Kernel methods for hypothesis testing and inference

MLSS Tübingen, 2015

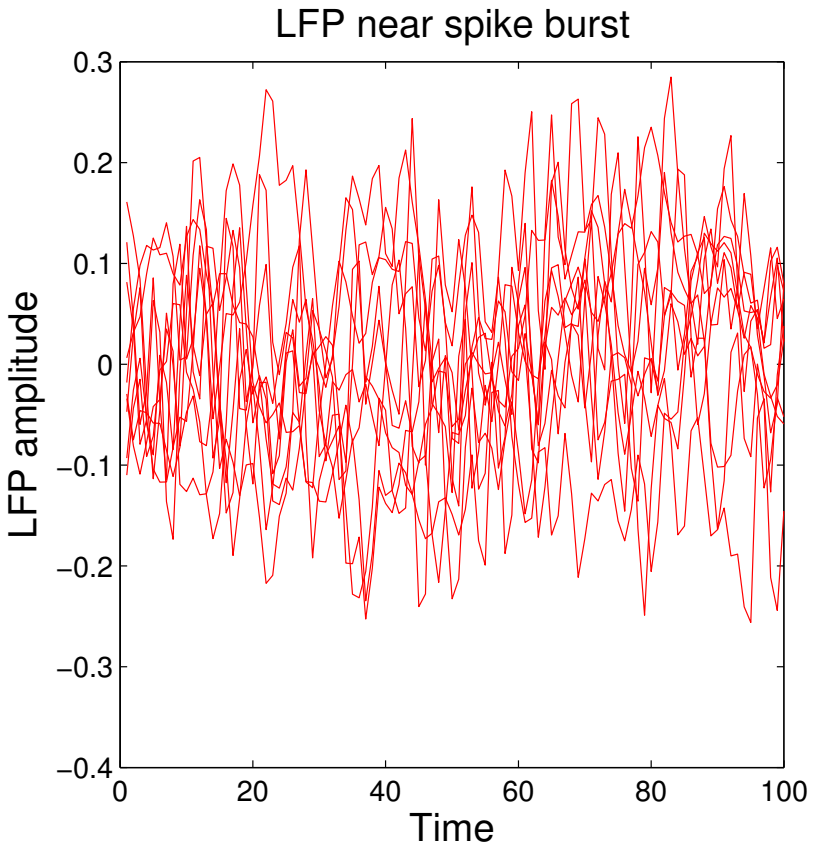
Arthur Gretton

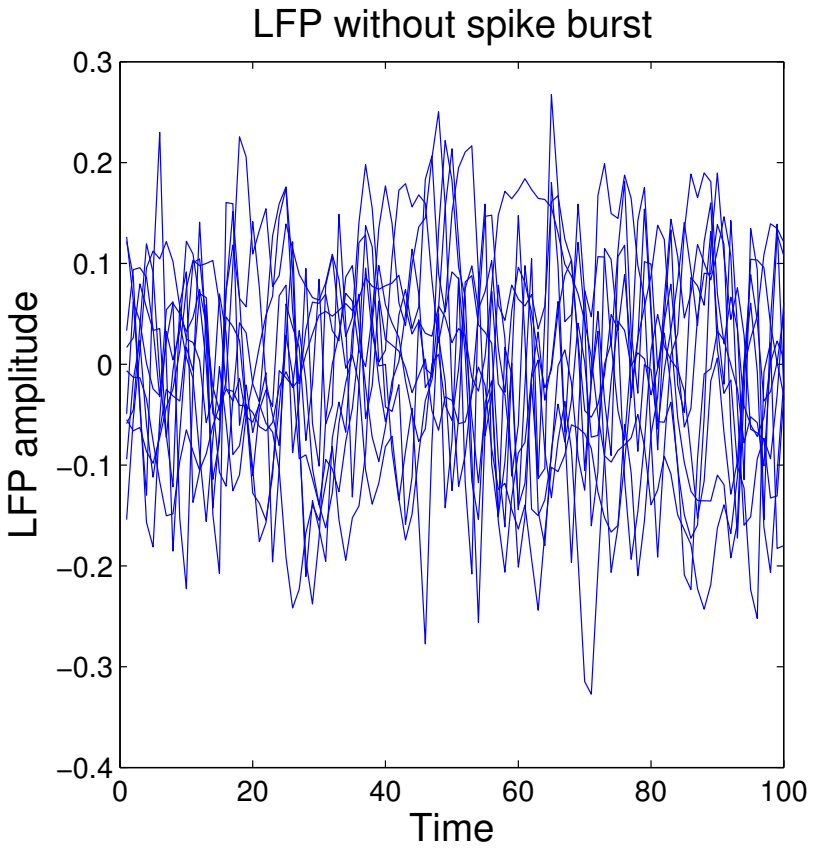
Gatsby Unit, CSML, UCL

Some motivating questions...

Detecting differences in brain signals

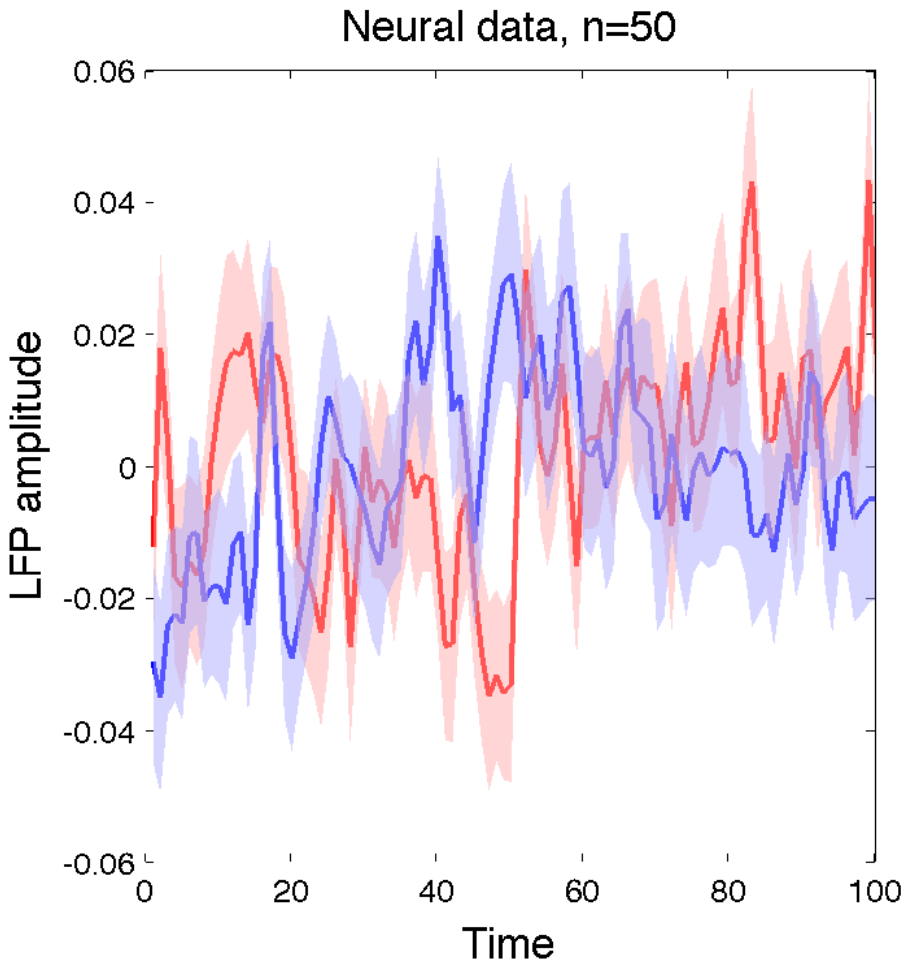
The problem: Do local field potential (LFP) signals change when measured near a spike burst?





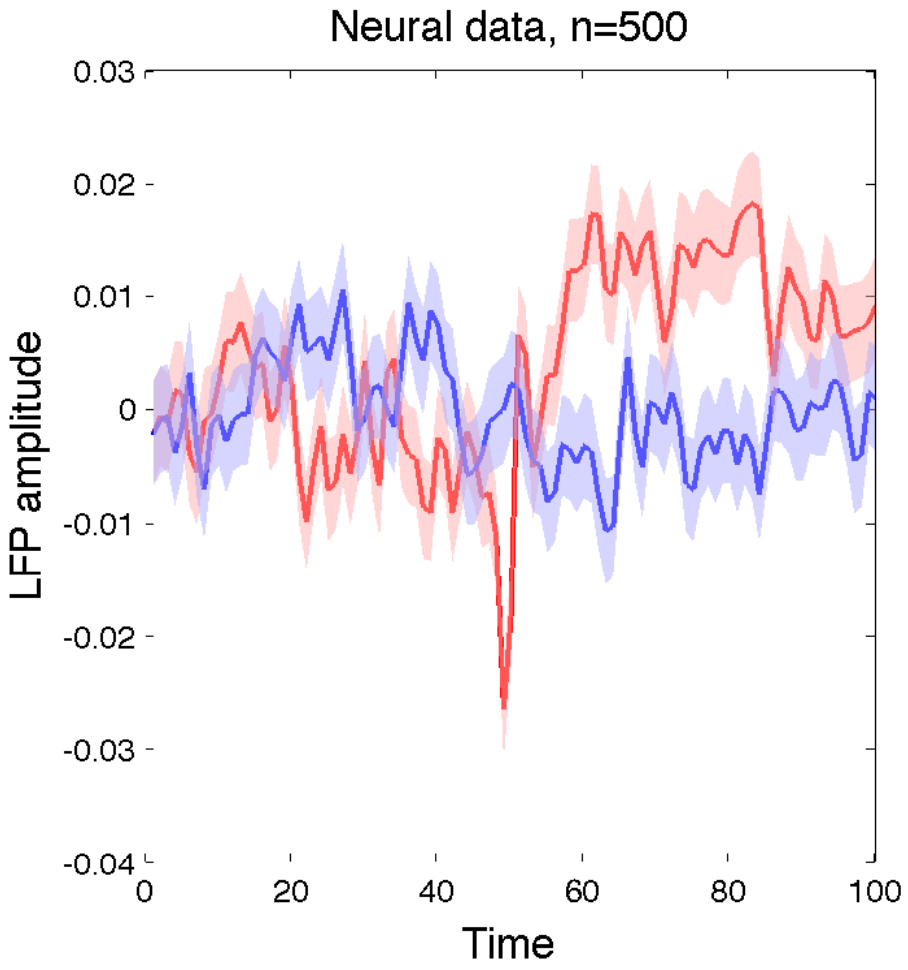
Detecting differences in brain signals

The problem: Do local field potential (LFP) signals change when measured near a spike burst?

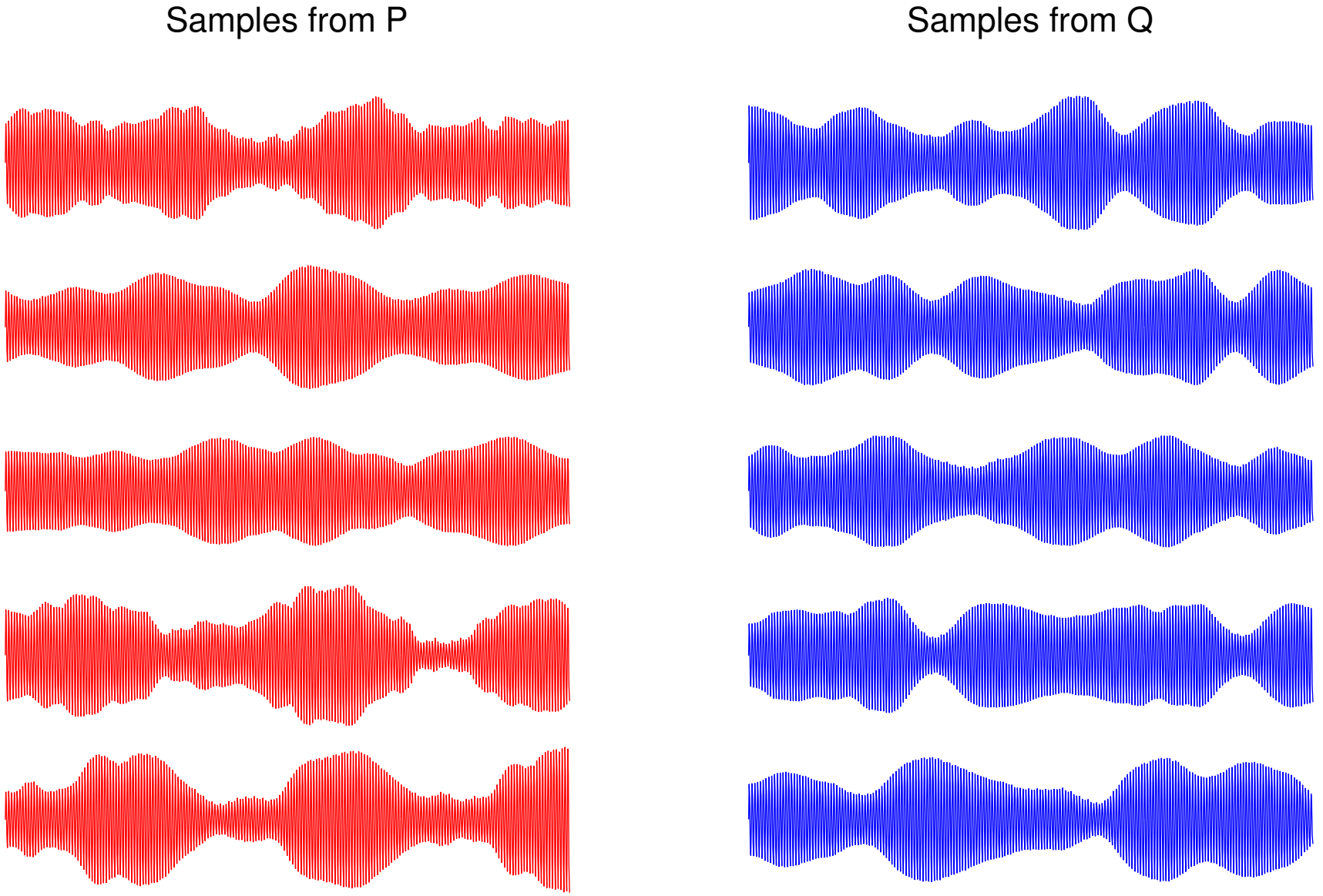


Detecting differences in brain signals

The problem: Do local field potential (LFP) signals change when measured near a spike burst?



Detecting differences in amplitude modulated signals



Adversarial training of deep neural networks

From ICML 2015:

Generative Moment Matching Networks

Yujia Li[1] YUJIALI@CS.TORONTO.EDU
Kevin Swersky[1] KSWERSKY@CS.TORONTO.EDU
Richard Zemel[1,2] ZEMEL@CS.TORONTO.EDU
[1]Department of Computer Science, University of Toronto, Toronto, ON, CANADA
[2]Canadian Institute for Advanced Research, Toronto, ON, CANADA

From UAI 2015:

Training generative neural networks via Maximum Mean Discrepancy optimization

Gintare Karolina Dziugaite
University of Cambridge

Daniel M. Roy
University of Toronto

Zoubin Ghahramani
University of Cambridge

Idea: In adversarial nets (Goodfellow et al. NIPS 2014), replace discriminator network with *maximum mean discrepancy*, a kernel distance between distributions.

Case of discrete domains

- How do you compare distributions. . .
- . . .in a discrete domain? [Read and Cressie, 1988]

Case of discrete domains

- How do you compare distributions...
- ...in a discrete domain? [Read and Cressie, 1988]

X_1: Now disturbing reports out of Newfoundland show that the fragile snow crab industry is in serious decline. First the west coast salmon, the east coast salmon and the cod, and now the snow crabs off Newfoundland.

X_2: To my pleasant surprise he responded that he had personally visited those wharves and that he had already announced money to fix them. What wharves did the minister visit in my riding and how much additional funding is he going to provide for Delaps Cove, Hampton, Port Lorne,

...

$$P_X \stackrel{?}{=} P_Y$$

Y_1: Honourable senators, I have a question for the Leader of the Government in the Senate with regard to the support funding to farmers that has been announced. Most farmers have not received any money yet.

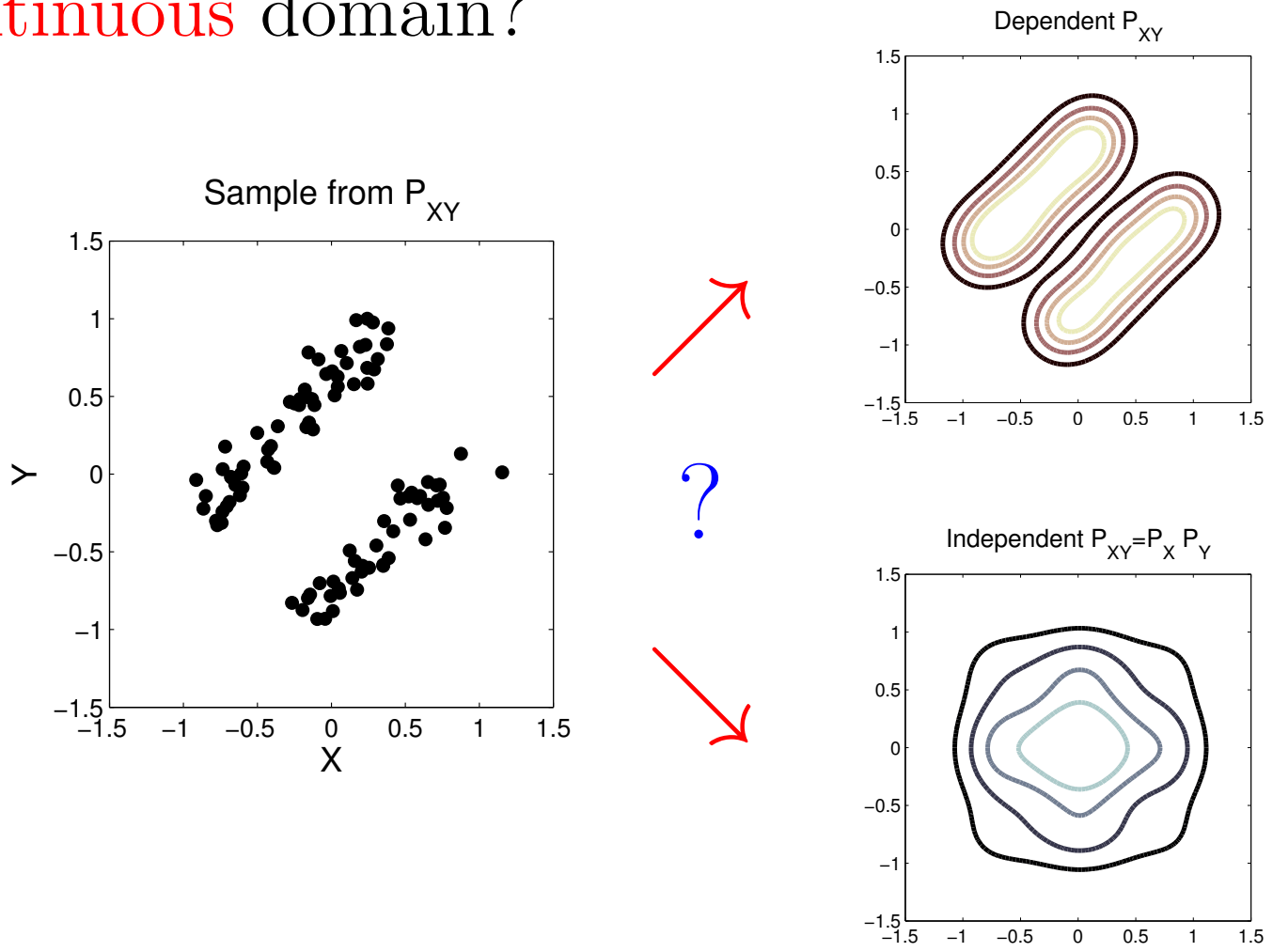
Y_2: On the grain transportation system we have had the Estey report and the Kroeger report. We could go on and on. Recently programs have been announced over and over by the government such as money for the disaster in agriculture on the prairies and across Canada.

...

Are the pink extracts from the same distribution as the gray ones?

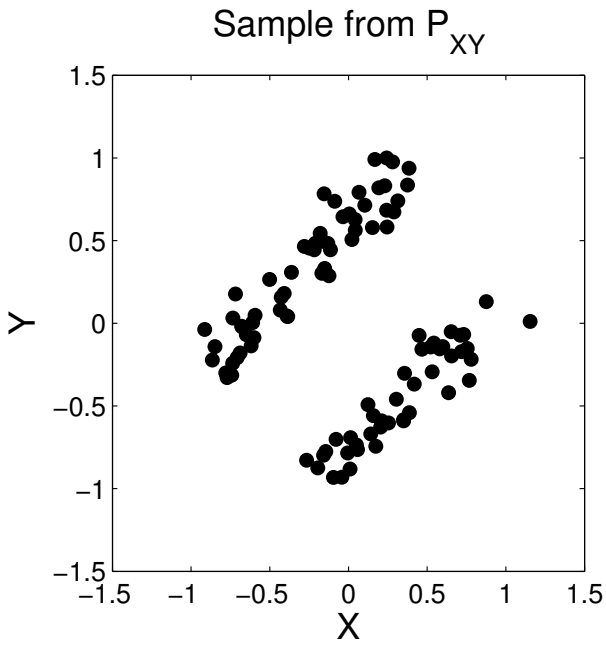
Detecting statistical dependence, continuous domain

- How do you detect dependence...
- ...in a continuous domain?

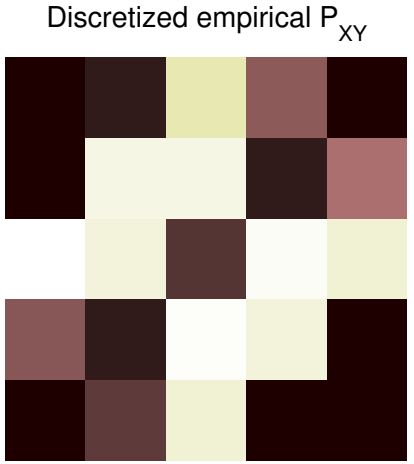


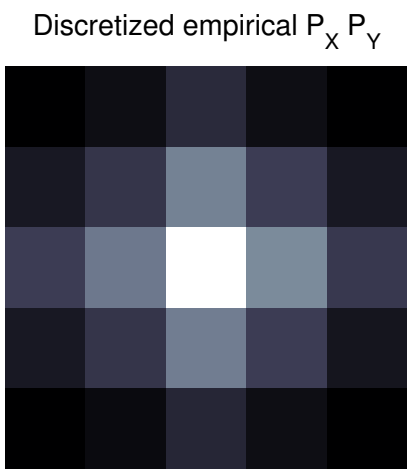
Detecting statistical dependence, continuous domain

- How do you detect dependence...
- ...in a continuous domain?



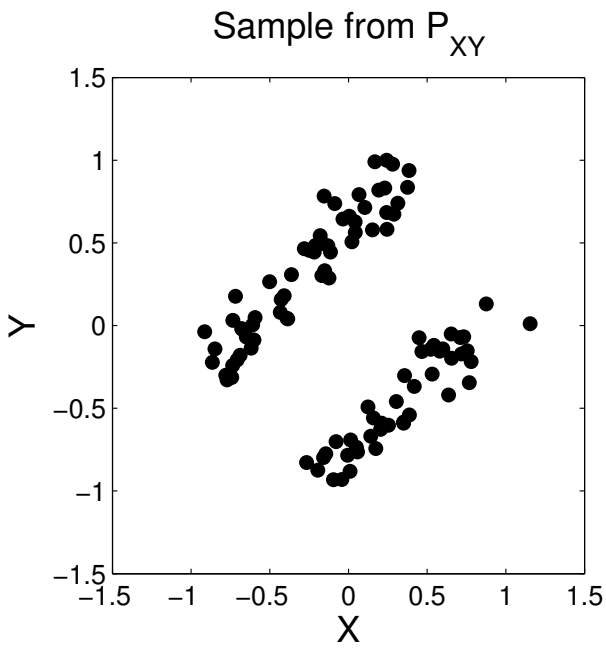
?





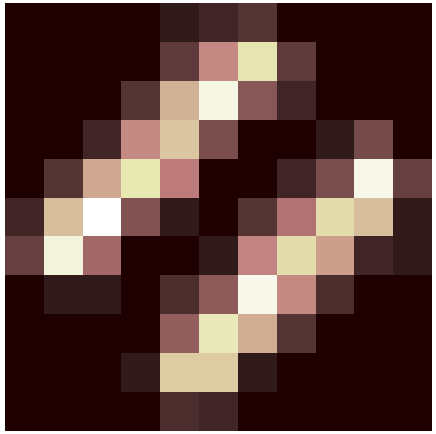
Detecting statistical dependence, continuous domain

- How do you detect dependence...
- ...in a continuous domain?

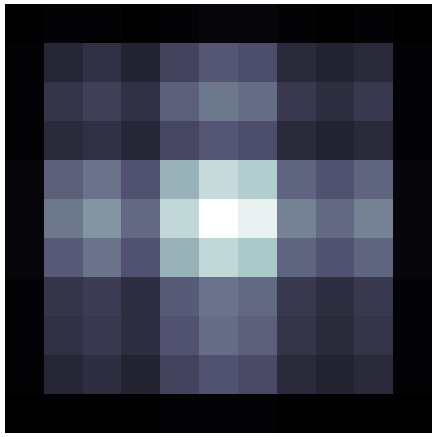


?

Discretized empirical P_{XY}



Discretized empirical $P_X P_Y$



Detecting statistical dependence, continuous domain

- How do you detect dependence. . .
- . . .in a continuous domain?
- Problem: fails even in "low" dimensions! [NIPS07a, ALT08]
  - X and Y in $\mathbb{R}^4$, statistic=Power divergence, samples= 1024, cases where dependence detected=0/500
- Too few points per bin

Detecting statistical dependence, discrete domain

- How do you detect dependence...
- ...in a discrete domain? [Read and Cressie, 1988]

X_1: Honourable senators, I have a question for the Leader of the Government in the Senate with regard to the support funding to farmers that has been announced. Most farmers have not received any money yet.

X_2: No doubt there is great pressure on provincial and municipal governments in relation to the issue of child care, but the reality is that there have been no cuts to child care funding from the federal government to the provinces. In fact, we have increased federal investments for early childhood development.

...

$$P_{XY} \stackrel{?}{=} P_X P_Y$$

Y_1: Honorables sénateurs, ma question s'adresse au leader du gouvernement au Sénat et concerne l'aide financiére qu'on a annoncée pour les agriculteurs. La plupart des agriculteurs n'ont encore rien reu de cet argent.

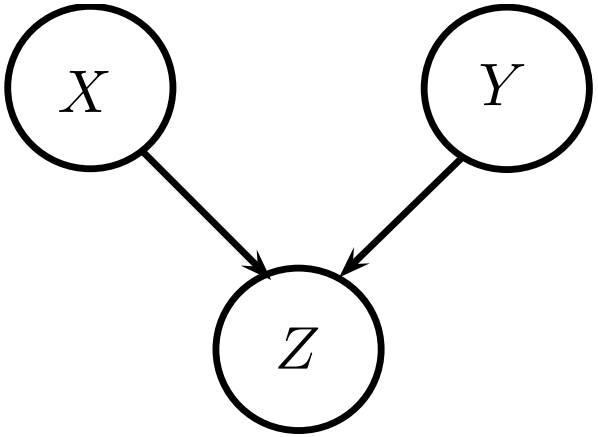
Y_2: Il est évident que les ordres de gouvernements provinciaux et municipaux subissent de fortes pressions en ce qui concerne les services de garde, mais le gouvernement n'a pas réduit le financement qu'il verse aux provinces pour les services de garde. Au contraire, nous avons augmenté le financement fédéral pour le développement des jeunes enfants.

...

Are the French text extracts translations of the English ones?

Detecting a higher order interaction

- How to detect V-structures with pairwise weak (or nonexistent) dependence?

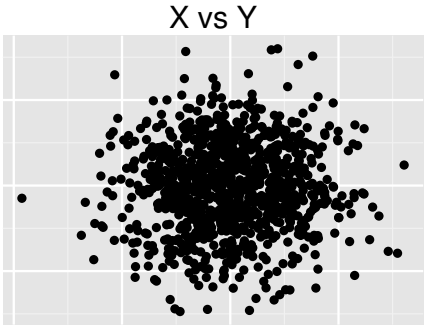


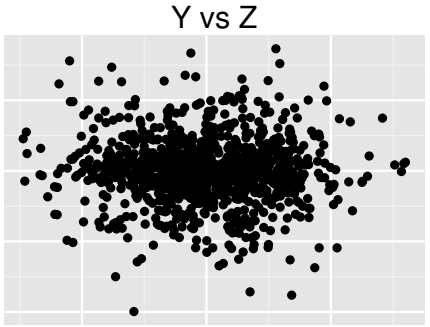
Detecting a higher order interaction

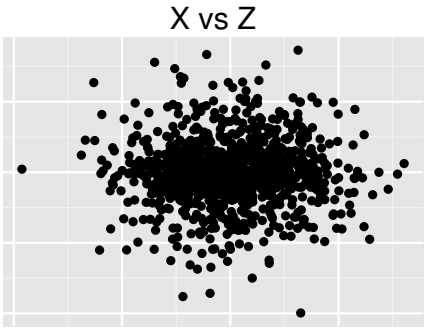
- How to detect V-structures with pairwise weak (or nonexistent) dependence?

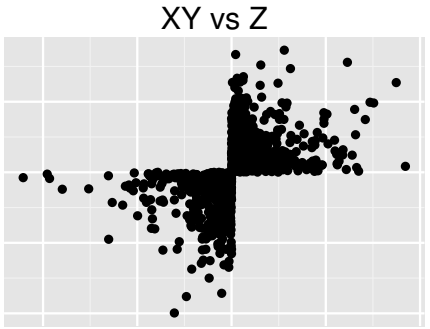
Detecting a higher order interaction

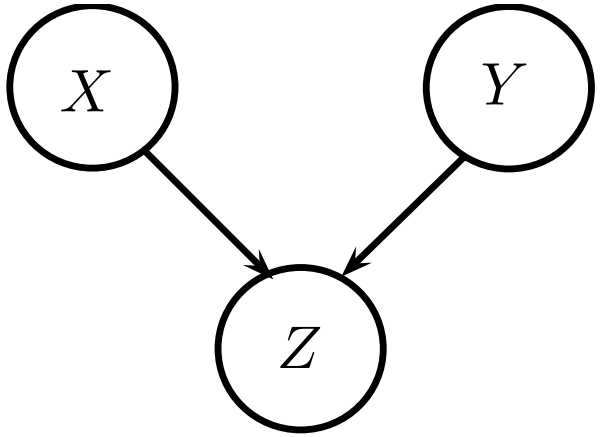
- How to detect V-structures with pairwise weak (or nonexistent) dependence?
- $X \perp\!\!\!\perp Y, Y \perp\!\!\!\perp Z, X \perp\!\!\!\perp Z$







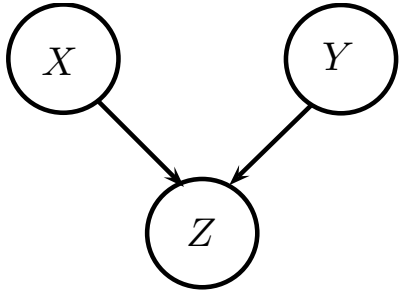




- $X, Y \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$,
- $Z|\ X, Y \sim \text{sign}(XY)Exp(\frac{1}{\sqrt{2}})$

Faithfulness violated here

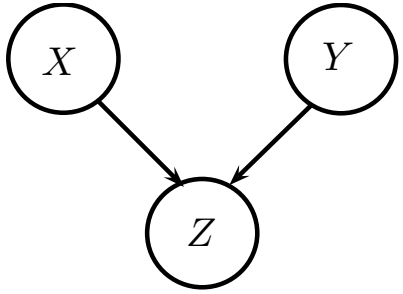
V-structure Discovery



Assume $X \perp\!\!\!\perp Y$ has been established. V-structure can then be detected by:

- CI test: $\mathbf{H_0} : X \perp\!\!\!\perp Y | Z$ (Zhang et al 2011) or

V-structure Discovery

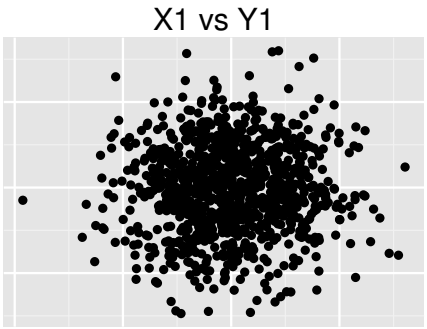


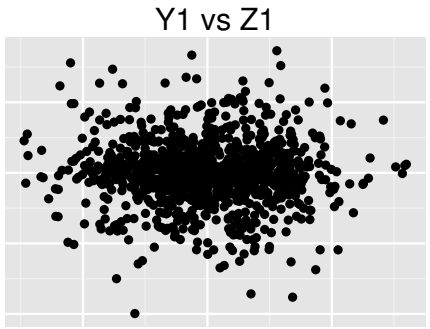
Assume $X \perp\!\!\!\perp Y$ has been established. V-structure can then be detected by:

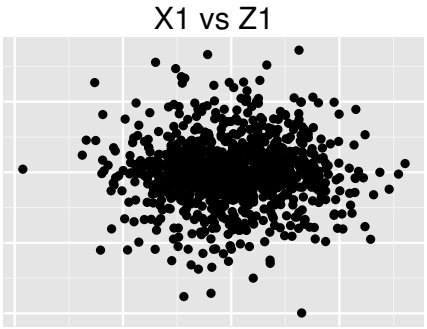
- CI test: $\mathbf{H_0} : X \perp\!\!\!\perp Y | Z$ (Zhang et al 2011) or
- Factorisation test: $\mathbf{H_0} : (X, Y) \perp\!\!\!\perp Z \vee (X, Z) \perp\!\!\!\perp Y \vee (Y, Z) \perp\!\!\!\perp X$ (multiple two-variable independence tests)
  - compute p-values for each of the marginal tests for $(Y, Z) \perp\!\!\!\perp X$, $(X, Z) \perp\!\!\!\perp Y$, or $(X, Y) \perp\!\!\!\perp Z$
  - apply Holm-Bonferroni (**HB**) sequentially rejective correction (Holm 1979)

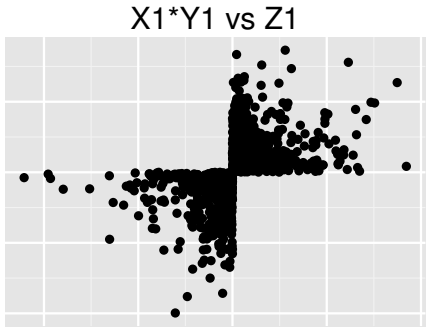
V-structure Discovery (2)

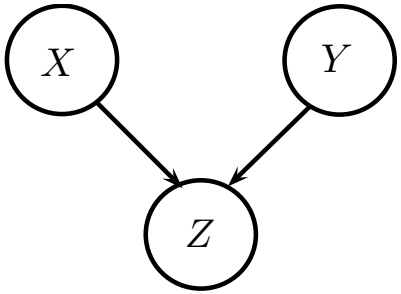
- How to detect V-structures with pairwise weak (or nonexistent) dependence?
- $X \perp\!\!\!\perp Y, Y \perp\!\!\!\perp Z, X \perp\!\!\!\perp Z$











- $X_1, Y_1 \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$,
- $Z_1 |\, X_1, Y_1 \sim \text{sign}(X_1 Y_1) Exp(\frac{1}{\sqrt{2}})$
- $X_{2:p}, Y_{2:p}, Z_{2:p} \overset{i.i.d.}{\sim} \mathcal{N}(0, \mathbf{I}_{p-1})$

Faithfulness violated here

V-structure Discovery (3)

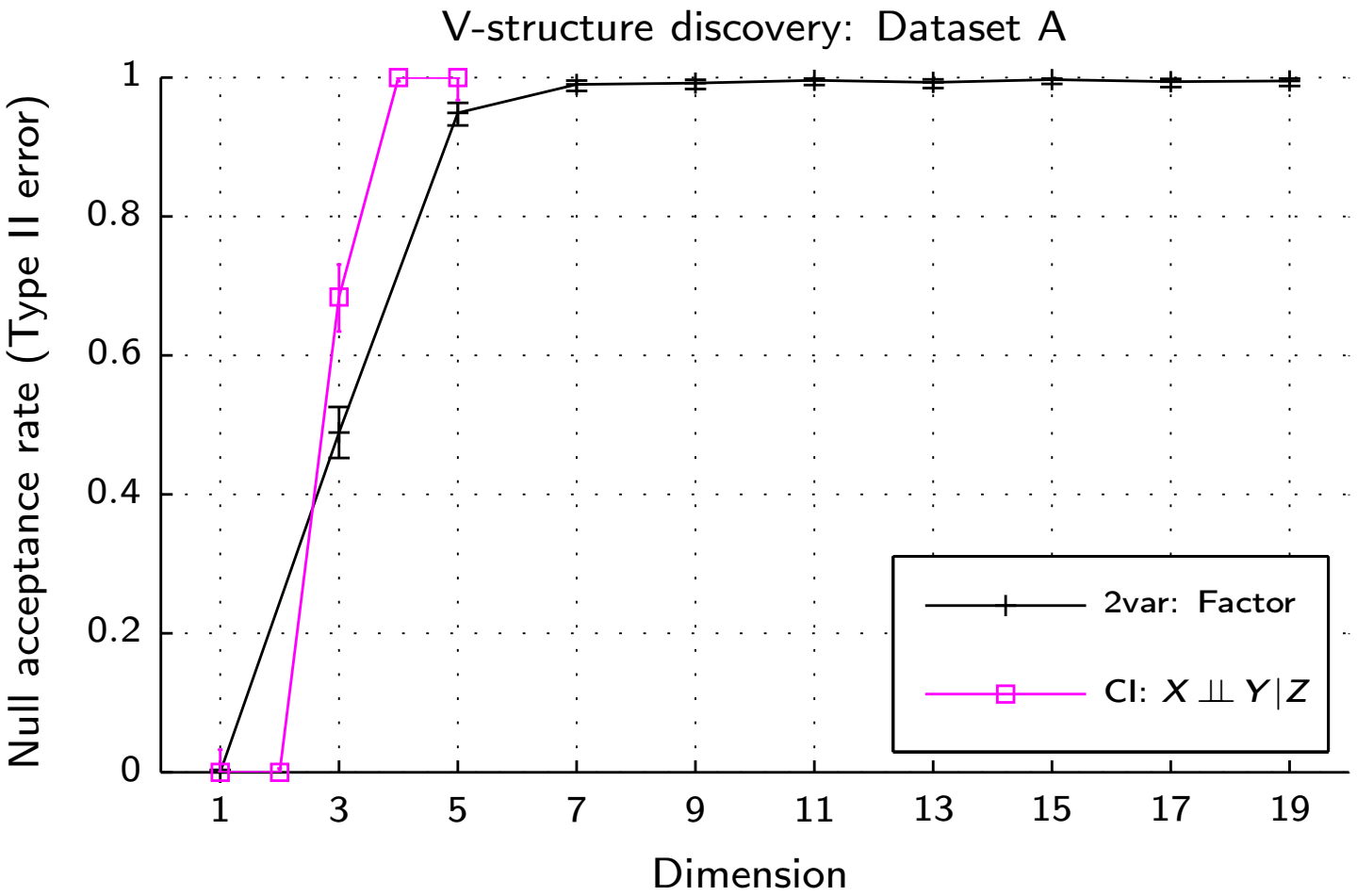


Figure 1: CI test for $X \perp\!\!\!\perp Y|Z$ from Zhang et al (2011), and a factorisation test with a **HB** correction, $n = 500$

Outline

- Intro to reproducing kernel Hilbert spaces (RKHS)
- An RKHS metric on the space of probability measures
  - Distance between means in space of features (RKHS)
  - Characteristic kernels: feature space mappings of probabilities unique
  - Nonparametric two-sample test
- Dependence detection
  - Covariance in feature space and test
- Relation with energy distance and distance covariance
- Advanced topics
  - Interactions with three (or more) variables, conditional indep. test
  - Optimal kernel choice
  - Bayesian inference without models

References

T. Read and N. Cressie. *Goodness-Of-Fit Statistics for Discrete Multivariate Analysis*. Springer-Verlag, New York, 1988.