

6. ノンパラメトリック推論への展開

正定値カーネルによるデータ解析
－ カーネル法の基礎と展開 －

福水健次

統計数理研究所／総合研究大学院大学

統計数理研究所 公開講座

2011年1月13,14日



概要

1. RKHS上の平均
2. RKHS上の共分散
3. 確率変数の依存性解析

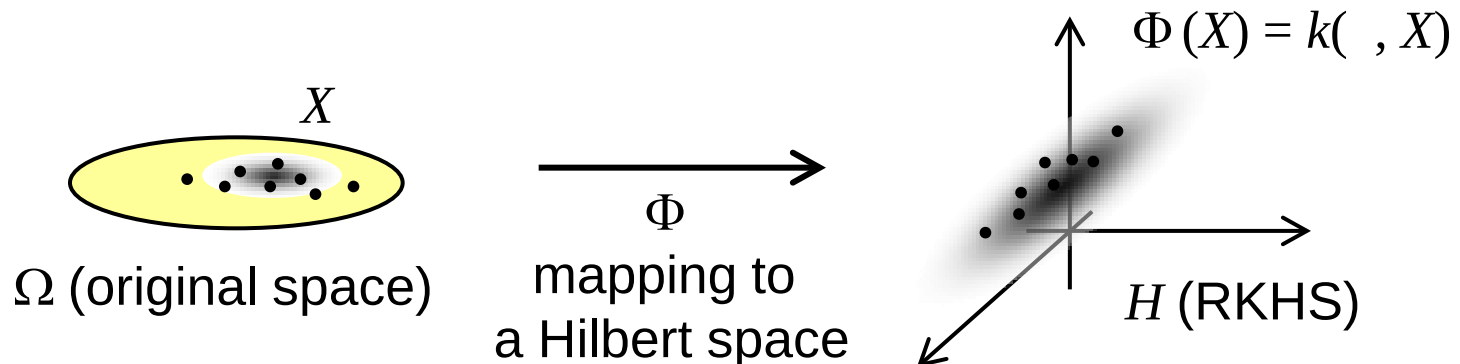
イントロ

- カーネル法(これまでの復習)

- 正定値カーネルは, 特徴写像を通じて, 計算効率の高い形で, データの非線形性／高次モーメントを抽出する.

e.g. Support vector machine, kernel PCA, kernel CCA, etc.

- Kernelization: データをRKHS に写像し, RKHS上で線形アルゴリズムを適用する.



- Consider more basic statistics!

- RKHS上で基本的な統計量(平均,共分散 ...) を考えると,元の確率変数／確率分布に関する推論が可能となる.

Basic statistics
on Euclidean space

Mean

Covariance

Conditional covariance

Basic statistics
on RKHS

Mean

Cross-covariance operator

Conditional-covariance operator

- 正定値カーネルによるノンパラメトリック推論が可能
 - 分布の均一性
 - 独立性／依存性
 - 条件付独立性／条件付依存性

概要

1. RKHS上の平均
2. RKHS上の共分散
3. 確率変数の依存性解析

RKHSにおける平均

$(\Omega, \mathcal{B})$: 測度空間.

X : Ω に値を取る確率変数.

k : Ω 上の可測な(実)正定値カーネル. H : 対応するRKHS.

– 特徴ベクトル

$\Phi(X) = k(\cdot, X)$: RKHSに値を取る確率変数 (random function)

– RKHS H における X の平均^{*)}:

$$m_X^k := E[\Phi(X)] \in H.$$

– 注意: m_X^k も H の元, すなわち Ω 上の関数.

^{*)} 厳密には $E \|\Phi(X)\| = E[\sqrt{k(X, X)}] < \infty$ の仮定の下で, 期待値の存在が保証される.

– 平均の再生性

$$\langle m_X^k, f \rangle = E[f(X)] \quad (\forall f \in H)$$

– 直観的な説明

Empirical mean $\hat{m}_X^k(u) = \frac{1}{N} \sum_{i=1}^N \Phi(X_i) = \frac{1}{N} \sum_{i=1}^N k(\cdot, X_i)$

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N f(X_i) &\longrightarrow E[f(X)] \\ = \langle \hat{m}_X^k, f \rangle &\longrightarrow \langle m_X^k, f \rangle \end{aligned}$$

– Explicit form:

$$m_X^k(u) = E[k(u, X)] = \int k(u, x) dP(x)$$

$$\because m_X(u) = \langle m_X, k(\cdot, u) \rangle = E[k(X, u)].$$

e.g.) For Gaussian kernel, $m_X^k(u) = \int \exp\left(-\frac{1}{2\sigma^2} \|x - u\|^2\right) p(x) dx$

平均と高次モーメント

- RKHSにおける平均は X の高次モーメントの情報を持つ.

例による直観的説明)

X : $\mathbf{R}$ -valued random variable. k : pos.def. kernel on $\mathbf{R}$.

k が $\mathbf{R}$ 上次のべき級数展開を持つと仮定する.

$$k(u, x) = c_0 + c_1(xu) + c_2(xu)^2 + \cdots \quad (c_i > 0)$$

$$\text{e.g.) } k(x, u) = \exp(xu)$$

H における平均 m_X^k はモーメント母関数の役割をする:

$$m_X^k(u) = E[k(u, X)] = c_0 + c_1 E[X]u + c_2 E[X^2]u^2 + \cdots$$

$$\frac{1}{c_\ell} \frac{\partial m_X^k(u)}{\partial u} \bigg|_{u=0} = E[X^\ell]$$

特性的な正定値カーネル

$\mathcal{P}$: 可測空間 $(\Omega, \mathcal{B})$ 上の確率測度全体

(H, k) : Ω 上の有界で可測な(実)正定値カーネルと, 対応するRKHS.

m_P : 確率 $P \in \mathcal{P}$ の H における平均

Def. 正定値カーネル k が($\mathcal{P}$ に関して) **特性的(characteristic)**であるとは, 写像

$$\mathcal{P} \rightarrow H, \quad P \mapsto m_P$$

が1対1であることをいう.

– 特性的なカーネルは, 確率測度を平均によって一意に定める.

$$m_P = m_Q \iff P = Q$$

i.e.

$$E_{X \sim P}[k(u, X)] = E_{X \sim Q}[k(u, X)] \iff P = Q$$

– 「確率分布に関する推論」 $\rightarrow$ 「RKHSにおける平均に関する推論」

– 「特性関数」の一般化

Fourier kernel $k_F(x, y) = \exp(\sqrt{-1} x^T y)$

$$\text{Ch.f.}_X(u) = E[k_F(X, u)].$$

- 特性関数は $\mathbf{R}^m$ 上のBorel 確率分布を一意に定める.
- 特性的なカーネルは, この性質の一般化.
注) Ω はユークリッド空間でなくてもよい.

– Examples: $\mathbf{R}^m$ 上の特性的な正定値カーネル

- Gaussian RBF kernel

$$k_G(x, y) = \exp\left(-\frac{1}{2\sigma^2} \|x - y\|^2\right)$$

- Laplacian kernel

$$k_L(x, y) = \exp\left(-\alpha \sum_{i=1}^m |x_i - y_i|\right)$$

- 特性的なカーネルの定めるRKHSは十分広い関数をカバーする.

Theorem (Fukumizu et al 2009)

k : 可測空間 $(\Omega, \mathcal{B})$ 上の有界で可測な正定値カーネル.

k が特性的であるための必要十分条件は, $(\Omega, \mathcal{B})$ の任意の確率分布 P に対して $H + \mathbf{R}$ が $L^2(P)$ で稠密なことである.

- 多項式カーネルは特性的でない.
 - $(\mathbf{x}^T \mathbf{y} + c)^d$ の定めるRKHS d 次以下の多項式
 - d 次より大きい次数のモーメントは考慮できない.

RKHSにおける平均の推定量

- RKHS上の標本平均

- RKHS上の統計量は, データによる推定は容易.
- $X^{(1)}, \dots, X^{(N)}$: i.i.d. sample
 $\rightarrow \Phi(X_1), \dots, \Phi(X_N)$: i.i.d. sample on RKHS

Empirical mean: $\hat{m}_X^{(N)} = \frac{1}{N} \sum_{i=1}^N \Phi(X_i) = \frac{1}{N} \sum_{i=1}^N k(\cdot, X_i)$

Theorem (strong $\sqrt{N}$ -consistency)

$E[k(X, X)] < \infty$ を仮定する. i.i.d. サンプル $X_1, \dots, X_N$ に対し

$$\|\hat{m}_X^{(N)} - m_X^k\| = O_p(1/\sqrt{N}) \quad (N \rightarrow \infty)$$

証明略: 福水2010を見よ.

分布の均一性への応用

- 2標本問題

Data: $X_1, \dots, X_n; Y_1, \dots, Y_m$

問題: 2つのサンプル(XとY)が同一の分布から発生したか?

- 帰無仮説: $P_X = P_Y$
- 対立仮説: $P_X \neq P_Y$

特性的なカーネルを用いて, 分布の問題を, RKHSにおける平均の問題に置き換える.

- 検定統計量

$$\|\hat{m}_X - \hat{m}_Y\|_H^2 = \frac{1}{n^2} \sum_{i,j=1}^n k(X_i, X_j) - \frac{2}{nm} \sum_{i=1}^n \sum_{a=1}^m k(X_i, Y_a) + \frac{1}{m^2} \sum_{a,b=1}^m k(Y_a, Y_b)$$

- 詳細は本講座では省略. 福水(2010 朝倉)または原論文参照.

概要

1. RKHS上の平均
2. RKHS上の共分散
3. 確率変数の依存性解析

RKHSにおける共分散

$(X, Y): \Omega_X \times \Omega_Y$ に値を取る確率変数.

$(H_X, k_X), (H_Y, k_Y): \Omega_X$ および Ω_Y 上の実RKHS.

Assume $E[k_X(X, X)] < \infty, E[k_Y(Y, Y)] < \infty$.

- Def. 相互共分散作用素 (Cross-covariance operator): $\Sigma_{YX} : H_X \rightarrow H_Y$
以下の関係式で定義

$$\begin{aligned} \langle g, \Sigma_{YX} f \rangle &= E[g(Y) f(X)] - E[g(Y)]E[f(X)] \\ &= \text{Cov}[f(X), g(Y)] \end{aligned} \quad \text{for all } f \in H_X, g \in H_Y$$

- c.f. Euclidean case

$$V_{YX} = E[YX^T] - E[Y]E[X]^T : \text{covariance matrix}$$

$$b^T V_{YX} a = \text{Cov}[b^T Y, a^T X]$$

– 積空間における平均による表現

$$\begin{aligned}\langle g, \Sigma_{YX} f \rangle &= \text{Cov}[g(Y), f(X)] \\ &= E_{P_{XY}}[f(X)g(Y)] - E_{P_X \otimes P_Y}[f(X)g(Y)] \\ &= \left\langle fg, m_{P_{XY}} - m_{P_X \otimes P_Y} \right\rangle_{H_X \otimes H_Y}\end{aligned}$$

$H_X \otimes H_Y$: $k_X k_Y$ で定義される積空間

$P_X \otimes P_Y$: 周辺分布の独立な積

- 線形写像とテンソル積を同一視すると

$$\Sigma_{YX} = m_{P_{XY}} - m_{P_X \otimes P_Y}$$

と考えてよい.

独立性の特徴づけ

- 相互共分散作用素と独立性

Theorem (Fukmizu et al 2004)

カーネルの積 $k_X k_Y$ が $\Omega_X \times \Omega_Y$ 上の特性的なカーネルであるとき,

$$X \text{ と } Y \text{ が独立} \quad \Leftrightarrow \quad \Sigma_{XY} = O$$

proof)

$$\begin{aligned} \Sigma_{XY} = O &\Leftrightarrow m_{P_{XY}} = m_{P_X \otimes P_Y} \\ &\Leftrightarrow P_{XY} = P_X \otimes P_Y \quad (\text{by characteristic assumption}) \end{aligned}$$

– c.f. ガウス変数の場合

$$X \perp\!\!\!\perp Y \quad \Leftrightarrow \quad V_{XY} = O \quad \text{i.e. 無相関}$$

– c.f. 特性関数

$$X \perp\!\!\!\perp Y \quad \Leftrightarrow \quad E_{XY}[e^{\sqrt{-1}(uX+vY)}] = E_X[e^{\sqrt{-1}uX}]E_Y[e^{\sqrt{-1}vY}]$$

相互共分散作用素の推定量

$(X_1, Y_1), \dots, (X_N, Y_N)$: i.i.d. sample on $\Omega_X \times \Omega_Y$

標本相互共分散作用素:

$$\hat{\Sigma}_{YX}^{(N)} = \frac{1}{N} \sum_{i=1}^N \{ \Phi_Y(Y_i) - \hat{m}_Y \} \{ \Phi_X(X_i) - \hat{m}_X \}$$

Theorem ($\sqrt{N}$ -consistency)

$$\left\| \hat{\Sigma}_{YX}^{(N)} - \Sigma_{YX} \right\|_{HS} = O_p(1/\sqrt{N}) \quad (N \rightarrow \infty)$$

Def. 作用素 $A: H_1 \rightarrow H_2$ が **Hilbert-Schmidt作用素** であるとは, H_1, H_2 の任意の正規直交基底 $\{\varphi_i\}, \{\psi_j\}$ に対し

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \left| \langle \psi_j, A\varphi_i \rangle \right|^2 < \infty$$

このとき $\|A\|_{HS}^2 := \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \left| \langle \psi_j, A\varphi_i \rangle \right|^2$. (行列のFrobeniusノルム)

概要

1. RKHS上の平均
2. RKHS上の共分散
3. 確率変数の依存性解析

依存性尺度

- 独立性／依存性尺度(HSIC, Hilbert-Schmidt Independence Criterion, Gretton et al 2005)

$$M_{YX} = \|\Sigma_{YX}\|_{HS}^2$$

$$M_{YX} = 0 \iff X \perp\!\!\!\perp Y \quad k_X k_Y : \text{characteristic}$$

- Empirical dependence measure

$$\hat{M}_{YX}^{(N)} = \|\hat{\Sigma}_{YX}^{(N)}\|_{HS}^2$$

M_{YX} および $\hat{M}_{YX}^{(N)}$ を依存性／独立性の尺度として用いることができる.

相互共分散作用素のHSノルム

– 推定量のGram行列表現

HS-norm は部分空間 $\text{Span}\{\Phi_X(X_i) - \hat{m}_X^{(N)}\}_{i=1}^N, \text{Span}\{\Phi_Y(Y_i) - \hat{m}_Y^{(N)}\}$ で計算すればOK.

$$\Rightarrow \hat{M}_{YX}^{(N)} = \frac{1}{N^2} \text{Tr}[G_X G_Y]$$

G_X, G_Y : 中心化Gram行列

展開すると

$$\begin{aligned} \hat{M}_{YX}^{(N)} = \|\hat{\Sigma}_{YX}^{(N)}\|_{HS}^2 &= \frac{1}{N^2} \sum_{i,j=1}^N k_X(X_i, X_j) k_Y(Y_i, Y_j) - \frac{2}{N^3} \sum_{i,j,k=1}^N k_X(X_i, X_j) k_Y(Y_i, Y_k) \\ &\quad + \frac{1}{N^4} \sum_{i,j=1}^N k_X(X_i, X_j) \sum_{k,\ell=1}^N k_Y(Y_k, Y_\ell) \end{aligned}$$

正規化された相互共分散作用素

- Normalized Cross-Covariance Operator

NOCCO $W_{YX} = \Sigma_{YY}^{-1/2} \Sigma_{YX} \Sigma_{XX}^{-1/2}$

- 独立性の特徴づけ

特性的なカーネルを用いると,

$$W_{YX} = O \quad \Leftrightarrow \quad X \perp\!\!\!\perp Y$$

W_{XY} が Hilbert-Schmidt であれば

– 依存性／独立性尺度として次を用いることができる.

$$\text{NOCCO} = \|W_{YX}\|_{HS}^2$$

カーネルに依存しない表現

Theorem (Fukumizu et al. NIPS 21, 2008)

Assume

P_{XY} have density $p_{XY}(x, y)$

$k_X k_Y$ is characteristic.

W_{YX} is Hilbert-Schmidt.

Then,

$$\|W_{YX}\|_{HS}^2 = \iint \left(\frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} - 1 \right)^2 p_X(x)p_Y(y) dx dy$$

- 左辺はカーネルを用いて定義したが、右辺は確率のみに依存する.
- 右辺は χ^2 -divergence (mean square contingency)としてよく知られた依存性尺度である.

サンプルによる推定量

- Cross-covariance operatorの推定量は既出: $\hat{\Sigma}_{YX}^{(N)}$
- 逆行列 $\rightarrow$ 正則化: $\Sigma_{XX}^{-1} \rightarrow (\hat{\Sigma}_{XX}^{(N)} + \varepsilon I)^{-1}$
- $W_{YX} = \Sigma_{YY}^{-1/2} \Sigma_{YX} \Sigma_{XX}^{-1/2}$ の各因子を推定量で置き換えると,

$$\text{NOCCO}_{emp} = \text{Tr}[R_X R_Y] \quad (\text{dependence measure})$$

$$\text{where } R_X \equiv G_X (G_X + N \varepsilon_N I_N)^{-1} \quad G_X : \text{中心化Gram行列}$$

- NOCCO_{emp} は χ^2 -divergence.の新しい推定量を与える.
一貫性を持つことが知られている (Fukumizu et al 2008)

独立性検定

- 正定値カーネルを用いたノンパラメトリックな独立性検定

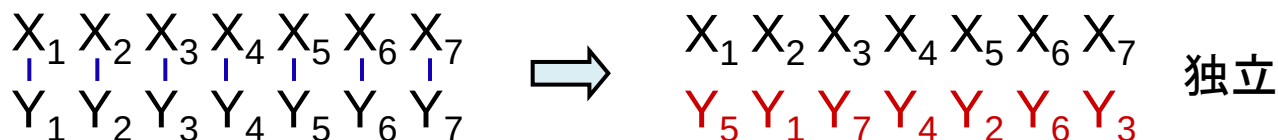
Data: $(X_1, Y_1), \dots, (X_N, Y_N)$ i.i.d.

- Null hypothesis H_0 : X and Y are independent
- Alternative H_1 : X and Y are **not** independent

$\hat{M}_{YX}^{(N)}$ および $\text{NOCCO}_{\text{emp}}$ を検定統計量として用いることができる.

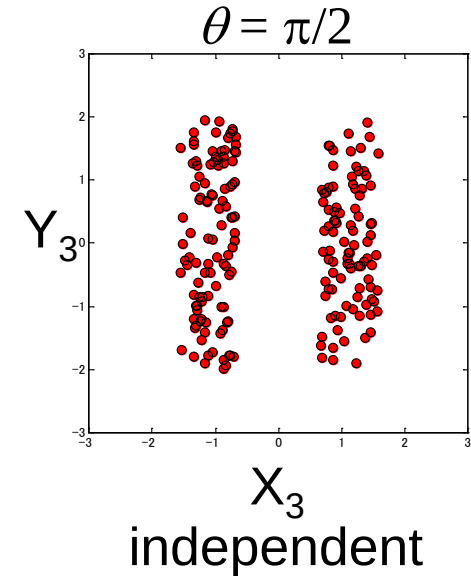
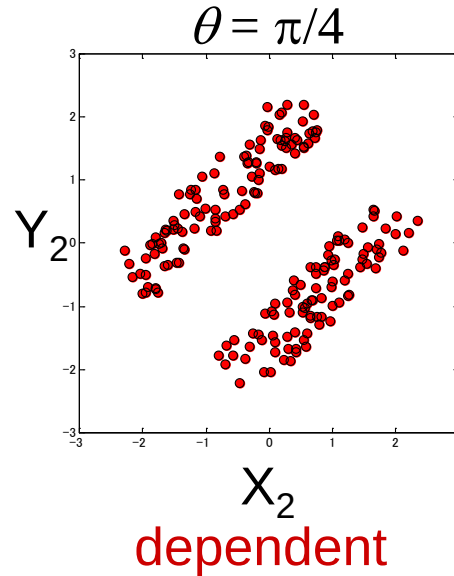
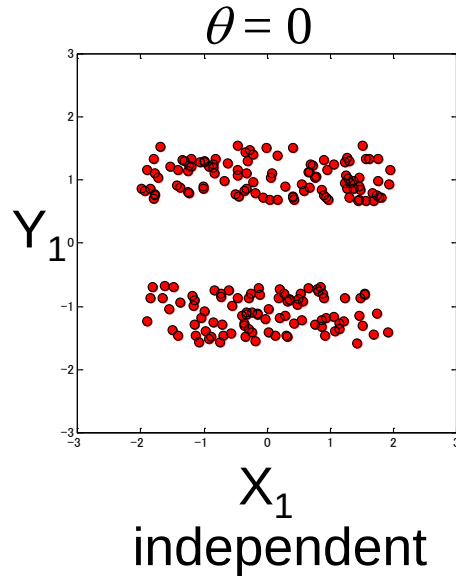
- $\hat{M}_{YX}^{(N)}$ に対しては, 帰無仮説, 対立仮説のもとでの漸近分布が知られている(Gretton et al 2008, 2010).
- $\text{NOCCO}_{\text{emp}}$ に対しては, 漸近分布は今のところ知られていない.
- どちらに対しても, 並べ替え検定*は可能.

* 並べ替え検定: (X_i, Y_i) に対して, Y の順序をランダムに変えることによって, 帰無仮説からのサンプルをシミュレートする方法.



独立性検定への応用

- Toy example



They are all uncorrelated, but dependent for $0 < \theta < \pi/2$

N = 200.

Permutation test is used.

Angle	indep.	—————→ more dependent				
	0.0	4.5	9.0	13.5	18.0	22.5
HSIC (Median)	93	92	63	5	0	0
HSIC (Asymp. Var.)	93	44	1	0	0	0
NOCCO ($\varepsilon = 10^{-4}$, Median)	94	23	0	0	0	0
NOCCO ($\varepsilon = 10^{-6}$, Median)	92	20	1	0	0	0
NOCCO ($\varepsilon = 10^{-8}$, Median)	93	15	0	0	0	0
NOCCO (Asymp. Var.)	94	11	0	0	0	0
MI (#NN = 1)	93	62	11	0	0	0
MI (#NN = 3)	96	43	0	0	0	0
MI (#NN = 5)	97	49	0	0	0	0
Power Diverg. (#Bins=3)	96	92	43	9	1	0
Power Diverg. (#Bins=4)	98	29	0	0	0	0
Power Diverg. (#Bins=5)	94	60	2	0	0	0

acceptance of independence out of 100 tests (significance level = 5%)

MI: mutual information estimated by the nearest neighbor method.

- Power Divergence (Ku&Fine05, Read&Cressie)

- 領域を分割 $\{A_j\}_{j \in J}$: 各次元を q 個の区間に分割 (各領域になるべく同数のデータが入るようにする).

- Power-divergence

$$T_N = 2I^\lambda(X, m) = N \frac{2}{\lambda(\lambda + 2)} \sum_{j \in J} \hat{p}_j \left\{ \left(\hat{p}_j / \prod_{k=1}^N \hat{p}_{j_k}^{(k)} \right)^\lambda - 1 \right\}$$

$I^0 = \text{MI}$

$I^2 = \text{Mean Square Conting.}$

$\hat{p}_j$: frequency in A_j

$\hat{p}_r^{(k)}$: marginal freq. in r -th interval

- 帰無仮説(独立)のもとでの漸近分布

$$T_N \Rightarrow \chi_{q^N - qN + N - 1}^2$$

テキストに対する独立性検定

- Data: Official records of Canadian Parliament in English and French.
 - Dependent data: 5 line-long parts from English texts and their French translations.
 - Independent data: 5 line-long parts from English texts and random 5 line-parts from French texts.
- Kernel: Bag-of-words and spectral kernel

Results of permutations test with HS measure

Topic	Match	BOW(N=10)	Spec(N=10)	BOW(N=50)	Spec(N=50)
Agri-culture	Random	0.94	0.95	0.93	0.95
	Same	0.18	0.00	0.00	0.00
Fishery	Random	0.94	0.94	0.93	0.95
	Same	0.20	0.00	0.00	0.00
Immig-ration	Random	0.96	0.91	0.94	0.95
	Same	0.09	0.00	0.00	0.00

Acceptance rate ($\alpha = 5\%$)

(Gretton et al. 2007)

独立性検定: 比較

- Brownian distance covariance (Székely & Rizzo AOAS 2010)
 - 特性関数に基づく独立性検定

$$X \perp\!\!\!\perp Y \quad \Leftrightarrow \quad \phi_{XY} = \phi_X \phi_Y$$

$$\phi_X(\omega) = E[e^{\sqrt{-1}X^T\omega}], \quad \phi_Y(\xi) = E[e^{\sqrt{-1}Y^T\xi}], \quad \phi_{XY}(\omega, \xi) = E[e^{\sqrt{-1}(X^T\omega + Y^T\xi)}].$$

- 重み付き積分による独立性尺度:

$$\int |\phi_{XY}(\omega, \xi) - \phi_X(\omega)\phi_Y(\xi)|^2 w(\omega, \xi) d\omega d\xi$$

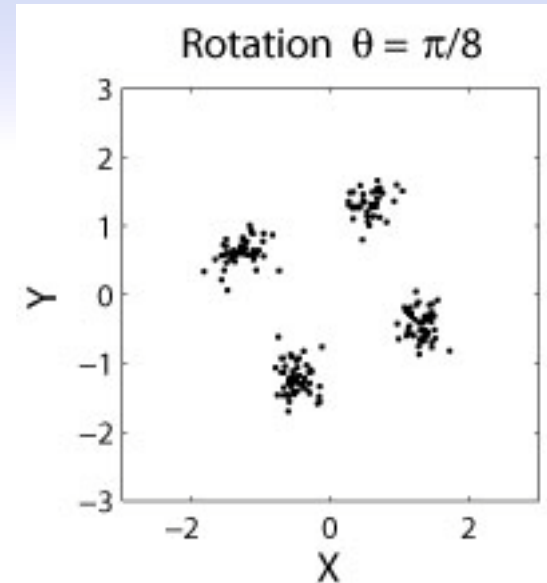
- 重み w を上手く選ぶと, 上の積分は, HSICのカーネルを
 $k(x_1, x_2) = \|x_1 - x_2\|$ で置き換えた形を持つ(正定値ではない).

比較実験

angle :		indep. $\longrightarrow$ more dependent			
		0	$\pi/12$	$\pi/6$	$\pi/4$
$d_X = d_Y = 2$	HS	0.94	0.77	0.48	0.42
$N = 128$	SR	0.95	0.83	0.66	0.65
$d_X = d_Y = 2$	HS	0.92	0.47	0.17	0.12
$N = 512$	SR	0.93	0.49	0.38	0.33
$d_X = d_Y = 4$	HS	0.92	0.60	0.35	0.23
$N = 1024$	SR	0.93	0.68	0.48	0.47
$d_X = d_Y = 4$	HS	0.92	0.44	0.15	0.12
$N = 2048$	SR	0.94	0.46	0.29	0.27

% of acceptance of indep. in permutation tests ($\alpha = 5\%$).

(Gretton, F, Sriperumbudur. 2010. AOAS Discussion)



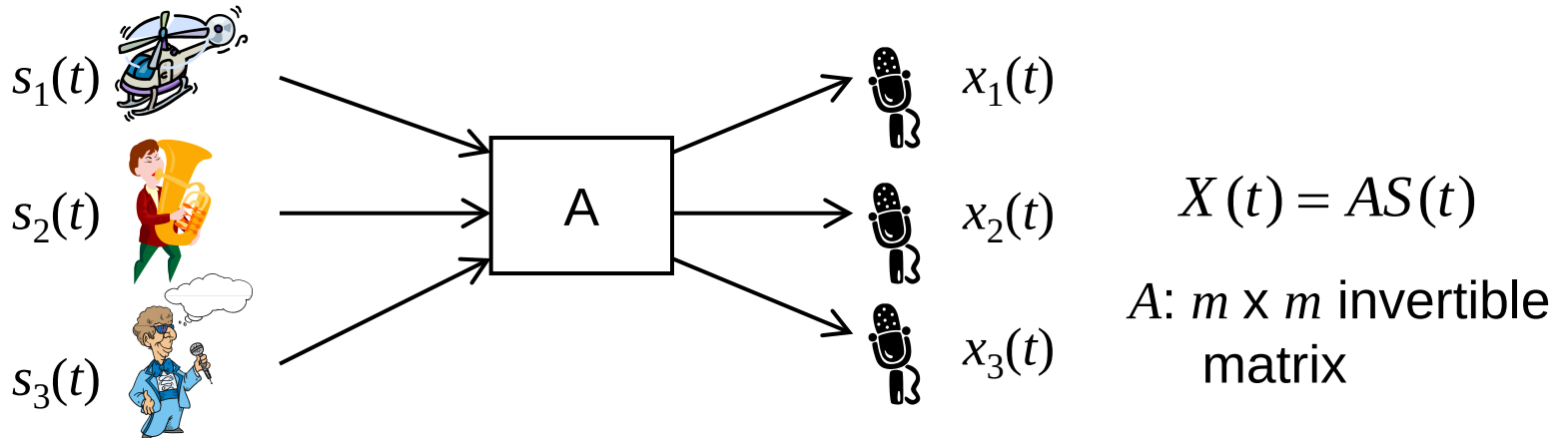
HS: Hilbert-Schmidt norm.
 Gaussian kernel
 $\sigma = \text{med}\{\|X_i - X_i\|\}$
 SR: Székely & Rizzo

独立成分分析への応用

- 独立成分分析 (Independent Component Analysis, ICA)

- 設定

- m 個の独立な信号源
- m 個の線形に重なり合った信号を測定



- 問題

- 測定値 X を用いて独立な信号 S を復元せよ.

$$\hat{S} = BX$$

B : $m \times m$ orthogonal matrix

- HSICを用いたICA

$X^{(1)}, \dots, X^{(N)}$: i.i.d. observation (m-dimensional)

ペアごとの独立性を尺度にする.

$$\text{Minimize} \quad L(B) = \sum_{a=1}^m \sum_{b>a} \hat{M}(Y_a, Y_b) \quad Y = BX$$

目的関数は非凸関数.

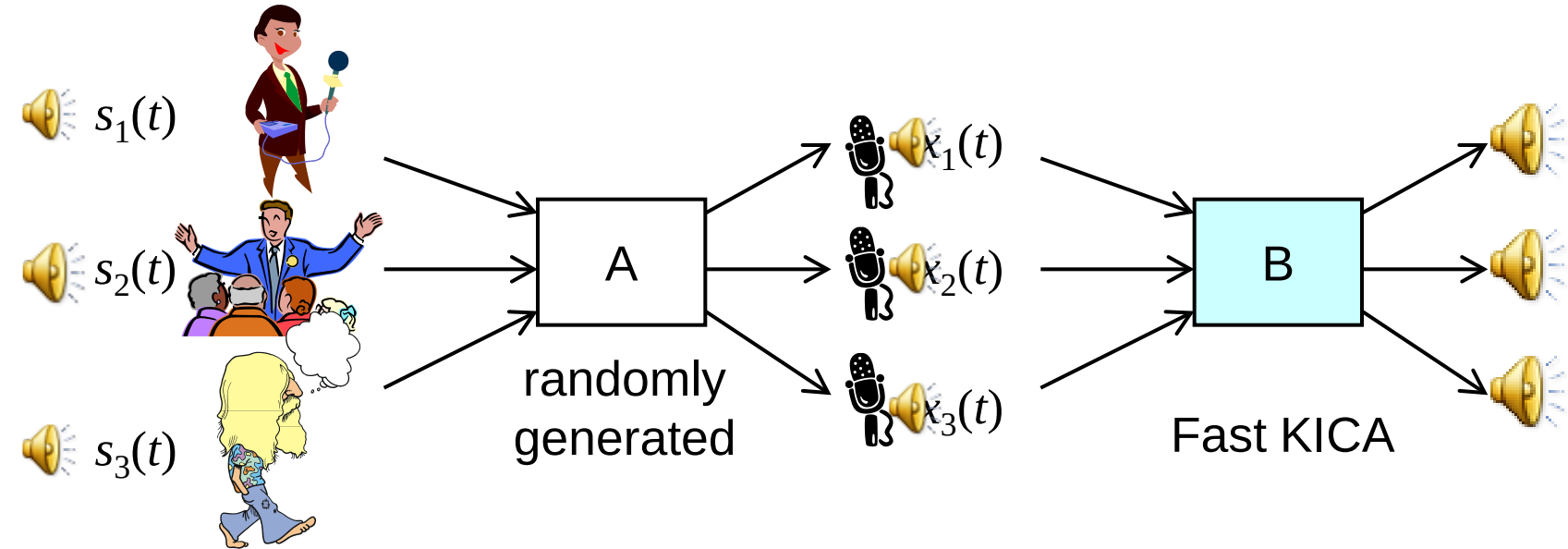
→ 近似的なNewton法による最小化

Fast Kernel ICA (FastKICA, Shen et al 2007)

- ICAの他の標準的方法

See, for example, Hyvärinen et al. (2001).

- Experiments (speech signal)



Three speech
signals

条件付独立性への発展

– アイデア:

RKHS上で, 条件付(相互)共分散作用素を考える.

$$\Sigma_{YX|Z} \equiv \Sigma_{YX} - \Sigma_{YZ} \Sigma_{ZZ}^{-1} \Sigma_{ZX} \quad : \quad H_X \rightarrow H_Y$$

$$\Sigma_{YY|Z} \equiv \Sigma_{YY} - \Sigma_{YZ} \Sigma_{ZZ}^{-1} \Sigma_{ZY} \quad : \quad H_Y \rightarrow H_Y$$

条件付独立性が特徴づけられる
(ガウス分布とのアナロジー)

– 応用

- 条件付独立性検定 (Fukumizu et al 2008)
- 因果性の推論 (Sun et al 2007, Mooij and Janzing 2010)
- 回帰問題における次元削減 (Fukumizu et al 2009)

– 詳細については, 福水(2010, 朝倉), 福水(2010, 統計数理)参照

セクション6のまとめ

- RKHSにおける平均

- 特性的なカーネルを用いると、確率分布が平均によって一意に定まる.
「分布に関する推論」 → 「RKHSでの平均に関する推論」

- RKHS上での共分散による依存性解析

- 相互共分散作用素により独立性の特徴づけが可能
- 作用素のノルムを用いて、独立性／依存性尺度を定義できる.
- 独立性尺度により、独立性検定が可能.

- さらなる話題

- 条件付独立性への発展
- Bayes推論への発展

Bayes事後確率のRKHSでの平均を計算することができる.

参考文献

- Fukumizu, K. Francis R. Bach and M. Jordan. (2009) Kernel dimension reduction in regression. *The Annals of Statistics*. 37(4), pp.1871-1905.
- Fukumizu, K., A. Gretton, X. Sun, and B. Schölkopf. (2008) Kernel Measures of Conditional Dependence. *Advances in Neural Information Processing Systems 21*, 489-496, MIT Press.
- Fukumizu, K., Bach, F.R., and Jordan, M.I. (2004) Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces. *Journal of Machine Learning Research*. 5(Jan):73-99.
- 福水 (2010) 正定値カーネルによるノンパラメトリック推論. 統計数理. 58(2)185—206
- Gretton, A., K. Fukumizu, C.H. Teo, L. Song, B. Schölkopf, and Alex Smola. (2008) A Kernel Statistical Test of Independence. *Advances in Neural Information Processing Systems 20*. 585-592. MIT Press.
- Gretton, A., K. M. Borgwardt, M. Rasch, B. Schölkopf and A. Smola. (2007) A Kernel Method for the Two-Sample-Problem. *Advances in Neural Information Processing Systems 19*, 513-520.
- Gretton, A., O. Bousquet, A. Smola and B. Schölkopf. (2005) Measuring Statistical Dependence with Hilbert-Schmidt Norms. Proc. Algorithmic Learning Theory (ALT2005), 63-78.

- Gretton, A., K. Fukumizu, Z. Harchaoui and B. Sriperumbudur. (2010) A Fast, Consistent Kernel Two-Sample Test. *Advances in Neural Information Processing Systems* 22, 673-681. MIT Press.
- Gretton, A., K. Fukumizu, Z. Harchaoui and B. Sriperumbudur. (2010) Discussion of: Brownian distance covariance. *Annals of Applied Statistics*. 3(4),1285-1294
- Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. Wiley.
- Mooij, J.M. and D. Janzing (2010) Distinguishing between cause and effect. *Journal of Machine Learning Research Workshop & Conference Proceedings* 6(Feb):147-156.
- Shen, H., S. Jegelka and A. Gretton: Fast Kernel ICA using an Approximate Newton Method. AISTATS 2007.
- Sriperumbudur, B., A. Gretton, K. Fukumizu, G. Lanckriet, and B. Schölkopf. (2008). Injective Hilbert Space Embeddings of Probability Measures. *Proc. 21st Annual Conference on Learning Theory (COLT 2008)*.
- Sun, X., Janzing, D. Schölkopf, B., and Fukumizu, K. (2007) A kernel-based causal learning algorithm. *Proc. 24th Annual International Conference on Machine Learning (ICML2007)*, 855-862.