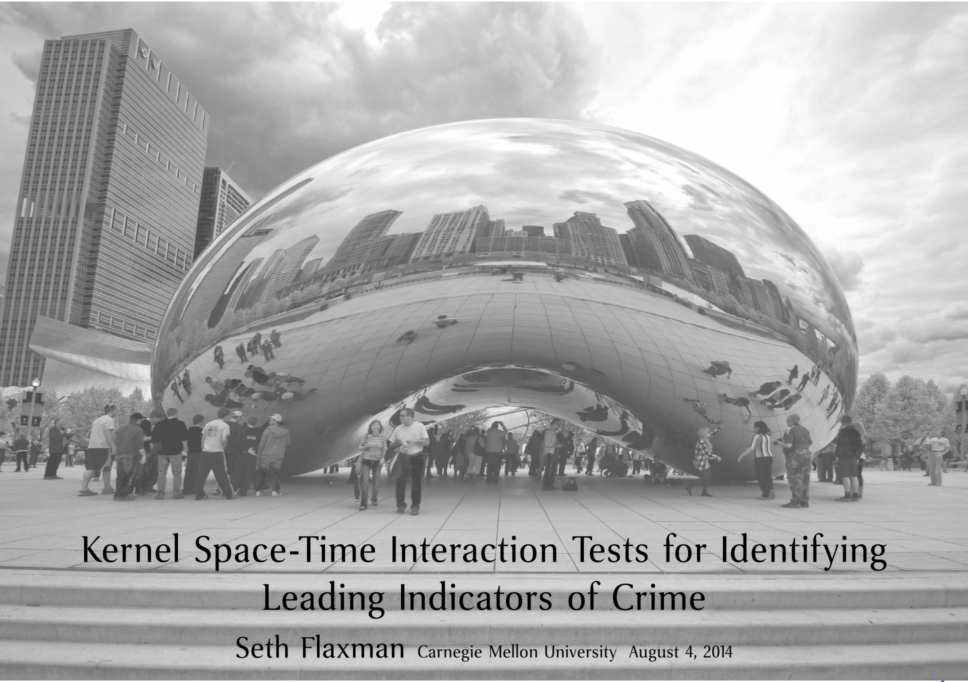


Kernel Space-Time Interaction Tests for Identifying Leading Indicators of Crime

Seth Flaxman Carnegie Mellon University August 4, 2014

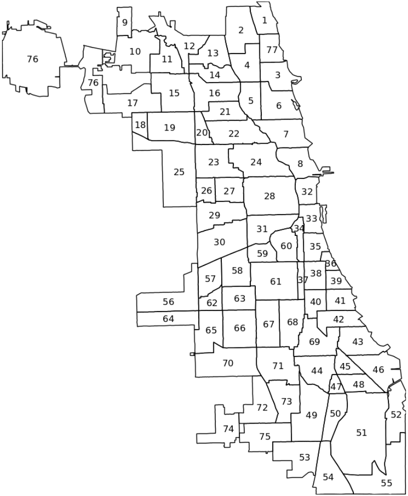
Joint work with....

Daniel Neill[1] and Alex Smola[2]
1. H. J. Heinz III College, Carnegie Mellon University
2. Machine Learning Department, School of Computer Science, Carnegie Mellon University

This work was partially supported by the National Science Foundation, grants IIS-0916345, IIS-0911032, and IIS-0953330.

- **Research questions:**
  1. What are the leading indicators of homicides?
  2. How to do feature selection with spatiotemporal data?
- **Background:** space-time interaction tests, statistical tests for independence
- **Methods:** "Kernel Space-Time" (KST) interaction test
- **Applications:** 911 call data, crime offense reports from Chicago

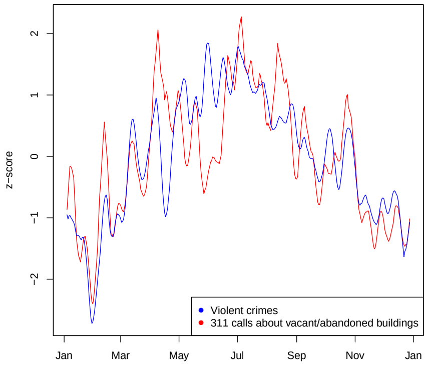
Chicago



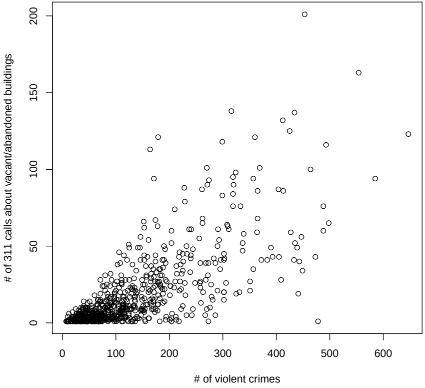
- Population: 2.7 million
- Area: 234 square miles
- CrimeScan (collaboration with CPD): use anomaly detection for prediction
- *Which types of calls to 911 are predictive of homicides and shootings nearby?*

Correlations in Space and Time

Time Series in 2011 (Pearson's R = .78)



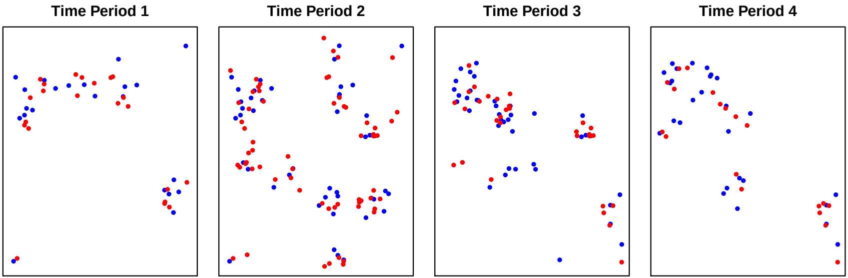
Spatial Distribution in 2011 (R = .73)



Space-time interaction

Point patterns

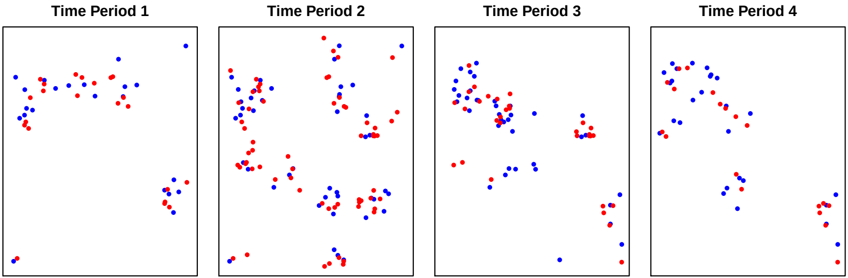
$$\mathcal{P}_1 = \{(x_i^1, y_i^1, t_i^1), i = 1, \ldots, n_1\}, \; \mathcal{P}_2 = \{(x_j^2, y_j^2, t_j^2), j = 1, \ldots, n_2\}$$



Space-time interaction

Point patterns

$$\mathcal{P}_1 = \{(x_i^1, y_i^1, t_i^1), i = 1, \ldots, n_1\},\ \mathcal{P}_2 = \{(x_j^2, y_j^2, t_j^2), j = 1, \ldots, n_2\}$$

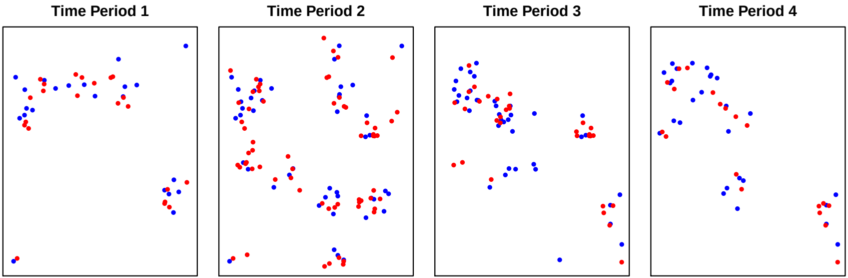


- Residual space-time dependence, after controlling for purely spatial and purely temporal dependence.

Space-time interaction

Point patterns

$$\mathcal{P}_1 = \{(x_i^1, y_i^1, t_i^1), i = 1, \ldots, n_1\}, \mathcal{P}_2 = \{(x_j^2, y_j^2, t_j^2), j = 1, \ldots, n_2\}$$



- Residual space-time dependence, after controlling for purely spatial and purely temporal dependence.
- When two events are close in space, are they also likely to be close in time?

Knox test [1964]

Put the $N = n_1 \cdot n_2$ pairs of points into a contingency table:

	close in space	far in space	
close in time	X	a	$= N_t$
far in time	b	c	
	$= N_s$		

Knox test [1964]

Put the $N = n_1 \cdot n_2$ pairs of points into a contingency table:

	close in space	far in space	
close in time	X	a	$= N_t$
far in time	b	c	
	$= N_s$		

Test statistic: $\frac{X}{N} - \frac{N_t}{N} \cdot \frac{N_s}{N}$

Mantel test [1967]

Put the N pairs of points into two matrices:

Mantel test [1967]

Put the N pairs of points into two matrices:

$$\text{space:} \quad K = \begin{pmatrix} 0 & \|s_1 - s_2\| & \dots & \|s_1 - s_n\| \\ \|s_2 - s_1\| & 0 & \dots & \|s_2 - s_n\| \\ & & \ddots & \\ \|s_n - s_1\| & \|s_n - s_2\| & \dots & 0 \end{pmatrix}$$

Mantel test [1967]

Put the N pairs of points into two matrices:

$$\text{space:} \quad K = \begin{pmatrix} 0 & \|s_1 - s_2\| & \dots & \|s_1 - s_n\| \\ \|s_2 - s_1\| & 0 & \dots & \|s_2 - s_n\| \\ & & \ddots & \\ \|s_n - s_1\| & \|s_n - s_2\| & \dots & 0 \end{pmatrix}$$

$$\text{time:} \quad L = \begin{pmatrix} 0 & |t_1 - t_2| & \dots & |t_1 - t_n| \\ |t_2 - t_1| & 0 & \dots & |t_2 - t_n| \\ & & \ddots & \\ |t_n - t_1| & |t_n - t_2| & \dots & 0 \end{pmatrix}$$

Mantel test [1967]

Put the N pairs of points into two matrices:

$$\text{space:} \quad K = \begin{pmatrix} 0 & \|s_1 - s_2\| & \dots & \|s_1 - s_n\| \\ \|s_2 - s_1\| & 0 & \dots & \|s_2 - s_n\| \\ & & \ddots & \\ \|s_n - s_1\| & \|s_n - s_2\| & \dots & 0 \end{pmatrix}$$

$$\text{time:} \quad L = \begin{pmatrix} 0 & |t_1 - t_2| & \dots & |t_1 - t_n| \\ |t_2 - t_1| & 0 & \dots & |t_2 - t_n| \\ & & \ddots & \\ |t_n - t_1| & |t_n - t_2| & \dots & 0 \end{pmatrix}$$

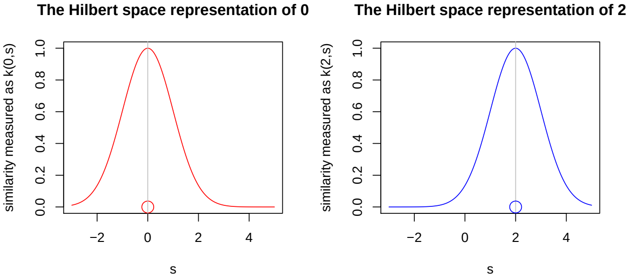
Test statistic: $\sum_{i,j} K_{i,j} L_{i,j}$

Shortcomings

- Knox: discretizes using pre-specified cutoffs
- Mantel: linear measure of independence (correlation)
- Focus is exclusively on interpoint (Euclidean) distances
- No way to include covariates, more spatial or temporal structure

Mercer Kernels

- A kernel is a real-valued paired similarity function: $k(x, y) \in \mathcal{R}$. Larger values $\Rightarrow$ more similar.
- Example: Gaussian $k(x, y) = e^{-\|x-y\|^2}$
- Mathematical theory: kernels turn points into infinite dimensional vectors, i.e. functions:



Theory

- Given points $\mathcal{P} = \{p_i = (s_i, t_i)\}$ we have two ways of measuring similarity:

$$k(p_i, p_j) := k(s_i, s_j) \quad \text{(similarity in space)}$$

$$\ell(p_i, p_j) := \ell(t_i, t_j) \quad \text{(similarity in time)}$$

Theory

- Given points $\mathcal{P} = \{p_i = (s_i, t_i)\}$ we have two ways of measuring similarity:

$$k(p_i, p_j) := k(s_i, s_j) \ \ \text{(similarity in space)}$$

$$\ell(p_i, p_j) := \ell(t_i, t_j) \ \ \text{(similarity in time)}$$

- Are these two notions of similarity independent?

Theory

- Given points $\mathcal{P} = \{p_i = (s_i, t_i)\}$ we have two ways of measuring similarity:

$$k(p_i, p_j) := k(s_i, s_j) \quad \text{(similarity in space)}$$

$$\ell(p_i, p_j) := \ell(t_i, t_j) \quad \text{(similarity in time)}$$

- Are these two notions of similarity independent?
- In Hilbert space we have vectors $\phi(s) := k(s, \cdot)$ and $\psi(t) = \ell(t, \cdot)$

Theory

- Given points $\mathcal{P} = \{p_i = (s_i, t_i)\}$ we have two ways of measuring similarity:

$$k(p_i, p_j) := k(s_i, s_j) \;\; \text{(similarity in space)}$$

$$\ell(p_i, p_j) := \ell(t_i, t_j) \;\; \text{(similarity in time)}$$

- Are these two notions of similarity independent?
- In Hilbert space we have vectors $\phi(s) := k(s, \cdot)$ and $\psi(t) = \ell(t, \cdot)$
- Consider $\phi(s)$ and $\psi(t)$ as random variables and ask:

$$\text{Is } \phi(s) \perp\!\!\!\perp \psi(t)?$$

Kernel measures of independence

- Hilbert-Schmidt Independence Criterion [Gretton et al 2012]

Kernel measures of independence

- Hilbert-Schmidt Independence Criterion [Gretton et al 2012]
- Given observations $\sim X \times Y$, is

$$Pr(X, Y) = Pr(X)Pr(Y)?$$

Kernel measures of independence

- Hilbert-Schmidt Independence Criterion [Gretton et al 2012]
- Given observations $\sim X \times Y$, is

$$Pr(X, Y) = Pr(X)Pr(Y)?$$

- HSIC works in the RKHS: given a feature space embedding $\phi(x) = k(x, \cdot)$ we can embed a distribution: $\mu_x := E_x[\phi(x)]$

Kernel measures of independence

- Hilbert-Schmidt Independence Criterion [Gretton et al 2012]
- Given observations $\sim X \times Y$, is
$$Pr(X, Y) = Pr(X)Pr(Y)?$$
- HSIC works in the RKHS: given a feature space embedding $\phi(x) = k(x, \cdot)$ we can embed a distribution: $\mu_x := E_x[\phi(x)]$
- $\mu_X \in H_{\mathcal{X}}, \mu_Y \in H_{\mathcal{Y}}, \mu_{XY} \in H_{\mathcal{X}} \otimes H_{\mathcal{Y}}$

Kernel measures of independence

- Hilbert-Schmidt Independence Criterion [Gretton et al 2012]
- Given observations $\sim X \times Y$, is

$$Pr(X, Y) = Pr(X)Pr(Y)?$$

- HSIC works in the RKHS: given a feature space embedding $\phi(x) = k(x, \cdot)$ we can embed a distribution: $\mu_x := E_x[\phi(x)]$
- $\mu_X \in H_{\mathcal{X}}, \mu_Y \in H_{\mathcal{Y}}, \mu_{XY} \in H_{\mathcal{X}} \otimes H_{\mathcal{Y}}$
- HSIC measures distance between embedding of joint distribution and marginal distributions in Hilbert space:

$$\text{HSIC} = \|\mu_{XY} - \mu_X \mu_Y\|_{HS}^2$$

Kernel measures of independence

- If kernels are characteristic / universal[1]:

  Theorem [Gretton et al 2012].

  $\mathrm{HSIC} = 0 \iff Pr(X, Y) = Pr(X)Pr(Y)$

[1] Sriperumbudur 2011 discusses relationship between these notions

Kernel measures of independence

- If kernels are characteristic / universal[1]:

  Theorem [Gretton et al 2012].

  $\text{HSIC} = 0 \iff Pr(X, Y) = Pr(X)Pr(Y)$
- Simple estimators available:

[1] Sriperumbudur 2011 discusses relationship between these notions

Kernel measures of independence

- If kernels are characteristic / universal[1]:

  Theorem [Gretton et al 2012].

  HSIC $= 0 \iff Pr(X, Y) = Pr(X)Pr(Y)$

- Simple estimators available:

$$\widehat{\mathrm{HSIC}} = \frac{1}{n}\mathrm{tr} HKHL$$

where $K_{ij} = k(x_i, x_j)$, $L_{ij} = \ell(y_i, y_j)$, $H = (I - \frac{1}{n}11^T)$.

[1]Sriperumbudur 2011 discusses relationship between these notions

Kernel Space-time interaction statistics

- Recall: space-time interaction means interpoint spatial and temporal "distances" are independent.

Kernel Space-time interaction statistics

- Recall: space-time interaction means interpoint spatial and temporal "distances" are independent.
- Our definition: work in Hilbert space and ask whether for $p = (s, t)$:

$$\phi(s) \sim X \perp\!\!\!\perp \psi(t) \sim Y$$

Kernel Space-time interaction statistics

- Recall: space-time interaction means interpoint spatial and temporal "distances" are independent.
- Our definition: work in Hilbert space and ask whether for $p = (s, t)$:

$$\phi(s) \sim X \perp\!\!\!\perp \psi(t) \sim Y$$

- Measure distance in the RKHS!

$$\mathrm{KST} = \|\mu_{XY} - \mu_X \mu_Y\|_{HS}^2$$

Kernel Space-time interaction statistics

- Recall: space-time interaction means interpoint spatial and temporal "distances" are independent.
- Our definition: work in Hilbert space and ask whether for $p = (s, t)$:

$$\phi(s) \sim X \perp\!\!\!\perp \psi(t) \sim Y$$

- Measure distance in the RKHS!

$$\text{KST} = \|\mu_{XY} - \mu_X \mu_Y\|_{HS}^2$$

- If kernels are stationary, then they're a similarity metric based on interpoint distances.

Kernel Space-time interaction statistics

- Recall: space-time interaction means interpoint spatial and temporal "distances" are independent.
- Our definition: work in Hilbert space and ask whether for $p = (s, t)$:

$$\phi(s) \sim X \perp\!\!\!\perp \psi(t) \sim Y$$

- Measure distance in the RKHS!

$$\text{KST} = \|\mu_{XY} - \mu_X \mu_Y\|_{HS}^2$$

- If kernels are stationary, then they're a similarity metric based on interpoint distances.
- But we could use non-stationary kernels, or domain appropriate kernels from geostatistics, time series literature

Comparison with Mantel

- Mantel:

$$\sum_{i,j} K_{ij} L_{ij}$$

- $\widehat{\mathrm{KST}}$:

$$\frac{1}{n^2} \sum_{i,j} K_{ij} L_{ij} - \frac{1}{n^3} \sum_{i,j,r} K_{ij} L_{ir} - \frac{1}{n^3} \sum_{i,j,r} K_{ij} L_{rj} + \frac{1}{n^4} \sum_{i,j,q,r} K_{ij} L_{qr}$$

- Notice: missing terms! Mantel is almost right, but the centering is wrong. With Euclidean distances, and correct centering, Mantel becomes dcor (Szekely and Rizzo, 2014)

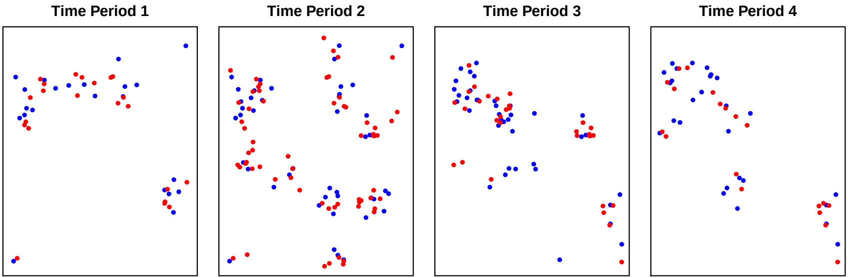
Our Contributions

- New, more general way of thinking about space-time interaction $\Rightarrow$ new test for space-time interaction
- Extensions to bivariate (KST_{12}) / forward in time cases ($KST_{1\rightarrow 2}$)
- Interesting connections with Mantel test, showing its shortcomings and fix
- More flexible test: kernels can encode more than just distance between points. KST tests for non-linear dependencies.

Experimental Setup

Synthetic data: draw $n = 40$ random cluster center parents, draw $k = 5$ children with locations displaced $N(0, \sigma)$ from parent in every direction.

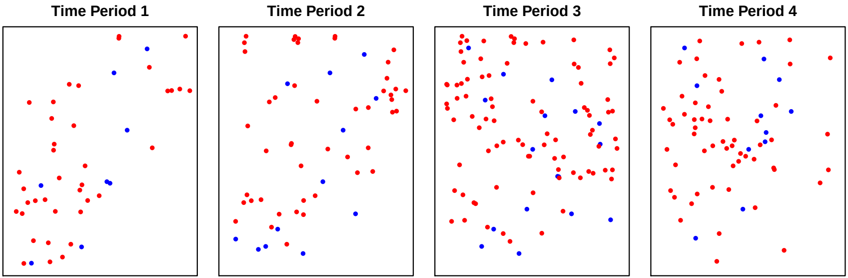
Easy example: $\sigma = .05$



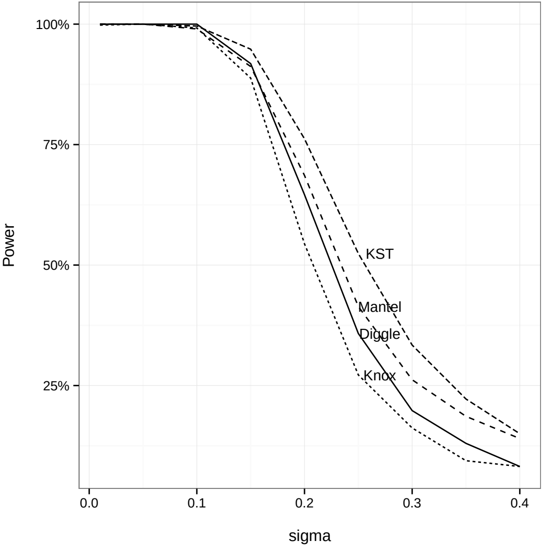
Experimental Setup

Synthetic data: draw $n = 40$ random cluster center parents, draw $k = 5$ children with locations displaced $N(0, \sigma)$ from parent in every direction.

Hard example: $\sigma = .2$



Synthetic Data: Results



Experimental Setup: Crime Data

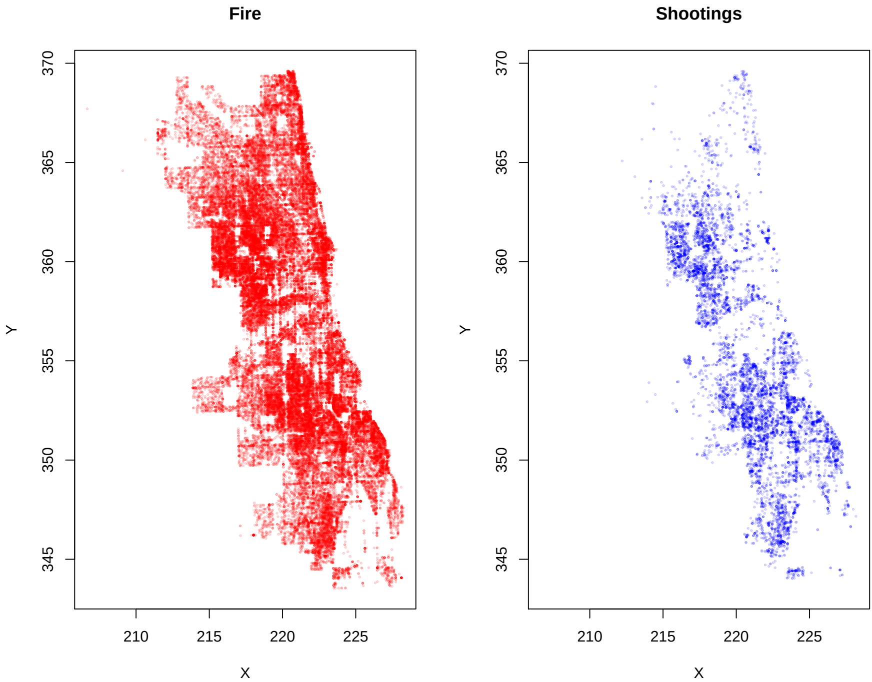
Question: which types of calls to 911 predict homicides and aggravated battery with a handgun ("shootings")?

Experimental Setup: Crime Data

Question: which types of calls to 911 predict homicides and aggravated battery with a handgun ("shootings")?

Data:

- Dispatcher calls from January 2007-May 2010, coded by one of 271 types ($\approx$ 9 million):
  ```
  "01-01-2010","12:25:00","ARSON",1172456,1834562
  "01-02-2010","19:55:00","THEFT",1173123,1831123
  ```
- All shootings / homicides from January 2007-May 2010 (9,087 total):
  ```
  "01-01-2010","19:00:37","HOMICIDE",1172001,1834023
  "01-07-2010","19:55:00","HOMICIDE",1173934,1831384
  ```



Fire
Shootings
370
365
360
355
350
345
Y
210
215
220
225
X

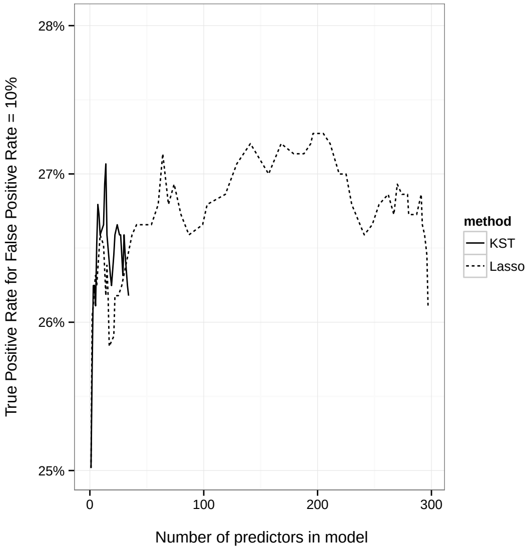
Experimental Setup

- Calculate p-values for $\mathrm{KST}_{1\to 2}$ between each 911 call type and shootings
- Use Gaussian RBF kernels: bandwidth $\frac{1}{4}$ mile, 14 days
- Permutation testing 500 times to calculate p-values
- Evaluated TPR on held-out data from an L_1-regularized logistic regression model with features pre-selected by $\mathrm{KST}_{1\to 2}$

Results

p-value	911 call type	p-value	911 call type
0.002	STREETS & SAN PINK CARD	0.002	MENTAL UNAUTH ABSENCE
0.002	PERSON SHOT	0.002	DEATH REMOVAL
0.002	WALK DOWN	0.002	SHOTS FIRED (OV)
0.002	ASSAULT IP	0.002	CRIMINAL TRES. (OV)
0.002	EVIDENCE TECHNICIAN (PRI. 1)	0.002	ARSON REPORT
0.002	AUTO THEFT IP	0.002	TASTE OF CHICAGO
0.002	EVIDENCE TECHNICIAN (PRI. 3)	0.002	AMBER ALERT
0.002	PERSON WITH A GUN	0.004	DETAIL
0.002	MISSION	0.004	GANG DISTURBANCE
0.002	PERSON WANTED	0.004	PURSUIT FOOT (OV)
0.002	PERSON STABBED	0.006	BATTERY IP
0.002	SHOTS FIRED	0.006	NOTIFY
0.002	EVIDENCE TECHNICIAN (PRI. 2)	0.008	ON VIEW
0.002	PLAN 1-5	0.008	CRIM DAM. TO PROP IP
0.002	K9 REQUEST	0.008	RECOVERED STOLEN AUTO
0.002	OUTDOOR ROLL CALL	0.010	THEFT IP
0.002	CRIM DAM. TO PROP RPT	0.010	CRIM DAM. TO PROP (OV)
0.002	HOLDING OFFENDER (CITZ.)	0.010	MUNICIPAL ORD. VIOLATION

Results



28%
27%
26%
25%
True Positive Rate for False Positive Rate = 10%
0
100
200
300
Number of predictors in model
method
KST
Lasso

Results

- KST + Lasso:

  shots fired, shooting, officer pursuing someone on foot, officer heard shots fired, narcotics loitering, officer station assignment, person shot, meeting of the police beat unit, support unit request, gang loitering.

- Lasso only:

  shots fired, domestic disturbance, person with a gun, shootings (the lagged version of the dependent variable), officer eating lunch, vicious animal, parking violation, gang disturbance, gambling, battery in progress.

Conclusions

- New data-driven formulation of “leading indicators” question as space-time interaction between pairs of point processes
- Defined a new kernel-based space-time interaction test
- Outperformed classical tests
- Applied to large, real, and important dataset: shootings in Chicago

Thank you! Questions?[2]

Seth Flaxman

flaxman@cmu.edu

www.sethrf.com (preprint available)

[2]Thank you to the Chicago Police Department for sharing data. Points of view or opinions contained within this presentation are those of the author and do not necessarily represent the official position or policies of the Chicago Police Department. Title page photo by Palsson on Flickr.

Extensions

- Bivariate case: for test statistic, restrict sums to pairs of points of different types:

Extensions

- Bivariate case: for test statistic, restrict sums to pairs of points of different types:

$$\frac{1}{n^2}\sum_{i,j} k(s_i^1, s_j^2)\ell(t_i^1, t_j^2) - \frac{2}{n^3}\sum_{i,j,r} k(s_i^1, s_j^2)\ell(t_i^1, t_r^2) + \frac{1}{n^4}\sum_{i,j,q,r} k(s_i^1, s_j^2)\ell(t_q^1, t_r^2)$$

- Interesting interpretation in Hilbert space

Extensions

- Bivariate case: for test statistic, restrict sums to pairs of points of different types:

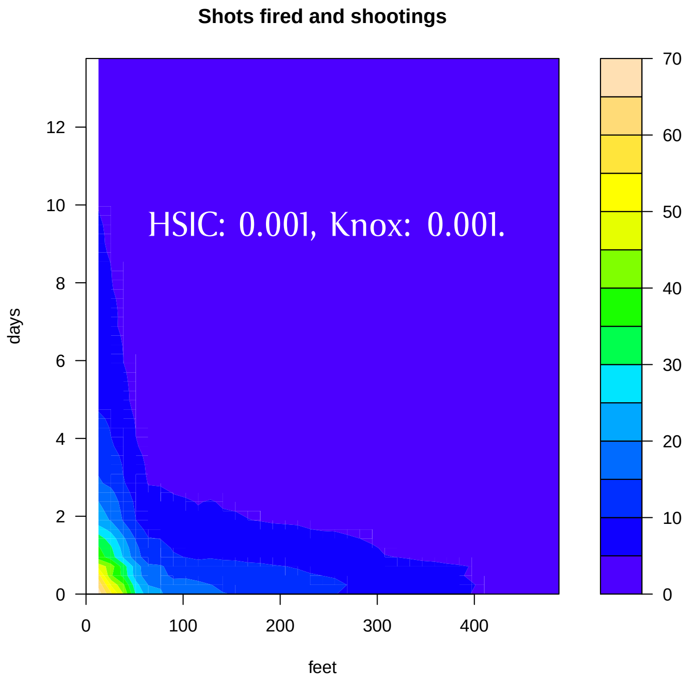
$$\frac{1}{n^2}\sum_{i,j} k(s_i^1, s_j^2)\ell(t_i^1, t_j^2) - \frac{2}{n^3}\sum_{i,j,r} k(s_i^1, s_j^2)\ell(t_i^1, t_r^2) + \frac{1}{n^4}\sum_{i,j,q,r} k(s_i^1, s_j^2)\ell(t_q^1, t_r^2)$$

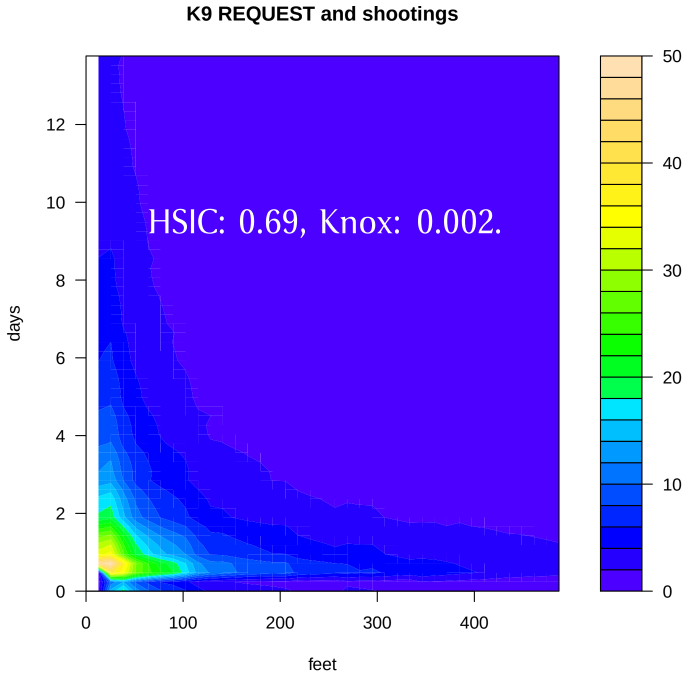
- Interesting interpretation in Hilbert space
- Only predict forward in time: restrict sums to pairs of points where $t_i < t_j$.

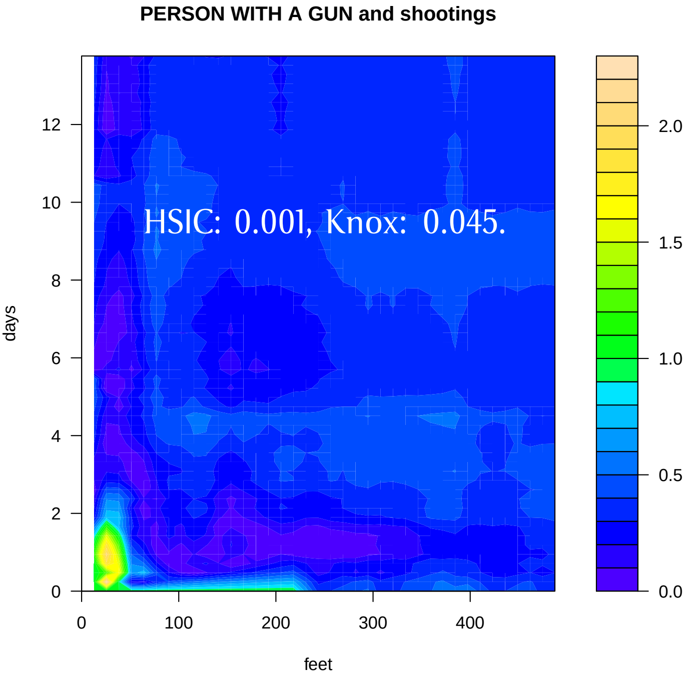
Excess risk attributable to space-time interaction

$$D(s,t) = \frac{F_{S,T}(s,t) - F_S(s)F_T(t)}{F_S(s)F_T(t)}$$

Given that we see an event of type 1, proportional increase (excess risk) of seeing an event of type 2.


Shots fired and shootings
HSIC: 0.001, Knox: 0.001.
days
feet
0
100
200
300
400
0
2
4
6
8
10
12
0
10
20
30
40
50
60
70

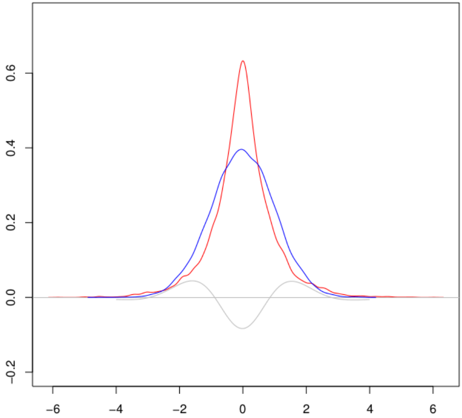

K9 REQUEST and shootings
HSIC: 0.69, Knox: 0.002.
days
feet
0
2
4
6
8
10
12
0
100
200
300
400
0
10
20
30
40
50


PERSON WITH A GUN and shootings
HSIC: 0.001, Knox: 0.045.
days
feet
0
2
4
6
8
10
12
100
200
300
400
0.0
0.5
1.0
1.5
2.0

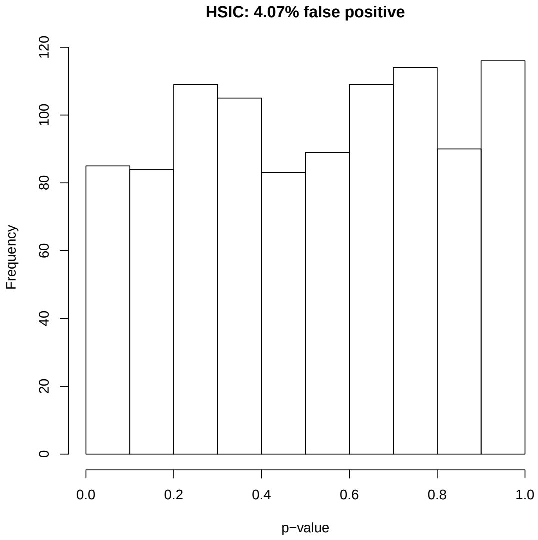
Maximum Mean Discrepancy (Gretton et al. 2012)

"Witness" $\hat{f}^*$:

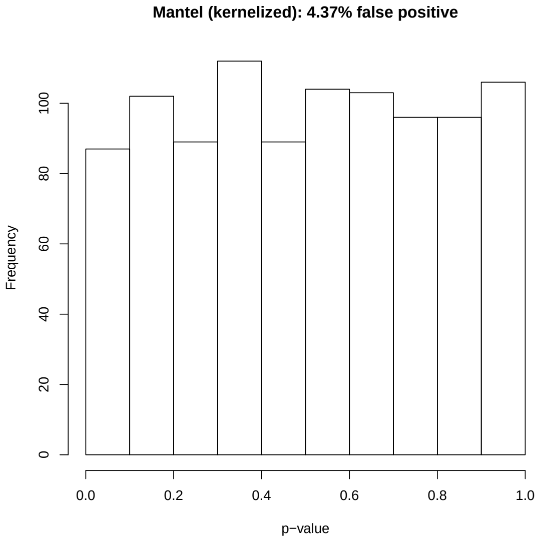
$$\hat{f}^*(x, y) \propto \sum_i k(x, x_i) - \sum_j k(y, y_j)$$



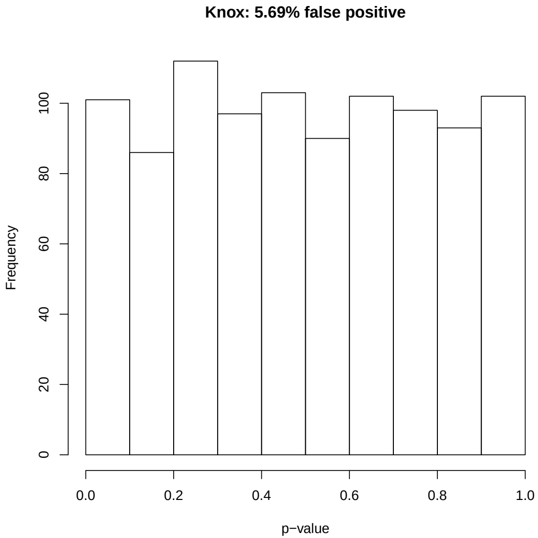
False positive rate

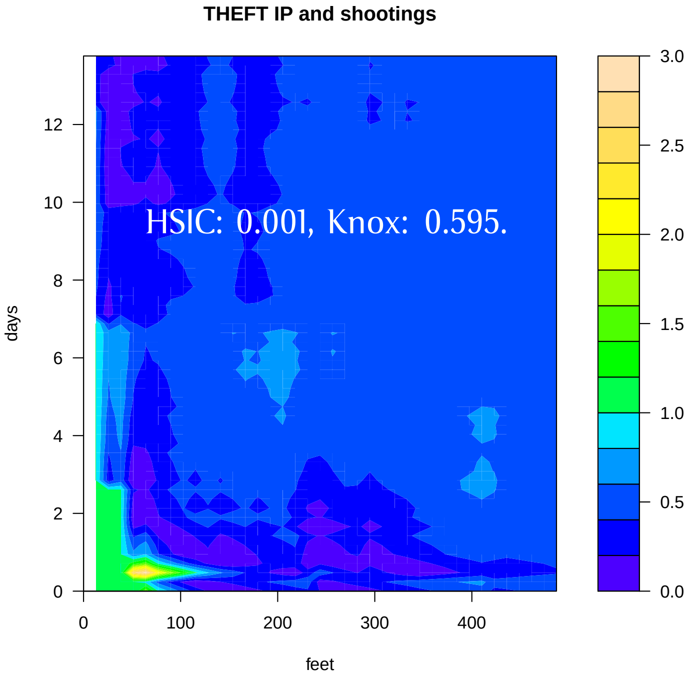


False positive rate



False positive rate




THEFT IP and shootings
HSIC: 0.001, Knox: 0.595.
days
feet
0
100
200
300
400
0
2
4
6
8
10
12
0.0
0.5
1.0
1.5
2.0
2.5
3.0