カーネル法によるノンパラメトリックなベイズ推論とその応用

福水　健次　　数理・推論研究系　教授　／　統計的機械学習研究センター長

■ カーネル法

正定値カーネルを用いた，データの非線形性，高次モーメントを取り込むための新しい方法論．効率的な計算を重視．

- **定義：正定値カーネル**　Ω：集合．$k:\Omega\times\Omega\to\mathbf{R}$ が正定値であるとは，k が対称で任意の $n\in\mathbf{N}$, $x_1,\ldots x_n\in\Omega$ に対し，

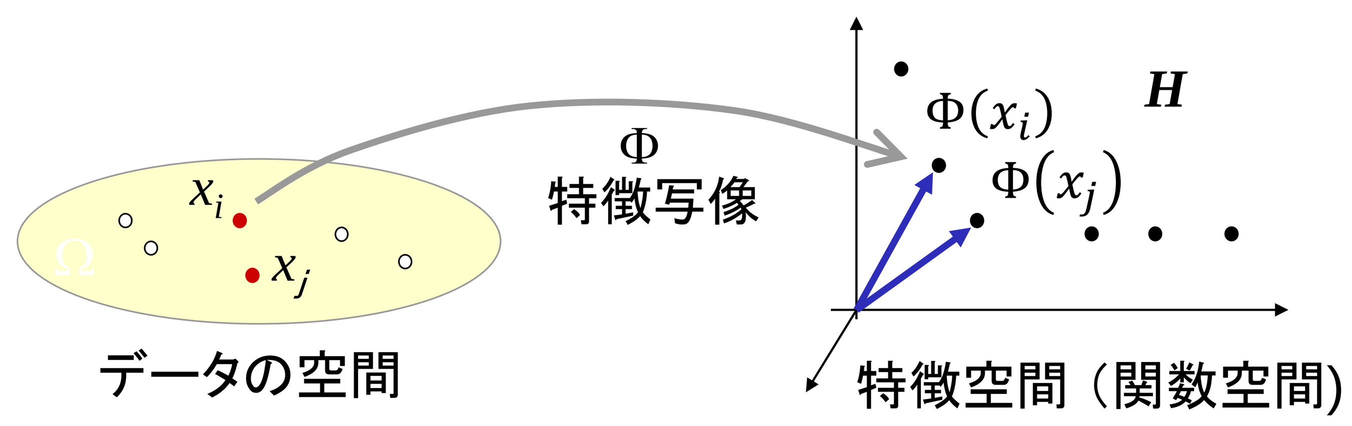
$$\text{グラム行列}\quad \begin{pmatrix} k(x_1,x_1) & \cdots & k(x_1,x_n) \\ \vdots & \ddots & \vdots \\ k(x_n,x_1) & \cdots & k(x_n,x_n) \end{pmatrix} \quad \text{が半正定値．}$$

例：ガウスカーネル　$\exp(-||x-y||^2/(2\sigma^2))$

- **再生核ヒルベルト空間（RKHS）**
  - 正定値カーネルに対しヒルベルト空間が一意に定まる．
  - Ω 上の関数からなる関数空間．一般に無限次元．
  - 特殊な内積を持つ．

$$\langle k(\cdot,x), f\rangle_{\mathcal{H}} = f(x)\ \forall f\in\mathcal{H}, x\in\Omega. \quad \text{（再生性）}$$

- **正定値カーネルとRKHSを用いたデータ解析**



- 特徴写像と特徴ベクトル

$$\Phi:\Omega\to H,\qquad x\mapsto\Phi(x):=k(\cdot,x).$$

- カーネルトリック：容易な内積計算

$$\langle\Phi(x),\Phi(y)\rangle = \langle k(\cdot,x),k(\cdot,y)\rangle = k(x,y).$$

- 特徴空間上で線形の解析手法を適用
  （注：多くの手法は内積計算が本質的）
  → カーネルPCA，サポートベクターマシン etc
- 非線形カーネル → 非線形／高次情報の抽出

■ カーネル平均による分布の表現

- **カーネル平均**：特徴ベクトル $\Phi(X)$ の期待値

$$m_P = E_{X\sim P}[\Phi(X)] = \int k(\cdot,x)dP(x)$$

P の積分変換

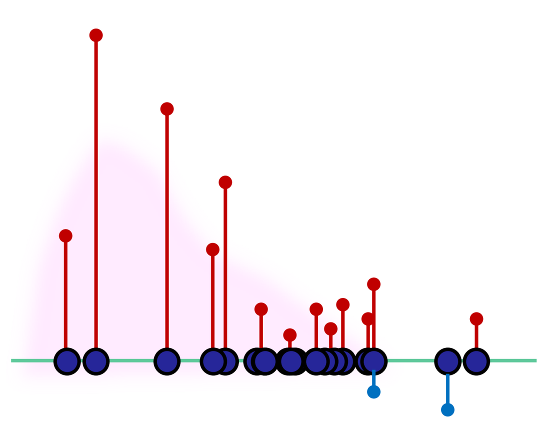
- **特性的なカーネル**
  m_P により P を一意に決める（例：ガウスカーネル）
- **カーネル平均によるノンパラメトリック推論**

> 原理：特性的なカーネルを用いると，
> 分布に関する推論　→　カーネル平均に関する推論

例：　2標本問題　$m_X = m_Y$？ (Gretton et al. NIPS'06)，
　　独立性検定　$m_{(X,Y)} = m_X\otimes m_Y$？ (Gretton et al. NIPS'07)

- **重み付サンプルによる推論**
  サンプルによる表現

$$\hat{m}_X = \sum_{i=1}^N w_i\Phi(X_i)$$



重み付サンプル (X_i,w_i) による表現とみなせる
重みは負も現れる．

■カーネル法による粒子フィルタ

観測モデルの密度関数（尤度）が書けない場合の粒子フィルタ
サンプリングは可能とする．

例）α-stable Stochstic Volatility model

$$x_t = \phi x_{t-1} + v_t,\quad v_t\sim N(0,\sigma_s^2)\qquad x_t\text{: log volatility}$$
$$y_t = e^{x_t/2}w_t,\quad w_t\sim S(\alpha,0,\sigma_o)\qquad y_t\text{: return}$$

$y_1,\ldots,y_t$：観測系列

- 事後確率 $p(X_t|y_1\ldots,y_t)$ のパーティクル表現：

$$\frac{1}{N}\sum_{i=1}^N\Phi(Z_i^t)$$

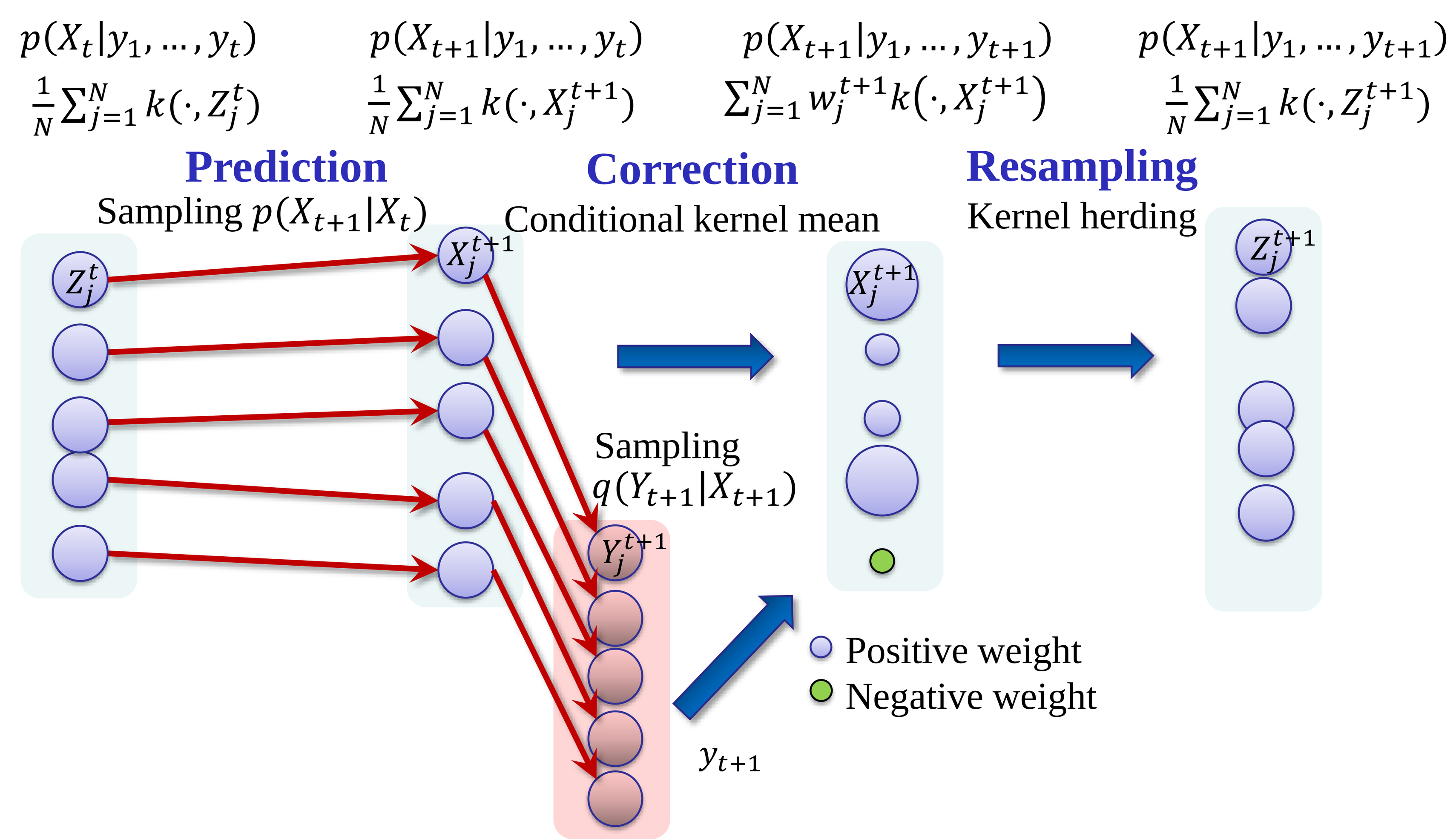
- Prediction step:
  サンプリング $X_i^{t+1}\sim p(x_{t+1}|x_t=Z_i^t)$
  $Y_i^{t+1}\sim p(y_{t+1}|x_{t+1}=X_i^{t+1})$　　$(i=1,\ldots,N)$
- Correction step:
  Importance weightingのカーネル化（線形演算）
  π: prior, $q(y|x)$: likelihood $\Rightarrow$　posterior $\propto q(y_o|x)\pi(x)$
  y_o：観測値

IW		カーネル版
$X_1,\ldots,X_N\sim Z\pi$	➡	$X_1,\ldots,X_N\sim\pi$
IW: $w_i = q(y_o\|X_i)$		$Y_1,\ldots,Y_N\sim q(y\|X_i)$
		$w=(G_Y+\lambda I_N)^{-1}\mathbf{k}(y_o)$
(X_i,w_i)		$\sum_{i=1}^N w_i^{t+1}\Phi(X_i^{t+1})$

- Resampling step:　Kernel herding による「リサンプリング」

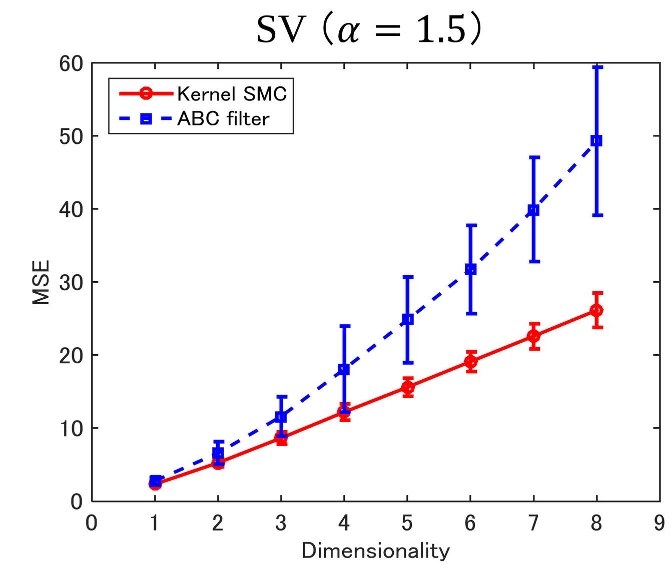
$$\sum_{i=1}^N w_i^{t+1}\Phi(X_i^{t+1})\ \Rightarrow\ \frac{1}{N}\sum_{i=1}^N\Phi(Z_i^{t+1})\quad \text{で近似}$$

一様重み

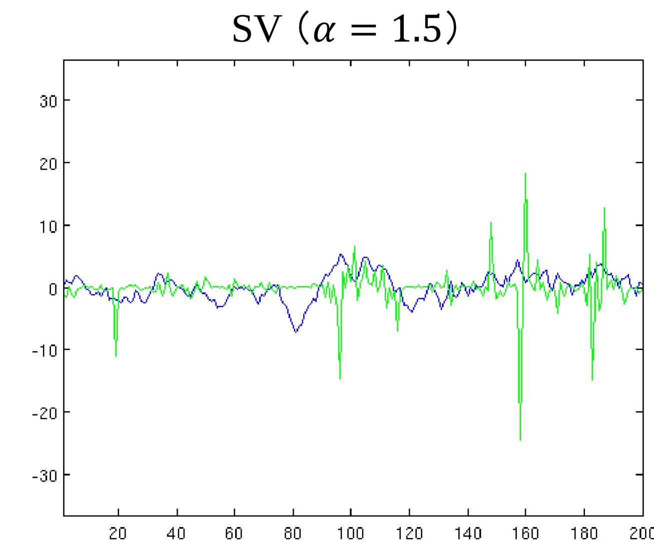


- 多変量Stochastic Vloatility model への応用
  $\alpha = 1.5, 2.0$ (Gauss)
  ABC filter (Calvet&Czeller 2014),
  標準的 Seq. MC (SIR) との比較

SV（$\alpha=1.5$）



SV（$\alpha=1.5$）



SV（α =2.0, Gauss）