhofer研究所提供

# 実験結果 (識別誤差)

被験者	試行	従来法	提案法	KL
1	1	9.3 %	10.0 %	0.76
	2	8.8 %	8.8 %	1.11
	3	4.3 %	4.3 %	0.69
2	1	40.0 %	40.0 %	0.97
	2	39.3 %	38.7 %	1.05
	3	25.5 %	25.5 %	0.43
3	1	36.9 %	34.4 %	2.63
	2	21.3 %	19.3 %	2.88
	3	22.5 %	17.5 %	1.25
4	1	21.3 %	21.3 %	9.23
	2	2.4 %	2.4 %	5.58
	3	6.4 %	6.4 %	1.83
5	1	21.3 %	21.3 %	0.79
	2	15.3 %	14.0 %	2.01

- 訓練時とテスト時で精神状態が異なるため、脳波の分布が変化.
- KL: 訓練・テスト入力分布のKL情報量の推定値
- 識別性能が向上
  - 密度比重み付き線形判別分析
  - 密度比重み付き交差確認法

# 共変量シフト適応：話者識別

45

- 音声から話者を特定する.
- 音声データの分布は一定でない
  - マイクや部屋の特徴
  - 話者の心理状況
- 識別性能が向上
  - 密度比重み付きロジスティック回帰
  - 密度比重み付き交差確認法

Yamada & Sugiyama (DMSS2008)



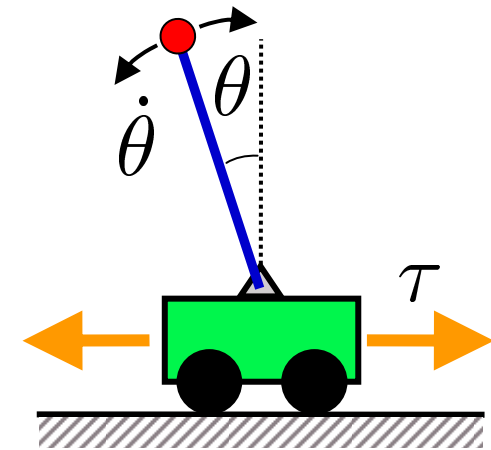
	従来法 (1.5s)	提案法 (1.5s)	従来法 (4.5s)	提案法 (4.5s)
3カ月後	13.9 %	<b>13.2 %</b>	7.7 %	<b>7.4 %</b>
6カ月後	18.0 %	<b>16.1 %</b>	7.3 %	<b>6.3 %</b>
9カ月後	8.3 %	<b>8.0 %</b>	0.3 %	<b>0.1 %</b>

# 共変量シフト適応： 強化学習によるロボット制御

Hachiya, Akiyama, Sugiyama & Peters (AAAI2008)

## ■ 振り子ロボット

- 状態  $s$  : 振り子の角度, 角速度
- 行動  $a$  : 左に加速, 右に加速



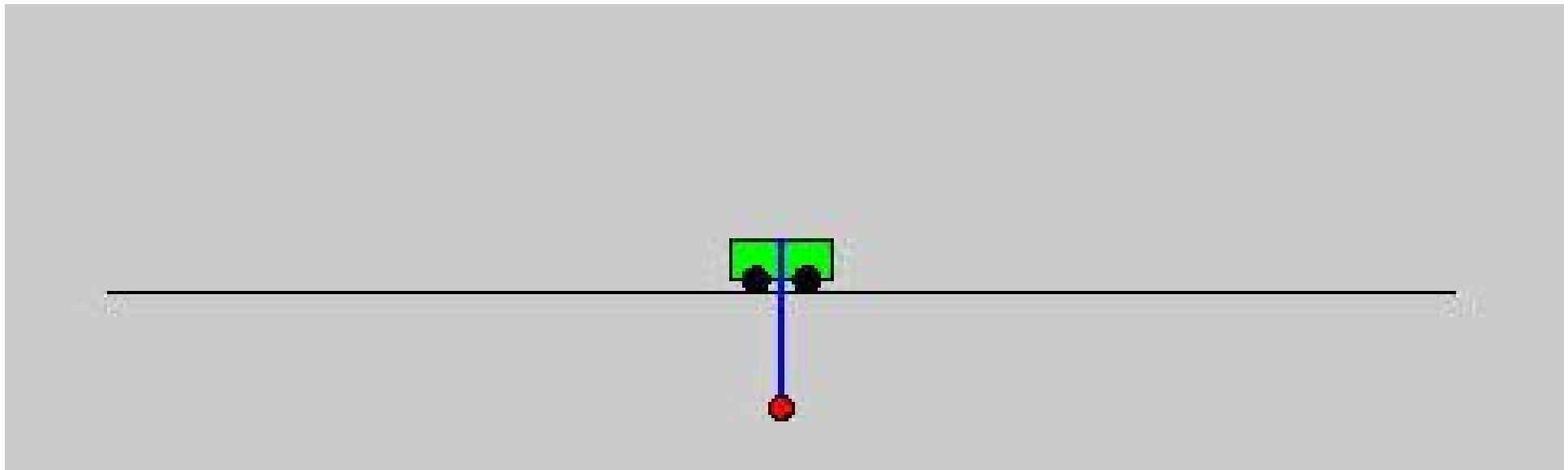
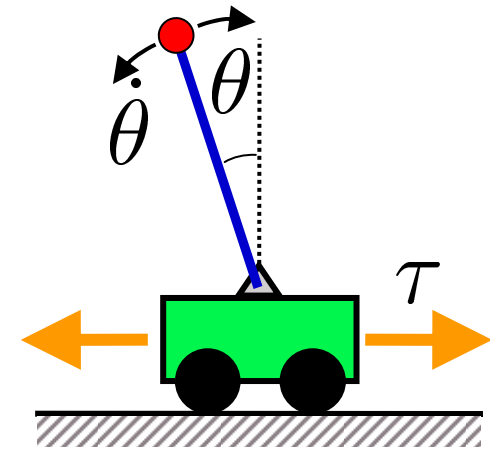
- ## ■ ゴール
- 車を左右に動かして振り子を振り上げ、  
垂直な状態を維持する制御則を獲得

$$\pi(a|s)$$

# 強化学習

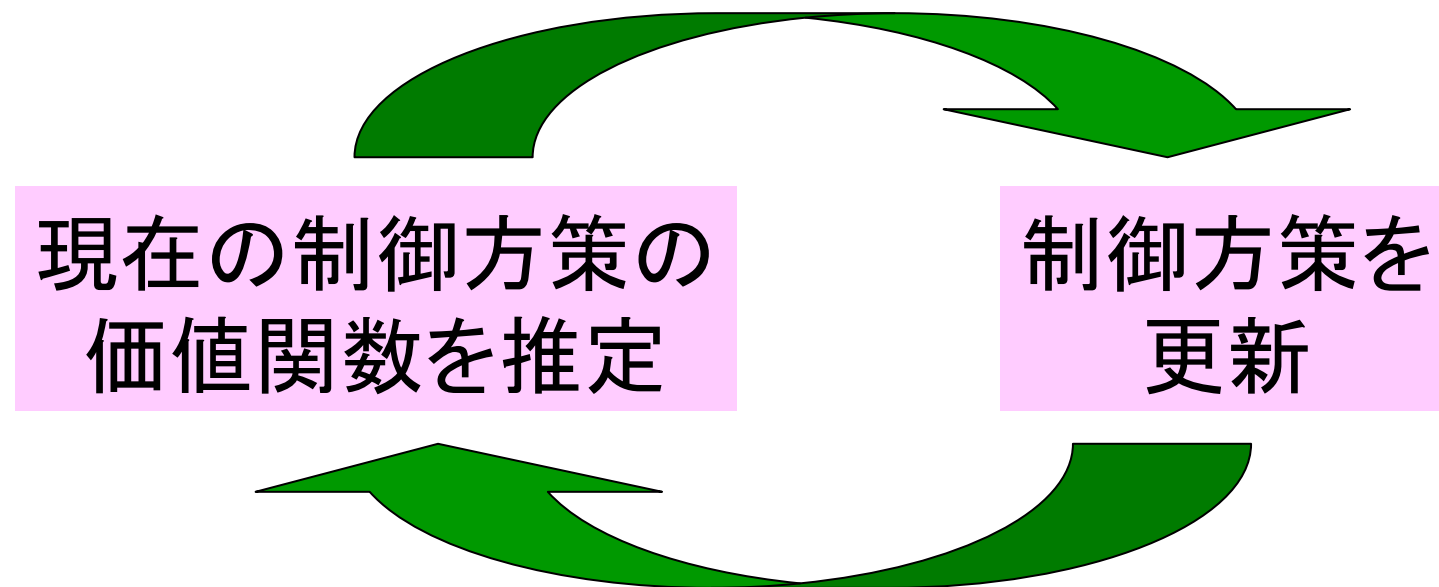
47

- **強化学習**: 得られる報酬が最大になるように制御方策  $\pi(a|s)$  を学習
- **報酬**: 振り子の角度  $\cos \theta$   
(振り子が上にあるほど報酬が大きい)



# 強化学習における共変量シフト 48

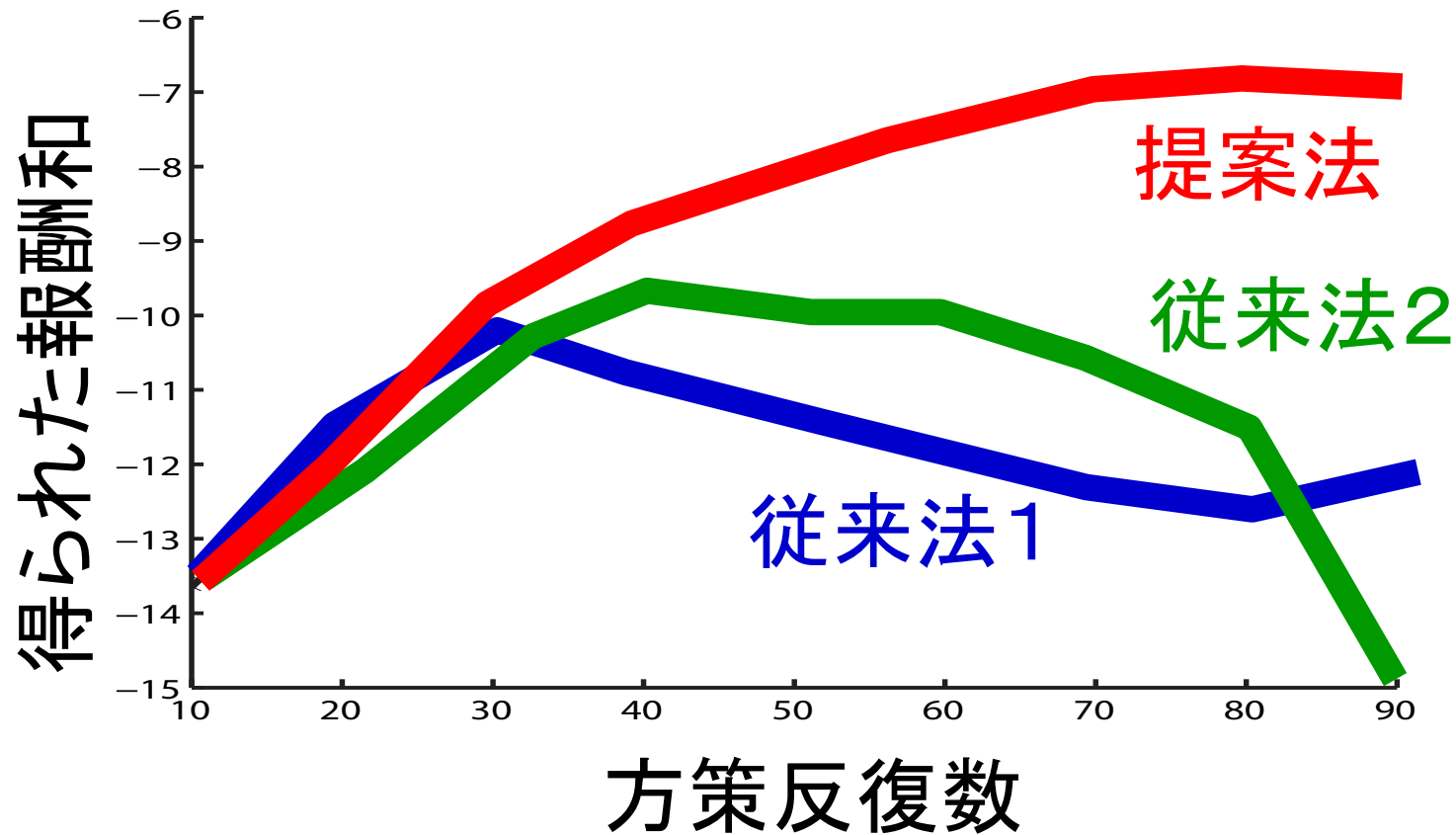
- **価値関数**  $Q^\pi(s, a)$  : 状態  $s$  で行動  $a$  を取り, その後  $\pi$  に従って行動して得られる報酬和
- **方策反復法** :



- 方策を更新すると,  $(s, a)$  の分布が変わる!



# 実験結果



## ■ 識別性能が向上

- 密度比重み付き最小二乗法
- 密度比重み付き交差確認法

# ドメイン適応：日本語単語分割 50

Tsuboi, Kashima, Hido, Bickel & Sugiyama (submitted)

## ■ 適応元データ：会話辞典の例文

- (例) こんな／失敗／は／ご／愛敬／だ／よ／ .

## ■ 適応先データ：医療マニュアル

- (例) 細胞膜には受容体があり、これによって細胞を識別することができます。

手法	従来法	提案法	適応先ラベル有
F値	92.30	94.46	94.43

## ■ 単語分割性能が向上

- 密度比重み付き条件付き確率場(CRF)
- 密度比重み付き交差確認法



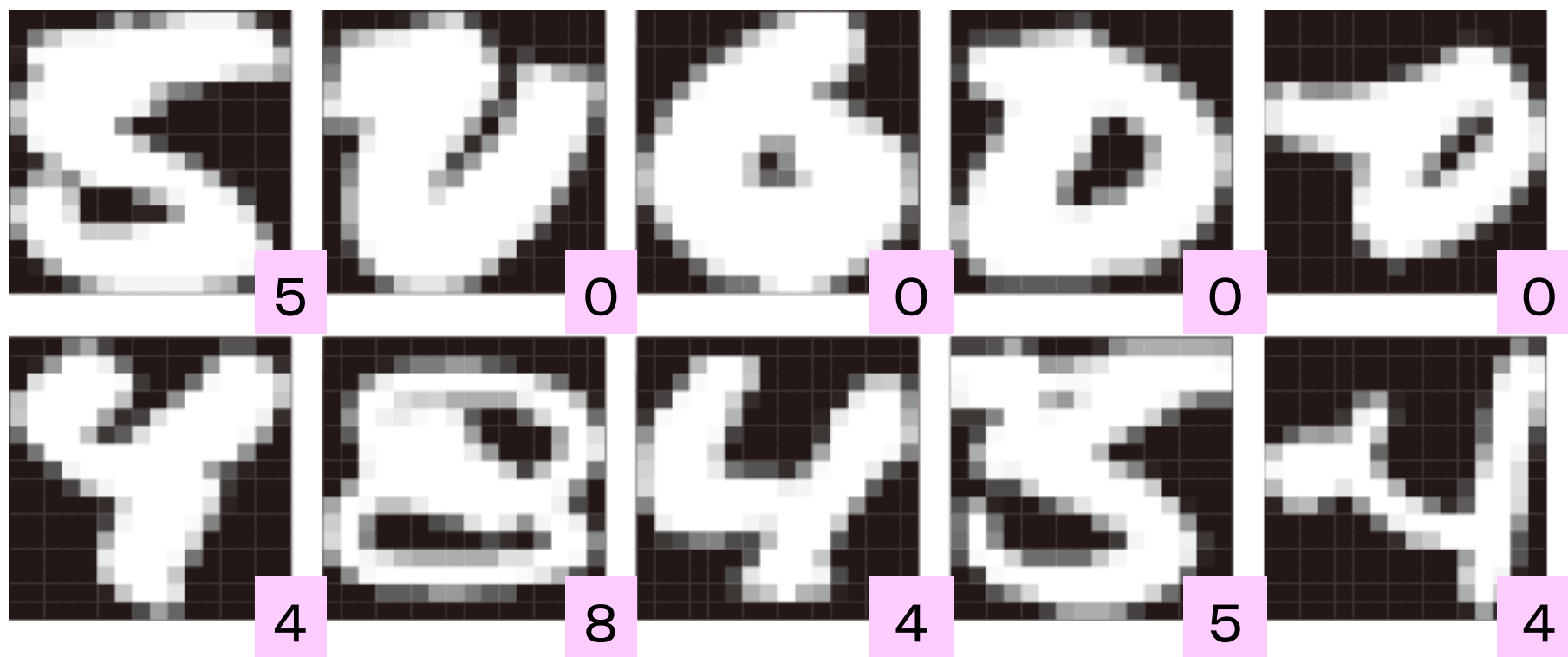
# 本発表の構成

51

1. 密度比は, 重点サンプリングだけでなく, 様々な統計的データ解析に役立つ!
2. 密度比推定の手法
3. 実験
  - 共変量シフト適応
  - 外れ値検出
  - 特徴選択
  - 条件付き確率推定

# 異常値検出: USPS手書き数字 52

Hido, Tsuboi, Kashima & Sugiyama (ICDM2008)



- 人間でも読みにくい数字がUSPSのテストデータには含まれている！

# 異常値検出:ハードディスク故障検出

Hido, Tsuboi, Kashima & Sugiyama (ICDM2008)

## ■ Self-Monitoring And Reporting Technology (SMART):ハードディスクの故障検知

手法	提案法	One-class SVM	LOF	
			近傍数5	近傍数30
AUC	0.881	0.843	0.847	0.924

- LOFは近傍数をうまく決めれば性能が良いが、モデル選択法がない
- 提案法はモデル選択ができる！

One-class SVM:

Schölkopf, Platt, Shawe-Taylor, Smola & Williamson (NeCo2001)

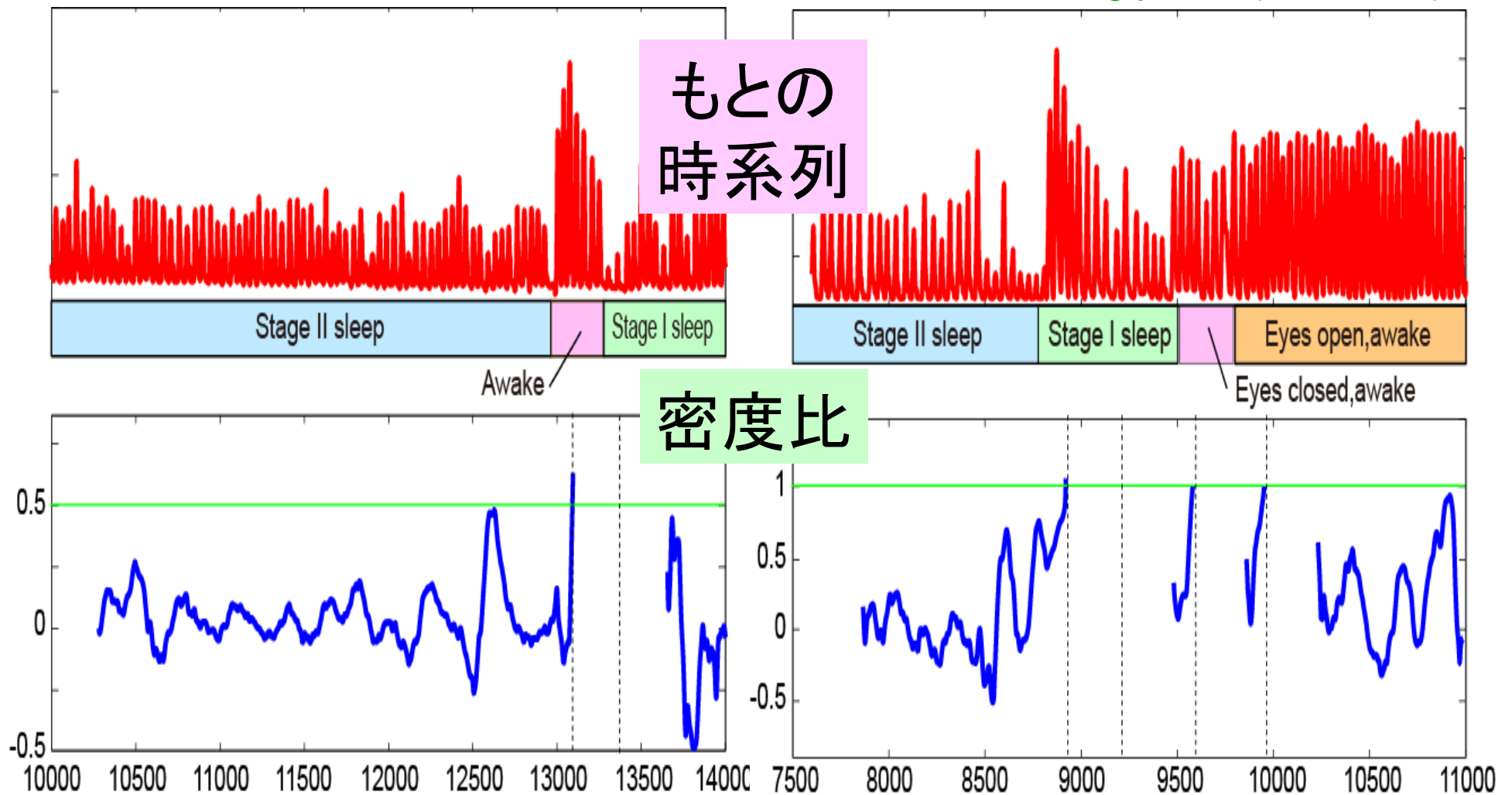
LOF: Local outlier factor

Breunig, Kriegel, Ng & Sander (SIGMOD2000)

# 変化点検知:呼吸データ

54

Kawahara & Sugiyama (IBIS2008)



■ 密度比によって変化点が検知できている



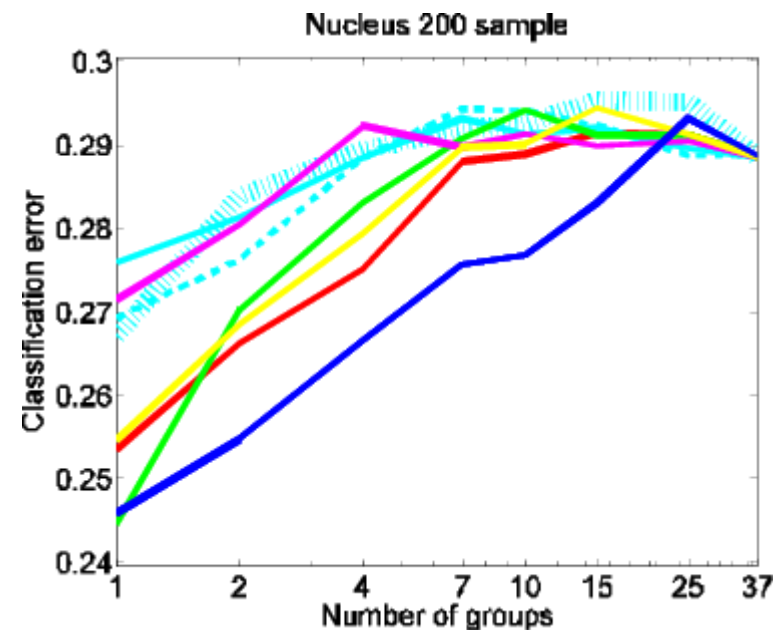
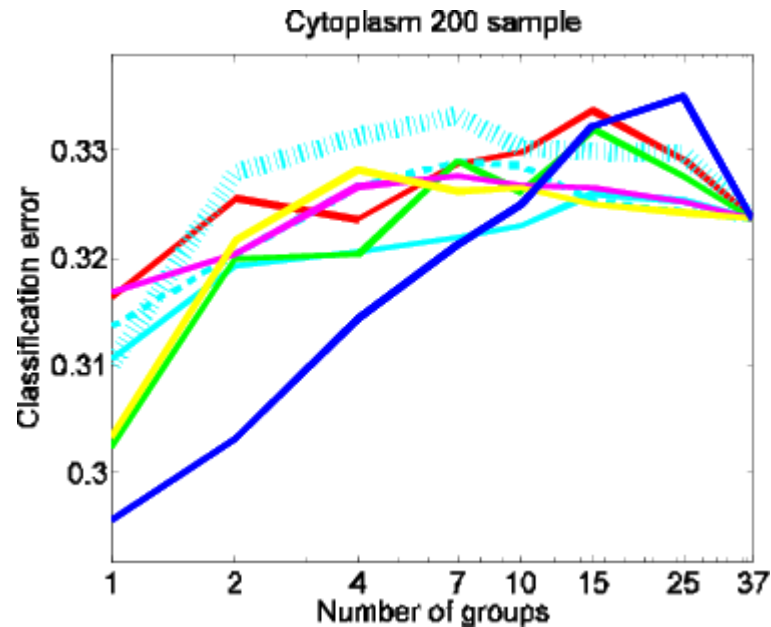
# 本発表の構成

55

1. 密度比は, 重点サンプリングだけでなく, 様々な統計的データ解析に役立つ!
2. 密度比推定の手法
3. 実験
  - 共変量シフト適応
  - 外れ値検出
  - 特徴選択
  - 条件付き確率推定

# 特徴選択：イースト菌細胞の分類 <sup>56</sup>

Suzuki, Sugiyama, Kanamori & Sese (BMC Bioinformatics 2009)



**KDE**: Kernel density estimation

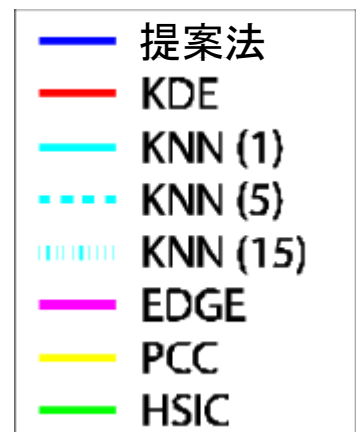
**KNN**: k-nearest neighbor

**EDGE**: Edgeworth expansion

**PCC**: Pearson correlation

**KNN**: k-nearest neighbor method

**HSIC**: Hilbert-Schmidt independence criterion







# 本発表の構成

57

1. 密度比は, 重点サンプリングだけでなく, 様々な統計的データ解析に役立つ!
2. 密度比推定の手法
3. 実験
  - 共変量シフト適応
  - 外れ値検出
  - 特徴選択
  - 条件付き確率推定

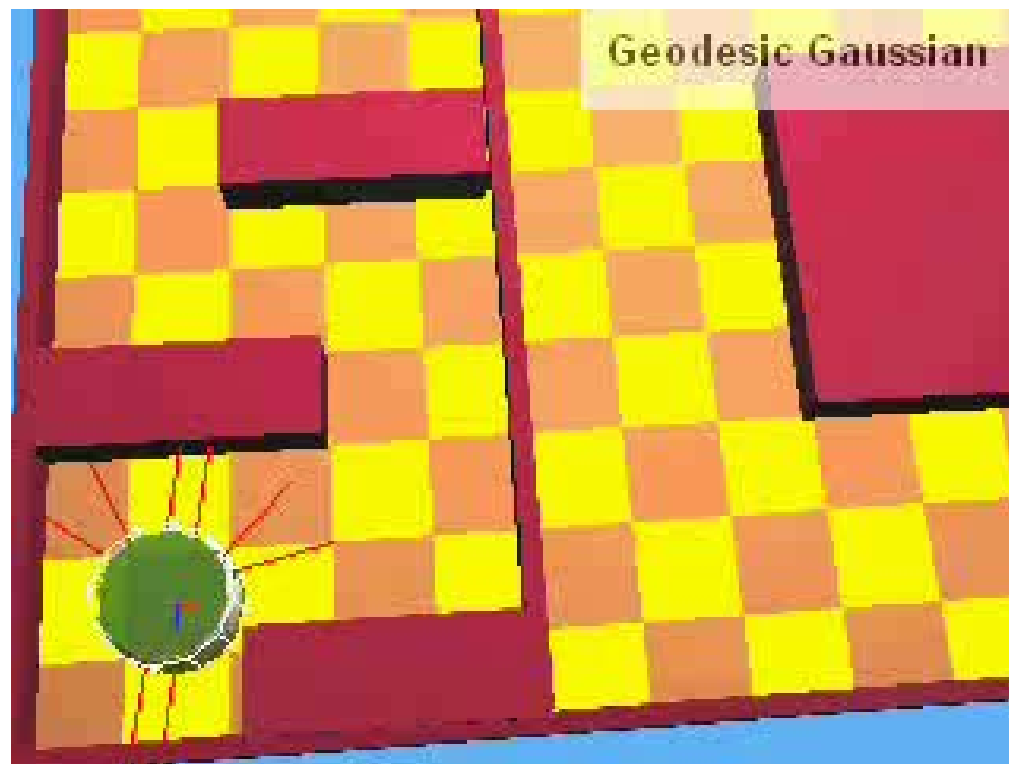
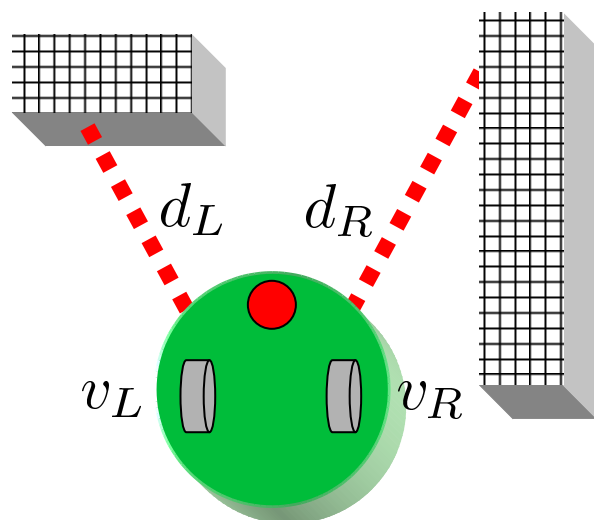
# 条件付き確率推定： ロボットの状態遷移推定

Sugiyama, Takeuchi, Suzuki, Kanamori & Hachiya (in preparation)

- **状態遷移確率**  $p(s'|s, a)$ : 状態  $s$  で行動  $a$  をとったときの行き先  $s'$  の確率分布

## Khereraロボット

- **状態**: 左右の赤外線センサの値
- **行動**: 左右のタイヤの回転速度



# 条件付き確率推定： ロボットの状態遷移推定

## ■ 相対対数尤度 (大きいほうが良い)

データ	提案法	KDE	MDN
Khepera1	1.07	0.80	0
Khepera2	1.02	0.78	0
振り子1	1.92	1.64	0
振り子2	1.92	1.62	0
平均	1.48	1.21	0

**KDE**: Kernel density estimation

**MDN**: Mixture density network      Bishop (TechRep1994)

## ■ 密度比推定: 統計的データ処理の新しい枠組み

- パターン認識, 非定常環境下での適応学習, ドメイン適応, マルチタスク学習, 異常値検出, 時系列の変化点検知, 特徴選択, 次元削減, 独立成分分析, 条件付き確率推定, 二標本検定など

## ■ 密度比を直接推定する手法

- KLIEP, LSIF, uLSIF
- 必要最小限の問題を解いている (Vapnikの原理)

## ■ 様々な応用例

- ブレイン・コンピュータインターフェース, ロボット制御, 音声認識, 自然言語, 故障検知, バイオデータ解析

# 関連書籍

61

- Quiñonero-Candela, Sugiyama, Schwaighofer & Lawrence (Eds.), **Dataset Shift in Machine Learning**, MIT Press, 2008.



- 八谷大岳, 杉山将  
**強くなるロボティック・  
ゲームプレイヤーの作り方,**  
毎日コミュニケーションズ, 2008

