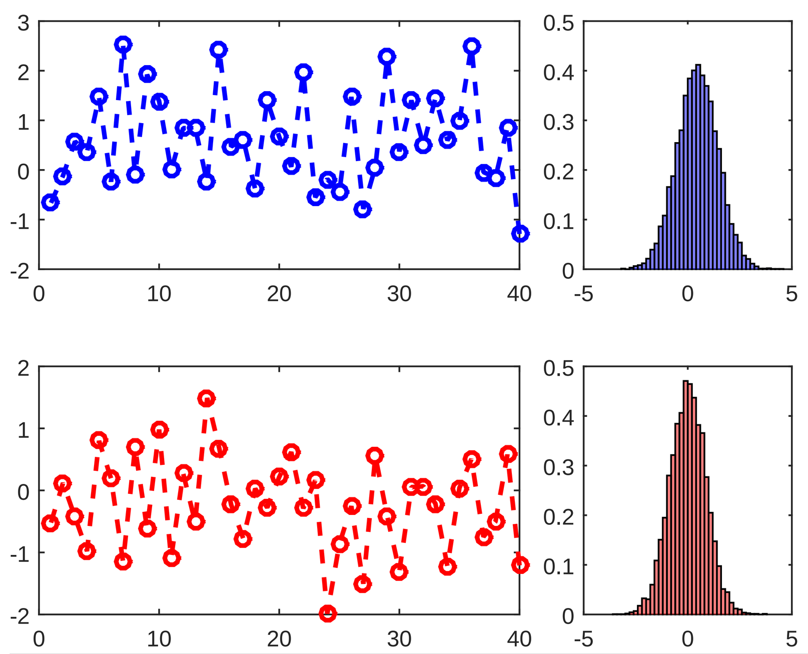
A Wild Bootstrap for Degenerate Kernel Tests

Kacper Chwialkowski[2], Dino Sejdinovic,[1] ,Arthur Gretton,[1]

[1]Gatsby Unit, CSML, UCL, UK; and [2]CS, UCL, UK;

Is P the same distribution as Q ?



Where one can use Maximum Mean Discrepancy ?

- Markov chains diagnostics
- Change point detection

Other tests

- Hilbert Schmidt Independence Criterion
  - Dependency structure in financial markets
  - Brain region activation
- Three Variables Interaction

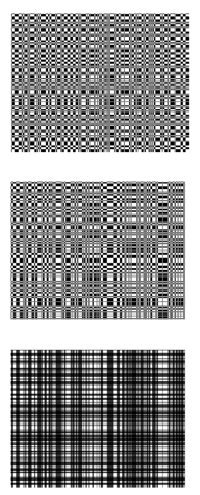
Background

Similarity

$$K_{P,P} = \begin{bmatrix} k(X_1,X_1) & \dots & k(X_1,X_n) \\ \vdots & \ddots & \vdots \\ k(X_n,X_1) & \dots & k(X_n,X_n) \end{bmatrix} =$$

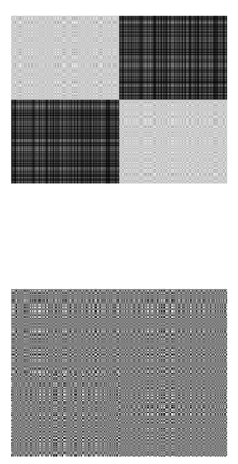
$$K_{Q,Q} = \begin{bmatrix} k(Y_1,Y_1) & \dots & k(Y_1,Y_n) \\ \vdots & \ddots & \vdots \\ k(Y_n,Y_1) & \dots & k(Y_n,Y_n) \end{bmatrix} =$$

$$K_{P,Q} = \begin{bmatrix} k(X_1,Y_1) & \dots & k(X_1,Y_n) \\ \vdots & \ddots & \vdots \\ k(X_n,Y_1) & \dots & k(X_n,Y_n) \end{bmatrix} =$$

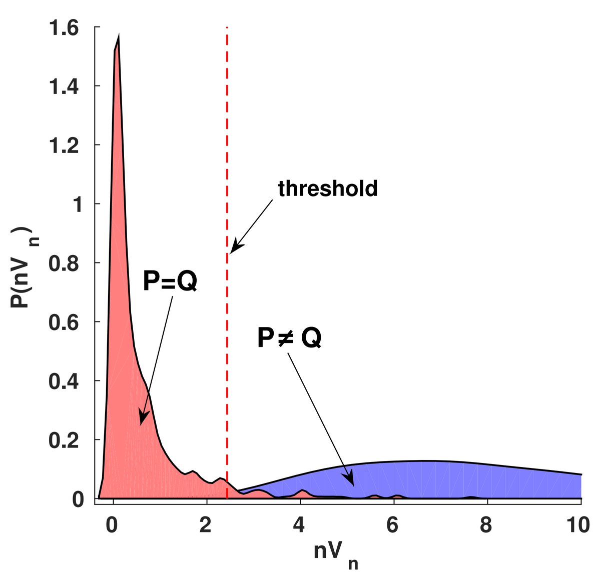


$$P \neq Q \Rightarrow \begin{bmatrix} K_{P,P} & K_{P,Q} \\ K_{Q,P} & K_{Q,Q} \end{bmatrix} =$$

$$P = Q \Rightarrow \begin{bmatrix} K_{P,P} & K_{P,Q} \\ K_{Q,P} & K_{Q,Q} \end{bmatrix} =$$



Quantifying Similarity



The V-statistics quantifies the concept of similarity.

$$V_n = \overline{K_{P,P}} + \overline{K_{Q,Q}} - 2\overline{K_{P,Q}}.$$

Explicitly

$$V_n = \frac{1}{n^2} \sum_{1 \le i,j \le n} k(X_i,X_j) + k(Y_i,Y_j) - k(X_i,Y_j) - k(X_j,Y_i).$$

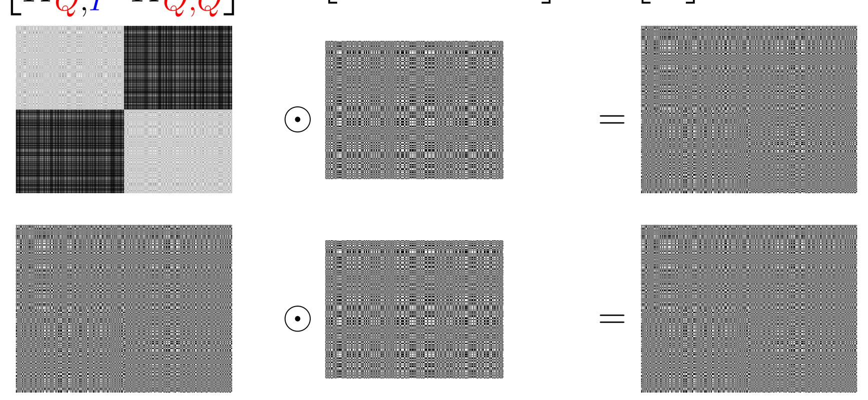
The degeneracy: if $P = Q$, then for $i \neq j$

$$\mathcal{E}\left[k(X_i,X_j) + k(Y_i,Y_j) - k(X_i,Y_j) - k(X_j,Y_i)\right] = 0.$$

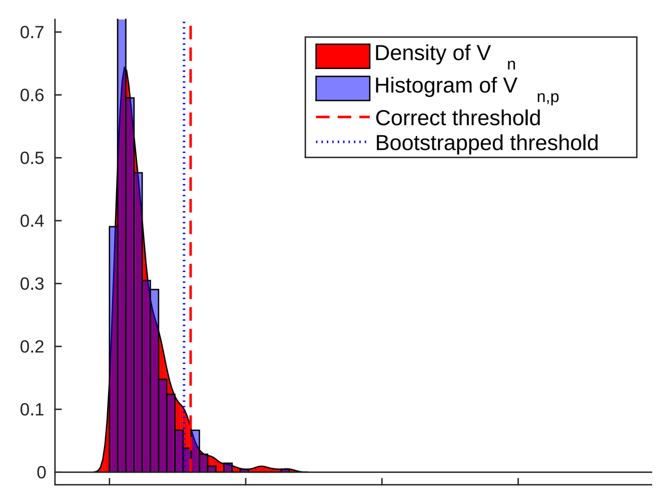
Estimation of V_n via Permutation

$$W = (-1,1,1, \quad \cdots \quad ,-1,-1,-1)$$

$$\begin{bmatrix} K_{P,P} & K_{P,Q} \\ K_{Q,P} & K_{Q,Q} \end{bmatrix} \odot [W^\top W] = [?]$$

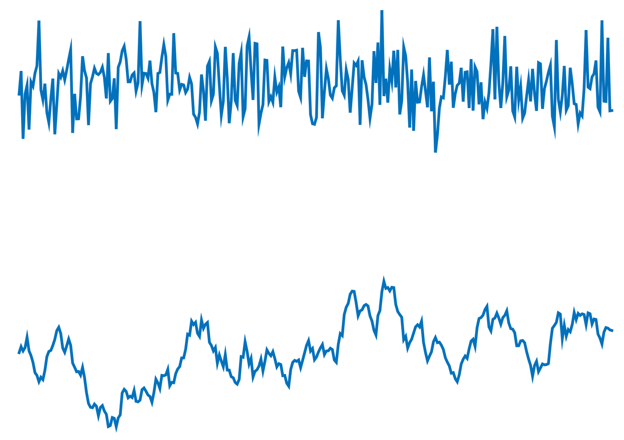


$$V_{n,p} = \overline{\begin{bmatrix} K_{P,P} & K_{P,Q} \\ K_{Q,P} & K_{Q,Q} \end{bmatrix} \odot [W^\top W]}$$



Permutation Tests for Random Processes

Memory of the Processes

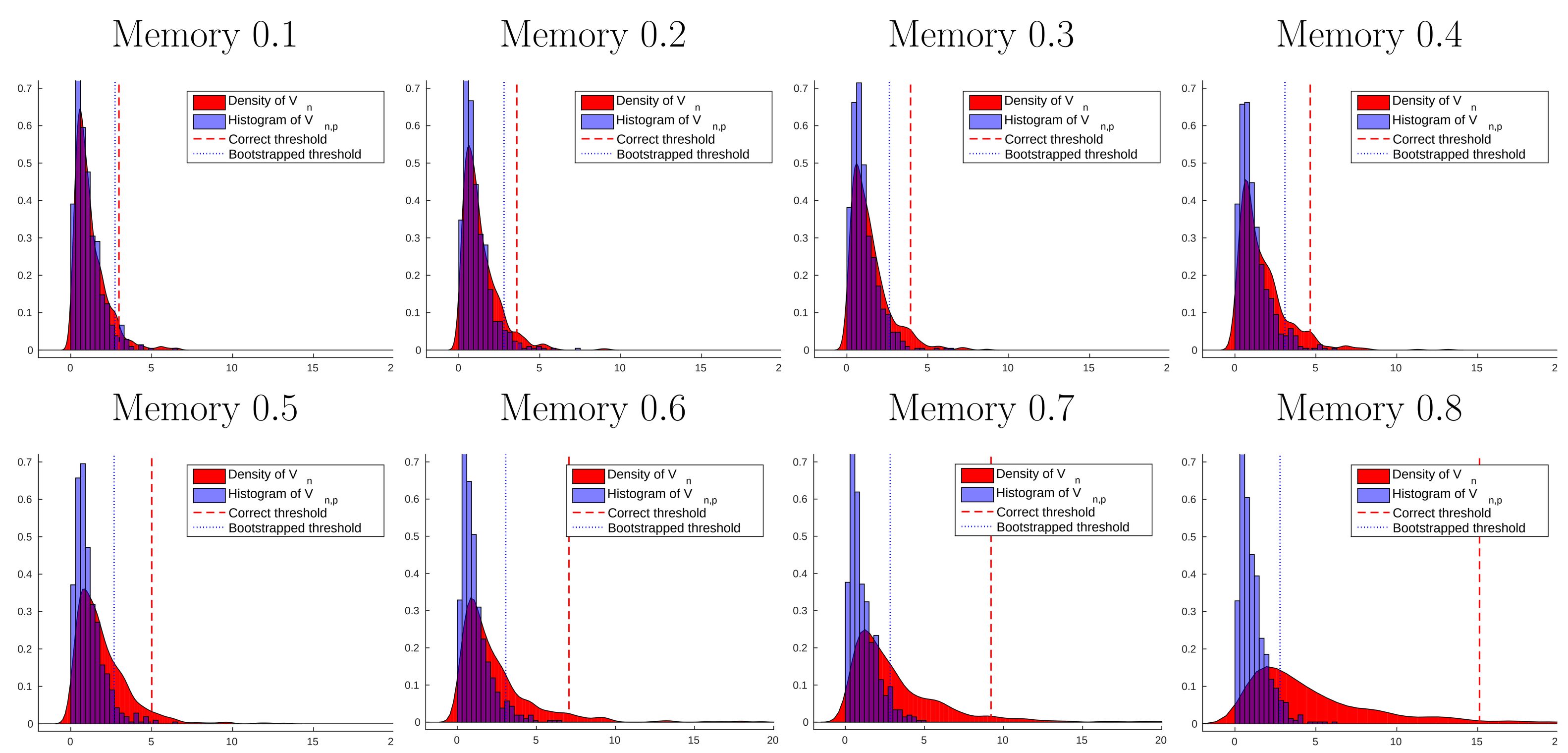


$$Q_t = \mathbf{0.14} Q_{t-1} + 0.98\epsilon_t$$

Processes with different memory

$$Z_t = \mathbf{0.97} Z_{t-1} + 0.22\epsilon_t.$$

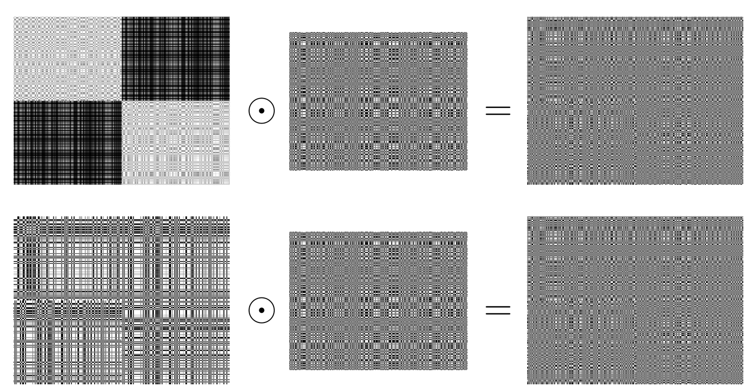
Estimation of V_n via Permutation Fails



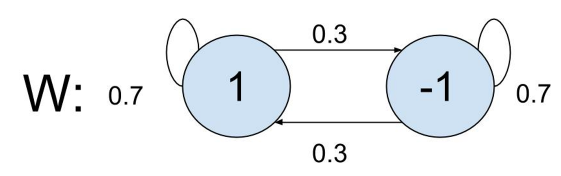
The Reason why Permutation Test Fails

$$W = (-1, 1 \cdots, -1, -1, -1)$$

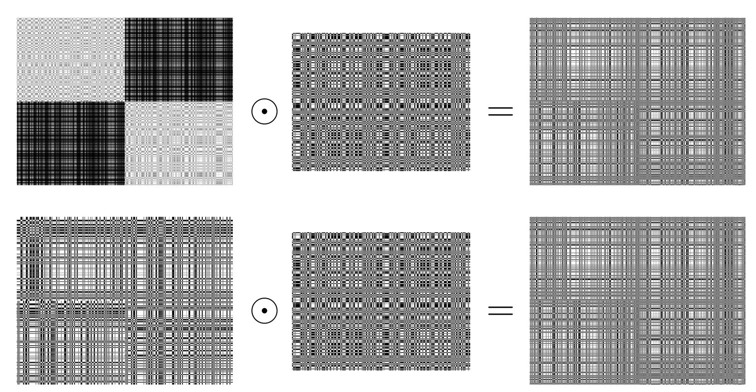
$$\begin{bmatrix} K_{P,P} & K_{P,Q} \\ K_{Q,P} & K_{Q,Q} \end{bmatrix} \odot [W^\top W] = [?]$$



Wild Bootstrap for Random Processes



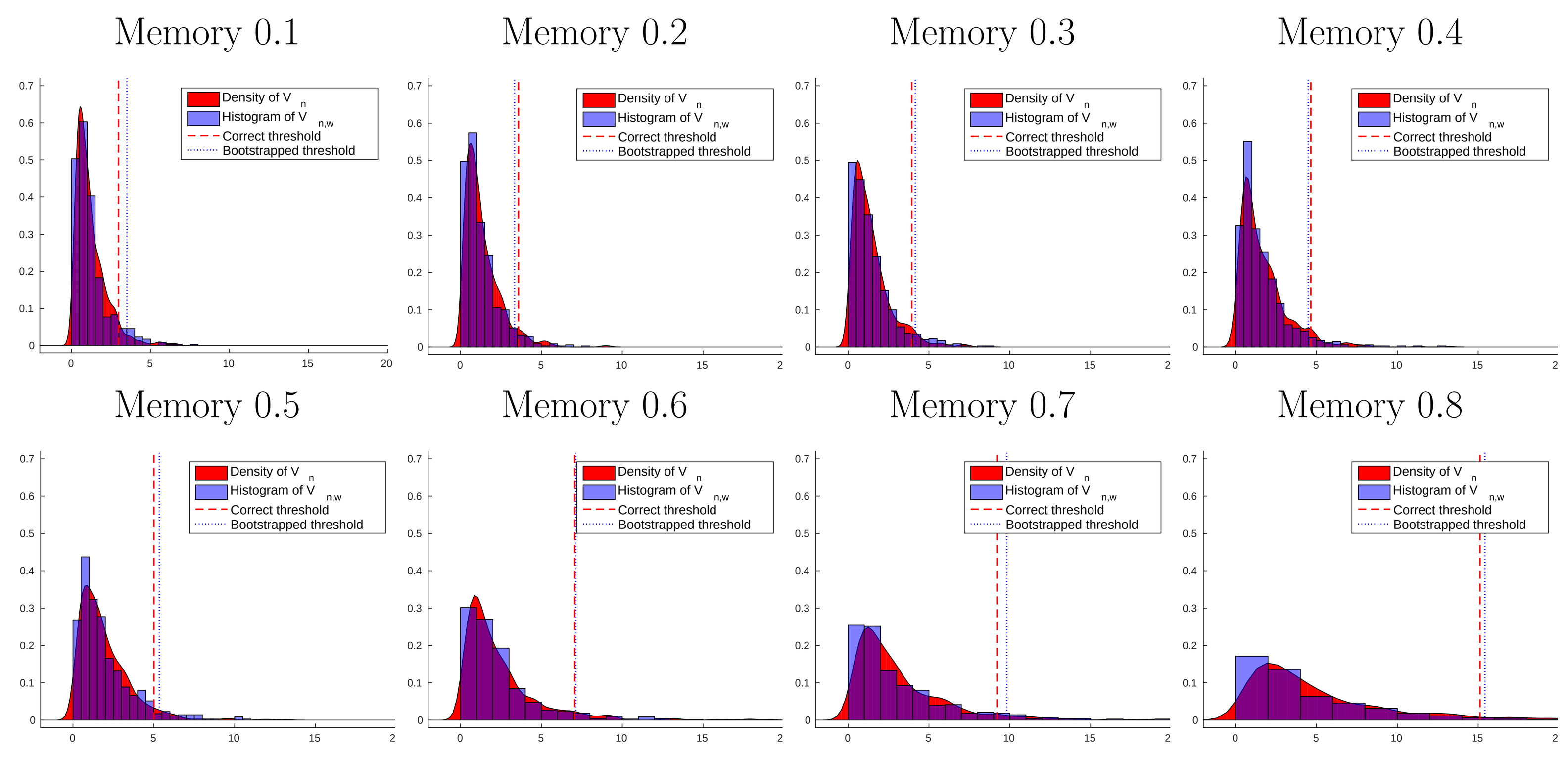
$$\begin{bmatrix} K_{P,P} & K_{P,Q} \\ K_{Q,P} & K_{Q,Q} \end{bmatrix} \odot \begin{bmatrix} W^\top W & -W^\top W \\ -W^\top W & W^\top W \end{bmatrix} = [?]$$



Estimation of V_n via Wild Bootstrap

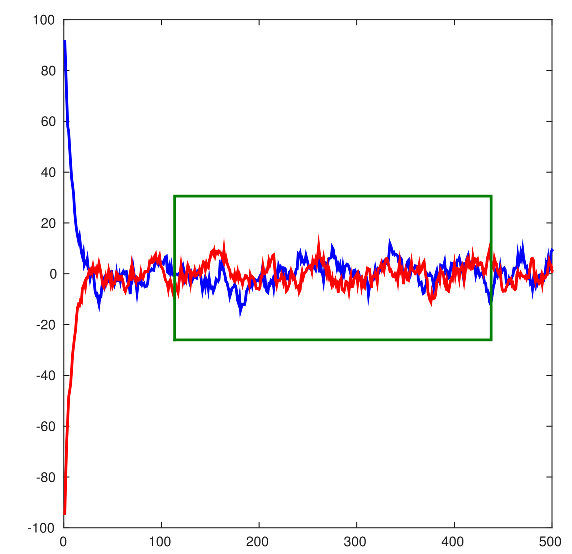
$$V_{n,w} = \overline{\begin{bmatrix} K_{P,P} & K_{P,Q} \\ K_{Q,P} & K_{Q,Q} \end{bmatrix} \odot \begin{bmatrix} W^\top W & -W^\top W \\ -W^\top W & W^\top W \end{bmatrix}}$$

Estimation of V_n via Permutation



Experiments

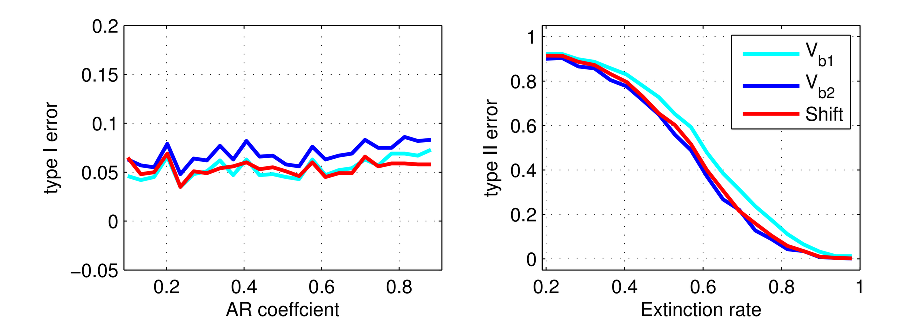
MCMC M.D.



Indie Pop Group Predicts the Volume of the Dow Jones!

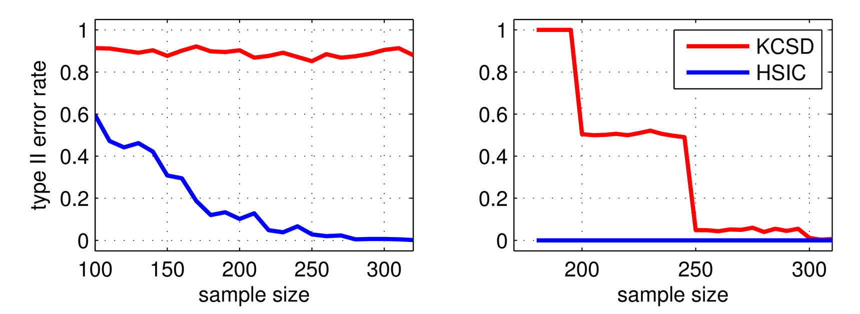
Test - MMD	Samples are different
Permutation	68 %
Wild Bootstrap	6 %

Test	p-value	Dependent
Permutation HSIC	0.003	Yes !
Wild Bootstrap HSIC	0.231	No



Comparison of Shift-HSIC [2] and tests based on the wild bootstrap. The left panel shows the performance under the null hypothesis, where a larger AR coefficient implies a stronger temporal dependence. The right panel show the performance under the alternative hypothesis, where a larger extinction rate implies a greater dependence between processes.

The Kernel Cross-Spectral Density [1] test is, to our knowledge, the only test procedure to reject the null hypothesis if there exist t,t' such that X_t and $Y_{t'}$ are dependent. In the experiments, we compare lag-HSIC with KCSD on two kinds of processes: one inspired by econometrics and one from [1]. In both panels Type II error is plotted.



References

[1] M. Besserve, N.K. Logothetis, and B. Schlkopf. Statistical analysis of coupled time series with kernel cross-spectral density operators. In *NIPS*, pages 2535–2543. 2013.

[2] K. Chwialkowski and A. Gretton. A kernel independence test for random processes. In *ICML*, 2014.