Relations to Other Statistical Methods

Statistical Data Analysis with Positive Definite Kernels

Kenji Fukumizu

Institute of Statistical Mathematics, ROIS

Department of Statistical Science, Graduate University for Advanced Studies

October 6-10, 2008, Kyushu University

Outline

1. Spline smoothing and RKHS

2. Relation to random process

Outline

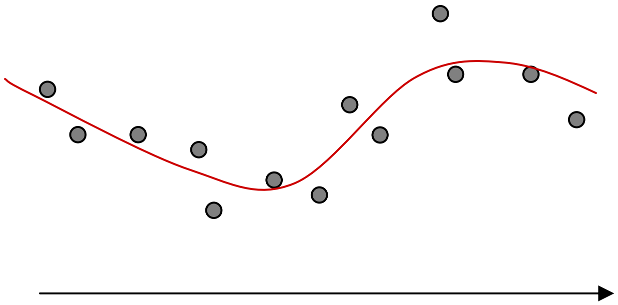
1. Spline smoothing and RKHS

2. 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 $|f(x)|, |g(x)| \to 0 \ (x \to \infty)$

$$\int \Delta f(x) g(x) dx = \int f(x) \Delta g(x) dx$$ [partial integral]

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.

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

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

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

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

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

Two cases

- Case $a_0 \neq 0$

  e.g. $$\int |f(x)|^2\,dx + \int \left(\frac{\partial f(x)}{\partial x}\right)^2 dx = (f, f)_{L^2} + (f, -\Delta f)_{L^2}$$

  – The Green function is a positive definite kernel.

  – The penalty term is equal to the squared RKHS norm.

- Case $a_0 = 0$

  e.g. $$\int \left(\frac{\partial^2 f(x)}{\partial x^2}\right)^2 dx = (f, \Delta^2 f)_{L^2}$$

  – Spline smoothing

  – 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

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

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

The solution:

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

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 ($G(x, y) = G(x\text{-}y)$). 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.

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

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

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

$a_0 = 0$: 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

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

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

$\Leftrightarrow$

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

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:

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

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

Outline

1. Spline smoothing

2. 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 $(X_{t_1},\dots,X_{t_n})$ is a Gaussian random vector.

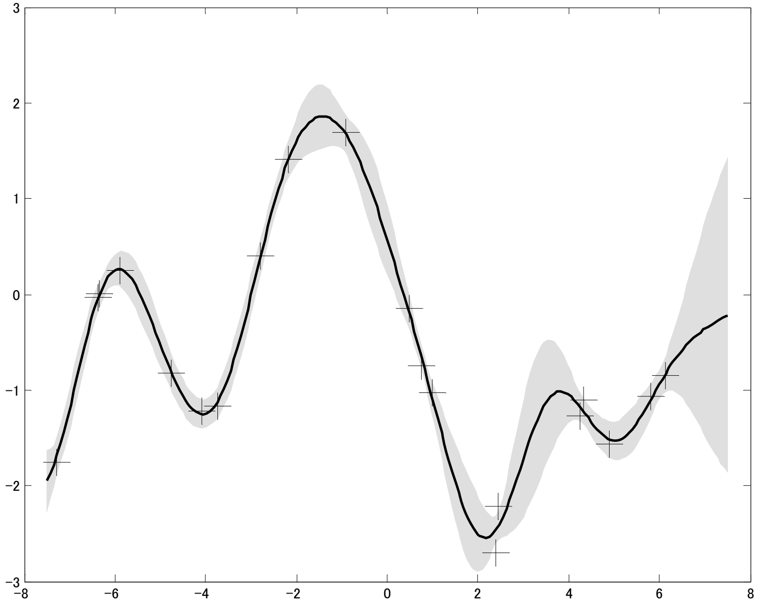
- Mean function $\mu(t) = E[X_t]$
- Covariance function $R(t,s) = \mathrm{Cov}[X_t, X_s]$

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

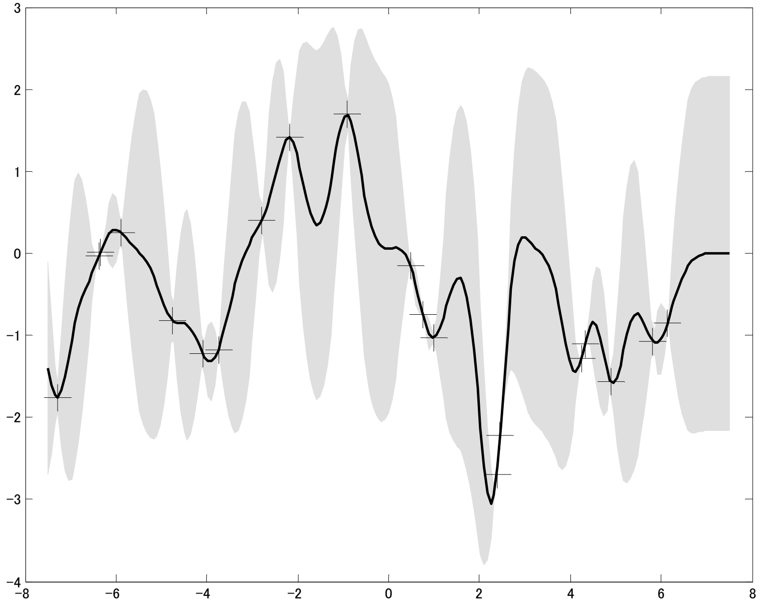
$$\mathbf{X} = (X_{t_1},\dots,X_{t_n}) \sim N(\mu_{\mathrm{X}}, \Sigma_{\mathrm{X}})$$

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

– 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.

∵) 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 = $k(s,t)$.

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.)

References

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.

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.