Other Topics in Kernel Method

Statistical Inference with Reproducing Kernel Hilbert Space

Kenji Fukumizu

Institute of Statistical Mathematics, ROIS

Department of Statistical Science, Graduate University for Advanced Studies

September 26, 2008 / Statistical Learning Theory II

Outline

1. Relation to functional data analysis
2. Spline smoothing
3. Relation to random process

Outline

1. Relation to functional data analysis
2. Spline smoothing
3. Relation to random process

Functional data analysis

For functional data analysis, see Ramsay & Silverman (2005)

- What are functional data?

Data: $f_1(t),\ f_2(t),\ \ldots,\ f_N(t)$ -- functions on an interval $[a, b]$

Example: Berkeley Growth Study

See http://www.psych.mcgill.ca/misc/fda/index.html

Converting raw data into functional form

- Data are often given by a set of $\{(t_j,\ y_i(t_j)) \mid i = 1, \ldots, N,\ j = 1,\ldots, m_i\}$.
- Converting data by smoothing

  For each i, fit a curve $\phi_i(t)$ to individual data $\{(t_j,\ y_i(t_j)) \mid j = 1,\ldots, m_i\}$ by smoothing (e.g. B-spline)
- The converted data are of the form

$$\phi(t) = c_1\theta_1(t) + \cdots + c_\ell\theta_\ell(t), \qquad \theta_1(t),\ldots,\theta_\ell(t): \text{basis functions}$$

Analysis on functional data

- Apply linear methods to the “converted data” in a function space (typically L^2).
- Examples:
  - Functional PCA
  - Functional CCA
  - Functional linear modeling, etc.

Functional PCA

Functional data: $\phi_1(t), \ldots, \phi_N(t)$ (already converted)

Find a function to maximize

$$\mathrm{Var}_{emp}\left[\int \phi_i(t) f(t) dt\right] \qquad \text{subj.to} \quad \int f(t)^2 dt = 1.$$

Variance of the projections
on the direction of *f*

If basis functions $\theta_1(t), \ldots, \theta_\ell(t)$ are used,

$$\phi_i(t) = \sum_{j=1}^{\ell} c_{ij}\theta_j(t), \qquad f(t) = \sum_{j=1}^{\ell} \beta_j \theta_j(t)$$

Solve: $\max_\beta \beta^T V \beta \qquad \text{subj.to} \quad \beta^T R \beta = 1.$

where $V_{jk} = \frac{1}{N}\sum_{i=1}^{N}\sum_{s,t=1}^{\ell} c_{is}c_{it}R_{js}R_{kt}, \quad R_{jk} = \int \theta_j(t)\theta_k(t)dt.$

The integral in R is computed numerically, or by the property of the basis

Kernel method v.s. functional data analysis

- Similarity
  - Both the methods extends linear methods to “functional data”.

- Difference
  - In kernel methods, the data conversion is given by a positive definite kernel, while in FDA the data are assumed to be functional.
  - Kernel methods use RKHS as a function space, while the FDA uses L2 space in principle.

- Roughness penalty in FDA
  - In FDA, smoothness is sometimes imposed on the solution.

$$\int f(t)^2 = 1 \quad \Longrightarrow \int f(t)^2 + \lambda \int |Df(t)|^2 dt = 1$$

This is essentially the Sobolev norm (RKHS).

With roughness penalty, FDA is more similar to kernel methods.

Outline

1. Relation to functional data analysis

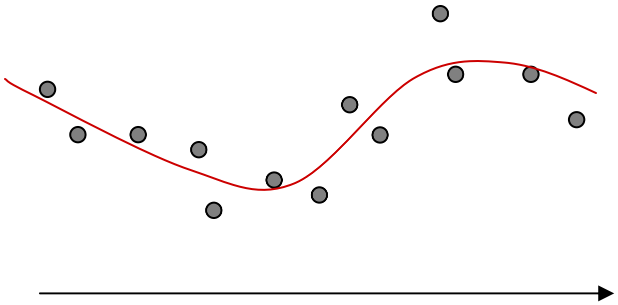
2. Spline smoothing

3. Relation to random process

Spline smoothing

$(X_1, Y_1), \ldots, (X_N, Y_N) \ : \ X_i \in \mathbf{R}^n, \ Y_i \in \mathbf{R}$

P: differential operator on $\mathbf{R}^n$



Spline smoothing:

$$\min_f \sum_{i=1}^{N} \left(Y^i - f(X^i) \right)^2 + \lambda \int | Pf(x) |^2 \, dx$$

Roughness penalty

Laplacian and Green function

- **Laplacian** $$\Delta f = \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}$$

Self-adjoint: if $|f(x)|, |g(x)| \rightarrow 0 \ (x \to \infty)$

$$\int \Delta f(x) g(x) dx = \int f(x) \Delta g(x) dx$$

- **Green function for Laplacian**

$$\Delta G(x,\xi) = \delta(x-\xi)$$

i.e. $$\int \Delta G(x,\xi) f(x) d\xi = f(\xi)$$

– Green function solves a differential equation: $\Delta f = \varphi$ given φ.

$\Longrightarrow$ $$f(x) = \int G(x,y)\varphi(y) dy$$

$\because$) $f(\xi) = \int f(x)\Delta G(x,\xi)dx = \int \Delta f(x) G(x,\xi)dx = \int \varphi(x) G(x,\xi) dx$

Smoothing penalty

- Regularization term

Consider functions on $\mathbf{R}^n$ for simplicity (no boundary)

$$J_m^n(f) = \sum_{\alpha_1+\cdots+\alpha_n=m} \frac{m!}{\alpha_1!\alpha_2!\cdots\alpha_n!} \left\| D^\alpha f \right\|_{L^2}^2$$

L^2 norm of m-th derivative

$$= \sum_{\alpha_1+\cdots+\alpha_n=m} \frac{m!}{\alpha_1!\alpha_2!\cdots\alpha_n!} \int \left| \frac{\partial^m f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_n^{\alpha_n}} \right|^2 dx$$

– example (n = m = 2)

$$J_2^2(f) = \int \left\{ \left| \frac{\partial^2 f}{\partial x_1^2} \right|^2 + 2 \left| \frac{\partial^2 f}{\partial x_1 \partial x_2} \right|^2 + \left| \frac{\partial^2 f}{\partial x_2^2} \right|^2 \right\} dx$$

Smoothing

$$\min_f \sum_{i=1}^{N} \left(Y^i - f(X^i) \right)^2 + \lambda \sum_{m=0}^{\infty} a_m J_m^n(f) \qquad (a_m \geq 0)$$

Expression by Laplacian

Partial integral shows

$$J_m^n(f) = (-1)^m \left(f, \Delta^m f \right)_{L^2}$$

The smoothing problem is expressed by

$$\min_f \sum_{i=1}^{N} \left(Y^i - f(X^i) \right)^2 + \lambda \left(f, Af \right)_{L^2}$$

where $A = \sum_{m=0}^{\infty} (-1)^m a_m \Delta^m$

Two cases

- Case $a_0 \neq 0$
  - The Green function is a positive definite kernel.
  - The penalty term is equal to the squared RKHS norm.

- Case $a_0 = 0$
  - Spline smoothing
  - The Green functions is conditionally positive definite.
  - The functional space is RKHS + polynomial of some order
  - The penalty term is equal to the squared RKHS norm of the projection of f onto the RKHS.

$a_0 \neq 0$: RKHS regularization

- Solution

$$\min_f \sum_{i=1}^{N}\left(Y^i - f(X^i)\right)^2 + \lambda(f, Af)_{L^2}$$

Variational calculus

$$\sum_{i=1}^{N}\left(Y^i - f(x)\right)\delta(x - X^i) + \lambda Af = 0$$

$$Af = -\frac{1}{\lambda}\sum_{i=1}^{N}\left(Y^i - f(x)\right)\delta(x - X^i)$$

If we have the Green function G for A $\;i.e.\; AG = \delta$

$$f(\xi) = -\frac{1}{\lambda}\sum_{i=1}^{N}\int (Y^i - f(x))\delta(x - X^i)G(x,\xi)dx$$

$$= -\frac{1}{\lambda}\sum_{i=1}^{N}(Y^i - f(X^i))G(\xi, X^i)$$

Note: $f(X_i)$ unknown

The solution is to have the form:

$$f = \sum_{i=1}^{N} c_i G(\cdot, X^i)$$

Plug it into the original problem:

$$\min_{c \in \mathbf{R}^N} \quad \sum_{i=1}^{N} \left(Y^i - \sum_{j=1}^{N} c_j G(X^i, X^j) \right)^2 + \lambda \sum_{i,j=1}^{N} c_i c_j G(X^i, X^j)$$

$$\because) \quad (Af, f)_{L^2} = \sum_{i.j} c_i c_j (AG(\cdot, X_i), G(\cdot, X_j))_{L^2} = \sum_{i.j} c_i c_j G(X_i, X_j)$$

By differentiation,

$$c = (G + \lambda I)^{-1} \mathbf{Y}$$

where $\quad G_{ij} = G(X^i, X^j) \quad \mathbf{Y} = (Y^1, \ldots, Y^N)^T$

The solution:

$$f(x) = \mathbf{Y}^T (G + \lambda I)^{-1} g(x)$$ where $g_i(x) = G(x, X^i)$

Green function

Theorem

If $a_0 \neq 0, a_j \neq 0 (\exists j \geq 1)$, the Green function of A is a positive definite kernel.

Proof.

Since A is shift invariant, so is G. Thus,

$$\sum_{m=0}^{\infty} (-1)^m a_m \Delta^m G(z) = \delta(z)$$

By Fourier transform

$$\sum_{m=0}^{\infty} a_m \|u\|^{2m} \widehat{G}(u) = \frac{1}{(2\pi)^{n/2}}$$

$$\widehat{G}(u) = \frac{1}{(2\pi)^{n/2}(a_0 + \sum_{m=1}^{\infty} a_m \|u\|^{2m})}$$

If $a_0 \neq 0, a_j \neq 0 (\exists j \geq 1)$, the Fourier inversion is possible.
Use Bochner's theorem.

Regularization by RKHS norm

Assume $a_0 \neq 0, a_1 \neq 0$

G: Green function of A.

H_G: RKHS w.r.t. G.

$$\min_f \quad \sum_{i=1}^{N} \left(Y^i - f(X^i) \right)^2 + \lambda \sum_{m=0}^{\infty} a_m J_m^n(f)$$

The solution is given by $f = \sum_{i=1}^{N} c_i G(\cdot, X^i)$

The penalty term is, then,

$$\sum_{m=0}^{\infty} a_m J_m^n(f) = \sum_{i,j} c_i c_j G(X_i, X_j) = \|f\|_{H_G}^2.$$

The above regularization is equivalent to the kernel ridge regression

$$\min_f \quad \sum_{i=1}^{N} \left(Y^i - f(X^i) \right)^2 + \lambda \| f \|_{H_G}^2$$

$a_0 = 0$: Spline smoothing

- Thin-plate spline

$$\min_f \sum_{i=1}^{N} \left(Y^i - f(X^i)\right)^2 + \lambda J_m^n(f)$$

$$J_m^n(f) = \sum_{\alpha_1+\cdots+\alpha_n=m} \frac{m!}{\alpha_1!\alpha_2!\cdots\alpha_n!} \left\| D^\alpha f \right\|_{L^2}^2$$

  – The Green function of J_m^n is not necessarily positive definite, but conditionally positive definite

  – The function space for f is

$$B_m^n: \quad D^\alpha f \in L^2(\mathbf{R}^n) \quad (|\alpha| = m)$$

and

$$J_m^n(f) = 0 \quad \Leftrightarrow \quad f \in \mathcal{P}_{m-1}$$

$\mathcal{P}_{m\text{-}1}$: Polynomials of degree at most m - 1

Let $B_m^n = \mathcal{P}_{m-1} \oplus H_*$ be decomposition by direct sum.

<u>Theorem</u> (Meinguet 1979)

If $m > n/2$, the subspace H_* is a RKHS with inner product

$$\langle f, g \rangle_{H_*} = \sum_{|\alpha|=m} \frac{m!}{\alpha_1! \cdots \alpha_n!} \left(D^\alpha f, D^\alpha g \right)_{L^2} = \left((-1)^m \Delta^m f, g \right)_{L^2}$$

In particular, the norm is given by

$$\| f \|_{H_*}^2 = J_m^n(f)$$

$$\min_f \quad \sum_{i=1}^N \left(Y^i - f(X^i) \right)^2 + \lambda J_m^n(f)$$

$\Longleftrightarrow$

$$\min_{g \in H_*, p \in \mathcal{P}_{m-1}} \quad \sum_{i=1}^N \left(Y^i - (g(X^i) + p(X^i)) \right)^2 + \lambda \| g \|_{H_*}^2$$

Solution of spline smoothing

By the representer theorem, the solution is to be of the form:

$$f(x) = \sum_{i=1}^{N} c_i K(x - X_i) + \sum_{\ell=1}^{M} b_\ell \phi_\ell (x)$$

By plugging it,

$$\min_{c,b} \quad (Y - Kc - Hb)^T (Y - Kc - Hb) + \lambda c^T K c$$

The solution:

$$(K + \lambda I)c + Hb = Y, \qquad H^T c = 0.$$

⇨

$$\begin{cases} c = (I_N - H(H^T H)^{-1} H^T)(K + \lambda I)^{-1} Y \\ b = (H^T H)^{-1} H^T (K + \lambda I)^{-1} Y \end{cases}$$

Conditionally positive definite

Definition. $K(x,y) : \Omega \times \Omega \rightarrow \mathbf{R}$ is said to be conditionally positive definite of order m if

1. $K(x,y) = K(y,x)$
2. If points $x_1, \ldots, x_n$ in Ω and real numbers $c_1,\ldots, c_n$ satisfy

$$\sum_{i=1}^{n} c_i p(x_i) = 0$$

for any polynomial $p(x) \in \mathcal{P}_{m-1}$ (generalized increment of order m), then

$$\sum_{i,j=1}^{n} c_i c_j K(x_i, x_j) \geq 0$$

– A positive definite kernel is conditionally positive definite of order 0.
– A negative definite kernel is negation of a conditionally positive definite kernel of order 1.
– Intuition: the above $c_1,\ldots, c_n$ is a generalization of the m-th order difference. Thus, the definition intuitively says that the m-th derivative of K is positive definite.

1st order diff.: $\dfrac{f(t_{i+1}) - f(t_i)}{t_{i+1} - t_i}$

Coeff. of $f(t_i)$

c_1	c_2	c_3	
$\frac{-1}{t_2 - t_1}$	$\frac{1}{t_2 - t_1}$	0	...
0	$\frac{-1}{t_3 - t_2}$	$\frac{1}{t_3 - t_2}$	

$c_1 + c_2 + \cdots + c_n = 0$ → coefficients of 1st order difference

2nd order diff.: $\left\{ \dfrac{f(t_{i+2}) - f(t_{i+1})}{t_{i+2} - t_{i+1}} - \dfrac{f(t_{i+1}) - f(t_i)}{t_{i+1} - t_i} \right\} \Big/ t_{i+2} - t_{i+1}$

Coeff. of $f(t_i)$

c_1	c_2	c_3	c_4	
$\frac{1}{t_2 - t_1} \Big/ t_3 - t_2$	$\frac{-1}{t_2 - t_1} + \frac{-1}{t_3 - t_2} \Big/ t_3 - t_2$	$\frac{1}{t_3 - t_2} \Big/ t_3 - t_2$	0	...
0	$\frac{1}{t_3 - t_2} \Big/ t_4 - t_3$	$\frac{-1}{t_3 - t_2} + \frac{-1}{t_4 - t_3} \Big/ t_4 - t_3$	$\frac{1}{t_4 - t_3} \Big/ t_4 - t_3$	

$$\begin{cases} c_1 + c_2 + \cdots + c_n = 0 \\ c_1 \times t_1 + c_2 \times t_2 + \cdots + c_n \times t_n = 0 \end{cases}$$

→ coefficients of 2nd order difference

Outline

1. Relation to functional data analysis
2. Spline smoothing
3. Relation to random process

Gaussian process

– A Gaussian process is a random process $\{X_t\}_{t\in\Omega}$ (random variables with index Ω) such that for any finite subset $\{t_1, ..., t_n\}$ of Ω, the random vector $(X_{t_1},\ldots,X_{t_n})$ is a Gaussian random vector.

– Mean function $\mu(t) = E[X_t]$

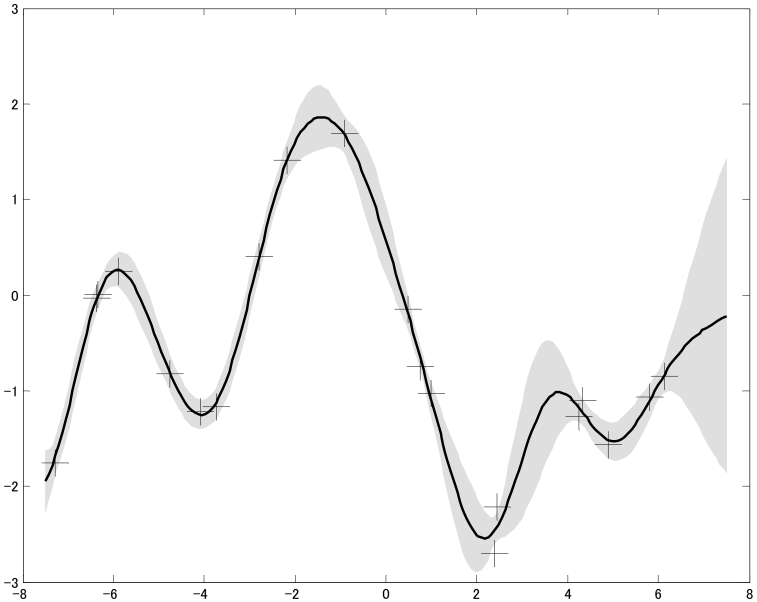
– Covariance function $R(t,s) = \mathrm{Cov}[X_t, X_s]$

– A Gaussian process is uniquely determined by the mean and covariance function.

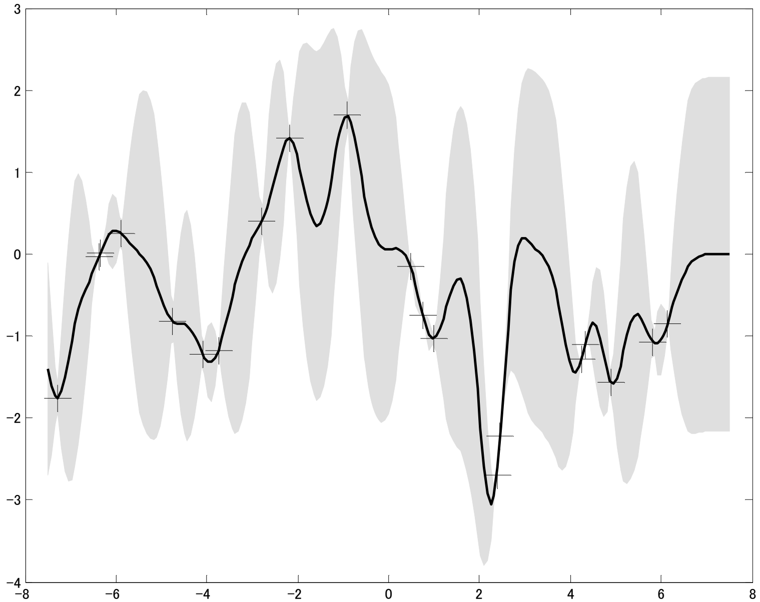
$$\mathbf{X} = (X_{t_1},\ldots,X_{t_n}) \sim N(\mu_{\mathrm{X}}, \Sigma_{\mathrm{X}})$$

$$\mu_{\mathbf{X}} = (\mu(t_1),\ldots,\mu(t_n)), \qquad \Sigma_{\mathbf{X}} = \begin{pmatrix} R(t_1,t_1) & R(t_1,t_2) & \cdots & R(t_1,t_n) \\ R(t_2,t_1) & R(t_2,t_2) & \cdots & R(t_2,t_n) \\ \vdots & \vdots & \ddots & \vdots \\ R(t_n,t_1) & R(t_n,t_2) & \cdots & R(t_n,t_n) \end{pmatrix}$$

– Examples



$\sigma = 1$



$\sigma = 0.3$

mean zero

covariance function $R(s,t) = \exp\left(-\frac{1}{2\sigma^2}(s-t)^2\right)$

Generated by Matlab gpml toolbox (Rasmussen and Williams)

Random process and positive definite kernel

- Covariance function is a positive definite kernel

<u>Theorem</u>

The covariance function $R(s,t)$ of a random process $\{X_t\}_{t\in\Omega}$ is a positive definite kernel.

$\because$) For simplicity, mean = 0.

$$\sum_{i,j=1}^n c_i c_j R(t_i,t_j) = \sum_{i,j=1}^n c_i c_j E[X_{t_i}, X_{t_j}]$$
$$= E\left[\sum_{i=1}^n c_i X_{t_i}, \sum_{j=1}^n c_j X_{t_j}\right] = E\left[\left(\sum_{i=1}^n c_i X_{t_i}\right)^2\right] \geq 0$$

– A random process on Ω determines a RKHS on Ω.

Positive definite kernel defines Gaussian process

$k(s,t)$: positive definite kernel on Ω.

For any finite subset $\mathbf{t} = (t_1, \ldots, t_n)$ of Ω, the Gram matrix $\Sigma_{\mathbf{t}} = (k(t_i, t_j))$ is always positive semidefinite.

By Kolmogorov extension theorem, there is a Gaussian process with index set Ω such that

$$\mathbf{X} = (X_{t_1}, \ldots, X_{t_n}) \quad \sim N(0, \Sigma_t)$$

The covariance function = $k(s,t)$.

RKHS by random process

$\{X_t\}_{t\in\Omega}$: random process on Ω with mean zero and finite 2nd moments.

$X_t : \Xi \to \mathbf{R}$ random variable defined by a probability space.

$\Rightarrow$

$$X_t \in L^2(\Xi, \mathcal{B}, P)$$

$$\overline{\mathcal{L}}(X) \equiv \overline{LH\{X_t \in L^2(\Xi) \mid t \in \Omega\}}$$ closed subspace of $L^2(\Xi)$

Hilbert space generated by $\{X_t\}_{t\in\Omega}$

Inner product

$$(U,V)_{\overline{\mathcal{L}}(\Xi)} = E[UV] \qquad U,V \in \overline{\mathcal{L}}(X)$$

(inner product of $L^2(\Xi)$)

RKHS and random process

<u>Theorem</u>

k : positive definite kernel on a set Ω

$\{X_t\}_{t\in\Omega}$: random process with mean 0 and covariance function k

⇨ $$\overline{\mathcal{L}}(X) \cong H_k \quad \text{(isomorphic as Hilbert space)}$$

$$X_t \leftrightarrow k(\cdot,t)$$

$$(U,V)_{\overline{\mathcal{L}}(\Xi)} = \langle f, g\rangle \qquad U \leftrightarrow f,\ V \leftrightarrow g$$

注） $$(X_t, X_s)_{\overline{\mathcal{L}}(\Xi)} = E[X_t X_s] = k(t,s) = \langle k(\cdot,t), k(\cdot,s)\rangle_{H_k}$$

(inner product) (cov) (reproducing)

Stationary process and shift-invariant kernel

- Stationary case

$\{X_t\}_{t\in\mathbb{R}^m}$: random process on $\mathbf{R}^m$

– stationary process

$$E[X_{t+h}X_{s+h}] = E[X_t X_s] \qquad (\forall t, s, h \in \mathbf{R}^m)$$

covariance function is given by

$$R(t,s) \equiv R(t-s)$$

– Positive definite kernel for a stationary process is given by

$$K(t,s) = K(t-s)$$

– Bochner's theorem ⇔ Wiener-Khinchine's theorem (covariance function of a stationary process on $\mathbf{R}^m$ is the inverse Fourier transform of the power spectral.)

Inference with random process

- Estimation of random process
  - Modeling by a random process

    X_t : random process on Ω with mean zero and finite 2nd moments

    $$Y_t = X_t + \varepsilon_t$$

    ε_t : noise indep. with X_t $E[\varepsilon_t] = 0,\ \mathrm{Cov}[\varepsilon_t, \varepsilon_s] = \sigma^2 \delta(t - s)$

    $R(t,s) = \mathrm{Cov}[X_t, X_s]$: known. σ^2 : known

  - Estimation

    Estimate X_{t_0} for t_0 given the observation $Y_{t_1}, \ldots, Y_{t_n}$

Minimizing mean square error

- Linear estimator for random process
  - Linear estimator

$$\hat{X}_{t_0} = \sum_{j=1}^{n} \alpha_j Y_{t_j}$$

  - Mean square error

$$\min E\,|X_{t_0} - \hat{X}_{t_0}|^2 = \min_{\alpha} E\,|X_{t_0} - \sum_{j=1}^{n} \alpha_j Y_{t_j}|^2$$

  - Least square error estimator

$$\hat{X}_{t_0} = \hat{\alpha}^T Y_{\mathbf{t}} = r^T (K + \sigma^2 I_n)^{-1} Y_{\mathbf{t}}$$

$$\because) \quad \min_{\alpha} \; \alpha^T (K + \sigma^2 I_n)\alpha - 2r^T \alpha$$

$$\Rightarrow \quad \hat{\alpha} = (K + \sigma^2 I_n)^{-1} r$$

Bayesian estimation of Gaussian process

– Joint probability

$$\begin{pmatrix} Y_{\mathbf{t}} \\ X_{t_0} \end{pmatrix} \sim N\left(0, \begin{pmatrix} K + \sigma^2 I_n & r \\ r^T & R(t_0, t_0) \end{pmatrix}\right)$$

where $K = (R(t_i, t_j)) \in \mathbf{R}^{n \times n}$ $\quad r = (R(t_i, t_0)) \in \mathbf{R}^{n}$

$\because)$ $E[Y_t, Y_s] = R(t,s) + \sigma^2 \delta(t-s),$ $\quad E[Y_t, X_s] = R(t,s)$

– Bayesian estimation = LSE estimation

$$E[X_{t_0} \mid Y_{\mathbf{t}}] = r^T (K + \sigma^2 I_n)^{-1} Y_{\mathbf{t}}$$

Gaussian process and regularization

- LSE estimation of a process ＝ Regularization with RKHS
  - Linear LSE estimator of a process (Bayesian estimator of Gaussian process)

$$\min E|X_{t_0} - \hat{X}_{t_0}|^2 = \min_{\alpha} E|X_{t_0} - \sum_{j=1}^{n} \alpha_j Y_{t_j}|^2$$

Sol. $$\hat{X}_{t_0} = r^T (K + \sigma^2 I_n)^{-1} Y_{\mathbf{t}}$$

  - Ridge regression on RKHS

$$\min_{f \in H} \sum_{i=1}^{N} (Y_i - f(t_i))^2 + \lambda \| f \|_H^2$$

Sol. $$f(t) = r(t)^T (K + \lambda I_N)^{-1} Y$$

identical

$\sigma^2 \Leftrightarrow \lambda$

Correspondence between RKHS and random process

RKHS	random process
Pos. def. kernel $K(\mathrm{t,s})$	Covariance fun. $K(\mathrm{t,s}) = \mathrm{E}[\mathrm{X_t},\mathrm{X_s}]$
$\sum c_i K(\cdot, t_i)$	$\sum c_i X_{t_i}$
$\lim \sum c_i K(\cdot, t_i)$ (completion)	$\lim \sum c_i X_{t_i}$ (closure)
Regularization (smoothing)	Linear estimation
$\min_{f \in H} \sum_{i=1}^{N} (Y_i - f(t_i))^2 + \lambda \Vert f \Vert_H^2$	$\min_{\alpha} E \lvert X_{t_0} - \sum_{j=1}^{n} \alpha_j Y_{t_j} \rvert^2$
$f(t) = r(t)^T (K + \lambda I_N)^{-1} Y$	$\hat{X}_{t_0} = r^T (K + \sigma^2 I_n)^{-1} Y_{\mathbf{t}}$
Shift-invariant kernel	Cov. fun. of a statinary process
$K(\mathrm{t,s}) = K(\mathrm{t-s})$	$K(\mathrm{t,s}) = K(\mathrm{t-s})$
Bochner's theorem	Wiener-Khinchine's theorem

Iterative random functions

m-IRF

A random process $\{X_t\}_{t\in\Omega}$ is said to be an m-iterative random functions (m-IRF) if for any finite subset $\mathbf{t} = (t_1, \ldots, t_n)$ of Ω and any generalized increment $c_1, \ldots, c_n$ of order m, the process

$$\left\{\sum_{i=1}^{n} c_i X_{t+t_i}\right\}_{t\in\Omega}$$

is second-order stationary.

– A stationary process is called (-1)-IRF in convention.

Modeling by non-stationary process

– Kriging is a modeling by 0-IRF.
The generalized covariance function $G(t\text{-}s)$ is used instead of covariance function $K(t, s)$ for the modeling.

Generalized covariance

<u>Theorem</u> (Matheron 1973)

$\{X_t\}_{t\in\Omega}$: continuous m-IRF.

There is a continuous function G_K such that for any finite subset $\mathbf{t} = (t_1, \ldots, t_n)$ of Ω and any two generalized increments $(c_1,\ldots, c_n)$ and $(d_1,\ldots, d_n)$ of order m,

$$\mathrm{Cov}\left[\sum_{i=1}^n c_i X_{t_i}, \sum_{i=1}^n d_i X_{t_i}\right] = \sum_{i,j=1}^n c_i d_i G_K(t_i - t_j).$$

– The function G_K is called generalized covariance.

– The generalized covariance is conditionally positive definite of order m (obvious by definition and above theorem).

– Matheron (1973) proves the converse, also. There is a correspondence between m-IRF and conditionally positive definite functions of order m.
(Generalization of the correspondence between the stationary processes and positive definite functions.)

References

Ramsay, J. and B. W. Silverman. *Functional Data Analysis* (2nd ed.). Springer, 2005.

Green, P.J. and Silverman, B.W. *Nonparametric Regression and Generalized Linear Models*. A Roughness Penalty Approach. Chapman & Hall. 1994.

Wahba, G. *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics 59. SIAM. 1990.

Meinguet, J. (1979) Multivariate Interpolation at Arbitrary Points Made Simple. *J. Applied Mathematics and Physics* (*ZAMP*) 30, 292−304.

Matheron, G. (1973) The intrinsic random functions and their applications. *Adv. Appl. Prob*., 5, pp.439-468.

Berlinet, A. and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics.* Kluwer Academic Publishers, 2003.

Gel'fand, I.M. and Vilenkin, N.Ya. *Generalized Functions Vol.4: Applications of Harmonic Analysis*. Academic Press. 1964.