Imperial College London

DataSig Seminar
14/10/2021

Kernel-based Statistical Methods for Functional Data

George Wynne (g.wynne18@imperial.ac.uk), Andrew B. Duncan,
Mikołaj Kasprzak, Stanislav Nagy

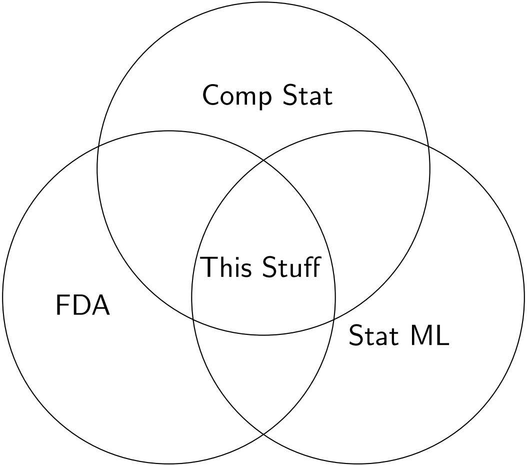
- A Kernel Two-Sample Test for Functional Data by **GW**, Andrew B. Duncan (`https://arxiv.org/pdf/2008.11095`)
- Statistical Depth Meets Machine Learning: Kernel Mean Embeddings and Depth in Functional Data Analysis by **GW**, Stanislav Nagy (`https://arxiv.org/pdf/2105.12778`)
- A Spectral View of Kernel Stein Discrepancy with Application to Goodness-of-Fit Tests for Measures on Hilbert Spaces by **GW**, Mikołaj J. Kasprzak, Andrew B. Duncan **Coming soon!**

Summary

- Functional data analysis
- Statistical kernel-based methods
  - Maximum mean discrepancy
  - Kernel Stein discrepancy
- Future thoughts

Talk in one slide

- Kernel-based methods can be adapted to Hilbert spaces
- So can apply to functional data
- Strong numerical performance
- Opens many questions


Comp Stat
This Stuff
FDA
Stat ML

Functional Data Analysis (FDA)

- Specific statistical challenges, distinct from finite dimensions
- Projection methods often employed, Hilbert view
- Gaussian processes, Gibbs measures, SDEs

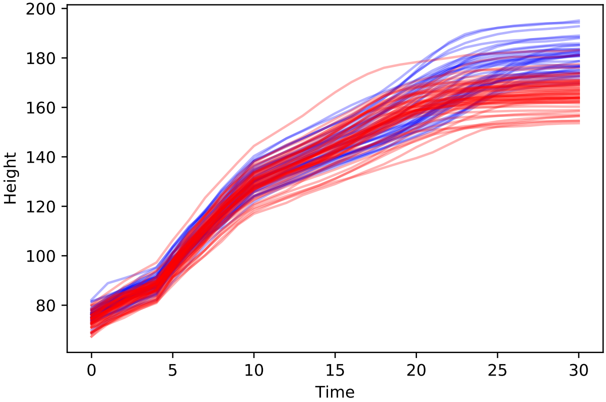
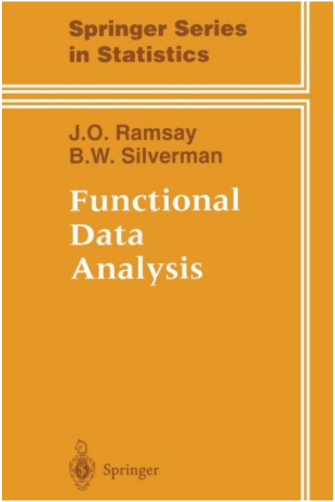
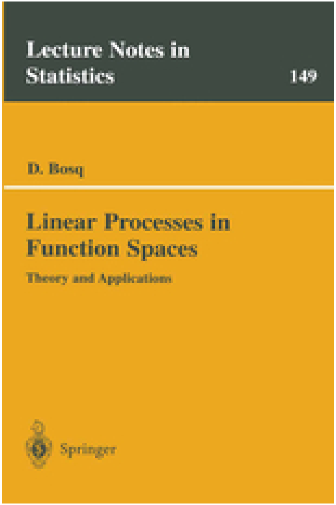


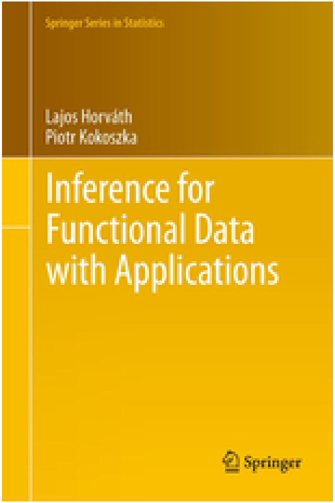
Figure 1: Height measures at different times for male and female children

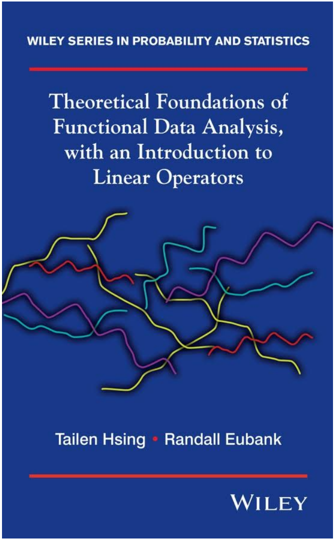
Textbook History of FDA





Textbook History of FDA



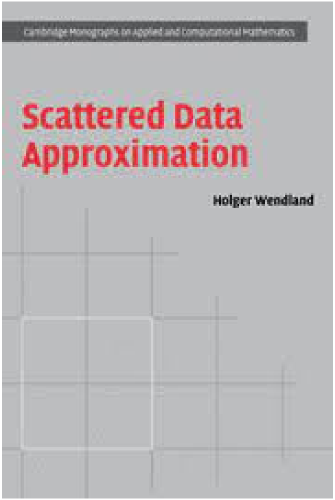


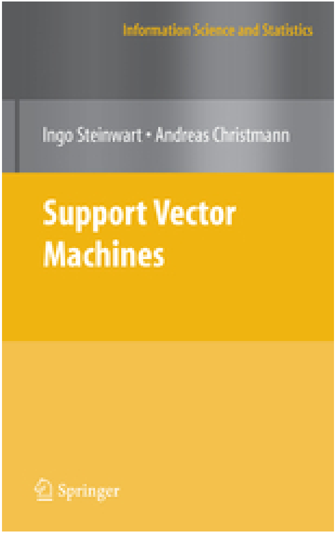
Statistical Kernel-Based Methods

(Brief) History of statistical kernel-based methods

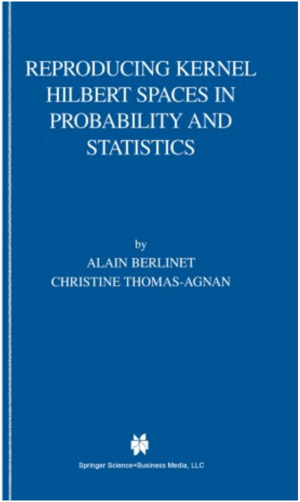
- Studied under a different name in 1970s by Guilbart and colleagues at Lille [Guilbart, 1978, Berlinet and Thomas-Agnan, 2004]
- Rose to prominence in statistical machine learning in mid 2000s [Gretton et al., 2012]
- Theory matured in 2010s with new applications beyond testing and different data types [Muandet et al., 2017]
- Kernel Stein discrepancy [Oates et al., 2016, Liu et al., 2016, Chwialkowski et al., 2016]
- Application to functional data [Chevyrev and Oberhauser, 2018, Wynne and Duncan, 2020, Hayati et al., 2020, Górecki et al., 2018, Jia et al., 2021]

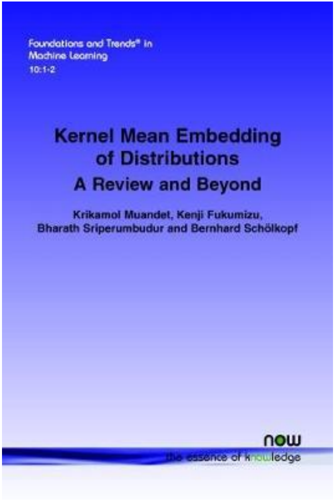
Textbook Using Kernels





Textbook Using KME





Notation

- Let $\mathcal{X}$ be a separable Hilbert space e.g. $L^2([0,1]^d), W_2^{\alpha}([0,1]^d)$
- $\mathcal{P}(\mathcal{X})$ Borel probability measures on $\mathcal{X}$
- $\widehat{P}(s) = \int_{\mathcal{X}} e^{i\langle s,x\rangle_{\mathcal{X}}} dP(x)$
- N_C Gaussian measure, mean zero, covariance operator C
- For some $P, Q \in \mathcal{P}(\mathcal{X})$ observe i.i.d. $\{X_i\}_{i=1}^N \sim P, \{Y_j\}_{j=1}^M \sim Q$

A kernel $k\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a symmetric, positive definite function

Example

$$\begin{aligned} \text{SE-}T \quad & k(x,y) = e^{-\|Tx - Ty\|^2/2} \\ \text{IMQ-}T \quad & k(x,y) = (\|Tx - Ty\|^2 + 1)^{-1/2} \end{aligned}$$

Theorem

For $\mu \in \mathcal{P}(\mathcal{X})$, $k(x,y) = \widehat{\mu}(x-y)$ is a kernel where $\widehat{\mu}(s) = \int_{\mathcal{X}} e^{i\langle s,z \rangle \mathcal{X}} d\mu(z)$ is the characteristic function (Fourier transform) of P.

RKHS

- A Hilbert space of functions H is called a reproducing kernel Hilbert space if there exists a kernel k such that
  1. $k(\cdot, x) \in H \, \forall x \in \mathcal{X}$
  2. $f(x) = \langle f, k(\cdot, x) \rangle_H \, \forall f \in H \, \forall x \in \mathcal{X}$
- Denote the RKHS of k by H_k
- Reproducing property gives closed forms

Maximum Mean Discrepancy

$$\mathrm{MMD}_k(Q,P) = \sup_{\|f\|_k \le 1} |\mathbb{E}_Q[f(X)] - \mathbb{E}_P[f(X)]|$$

$$\mathrm{MMD}_k(Q,P)^2 = \int_{\mathcal{X}} \int_{\mathcal{X}} k(x,y) d(P-Q)(x) d(P-Q)(y)$$

$$\mathrm{MMD}_k(Q,P)^2 = \int_{\mathcal{X}} \left| \widehat{P}(s) - \widehat{Q}(s) \right|_{\mathbb{C}}^2 d\mu(s)$$

when $k(x,y) = \widehat{\mu}(x-y)$, which is common.

Easiest proof of characteristicness

Theorem ([Sriperumbudur et al., 2010])

If $k(x, y) = \widehat{\mu}(x - y)$ and μ has full support then $MMD_k(Q, P) = 0$ if and only if $P = Q$.

Proof.

$$\begin{aligned} \mathrm{MMD}_k(Q, P) &= \int_{\mathcal{X}} \left| \widehat{P}(s) - \widehat{Q}(s) \right|_{\mathbb{C}}^2 d\mu(s) = 0 \\ &\iff \widehat{P} = \widehat{Q} \\ &\iff P = Q \end{aligned}$$

□

- In finite dimensions Bochner $k(x, y) = \widehat{\mu}(x - y) \iff k$ is continuous
- Minlos-Sazonov shows MUCH stronger continuity is needed in infinite dimensions

Example

$k(x, y) = e^{-\|Tx - Ty\|_{\mathcal{X}}^2/2} = \widehat{\mu}(x - y)$ for some μ if and only if $T = C^{1/2}$ for a trace class C, so $T = I_{\mathcal{X}}$ doesn't work. Such a T smooths the signal a lot.

Extending class of kernels

Theorem ([Wynne and Duncan, 2020])

Let $\mathcal{X}, \mathcal{Y}$ be separable Hilbert spaces, $T\colon \mathcal{X} \to \mathcal{Y}$ be injective then the SE-T and IMQ-T kernels

$$k_{SE}(x, y) = e^{-\frac{1}{2}\|Tx - Ty\|_{\mathcal{Y}}^2}$$
$$k_{IMQ}(x, y) = (\|Tx - Ty\|_{\mathcal{Y}}^2 + 1)^{-1/2}$$

are characteristic.

Compliments

- Topological properties of MMD can be investigated
- Estimation of MMD using reconstructions based on discretised data can be addressed
- The RKHS perspective provides a unification between other existing approaches in FDA e.g. ECF $=$ h-depth

Kernel Stein Discrepancy

- Compute a kernel-based distance using only one set of samples
- Derived by [Oates et al., 2016, Liu et al., 2016, Chwialkowski et al., 2016]
- All theory done intrinsically on $\mathbb{R}^d$
- **Very wide** applications in computational statistics and statistical machine learning

Call Γ and $\mathcal{F}$ a Stein operator and Stein class for $P \in \mathcal{P}(\mathcal{X})$ if

$$\mathbb{E}_Q[\Gamma f(X)] = 0 \, \forall f \in \mathcal{F} \iff P = Q$$

Example

- If $\mathcal{X} = \mathbb{R}, P = N(0,1)$ then $\mathcal{A}f(x) = f'(x) - xf(x)$ and $\mathcal{F} = C_b^1(\mathcal{X})$.
- Generators of Markov processes

If Γ acts on $f : \mathcal{X} \to \mathbb{R}$

$$\text{KSD}_{\Gamma,k}(Q,P) \coloneqq \sup_{\|f\|_k \leq 1} |\mathbb{E}_Q[\Gamma f(X)]||$$

If Γ acts on $f : \mathcal{X} \to \mathcal{X}$

$$\text{KSD}_{\Gamma,K}(Q,P) \coloneqq \sup_{\|f\|_K \leq 1} |\mathbb{E}_Q[\Gamma f(X)]||$$

where $K(x,y) = k(x,y)I_{\mathcal{X}}$ is an operator valued kernel. So RKHS if functions $f(x) = \sum_{n=1}^{\infty} e_n f_n(x)$, $f_n \in H_k$

In practice people use vectorised versions of generators

Example

$\mathcal{X} = \mathbb{R}^d$, P has density p

$$Bf(x) = \mathrm{Tr}(D^2 f(x)) + \langle \nabla \log p(x), Df(x) \rangle_{\mathbb{R}^d}$$
$$\mathcal{B}f(x) = \mathrm{Tr}(Df(x)) + \langle \nabla \log p(x), f(x) \rangle_{\mathbb{R}^d}$$

B is generator of Langevin diffusion
$\mathcal{B}$ is the most used Stein operator in $\mathbb{R}^d$

For target measures $P = e^{-U(x)} N_C$ we can use generators of infinite dimensional SDEs

$$Af(x) = \mathrm{Tr}(CD^2 f(x)) - \langle x + CDU(x), Df(x) \rangle_{\mathcal{X}}$$
$$\mathcal{A}f(x) = \mathrm{Tr}(CDf(x)) - \langle x + CDU(x), f(x) \rangle_{\mathcal{X}}$$

$$\mathcal{A}f(x) = \mathrm{Tr}(C Df(x)) - \langle x + C DU(x), f(x)\rangle_{\mathcal{X}}$$
$$\mathcal{B}f(x) = \mathrm{Tr}(Df(x)) + \langle \nabla \log p(x), f(x)\rangle_{\mathbb{R}^d}$$

For $\mathcal{X} = \mathbb{R}^d$ let P have density p and fix $\Sigma \in \mathbb{R}^{d\times d}$ then $P = e^{-U(x)} N_\Sigma$ where

$$U(x) = \langle \Sigma^{-1}x, x\rangle_{\mathbb{R}^d}/2 + \log p(x)$$
$$DU(x) = \Sigma^{-1}x + \nabla \log p(x)$$

Subbing into $\mathcal{A}$ gives

$$\mathcal{A}f(x) = \mathrm{Tr}(\Sigma Df(x)) - \langle \nabla \log p(x), \Sigma f(x)\rangle_{\mathbb{R}^d} = \mathcal{B}(\Sigma f)(x)$$

Assumption

$\mathcal{X}$ is an infinite dimensional, real, separable Hilbert space and the target probability measure P is defined as $\frac{dP}{dN_C}(x) = Z^{-1}e^{-U(x)}$, for a normalising constant Z, with $e^{-U(x)/2} \in W_C^{1,2}(\mathcal{X})$ and $C \in L_1^+(\mathcal{X})$ is injective and such that $\mathbb{E}_{N_C}[\|C^{1/2}DU(X)\|_{\mathcal{X}}^2] < \infty$.

Assumption

k is a real valued, bounded kernel on $\mathcal{X}$ such that D_1k, D_2k, D_2D_1k exist, are continuous and

$$\sup_{x,y\in\mathcal{X}} \|D_1k(x,y)\|_{\mathcal{X}}, \sup_{x,y\in\mathcal{X}} \|D_2k(x,y)\|_{\mathcal{X}} < \infty$$

$$\sup_{x,y\in\mathcal{X}} \|D_2D_1k(x,y)\|_{L(\mathcal{X}\times\mathcal{X},\mathbb{R})} < \infty.$$

Theorem

Let $\mathcal{X}, P, k$ satisfy Assumptions 1 and 2. Denote the eigensystem of C by $\{\lambda_i, e_i\}_{i=1}^{\infty}$ and define the Stein kernel h corresponding to k and $\mathcal{A}$ as

$$\begin{aligned} h(x,y) = &\, k(x,y)\langle x + CDU(x), y + CDU(y)\rangle_{\mathcal{X}} \\ &- D_1 k(x,y)(Cy + C^2 DU(y)) \\ &- D_2 k(x,y)(Cx + C^2 DU(x)) + \sum_{i=1}^{\infty} \lambda_i^2 D_2 D_1 k(x,y)(e_i, e_i), \end{aligned}$$

then for $Q \in \mathcal{P}(\mathcal{X})$ such that $\mathbb{E}_Q[\|X\|_{\mathcal{X}}], \mathbb{E}_Q[\|CDU(X)\|_{\mathcal{X}}] < \infty$,

$$KSD_{\mathcal{A},K}(Q,P) := \sup_{\|f\|_K \le 1} \mathbb{E}_Q[\mathcal{A}f(X)] = \mathbb{E}_Q[h(X,X')]^{1/2},$$

where $X, X' \sim Q$ are independent.

- No known conditions for KSD to separate in infinite dimensions
- No spectral view of KSD in finite or infinite dimensions
- Hard to see impact of hyper-parameters

For $k(x, y) = \widehat{\mu}(x - y)$

$$\begin{aligned}\mathrm{MMD}_k(Q, P) &= \sup_{\|f\|_k \leq 1} |\mathbb{E}_Q[f(X)] - \mathbb{E}_P[f(X)]| \\ &= \sup_{\|f\|_k \leq 1} |\mathbb{E}_Q[\Theta(f)(X)]|\end{aligned}$$

$$\begin{aligned}\mathrm{MMD}_k(Q, P)^2 &= \int_{\mathcal{X}} \left| \widehat{P}(s) - \widehat{Q}(s) \right|_{\mathbb{C}}^2 d\mu(s) \\ &= \int_{\mathcal{X}} \left| \mathbb{E}_Q \left[\Theta \left(e^{i\langle \cdot, s \rangle_{\mathcal{X}}} \right) (X) \right] \right|_{\mathbb{C}}^2 d\mu(s)\end{aligned}$$

where

$$\Theta(f)(x) := f(x) - \mathbb{E}_P[f(X)]$$

$$\mathrm{MMD}_k(Q,P) = \sup_{\|f\|_k \le 1} |\mathbb{E}_Q[\Theta(f)(X)]|$$

$$\overset{2}{=} \int_{\mathcal{X}} \left| \mathbb{E}_Q \left[\Theta \left(e^{i\langle \cdot, s\rangle_{\mathcal{X}}} \right) (X) \right] \right|_{\mathbb{C}}^2 d\mu(s)$$

$$\mathrm{KSD}_{\mathcal{A},K}(Q,P) = \sup_{\|f\|_K \le 1} \sum_{n=1}^{\infty} \mathbb{E}_Q[\mathcal{A}(e_n f_n)(X)]$$

$$\overset{2}{=} \sum_{n=1}^{\infty} \int_{\mathcal{X}} \left| \mathbb{E}_Q[\mathcal{A}(e_n e^{i\langle \cdot, s\rangle_{\mathcal{X}}})(X)] \right|_{\mathbb{C}}^2 d\mu(s)$$

$$\mathrm{KSD}_{A,k}(Q,P) = \sup_{\|f\|_k \le 1} \mathbb{E}_Q[Af(X)]$$

$$\overset{2}{=} \int_{\mathcal{X}} \left| \mathbb{E}_Q[A(e^{i\langle \cdot, s\rangle_{\mathcal{X}}})(X)] \right|_{\mathbb{C}}^2 d\mu(s)$$

Theorem

Under Assumption 1,2, $k(x,y) = \widehat{\mu}(x-y)$ *and* Q *such that* $\mathbb{E}_Q[\|X\|_{\mathcal{X}}], \mathbb{E}_Q[\|CDU(X)\|_{\mathcal{X}}] < \infty$

$$KSD_{\mathcal{A},K}(Q,P) = \sup_{\|f\|_K \le 1} \mathbb{E}_Q[\mathcal{A}f(X)] = \sup_{\|f\|_K \le 1} \sum_{n=1}^{\infty} \mathbb{E}_Q[\mathcal{A}(e_n f_n)(X)]$$

where $f(x) = \sum_{n=1}^{\infty} e_n f_n(x)$.

$$KSD_{\mathcal{A},K}(Q,P)^2 = \sum_{n=1}^{\infty} \int_{\mathcal{X}} \left| \mathbb{E}_Q[\mathcal{A}(e_n e^{i\langle \cdot, s\rangle_{\mathcal{X}}})(X)] \right|_{\mathbb{C}}^2 d\mu(s)$$

$$= \int_{\mathcal{X}} \left\| Cs\widehat{Q}(s) + D\widehat{Q}(s) + i \int_{\mathcal{X}} CDU(x) e^{i\langle s, x\rangle_{\mathcal{X}}} dQ(x) \right\|_{\mathcal{X}_{\mathbb{C}}}^2 d\mu(s)$$

Theorem

Under Assumption 1,2, $k(x,y) = \widehat{\mu}(x-y)$ *and* Q *such that* $\mathbb{E}_Q[\|X\|_{\mathcal{X}}^2], \mathbb{E}_Q[\|C^{1/2}DU(X)\|_{\mathcal{X}}^2] < \infty$ *if* μ *has full support then* $KSD_{\mathcal{A},K}(Q,P) = 0 \iff Q = P$.

Sketch Proof.

The integrand in the spectral representation is zero for all s if and only if $Q = P$ since it characterises the solution of a measure equation whose unique solution is P [Bogachev and Röckner, 1995, Albeverio et al., 1999]. □

Using the same limit argument as before

Theorem

Under Assumption 1,2, $T \in L(\mathcal{X})$ *is injective and* Q *such that* $\mathbb{E}_Q[\|X\|_{\mathcal{X}}^2], \mathbb{E}_Q[\|C^{1/2}DU(X)\|_{\mathcal{X}}^2] < \infty$ *then the SE-*T *and IMQ-*T *kernels ensure* $KSD_{\mathcal{A},K}(Q,P)^2 = 0 \iff Q = P$.

$$\mathrm{MMD}_k(Q,P) = \sup_{\|f\|_k \leq 1} |\mathbb{E}_Q[\Theta(f)(X)]|$$

$$\overset{2}{=} \int_{\mathcal{X}} \left| \mathbb{E}_Q \left[\Theta \left(e^{i\langle \cdot, s \rangle_{\mathcal{X}}} \right)(X) \right] \right|_{\mathbb{C}}^2 d\mu(s)$$

$$\mathrm{KSD}_{\mathcal{A},K}(Q,P) = \sup_{\|f\|_K \leq 1} \sum_{n=1}^{\infty} \mathbb{E}_Q[\mathcal{A}(e_n f_n)(X)]$$

$$\overset{2}{=} \sum_{n=1}^{\infty} \int_{\mathcal{X}} \left| \mathbb{E}_Q[\mathcal{A}(e_n e^{i\langle \cdot, s \rangle_{\mathcal{X}}})(X)] \right|_{\mathbb{C}}^2 d\mu(s)$$

$$\mathrm{KSD}_{A,k}(Q,P) = \sup_{\|f\|_k \leq 1} \mathbb{E}_Q[Af(X)]$$

$$\overset{2}{=} \int_{\mathcal{X}} \left| \mathbb{E}_Q[A(e^{i\langle \cdot, s \rangle_{\mathcal{X}}})(X)] \right|_{\mathbb{C}}^2 d\mu(s)$$

Let X_t be a Markov process with invariant measure P and generator A then

$$\mathbb{E}_Q[\Theta f(X)] = -\int_0^\infty \mathbb{E}[A f(X_t^Q)]dt$$

so

$$\begin{aligned} \mathrm{MMD}_k(Q,P)^2 &= \int_{\mathcal{X}} \left| \mathbb{E}_Q \left[\Theta \left(e^{i\langle \cdot, s \rangle_{\mathcal{X}}} \right) (X) \right] \right|_{\mathbb{C}}^2 d\mu(s) \\ &= \int_{\mathcal{X}} \left| \int_0^\infty \mathbb{E}[A(e^{i\langle \cdot, s \rangle_{\mathcal{X}}})(X_t^Q)]dt \right|_{\mathbb{C}}^2 d\mu(s) \\ \mathrm{KSD}_{A,k}(Q,P)^2 &= \int_{\mathcal{X}} \left| \mathbb{E}_Q[A(e^{i\langle \cdot, s \rangle_{\mathcal{X}}})(X)] \right|_{\mathbb{C}}^2 d\mu(s) \end{aligned}$$

where X_t^Q is the process with $X_0 \sim Q$.

Numerics

KSD Estimator

$\{X_n\}_{n=1}^N \sim Q$ samples and the target measure is $P = e^{-U(x)} N_C$

$$\widehat{\mathrm{KSD}}(Q, P) = \frac{1}{N-1} \sum_{1 \le i \neq j \le N} h(X_i, X_j)$$

Use bootstrap to calculate rejection threshold

Goodness-of-Fit

- P will be Brownian motion over $[0, 1]$
- Use SE-T and IMQ-T kernel
- $T_1 = I$, $T_2 x = \sum_{i=1}^{\infty} \eta_i \langle x, e_i \rangle_{\mathcal{X}} e_i$ where $\eta_i = \lambda_i^{-1}$ for $1 \leq i \leq 50$ and $\eta_i = 1$ for $i > 50$ with e_i, λ_i the eigensystem of Brownian motion
- T_2 increasingly penalises deviations in higher frequencies

1. $N = 50$ and Q is the law of Brownian motion
2. $N = 50$ and Q is the law of the Brownian motion clipped to 5 frequencies $\sum_{i=1}^{5} \lambda_i^{1/2} \xi_i e_i$ with $\xi_i \overset{i.i.d.}{\sim} N(0,1)$ and λ_i, e_i from the eigensystem of C as discussed above.
3. $N = 25$ and Q is the law of the Ornstein-Uhlenbeck process

$$dX(t) = 0.5(5 - X(t))dt + dB(t)$$

4. $N = 25$ and Q is the law of $2B(t)$
5. $N = 25$ and Q is the law of $B(t) + 1.5t(t-1)$

Experiment	SE-T_1	SE-T_2	IMQ-T_1	IMQ-T_2	SB
1	0.06	**0.05**	0.052	0.048	0.032
2	0.056	**1.0**	0.054	0.952	0.615
3	**1.0**	**1.0**	**1.0**	**1.0**	0.023

Table 1: Performance on Experiments 1-3, SB denotes the small-ball probability method of [Bongiorno et al., 2018].

Experiment	SE-T_1	SE-T_2	IMQ-T_1	IMQ-T_2	CvM SP	CvM GP
6	0.858	0.786	0.332	0.206	**0.895**	0.763
7	0.522	**0.99**	0.608	0.87	0.98	0.858

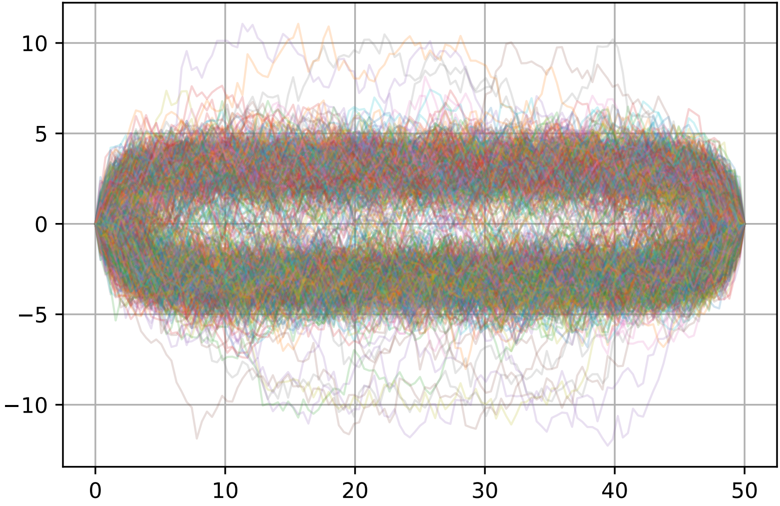
Table 2: Performance on Experiments 4-5, CvM SP denotes the Cramér von-Mises test based on spherical projections of [Ditzhaus and Gaigall, 2018] and CvM GP denotes the Cramér von-Mises test based on Gaussian process projections of [Bugni et al., 2009].

Conditioned Diffusion

- The paths are the following SDE conditioned to start and end at 0 over $[0, 30]$

$$dX_t = 0.7 \sin(X_t) dt + dW_t$$

- This is a Gibbs measure with N_C being Brownian bridge
- **Very hard to sample from**



10
5
0
−5
−10
0
10
20
30
40
50

Numerics

$N = 50$, 50 tests, using T_2 as before, samples come from

$$X(t) + \delta t/30$$

δ	SE-T_2	IMQ-T_2
0	0.06	0.04
0.5	0.12	0.10
1.0	0.42	0.38
1.5	0.68	0.7
2.0	0.98	0.98

Two-Sample Testing

- Estimate MMD using $\{X_i\}_{i=1}^N \overset{i.i.d.}{\sim} P, \{Y_i\}_{i=1}^N \overset{i.i.d.}{\sim} Q$

$$\widehat{MMD}(P,Q)^2 = \frac{1}{N(N-1)} \sum_{1 \le i \ne j \le N} h(Z_i, Z_j)$$

where $h(Z_i, Z_j) = k(X_i, X_j) + k(Y_i, Y_j) - k(X_i, Y_j) - k(X_j, Y_i)$

Experimental Setting

- $\mathcal{X} = L^2([0,1])$
- Use SE-T kernel for two choices of T
  - $T = I_{\mathcal{X}}$ (ID)
  - $Tx(t) = \int_0^1 x(s)k_0(s,t)ds$ for a cosine-exponential kernel k_0 (CEXP)
- $k(x,y) = \langle x,y \rangle_{\mathcal{X}}^2$ (COV)

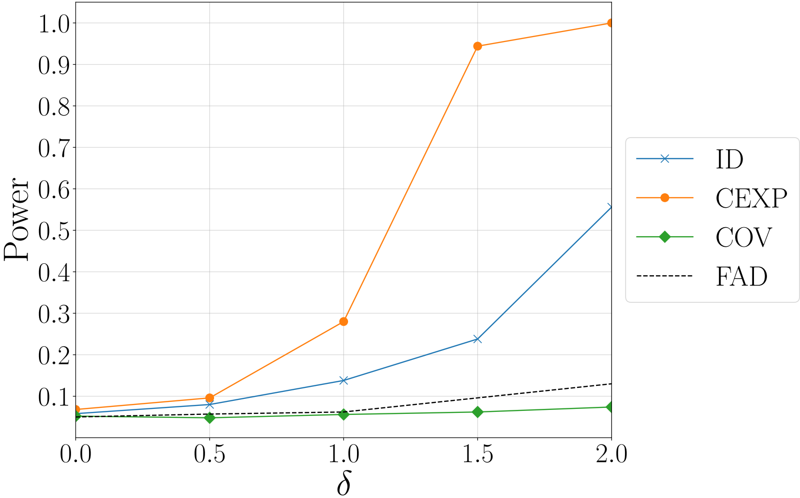
Mean Shift Experiment

Sample size $N = 100$, observed at 100 points on a grid over $[0, 1]$

$$X(t) \sim t + \xi_{10}^X \sqrt{2} \sin(2\pi t) + \xi_5^X \sqrt{2} \cos(2\pi t)$$
$$Y(t) \sim X(t) + \delta t^3$$

where $\xi_{10}^X \sim N(0, 10)$ and $\xi_5^X \sim N(0, 5)$.

Compare to Functional Anderson-Darling (FAD) test of [Pomann et al., 2016]


Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.5
1.0
1.5
2.0
δ
ID
CEXP
COV
FAD

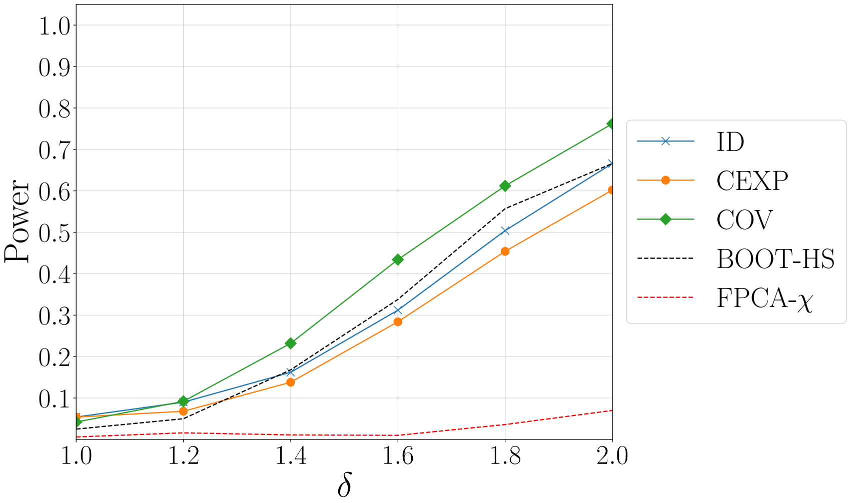
Variance Shift Experiment

Sample size if $N = 25$, observed at 500 points on a grid over $[0, 1]$

$$X(t) \sim \sum_{n=1}^{10} \xi_n^X \sqrt{2} \sin(\pi n t) + \eta_n^X \sqrt{2} \cos(\pi n t)$$
$$Y(t) \sim \delta X(t)$$

where $\xi_n^X, \eta_n^X \sim t_5$.

Compare to bootstrapped Hilbert-Schmidt norm (BOOT-HS) test of [Paparoditis and Sapatinas, 2016] and FPCA chi-squared (FPCA-χ^2) test of [Fremdt et al., 2012].



Power
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
1.0
1.2
1.4
1.6
1.8
2.0
δ
ID
CEXP
COV
BOOT-HS
FPCA-χ

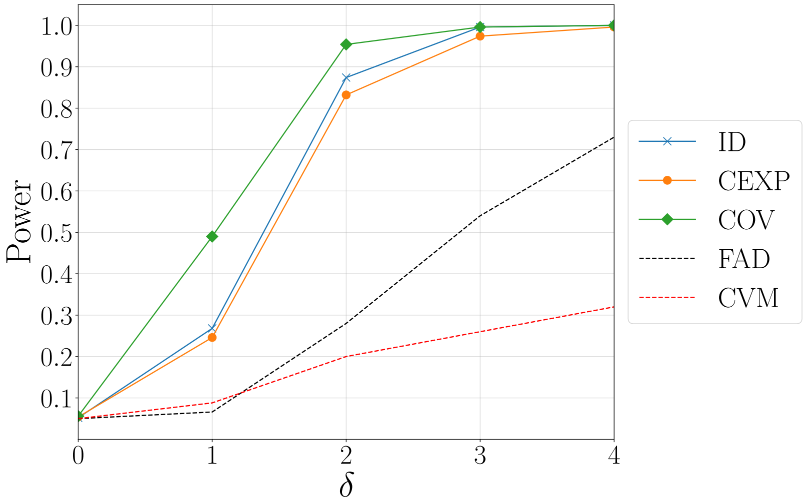
Higher Order Difference Experiment

Sample size is $N = 15$, observed at 20 random points on a grid over $[0, 1]$ with different sampling densities for X, Y. GP regression was used to reconstruct the paths.

$$X(t) \sim \sum_{n=1}^{15} e^{-n/2} \xi_n^X \psi_n(t)$$
$$Y(t) \sim X(t) + \delta n^{-2} \xi_n^Y \psi_n^*(t)$$

where $\xi_n^X, \xi_n^Y \sim N(0, 1)$, and ψ_n, ψ_n^* are trigonometric functions.

Compare to bootstrapped Cramér-von Mises (CVM) test of [Hall and Keilegom, 2007] and FAD test.



1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Power
0
1
2
3
4
δ
ID
CEXP
COV
FAD
CVM

Conclusion

- KSD and MMD can be adapted to Hilbert spaces
- KSD and MMD are linked through Markov view
- Many open questions
  - Different choices of Θ
  - Bounds using Markov theory
  - Stein-Malliavin?
  - Hyper-parameters
  - Non-Hilbert?

Thank you for listening!

S. Albeverio, V. Bogachev, and M. Röckner. On uniqueness of invariant measures for finite- and infinite-dimensional diffusions. *Communications on Pure and Applied Mathematics*, 52(3):325–362, 1999.

Alain Berlinet and Christine Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer US, 2004.

V.I. Bogachev and M. Röckner. Regularity of invariant measures on finite and infinite dimensional spaces and applications. *Journal of Functional Analysis*, 133(1):168–223, 1995.

Enea G. Bongiorno, Aldo Goia, and Philippe Vieu. Modeling functional data: a test procedure. *Computational Statistics*, 34(2):451–468, 2018.

Federico A. Bugni, Peter Hall, Joel L. Horowitz, and George R. Neumann. Goodness-of-fit tests for functional data. *Econometrics Journal*, 12:S1–S18, January 2009.

Ilya Chevyrev and Harald Oberhauser. Signature moments to characterize laws of stochastic processes. *arXiv:1810.10971*, 2018.

Kacper Chwialkowski, Heiko Strathmann, and Arthur Gretton. A kernel test of goodness of fit. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48, pages 2606–2615, 2016.

Marc Ditzhaus and Daniel Gaigall. A consistent goodness-of-fit test for huge dimensional and functional data. *Journal of Nonparametric Statistics*, 30(4):834–859, 2018.

Stefan Fremdt, Josef G. Steinebach, Lajos Horváth, and Piotr Kokoszka. Testing the equality of covariance operators in functional samples. *Scandinavian Journal of Statistics*, 40(1):138–152, 2012.

Tomasz Górecki, Mirosław Krzyśko, and Waldemar Wołyński. Independence test and canonical correlation analysis based on the alignment between kernel matrices for multivariate functional data. *Artificial Intelligence Review*, 53(1):475–499, 2018.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012.

Christian Guilbart. *Étude des produits scalaires sur l'espace des mesures: Estimation par projections tests à noyaux*. PhD thesis, Université des Sciences et Techniques de Lille, 1978.

P. Hall and I. Keilegom. Two-sample tests in functional data analysis starting from discrete data. *Statistica Sinica*, 17:1511–1531, 2007.

Saeed Hayati, Kenji Fukumizu, and Afshin Parvardeh. Kernel mean embedding of probability measures and its applications to functional data analysis. *arXiv:2011.02315*, 2020.

Junxiong Jia, Peijun Li, and Deyu Meng. Stein variational gradient descent on infinite-dimensional space and applications to statistical inverse problems. *arXiv:2102.09741*, 2021.

Qiang Liu, Jason Lee, and Michael Jordan. A kernelized Stein discrepancy for goodness-of-fit tests. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48, pages 276–284, 2016.

Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, and Bernhard Schölkopf. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends® in Machine Learning*, 10(1-2):1–141, 2017.

Chris J. Oates, Mark Girolami, and Nicolas Chopin. Control functionals for monte carlo integration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):695–718, 2016.

E. Paparoditis and T. Sapatinas. Bootstrap-based testing of equality of mean functions or equality of covariance operators for functional data. *Biometrika*, 103(3):727–733, 2016.

Gina-Maria Pomann, Ana-Maria Staicu, and Sujit Ghosh. A two-sample distribution-free test for functional data with application to a diffusion tensor imaging study of multiple sclerosis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(3):395–414, 2016.

Bharath K. Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert R.G. Lanckriet. Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11(50):1517–1561, 2010.

George Wynne and Andrew B. Duncan. A kernel two-sample test for functional data. *arXiv:2008.11095*, 2020.

$$\begin{aligned} h(x,y) = {} & k(x,y)\langle x + CDU(x), y + CDU(y)\rangle_{\mathcal{X}} \\ & - D_1 k(x,y)(Cy + C^2 DU(y)) \\ & - D_2 k(x,y)(Cx + C^2 DU(x)) \\ & + \sum_{i=1}^{\infty} \lambda_i^2 D_2 D_1 k(x,y)(e_i, e_i) \end{aligned}$$

For the SE-I kernel this gives

$$\begin{aligned} h(x,y) = k(x,y)\Big(& \langle x + CDU(x), y + CDU(y)\rangle \\ & - \langle C(x-y), x-y\rangle \\ & - \langle C(DU(x) - DU(y)), x-y\rangle \\ & + \mathrm{Tr}(C^2) - \|C(x-y)\|^2 \Big) \end{aligned}$$