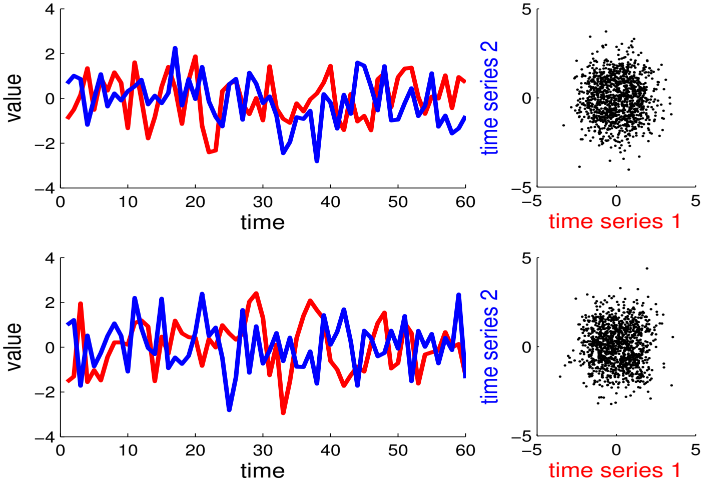
A Kernel Independence Test for Random Processes.

Kacper Chwialkowski, Arthur Gretton

June 14, 2014

The Problem

In which plot are the time series 1 and the time series 2 independent ?



Hilbert Schmidt Independence Criterion

$Z_t = (X_t , Y_t)$

Stationary processes X_t and Y_t together form a two-dimensional process.

Let k and m be

- characteristic,
- continuous,
- translation invariant,

kernels, vanishing at infinity.

HSIC function

$$h(z_a, z_b, z_c, z_d) = k(x_a, x_b)[m(y_a, y_b) + m(y_c, y_d) - 2m(y_b, y_c)].$$

The degenerate V-statistic

Theorem

For a stationary process $Z_t = (X_t, Y_t)$

$$X_t \perp\!\!\!\perp Y_t \Leftrightarrow \underset{t_1,t_2,t_3,t_4}{\forall} \mathcal{E}^* h(Z_{t_1}, Z_{t_2}, Z_{t_3}, Z_{t_4}) = 0,$$

where $\mathcal{E}^*$ is the expected value with respect to marginal distribution of $Z_{t_1}, Z_{t_2}, Z_{t_3}, Z_{t_4}$.

An empirical integral

$$V_n = \frac{1}{n^4} \sum_{1 \le t_1,t_2,t_3,t_4 \le n} h(Z_{t_1}, Z_{t_2}, Z_{t_3}, Z_{t_4}).$$

Result for observations with no temporal dependence

Theorem

If X_t and Y_t are independent, then, under some assumptions

$$\lim_{n\to\infty} n \cdot V_n \stackrel{D}{=} \sum_{j=1}^{\infty} \lambda_j \tau_j^2,$$

where $\| \lambda_j \|_1 \leq \infty$, $\lambda_j \geq 0$, $\{\tau_j\}_{j\in\mathbb{N}}$ *are i.i.d standard Gaussian variables.*

Result for processes

Theorem

If X_t and Y_t are independent, then, under some assumptions, for a stationary process $Z_t = (X_t, Y_t)$

$$\lim_{n\to\infty} n \cdot V_n \overset{D}{=} \sum_{j=1}^{\infty} \lambda_j \tau_j^2,$$

where $\| \lambda_j \|_1 \leq \infty$, $\lambda_j \geq 0$, $\{\tau_j\}_{j\in\mathbb{N}}$ are Gaussian, and

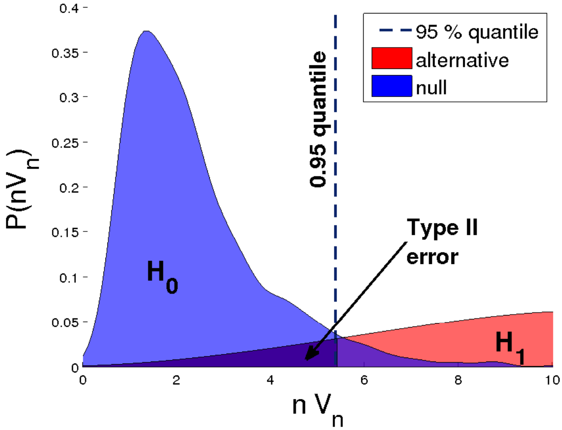
$$\mathcal{E}\tau_a\tau_b = \mathcal{E}e_a(Z_1)e_b(Z_1) + \sum_{j=1}^{\infty} [\mathcal{E}e_a(Z_1)e_b(Z_{j+1}) + \mathcal{E}e_b(Z_1)e_a(Z_{j+1})] .$$

with $\{e_i\}_{i\in\mathbb{N}}$ determined by h and the distribution of Z_t.

HSIC test is consistent

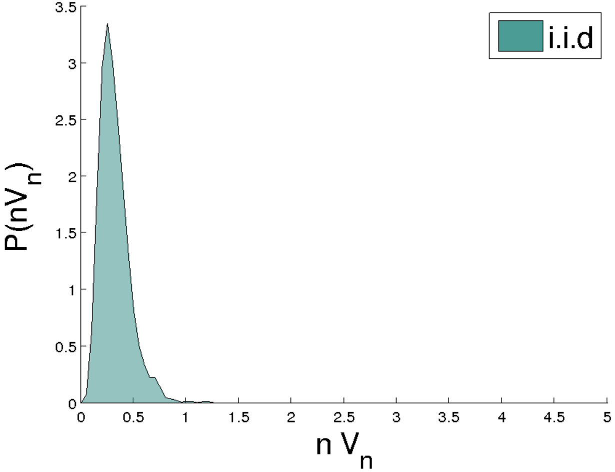
Theorem

If $X_t \not\perp\!\!\!\perp Y_t$ *then, under some assumptions* $\lim_{n\to\infty} V_n \overset{P}{=} q > 0$.



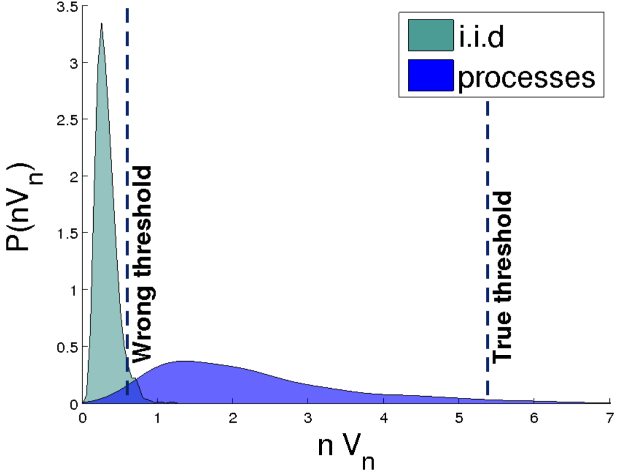
i.i.d. case, null scenario

X_t and Y_t are **i.i.d.** standard independent Gaussian

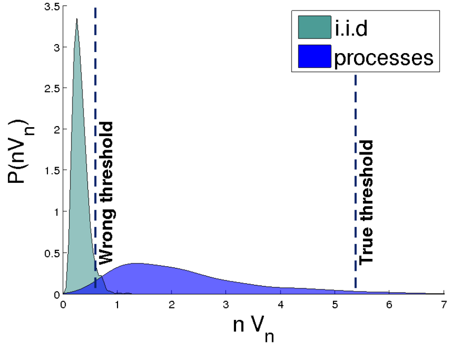


i.i.d vs. **processes**, null scenario

$$X_t = e^{-0.05}X_{t-1} + \sqrt{1 - e^{-0.1}}\epsilon_{1,t}, Y_t = e^{-0.05}Y_{t-1} + \sqrt{1 - e^{-0.1}}\epsilon_{2,t}$$



Resampling - bootstrap



Original time series:

X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10}
Y_1 Y_2 Y_3 Y_4 Y_5 Y_6 Y_7 Y_8 Y_9 Y_{10}

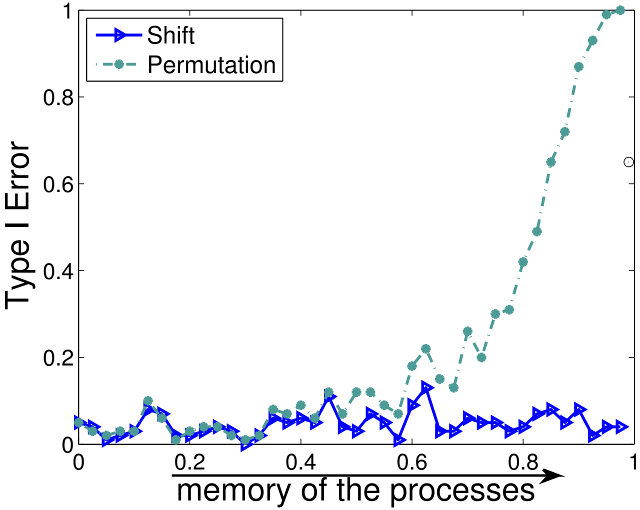
Permutation:

X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10}
Y_7 Y_3 Y_9 Y_2 Y_4 Y_8 Y_5 Y_1 Y_6 Y_{10}

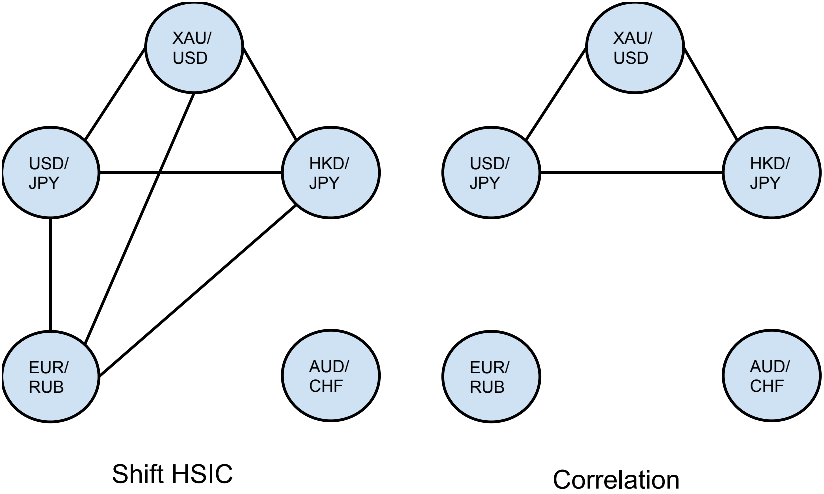
Shift:

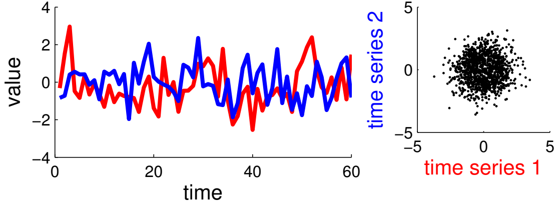
X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10}
Y_5 Y_6 Y_7 Y_8 Y_9 Y_{10} Y_1 Y_2 Y_3 Y_4

Memory vs Type I error

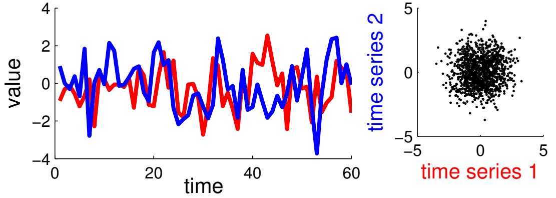


FOREX dependencies

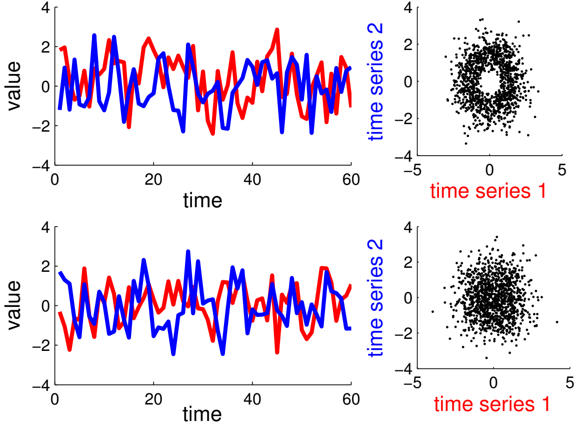




Dependent,
Type II Error
0%



Independent,
Type I Error
5%



Dependent,
Type II Error
0%

Independent,
Type I Error
5%