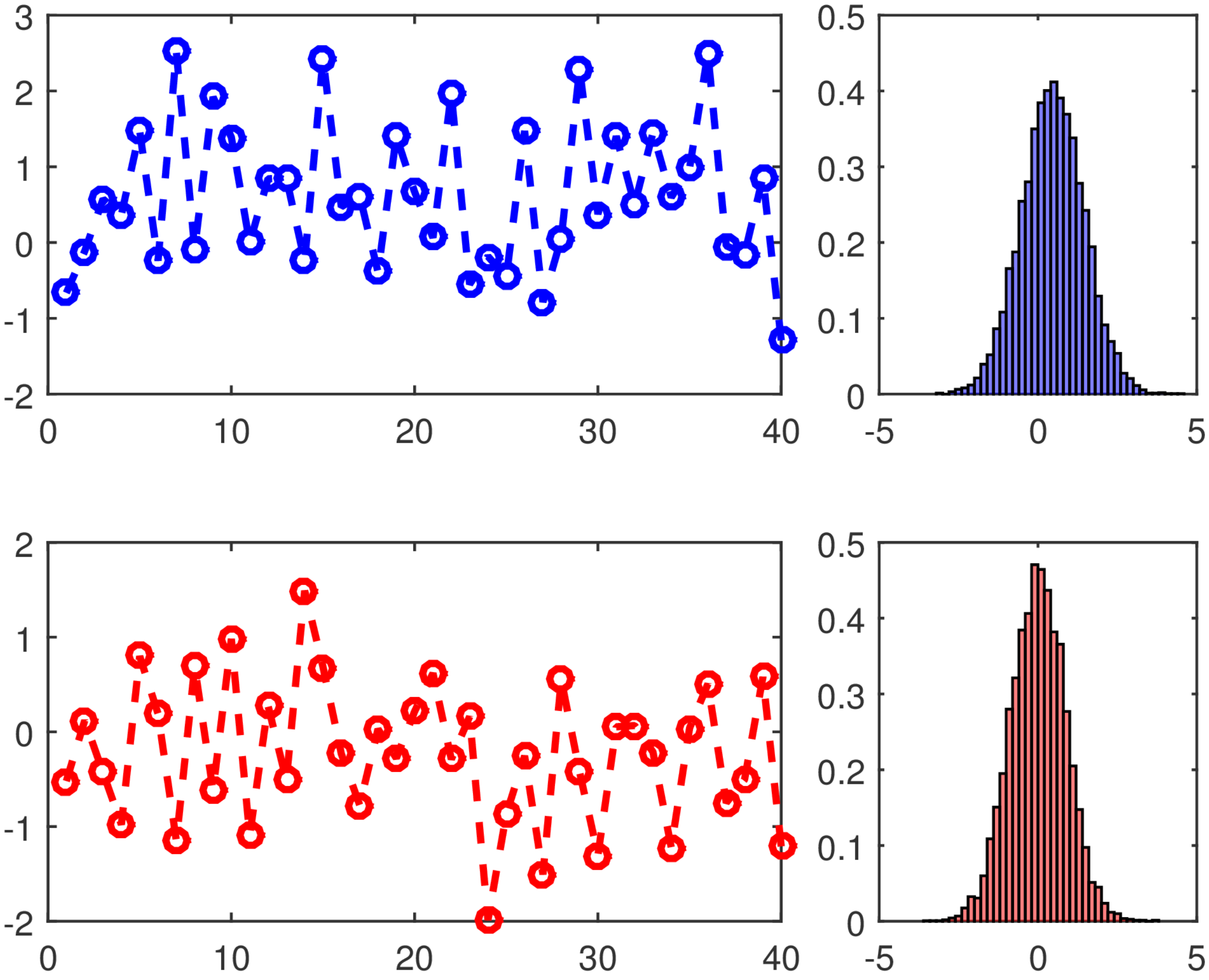
Kernel Statistical Tests for Random Processes

ISNPS, Avignon, 2016

Arthur Gretton

Gatsby Unit, UCL

Two-Sample Test for Random Processes



Is P the same distribution as Q?

Outline

Testing for differences in marginal distributions of random processes (MMD):

- Markov chain convergence diagnostics
- Change point detection

Testing for independence between random processes (HSIC)

- Dependency structure in financial markets
- Brain region activation

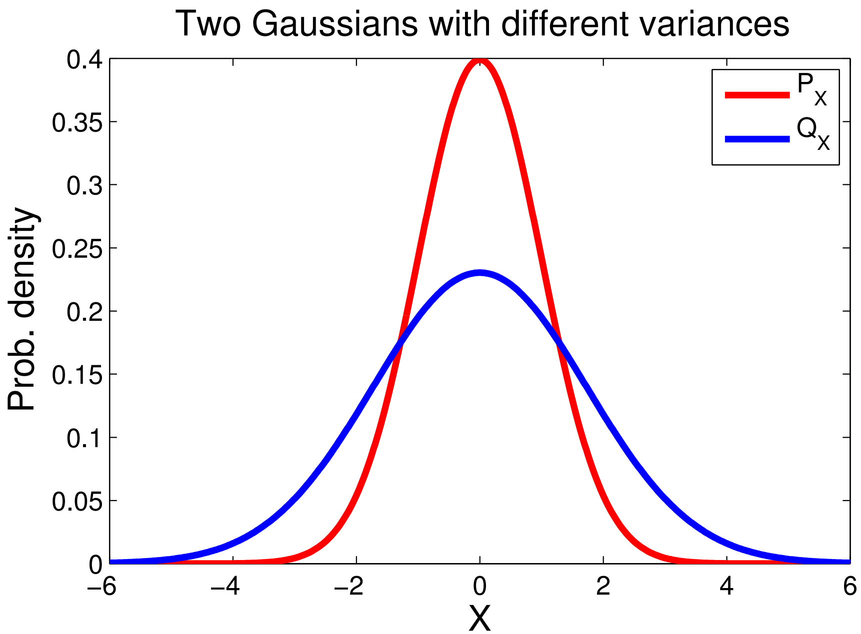
Why time series-based tests needed:

- Most real data (in the brain!) are time series
- MCMC diagnostics require tests on time series (or throwing out most of the data)

Maximum mean discrepancy, two-sample test

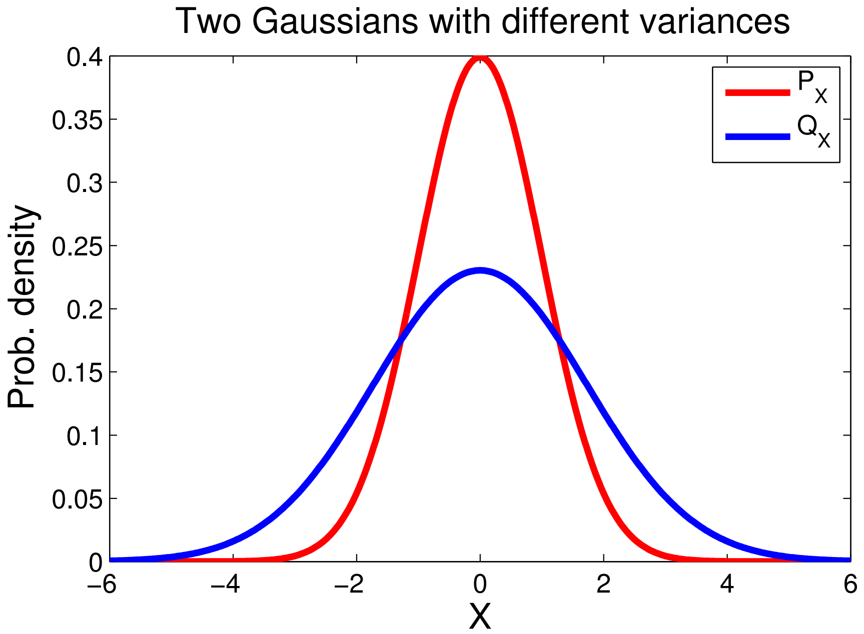
Feature mean difference

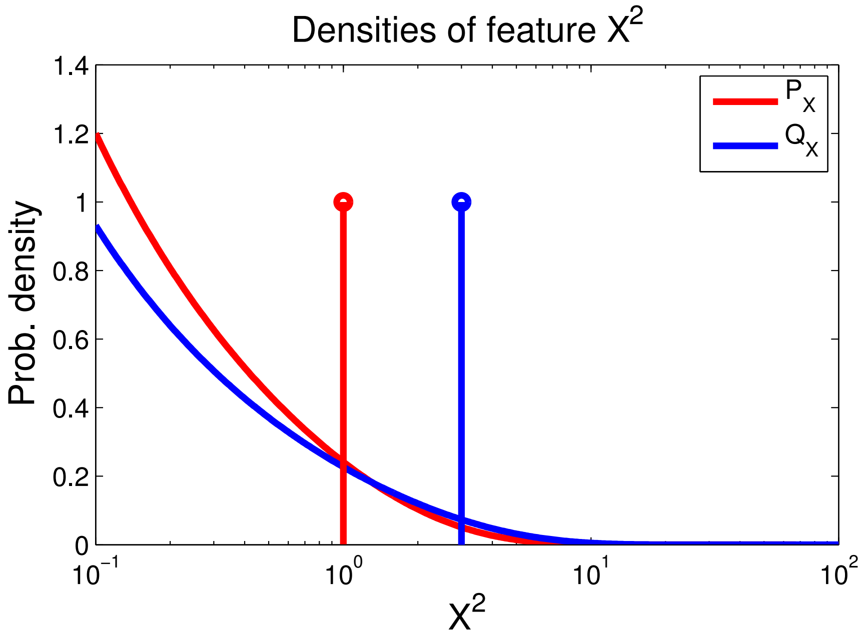
- Two Gaussians with same means, different variance
- Idea: look at difference in means of features of the RVs
- In Gaussian case: second order features of form $\varphi(x) = x^2$



Feature mean difference

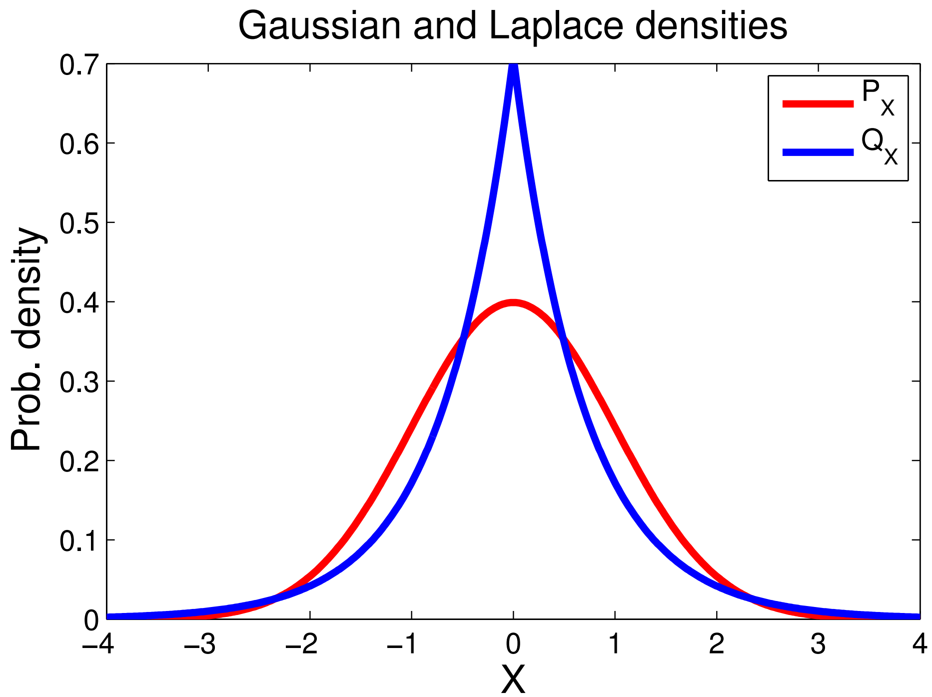
- Two Gaussians with same means, different variance
- Idea: look at difference in means of features of the RVs
- In Gaussian case: second order features of form $\varphi_x = x^2$





Feature mean difference

- Gaussian and Laplace distributions
- Same mean *and* same variance
- Difference in means using higher order features



. . . so let's explore feature representations!

Kernels: similarity between features

Kernel:

- We have two objects x and x' from a set $\mathcal{X}$ (documents, images, $\dots$).
  How similar are they?

Kernels: similarity between features

Kernel:

- We have two objects x and x' from a set $\mathcal{X}$ (documents, images, ...).
  How similar are they?
- Define features of objects:
  - φ_x are features of x,
  - $\varphi_{x'}$ are features of x'
- A kernel is the dot product between these features:

$$k(x, x') := \langle \varphi_x, \varphi_{x'} \rangle_{\mathcal{F}} \, .$$

Probabilities in feature space: the mean trick

The kernel trick

- Given $x \in \mathcal{X}$ for some set $\mathcal{X}$,
  define feature map $\varphi_x \in \mathcal{F}$,

$$\varphi_x = [\ldots e_i(x) \ldots]$$

- For kernel $k(x, x')$,

$$k(x, x') = \langle \varphi_x, \varphi_{x'} \rangle_{\mathcal{F}}$$

Probabilities in feature space: the mean trick

The kernel trick

- Given $x \in \mathcal{X}$ for some set $\mathcal{X}$, define feature map $\varphi_x \in \mathcal{F}$,

$$\varphi_x = [\ldots e_i(x) \ldots]$$

- For kernel $k(x, x')$,

$$k(x, x') = \langle \varphi_x, \varphi_{x'} \rangle_{\mathcal{F}}$$

The mean trick

- Given probability P define mean embedding $\mu_{\mathsf{P}} \in \mathcal{F}$

$$\mu_{\mathsf{P}} = [\ldots \mathbf{E}_{\mathsf{P}}[e_i(X)] \ldots]$$

- For kernel $k(x, x')$,

$$\langle \mu_{\mathsf{P}}, \mu_{\mathsf{Q}} \rangle_{\mathcal{F}} = \mathbf{E}_{\mathsf{P},\mathsf{Q}} k(X, Y)$$

  for $X \sim \mathsf{P}$ and $Y \sim \mathsf{Q}$.

Need to ensure Bochner integrability of φ_{x} for $\mathsf{x} \sim \mathsf{P}$: true for bounded kernels.

The maximum mean discrepancy

The maximum mean discrepancy is the distance between feature means:

$$
\begin{aligned}
MMD^2(\mathsf{P}, \mathsf{Q}) = \|\mu_{\mathsf{P}} - \mu_{\mathsf{Q}}\|_{\mathcal{F}}^2 &= \langle \mu_{\mathsf{P}}, \mu_{\mathsf{P}} \rangle_{\mathcal{F}} + \langle \mu_{\mathsf{Q}}, \mu_{\mathsf{Q}} \rangle_{\mathcal{F}} - 2 \langle \mu_{\mathsf{P}}, \mu_{\mathsf{Q}} \rangle_{\mathcal{F}} \\
&= \underbrace{\mathbf{E}_{\mathsf{P}} k(\mathsf{x}, \mathsf{x}')}_{\text{(a)}} + \underbrace{\mathbf{E}_{\mathsf{Q}} k(\mathsf{y}, \mathsf{y}')}_{\text{(a)}} - 2\underbrace{\mathbf{E}_{\mathsf{P},\mathsf{Q}} k(\mathsf{x}, \mathsf{y})}_{\text{(b)}}
\end{aligned}
$$

(a)= within distrib. similarity, (b)= cross-distrib. similarity

The maximum mean discrepancy

The maximum mean discrepancy is the distance between feature means:

$$MMD^2(\mathsf{P}, \mathsf{Q}) = \|\mu_{\mathsf{P}} - \mu_{\mathsf{Q}}\|^2_{\mathcal{F}} = \langle \mu_{\mathsf{P}}, \mu_{\mathsf{P}} \rangle_{\mathcal{F}} + \langle \mu_{\mathsf{Q}}, \mu_{\mathsf{Q}} \rangle_{\mathcal{F}} - 2 \langle \mu_{\mathsf{P}}, \mu_{\mathsf{Q}} \rangle_{\mathcal{F}}$$

$$= \underbrace{\mathbf{E}_{\mathsf{P}} k(\mathsf{x}, \mathsf{x}')}_{\text{(a)}} + \underbrace{\mathbf{E}_{\mathsf{Q}} k(\mathsf{y}, \mathsf{y}')}_{\text{(a)}} - 2\underbrace{\mathbf{E}_{\mathsf{P},\mathsf{Q}} k(\mathsf{x}, \mathsf{y})}_{\text{(b)}}$$

(a)= within distrib. similarity, (b)= cross-distrib. similarity

A biased empirical estimate (V-statistic):

$$\widehat{\text{MMD}}^2 = \frac{1}{n^2} \sum_{i,j}^{n} \left[k(x_i, x_j) - k(x_i, y_j) - k(y_i, x_j) + k(y_i, y_j) \right]$$

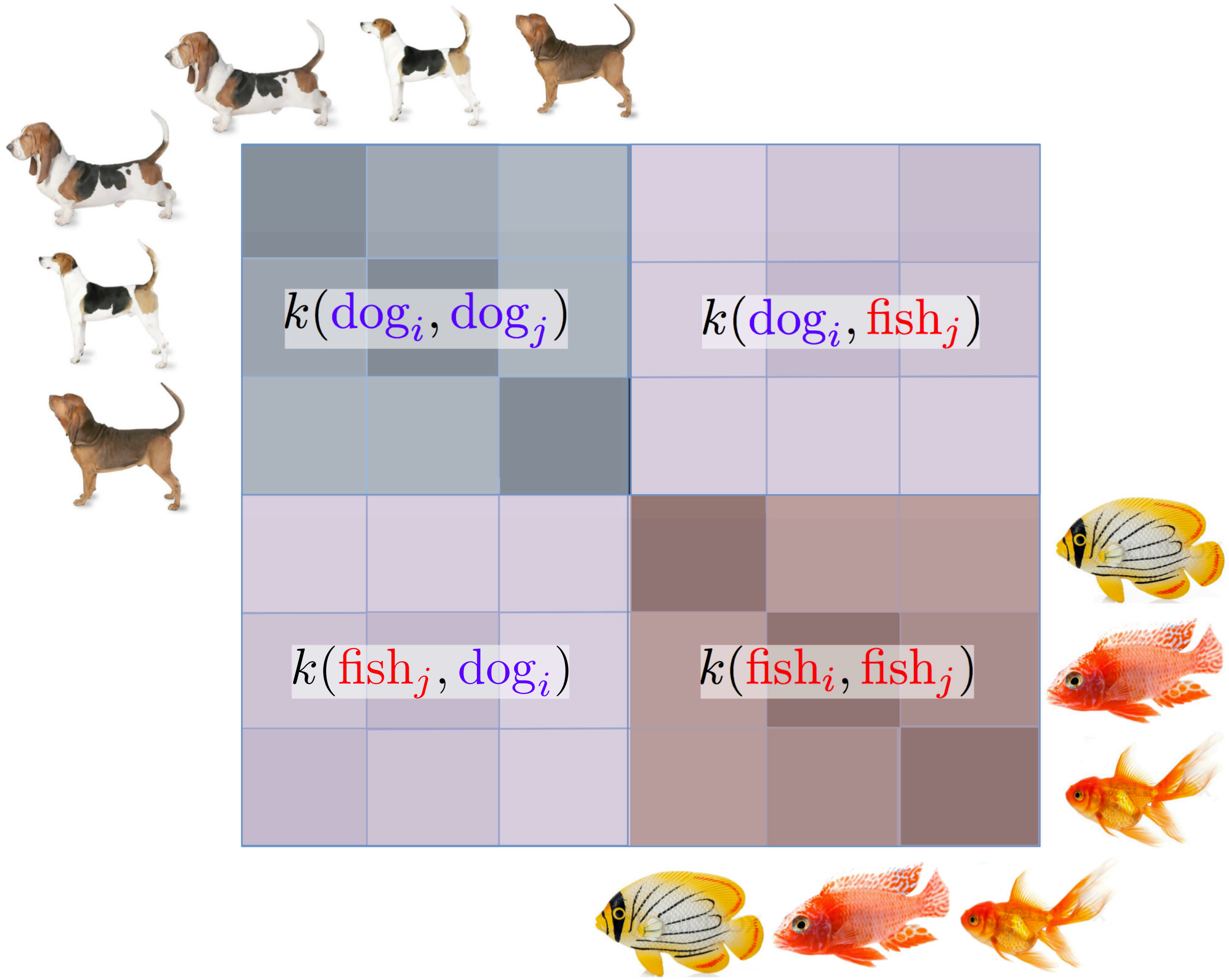
$$= \frac{1}{n^2} \sum_{i,j}^{n} \underbrace{\langle \varphi_{x_i} - \varphi_{y_i}, \varphi_{x_j} - \varphi_{y_j} \rangle_{\mathcal{F}}}_{\mathfrak{K}((x_i, y_i), (x_j, y_j))}$$

The maximum mean discrepancy

$$\sim P$$

$$\sim Q$$

The maximum mean discrepancy



$$\widehat{MMD}^2 = \overline{K_{P,P}} + \overline{K_{Q,Q}} - 2\overline{K_{P,Q}}$$

Statistical test using MMD

- Two hypotheses:
  - H_0: null hypothesis ($\mathbf{P} = \mathbf{Q}$)
  - H_1: alternative hypothesis ($\mathbf{P} \neq \mathbf{Q}$)
- Observe dependent samples $\boldsymbol{x} := \{x_1, \ldots, x_t, \ldots, x_n\}$ with marginal distribution $\mathbf{P}$, and $\boldsymbol{y} := \{y_1, \ldots, y_t, \ldots, y_n\}$ with marginal distribution $\mathbf{Q}$
- If empirical $\widehat{\mathrm{MMD}}^2$ is
  - "far from zero": reject H_0
  - "close to zero": accept H_0

Statistical test using MMD

- Two hypotheses:
  - H_0: null hypothesis ($\mathbf{P} = \mathbf{Q}$)
  - H_1: alternative hypothesis ($\mathbf{P} \neq \mathbf{Q}$)
- Observe dependent samples $\boldsymbol{x} := \{x_1, \ldots, x_t, \ldots, x_n\}$ with marginal distribution $\mathbf{P}$, and $\boldsymbol{y} := \{y_1, \ldots, y_t, \ldots, y_n\}$ with marginal distribution $\mathbf{Q}$
- If empirical $\widehat{\mathrm{MMD}}^2$ is
  - "far from zero": reject H_0
  - "close to zero": accept H_0
- Assumptions: $(X_t)_{t\in\mathbb{Z}}$ and $(Y_t)_{t\in\mathbb{Z}}$ are strictly stationary and τ-dependent with $\sum_{r=1}^{\infty} \sqrt{\tau(r)} < \infty$.

$$E \left\| X_r - \tilde{X}_r \right\|_1 < \tau(r),$$

where X_r is dependent on X_0, $\tilde{X}_r$ is a copy of X_r independent of X_0.

Statistical test using MMD

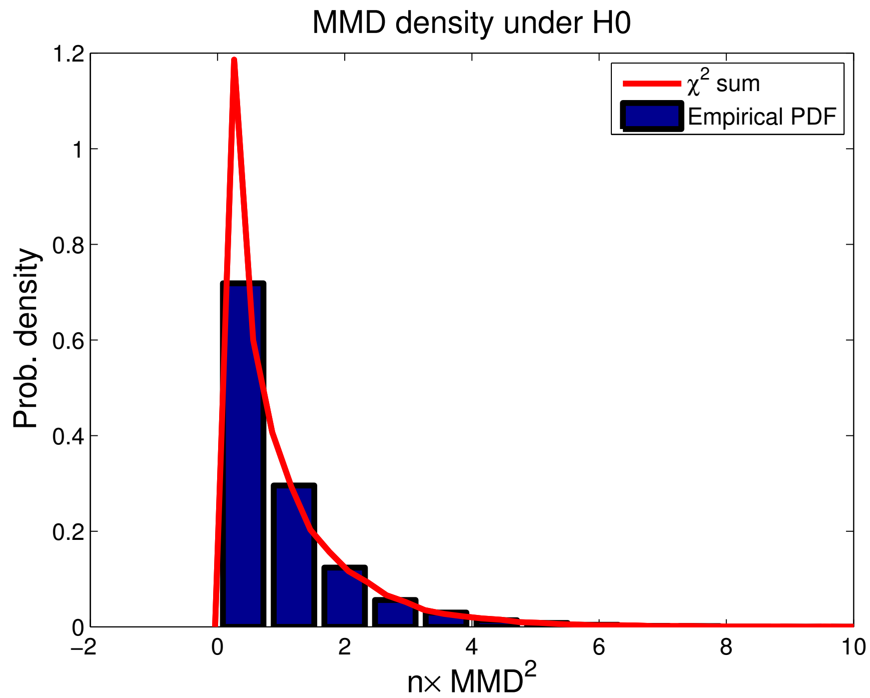
- "far from zero" vs "close to zero" - threshold?
- One answer: asymptotic distribution of $\widehat{\text{MMD}}^2$

When $\mathbf{P} = \mathbf{Q}$, asymptotic distribution is [Leucht and Neumann, 2013]

$$n\widehat{\text{MMD}}^2 \sim \sum_{l=1}^{\infty} \lambda_\ell Q_\ell^2 =: \mathcal{D}_{MMD}$$

- where
  - $\int_{\mathcal{X}} \mathfrak{K}(z, z')\psi_i(z) d\mathbf{P}_0(z) = \lambda_i \psi_i(z')$
  - $z := (x, y)$
  - $Q_\ell \sim \mathcal{N}(0,1)$ correlated, $\text{cov}(Q_u, Q_v)$

$$= \sum_{r=-\infty}^{\infty} \text{cov}\left[\psi_u(Z_0), \psi_v(Z_r)\right]$$



Asymptotics of $\widehat{MMD}^2$: proof idea

First define an order m truncation of $\mathfrak{K}(z, z')$:

$$\mathfrak{K}^{(m)}(z, z') = \sum_{\ell=1}^{m} \lambda_\ell \psi_\ell(z)\psi_\ell(z').$$

We can prove that as $m \to \infty$ the asymptotics of the truncation approach those of $\mathfrak{K}$.

The associated V-statistic is:

$$nV_n^{(m)} = \frac{1}{n}\sum_{s,t=1}^{n} \underbrace{\left(\sum_{\ell=1}^{m} \lambda_\ell \psi_\ell(Z_s)\psi_\ell(Z_t)\right)}_{\mathfrak{K}^{(m)}(z_s, z_t)}$$

$$= \sum_{\ell=1}^{m} \lambda_\ell \left(n^{-1/2}\sum_{t=1}^{n} \psi_\ell(Z_t)\right)^2$$

Asymptotics of $\widehat{MMD}^2$: proof idea

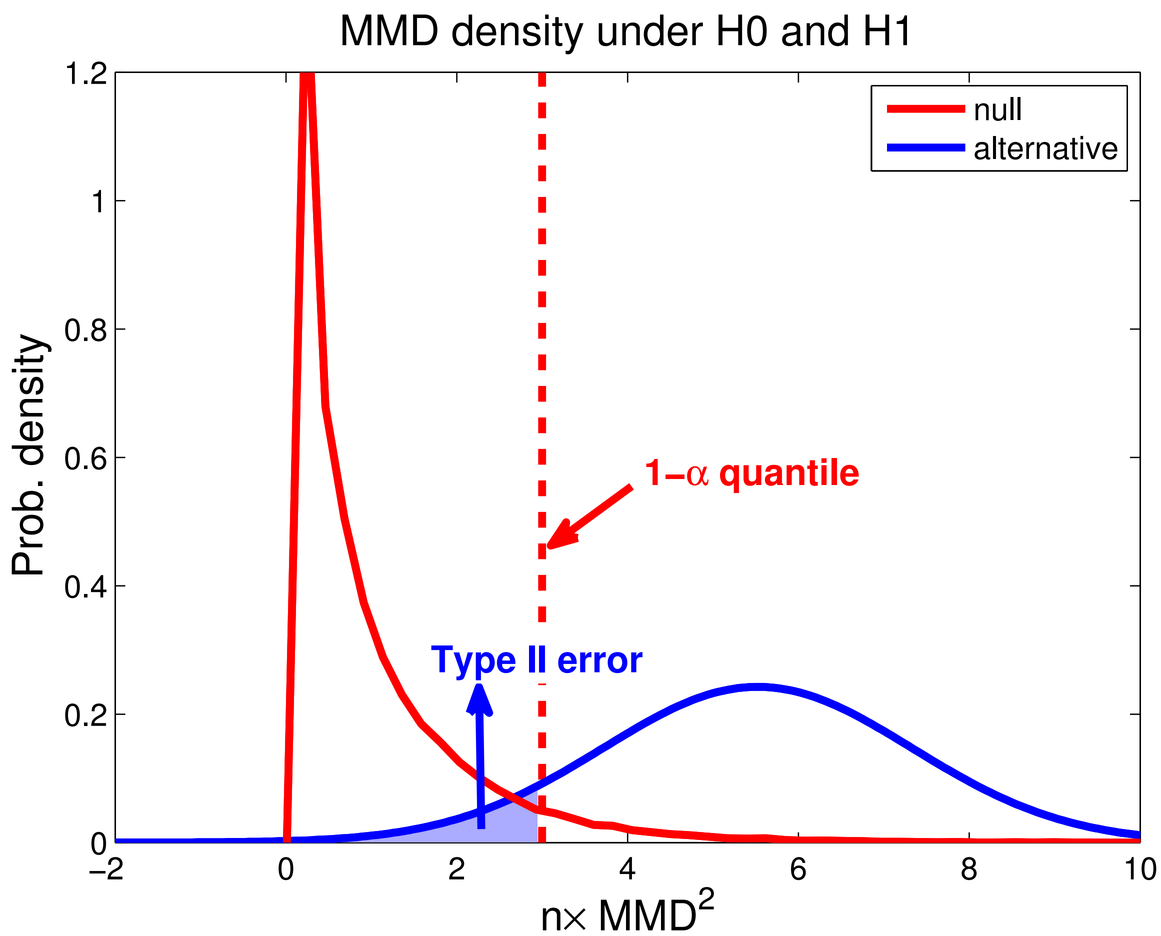
Under the assumptions on Z_t, we can apply a central limit theorem for weakly dependent random variables on the inner sum:

$$n^{-1/2}\sum_{t=1}^{n}\left[\begin{array}{ccc}\psi_1(Z_t) & \ldots & \psi_\ell(Z_t)\end{array}\right] \xrightarrow{d} \left[\begin{array}{ccc}Q_1 & \ldots & Q_\ell\end{array}\right]$$

Statistical test using MMD

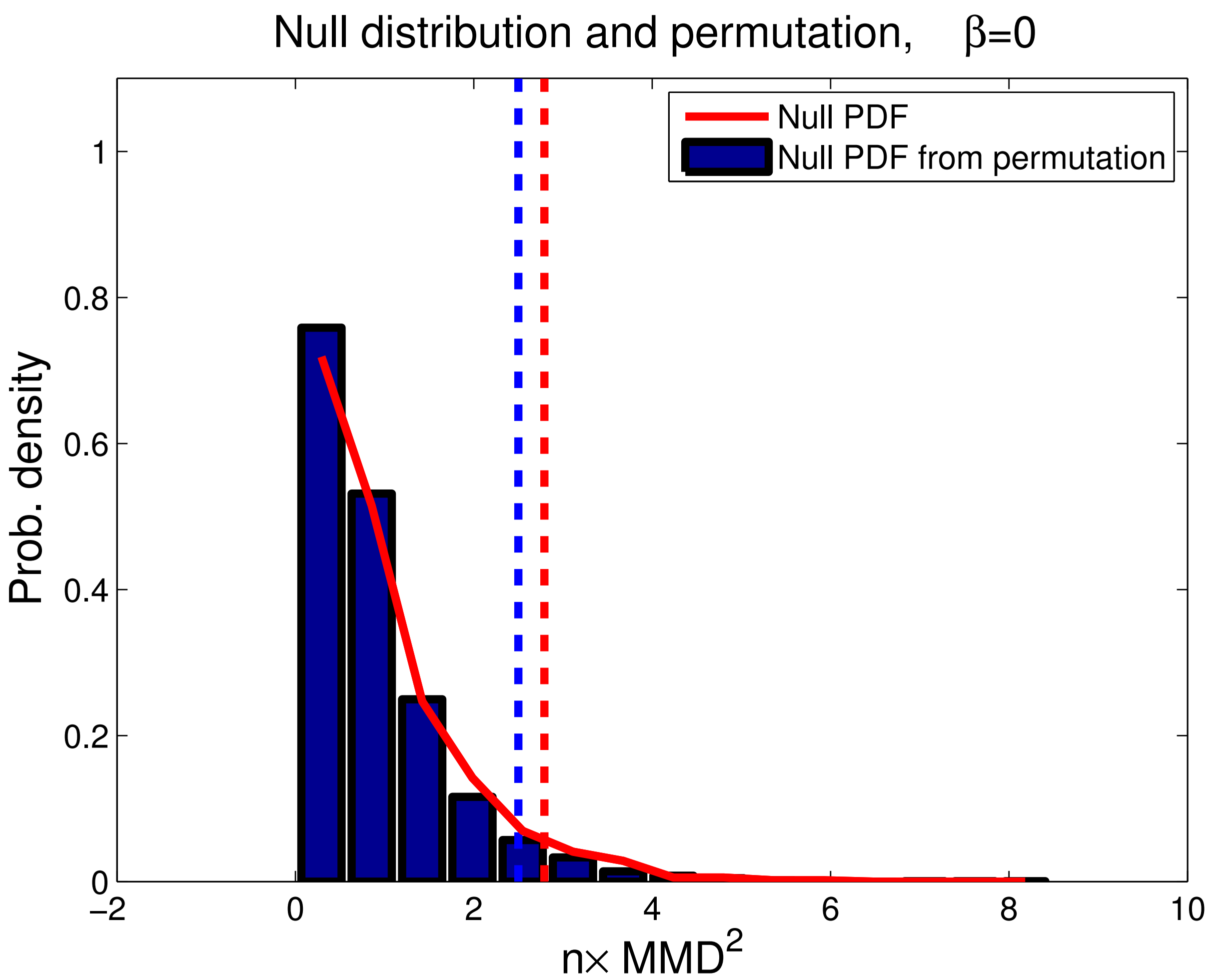
- Given **P** = **Q**, want threshold T such that $\mathbf{P}(\text{MMD} > T) \leq \alpha$

$$\widehat{MMD}^2 = \overline{K_{P,P}} + \overline{K_{Q,Q}} - 2\overline{K_{P,Q}}$$



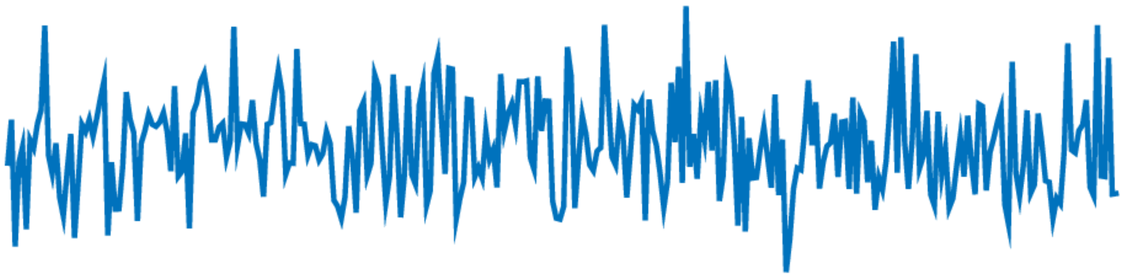
Statistical test using MMD

- Given $\mathbf{P} = \mathbf{Q}$, want threshold T such that $\mathbf{P}(\mathrm{MMD} > T) \leq 0.05$
- Permutation for empirical CDF [Arcones and Giné, 1992]



Memory of the Processes

$$X_t = \beta X_{t-1} + \epsilon_t \qquad \epsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

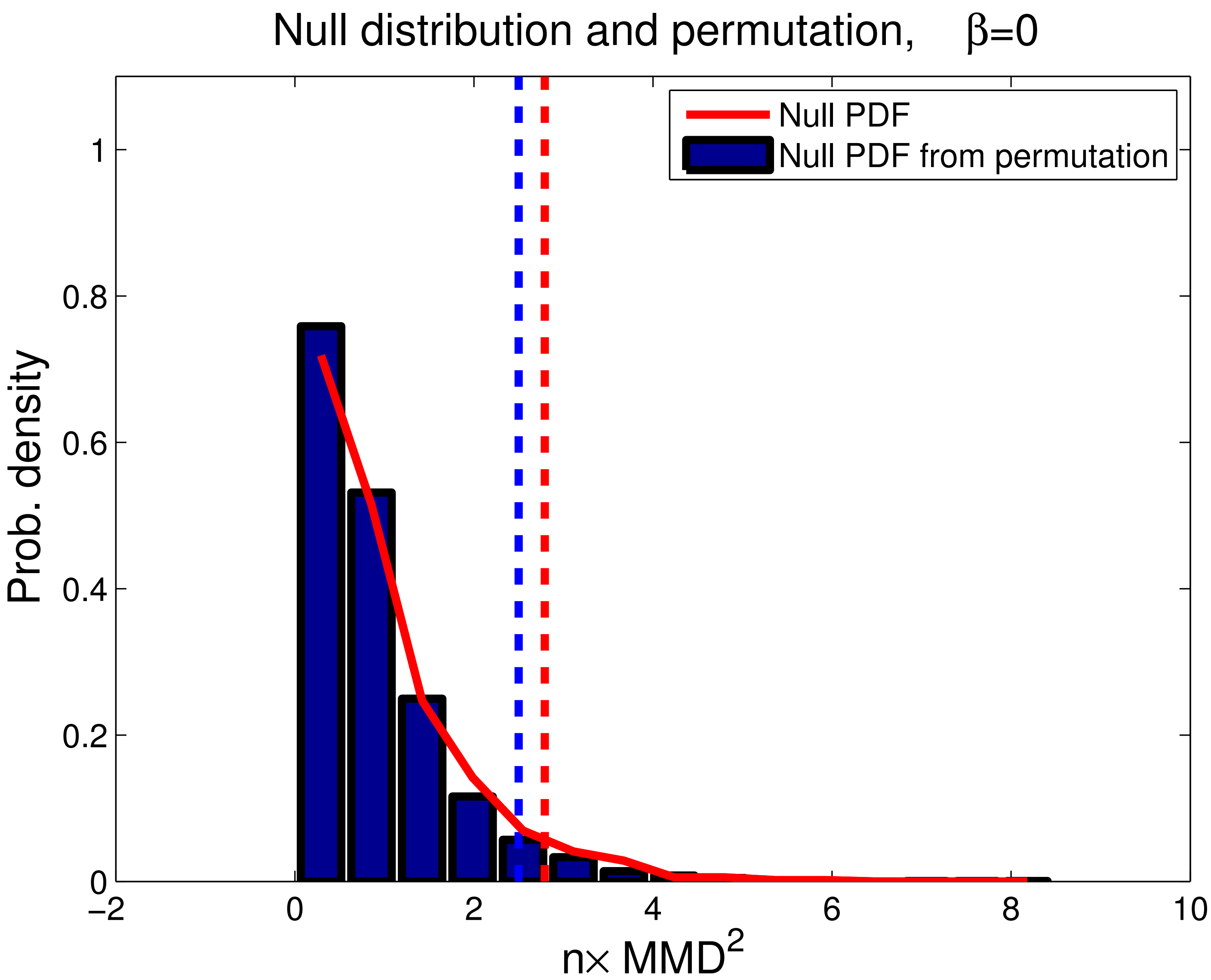


$\beta = 0.14$

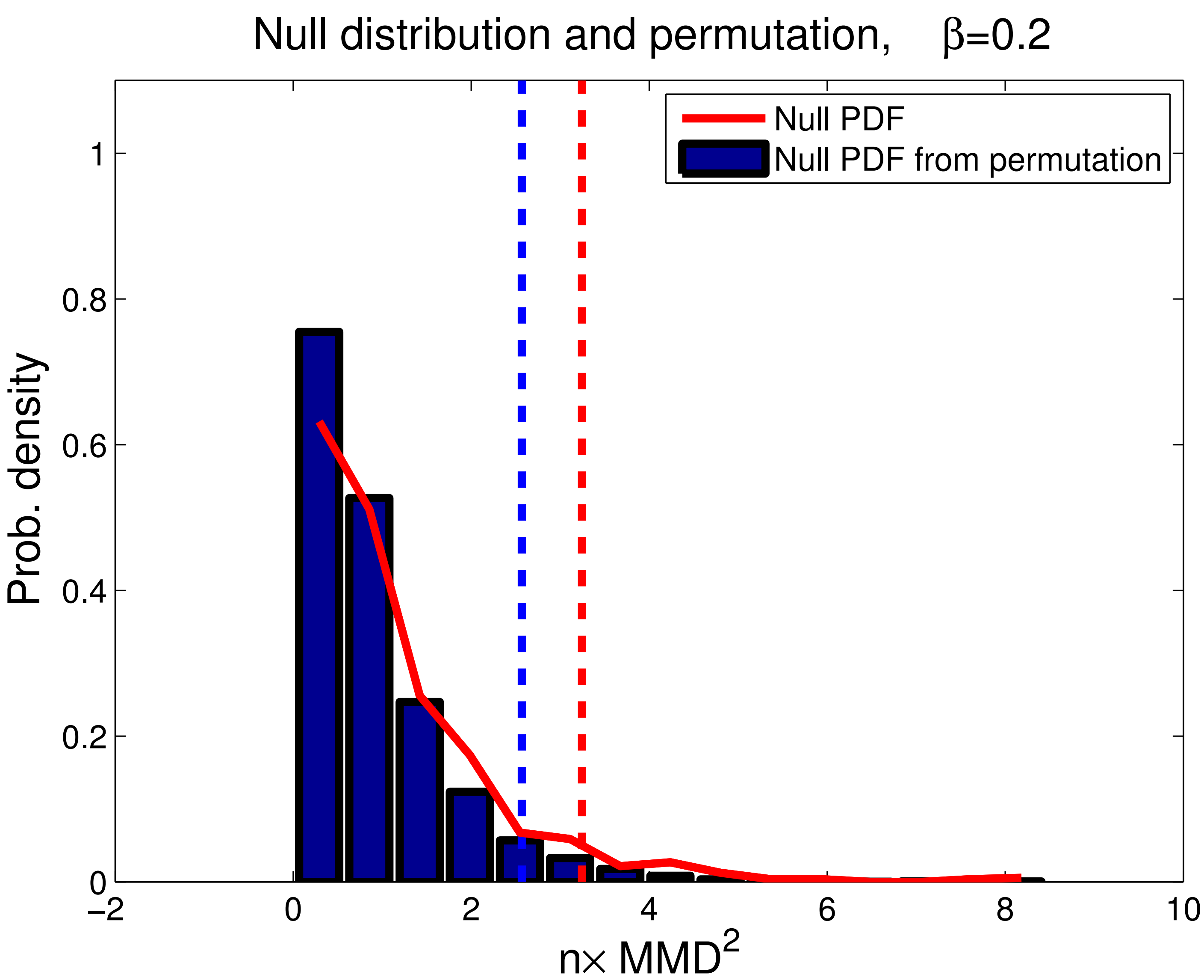
$\beta = 0.97$

The null distribution of the V-statistic is strongly affected by memory

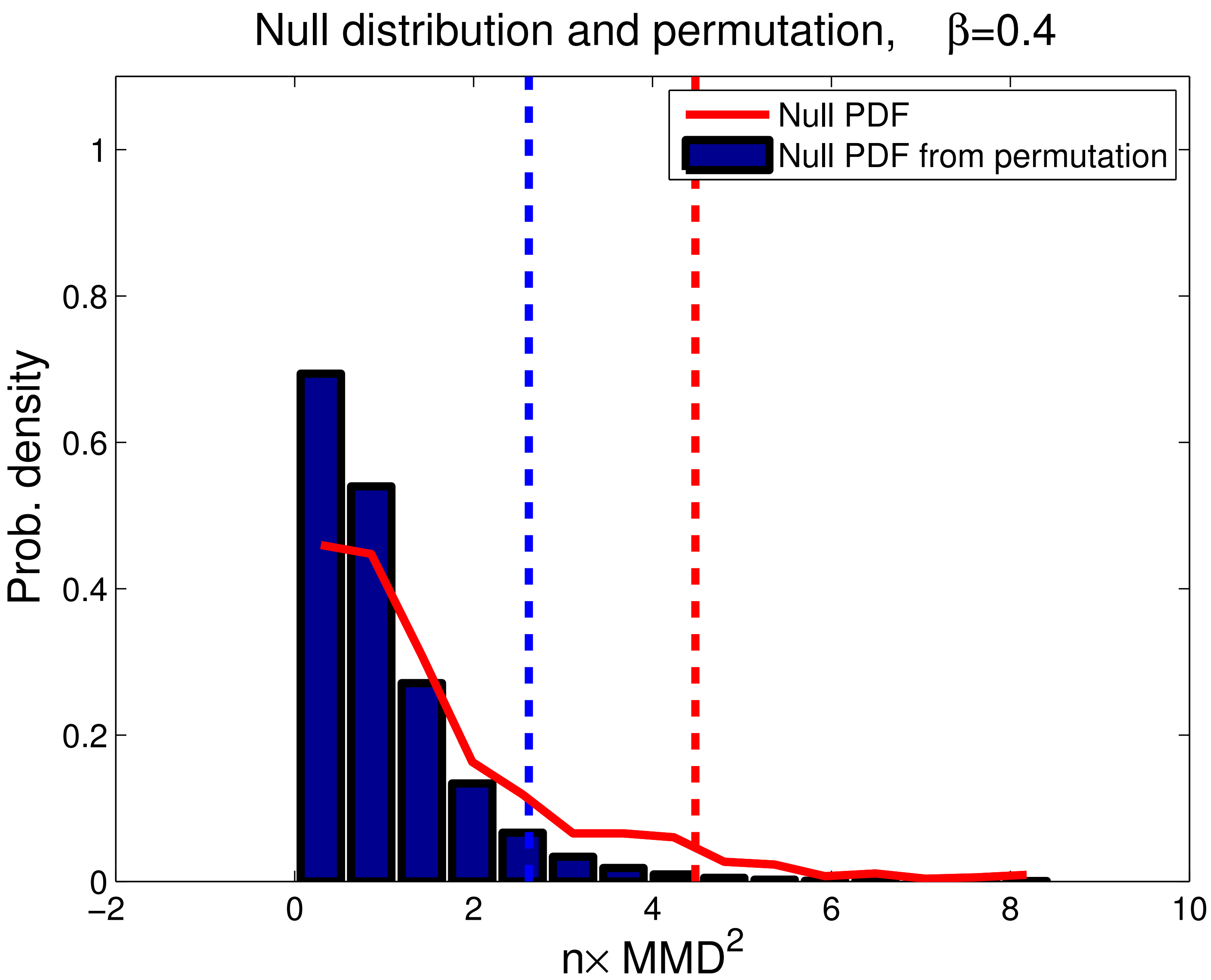
Memory $\beta = 0.0$, permutation for null



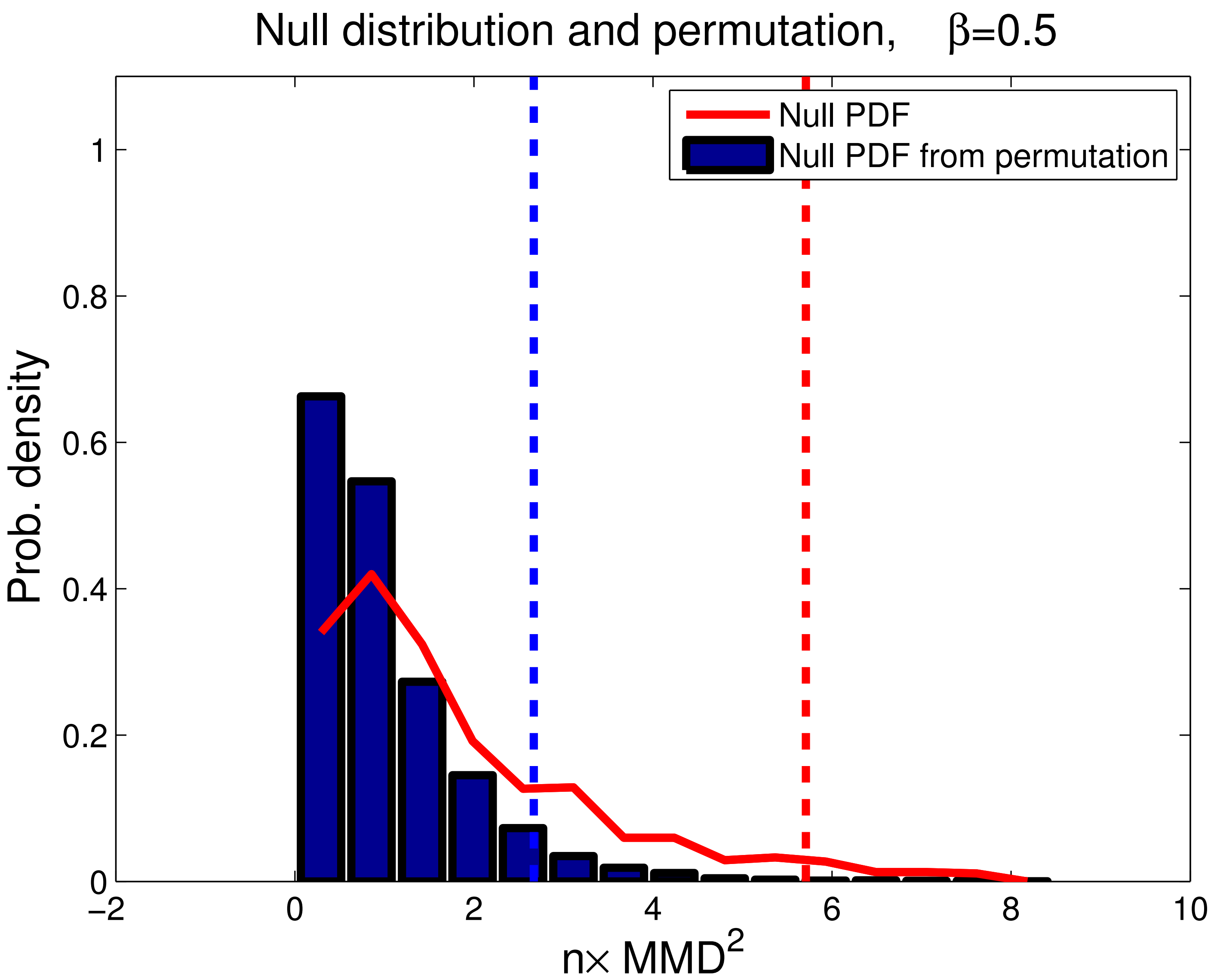
Memory $\beta = 0.2$, permutation for null



Memory $\beta = 0.4$, permutation for null



Memory $\beta = 0.5$, permutation for null



Wild bootstrap estimate of the asymptotic distribution

Define a new time series W_t^* with the property

$$\operatorname{cov}(W_s^*, W_t^*) = \rho\left(|s-t|/\ell_n\right),$$

where ℓ_n is a width parameter growing with n, and ρ is a window, e.g.

$$\operatorname{cov}(W_s^*, W_t^*) = \exp\left(-|s-t|/\ell_n\right).$$

X_t and Y_t τ-dependent with $\sum_{r=1}^{\infty} r^2 \sqrt{\tau(r)} < \infty$.

Wild bootstrap estimate of the asymptotic distribution

Define a new time series W_t^* with the property

$$\mathrm{cov}(W_s^*, W_t^*) = \rho\left(|s-t|/\ell_n\right),$$

where ℓ_n is a width parameter growing with n, and ρ is a window, e.g.

$$\mathrm{cov}(W_s^*, W_t^*) = \exp\left(-|s-t|/\ell_n\right).$$

X_t and Y_t τ-dependent with $\sum_{r=1}^{\infty} r^2\sqrt{\tau(r)} < \infty$.

Wild bootstrap estimate of the null:

$V_n^* := \frac{1}{n}\sum_{s,t=1}^{n} h((X_s,Y_s),(X_t,Y_t))W_s^*W_t^*$

As measured via Prokhorov metric d_p,

$$d_p\left(\mathcal{D}_{MMD}, \frac{1}{n}\sum_{s,t=1}^{n} \mathfrak{K}((X_s,Y_s),(X_t,Y_t))W_s^*W_t^*\right) \xrightarrow{p} 0 \quad \text{as} \quad n\to\infty.$$

How the proof works (1)

Again define a finite approximation,

$$V_n^{(m)*} = \frac{1}{n} \sum_{s,t=1}^{n} \mathfrak{K}^{(m)}(Z_s, Z_t) W_s^* W_t^*$$

$$= \sum_{k=1}^{m} \lambda_k \left(n^{-1/2} \sum_{t=1}^{n} \psi_k(Z_t) W_t^* \right)^2$$

which can be shown to converge as $m \to \infty$. Define

$$U_t^* := \left[\begin{array}{ccc} \psi_1(Z_t) & \ldots & \psi_m(Z_t) \end{array} \right] W_t^*$$

We need that in probability (as $n \to \infty$),

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t^* \xrightarrow{d} \mathcal{N}(0, \Sigma_m)$$

How the proof works (2)

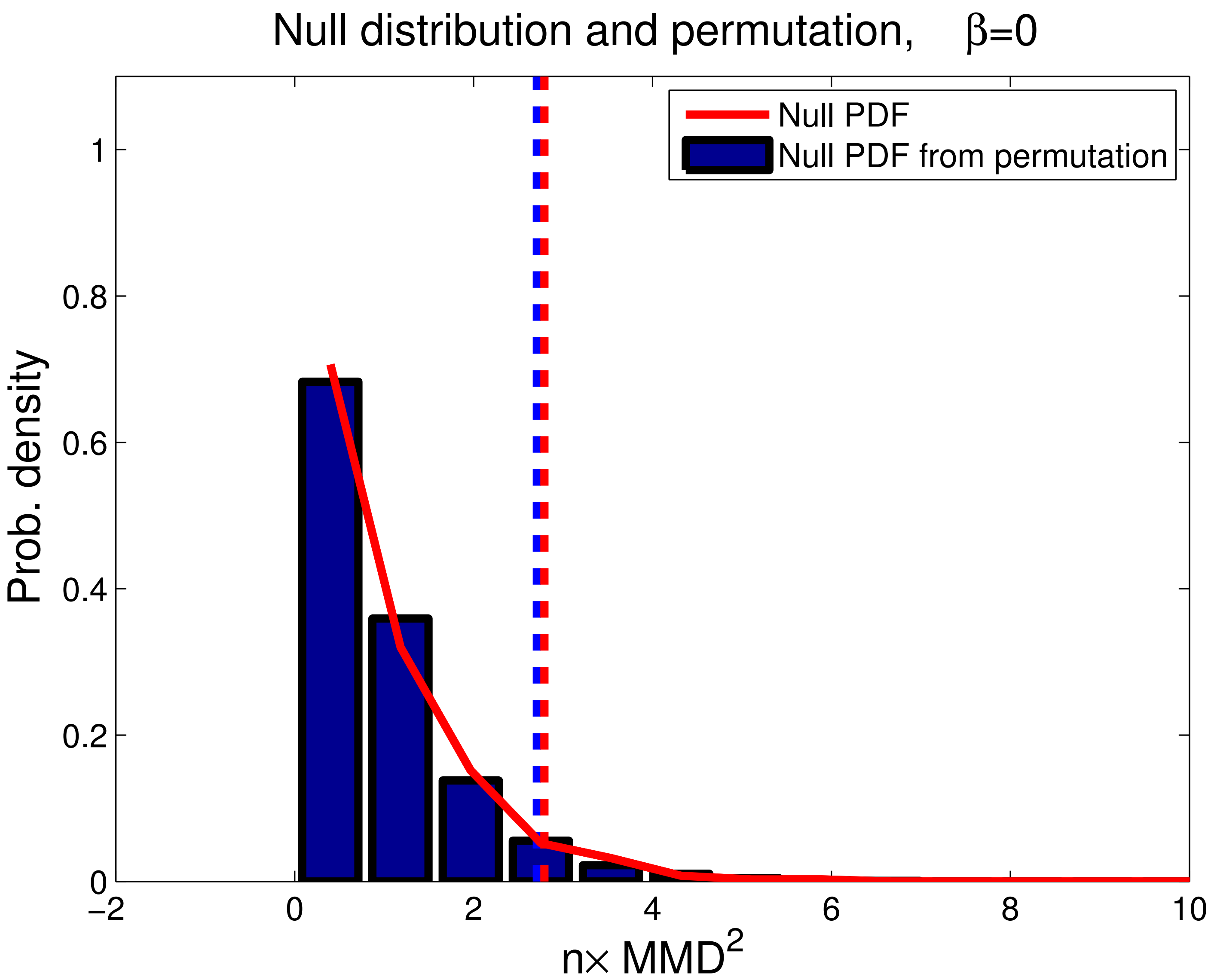
$$\operatorname{cov}\left(n^{-1/2}\sum_{s=1}^{n}\psi_j(Z_s)W_s^*, n^{-1/2}\sum_{t=1}^{n}\psi_k(Z_t)W_t^*\right)$$

$$= \frac{1}{n}\sum_{s,t=1}^{n}\psi_j(Z_s)\psi_k(Z_t)\rho(|s-t|\,\ell_n)$$

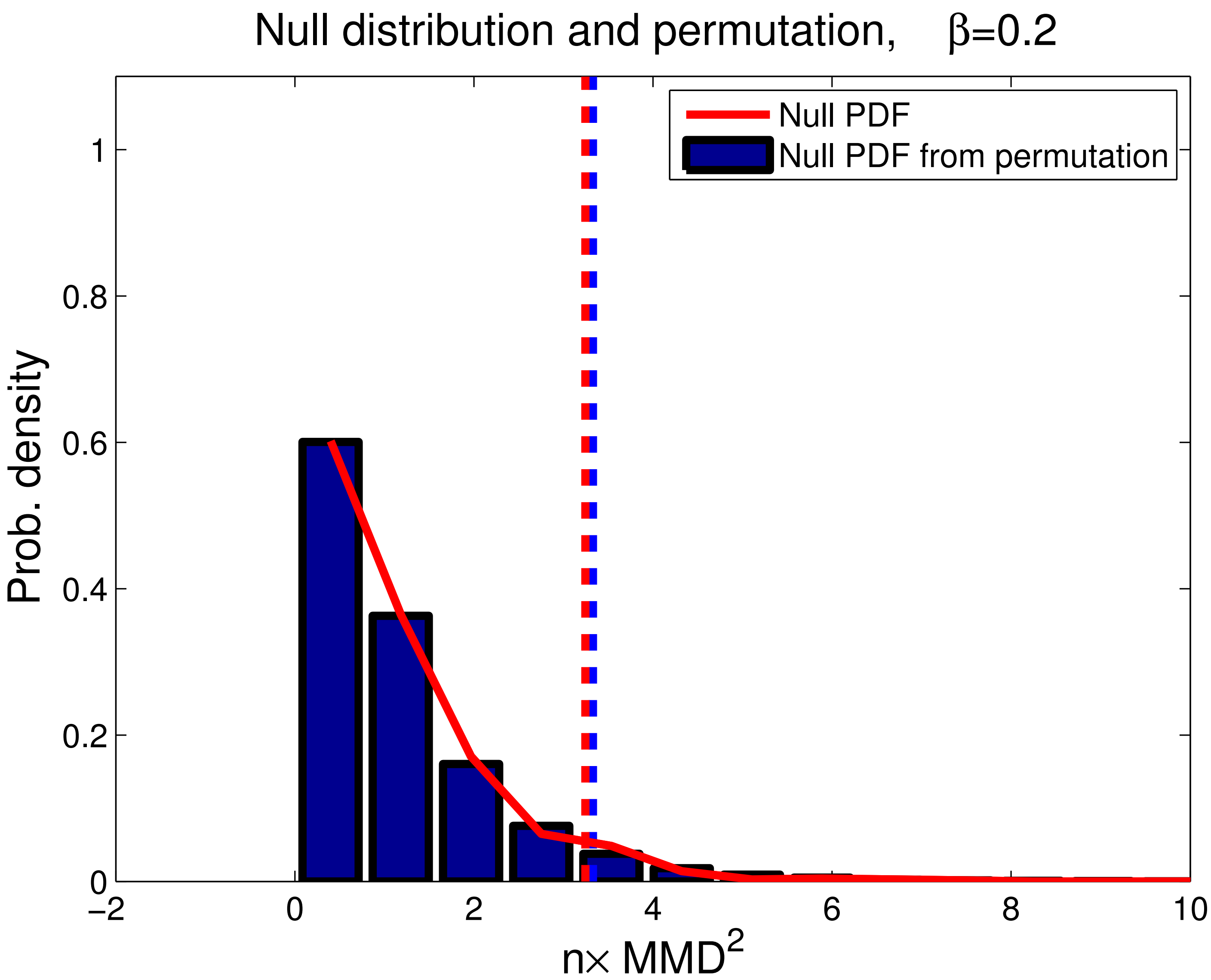
$$= \underbrace{\frac{1}{n}\sum_{s,t=1}^{n}\left(\psi_j(Z_s)\psi_k(Z_t)-E\left[\psi_j(Z_s)\psi_k(Z_t)\right]\right)\rho(|s-t|\,\ell_n)}_{\text{converges to } 0}$$

$$+\underbrace{\sum_{r=-\infty}^{\infty}E\left(\psi_j(Z_0)\psi_k(Z_r)\right)\rho(|r|/\ell_n)\max\left\{1-|r|/n,0\right\}}_{\text{converges to } (\Sigma_m)_{j,k}}$$

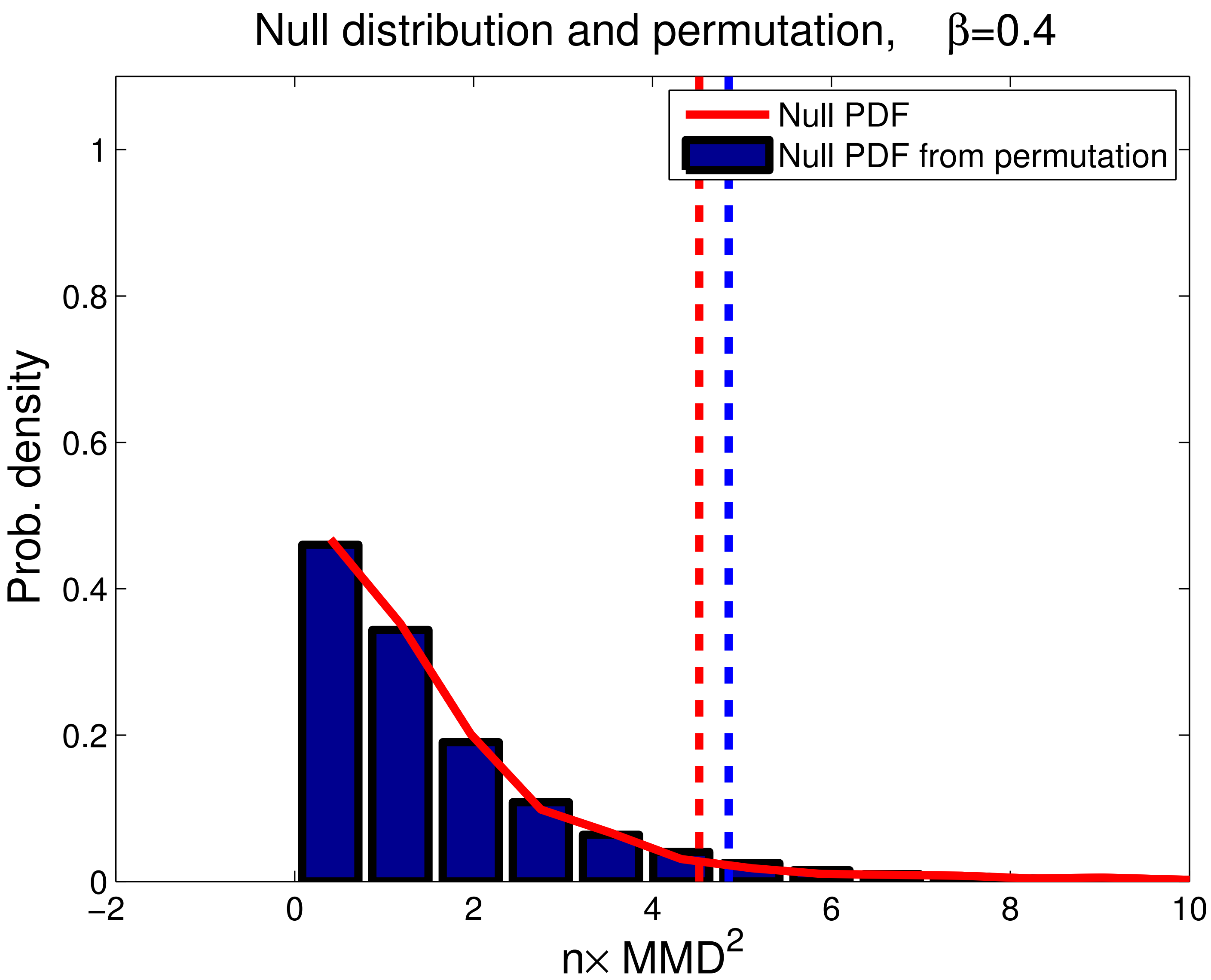
Memory $\beta = 0.0$, wild bootstrap for null



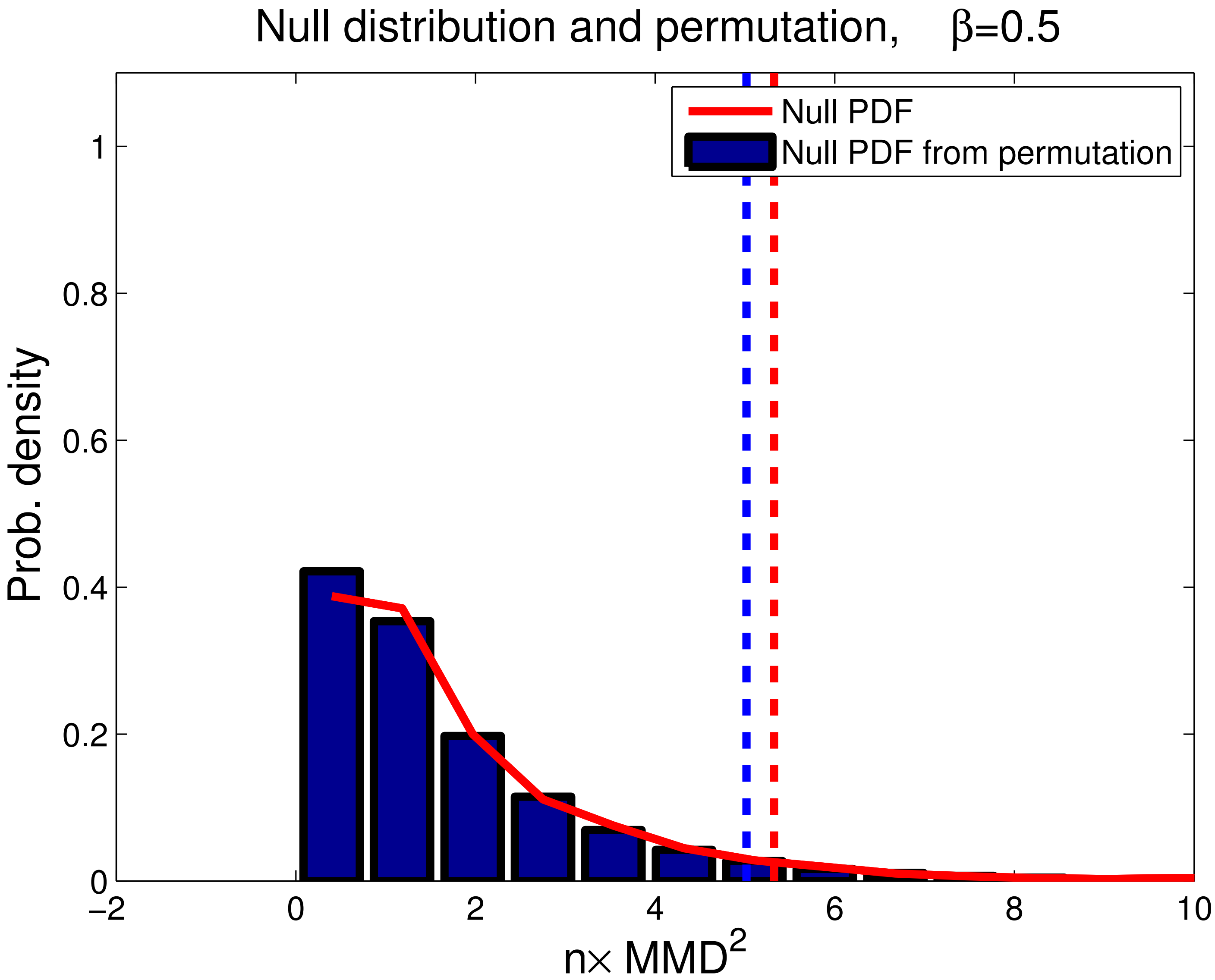
Memory $\beta = 0.2$, wild bootstrap for null



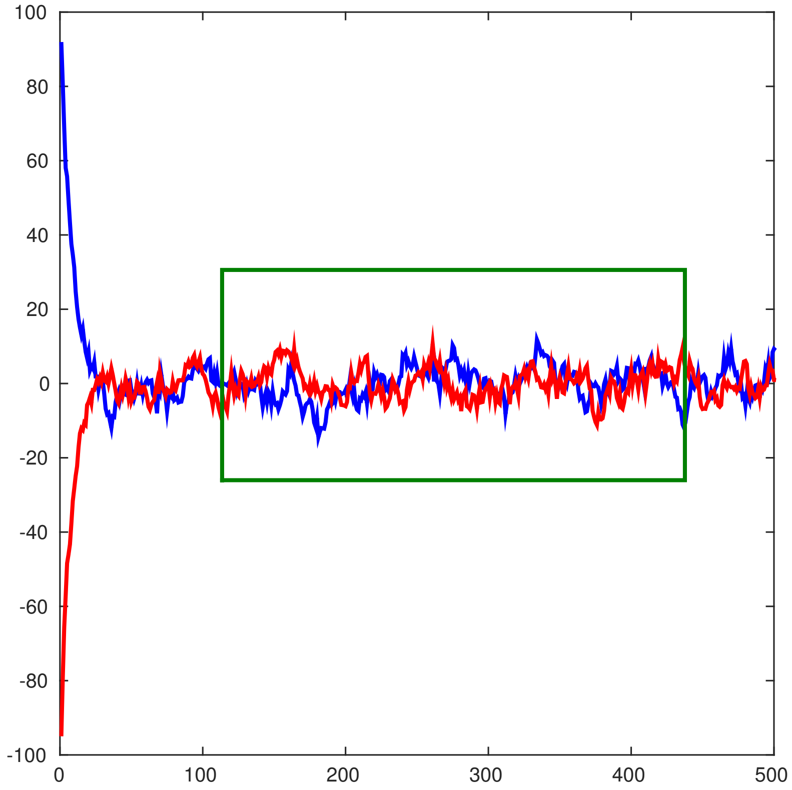
Memory $\beta = 0.4$, wild bootstrap for null



Memory $\beta = 0.5$, wild bootstrap for null



MCMC M.D. Experiment



Does P have the same marginal distribution as Q?

Test - MMD	Type one error
Permutation	68 %
Wild Bootstrap	6 %

Testing Independence and the Hilbert-Schmidt Independence Criterion

MMD for independence

- Dependence measure: the Hilbert Schmidt Independence Criterion [ALT05, NIPS07a, ALT07, ALT08, JMLR10]

  Related to [Feuerverger, 1993]and [Székely and Rizzo, 2009, Székely et al., 2007]

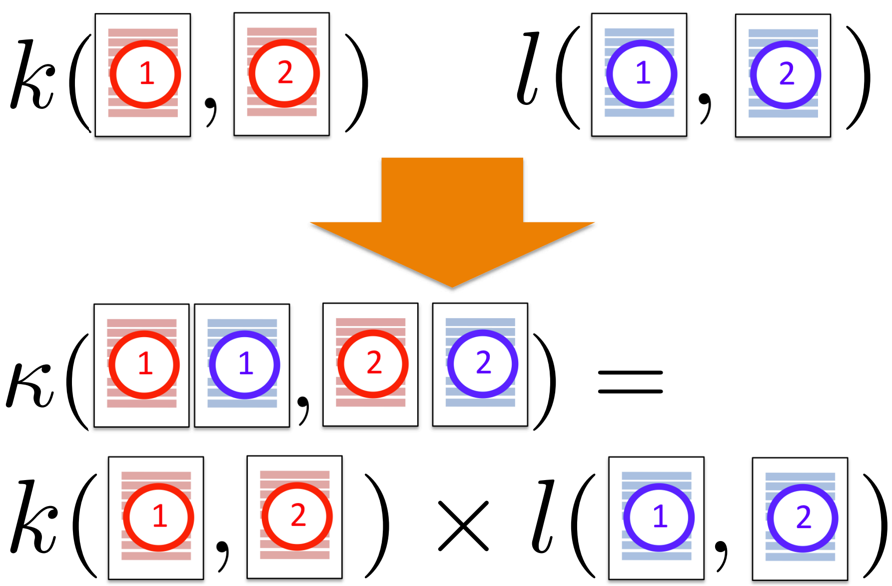
$$HSIC^2(\mathbf{P}_{XY}, \mathbf{P}_X\mathbf{P}_Y) := \|\mu_{\mathbf{P}_{XY}} - \mu_{\mathbf{P}_X\mathbf{P}_Y}\|^2$$

MMD for independence

- Dependence measure: the Hilbert Schmidt Independence Criterion [ALT05, NIPS07a, ALT07, ALT08, JMLR10]

  Related to [Feuerverger, 1993]and [Székely and Rizzo, 2009, Székely et al., 2007]

$$HSIC^2(\mathbf{P}_{XY}, \mathbf{P}_X\mathbf{P}_Y) := \|\mu_{\mathbf{P}_{XY}} - \mu_{\mathbf{P}_X\mathbf{P}_Y}\|^2$$



MMD for independence

- Dependence measure: the Hilbert Schmidt Independence Criterion [ALT05, NIPS07a, ALT07, ALT08, JMLR10]

  Related to [Feuerverger, 1993]and [Székely and Rizzo, 2009, Székely et al., 2007]

$$HSIC^2(\mathbf{P}_{XY}, \mathbf{P}_X\mathbf{P}_Y) := \left\|\mu_{\mathbf{P}_{XY}} - \mu_{\mathbf{P}_X\mathbf{P}_Y}\right\|^2$$

HSIC using expectations of kernels:

Define RKHS $\mathcal{F}$ on $\mathcal{X}$ with kernel k, RKHS $\mathcal{G}$ on $\mathcal{Y}$ with kernel l. Then

$$\begin{aligned}
&\mathrm{HSIC}^2(\mathbf{P}_{XY}, \mathbf{P}_X\mathbf{P}_Y) \\
&= \left\|\mathbf{E}_{XY}\left[(\varphi_X - \mu_{\mathbf{P}_X}) \otimes (\psi_Y - \mu_{\mathbf{P}_Y})\right]\right\|^2_{\mathcal{F}\times\mathcal{G}} \\
&= \mathbf{E}_{XY}\mathbf{E}_{X'Y'}k(\mathsf{x},\mathsf{x}')l(\mathsf{y},\mathsf{y}') + \mathbf{E}_X\mathbf{E}_{X'}k(\mathsf{x},\mathsf{x}')\mathbf{E}_Y\mathbf{E}_{Y'}l(\mathsf{y},\mathsf{y}') \\
&\qquad - 2\mathbf{E}_{X'Y'}\left[\mathbf{E}_X k(\mathsf{x},\mathsf{x}')\mathbf{E}_Y l(\mathsf{y},\mathsf{y}')\right].
\end{aligned}$$

HSIC: empirical estimate and intuition

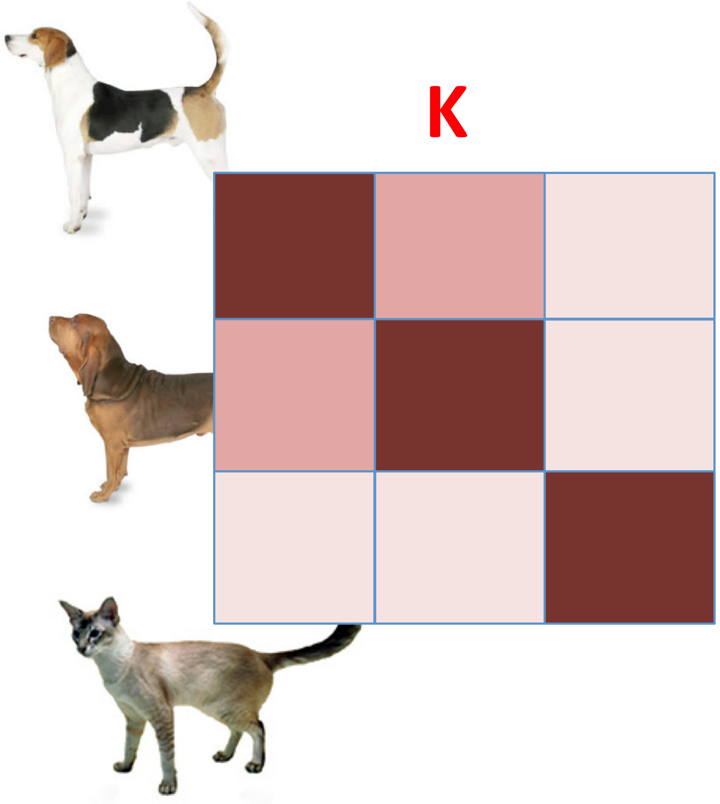
Their noses guide them through life, and they're never happier than when following an interesting scent. They need plenty of exercise, about an hour a day if possible.

A large animal who slings slobber, exudes a distinctive houndy odor, and wants nothing more than to follow his nose. They need a significant amount of exercise and mental stimulation.

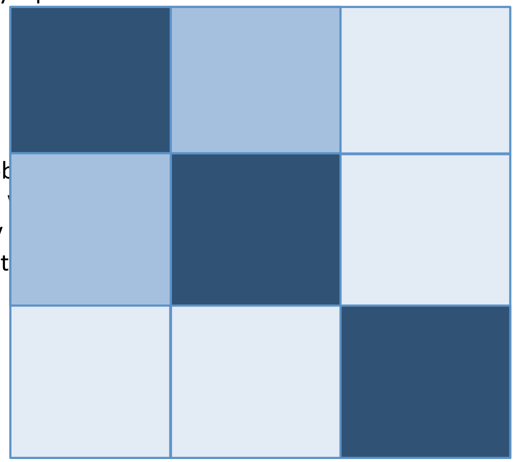
Known for their curiosity, intelligence, and excellent communication skills, the Javanese breed is perfect if you want a responsive, interactive pet, one that will blow in your ear and follow you everywhere.

Text from dogtime.com and petfinder.com

HSIC: empirical estimate and intuition



Their noses guide them through life, and they're never happier than when following an interesting scent. They need plenty of exercise, about an hour a day if possible.

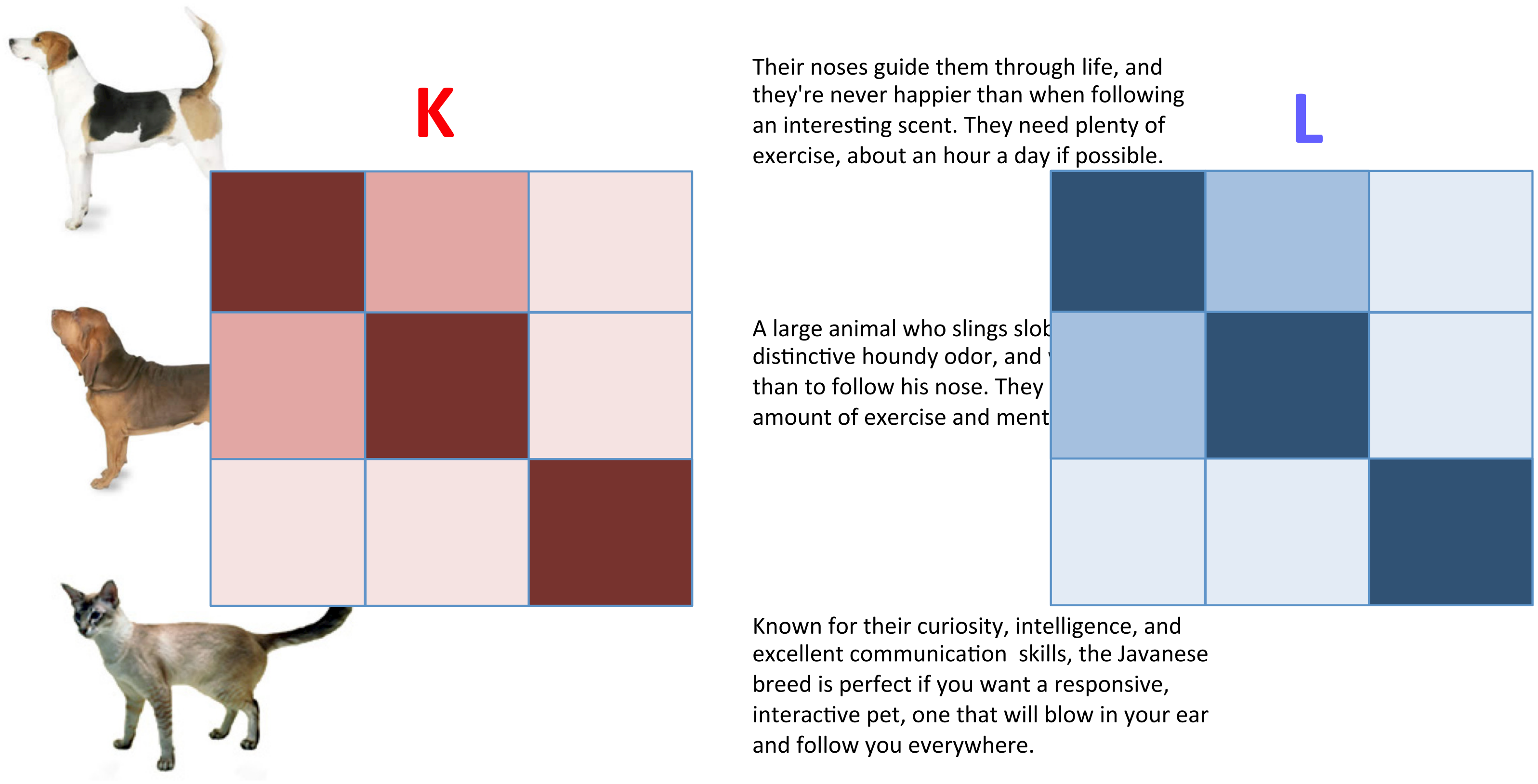


A large animal who slings slob
distinctive houndy odor, and
than to follow his nose. They
amount of exercise and ment

Known for their curiosity, intelligence, and excellent communication skills, the Javanese breed is perfect if you want a responsive, interactive pet, one that will blow in your ear and follow you everywhere.

Text from dogtime.com and petfinder.com

HSIC: empirical estimate and intuition



Text from dogtime.com and petfinder.com

Empirical $HSIC^2(\mathbf{P}_{XY}, \mathbf{P}_X\mathbf{P}_Y)$:

$$\frac{1}{n^2}\left(HKH \circ HLH\right)_{++}$$

HSIC and independence testing

Assume $Z_t := (X_t, Y_t)$ is β mixing with $\beta(r) = o(r^{-6})$. Then

$$\begin{aligned}
\widehat{\mathrm{HSIC}}^2 &= \left\| \frac{1}{n} \sum_{i=1}^{n} (\varphi_{x_i} - \hat{\mu}_{\mathbf{P}_x}) \otimes \left(\phi_{y_i} - \hat{\mu}_{\mathbf{P}_y}\right) \right\|_{\mathcal{F}\times\mathcal{G}}^2 \\
&= \left\| \frac{1}{n} \sum_{i=1}^{n} (\varphi_{x_i} - \mu_{\mathbf{P}_x}) \otimes \left(\phi_{y_i} - \mu_{\mathbf{P}_y}\right) \right\|_{\mathcal{F}\times\mathcal{G}}^2 + O_P(n^{-1}) \\
&= \frac{1}{n^2} \sum_{i,j=1}^{m} \underbrace{\tilde{k}(x_i, x_j)\tilde{l}(y_i, y_j)}_{\mathcal{K}((x_i,y_i),(x_j,y_j))} + O_P(n^{-1})
\end{aligned}$$

where $\tilde{k}(x_i, x_j) = \langle \varphi_{x_i} - \mu_{\mathbf{P}_X}, \varphi_{y_i} - \mu_{\mathbf{P}_Y} \rangle_{\mathcal{F}}$.

HSIC and independence testing

Assume $Z_t := (X_t, Y_t)$ is β mixing with $\beta(r) = o(r^{-6})$. Then

$$
\begin{aligned}
\widehat{\mathrm{HSIC}}^2 &= \left\| \frac{1}{n} \sum_{i=1}^{n} (\varphi_{x_i} - \hat{\mu}_{\mathbf{P}_x}) \otimes \left(\phi_{y_i} - \hat{\mu}_{\mathbf{P}_y}\right) \right\|_{\mathcal{F}\times\mathcal{G}}^2 \\
&= \left\| \frac{1}{n} \sum_{i=1}^{n} (\varphi_{x_i} - \mu_{\mathbf{P}_x}) \otimes \left(\phi_{y_i} - \mu_{\mathbf{P}_y}\right) \right\|_{\mathcal{F}\times\mathcal{G}}^2 + O_P(n^{-1}) \\
&= \frac{1}{n^2} \sum_{i,j=1}^{m} \underbrace{\tilde{k}(x_i, x_j)\tilde{l}(y_i, y_j)}_{\mathcal{K}((x_i,y_i),(x_j,y_j))} + O_P(n^{-1})
\end{aligned}
$$

where $\tilde{k}(x_i, x_j) = \langle \varphi_{x_i} - \mu_{\mathbf{P}_X}, \varphi_{y_i} - \mu_{\mathbf{P}_Y} \rangle_{\mathcal{F}}$.

Wild bootstrap estimate of null for $n\widehat{\mathrm{HSIC}}^2$ is

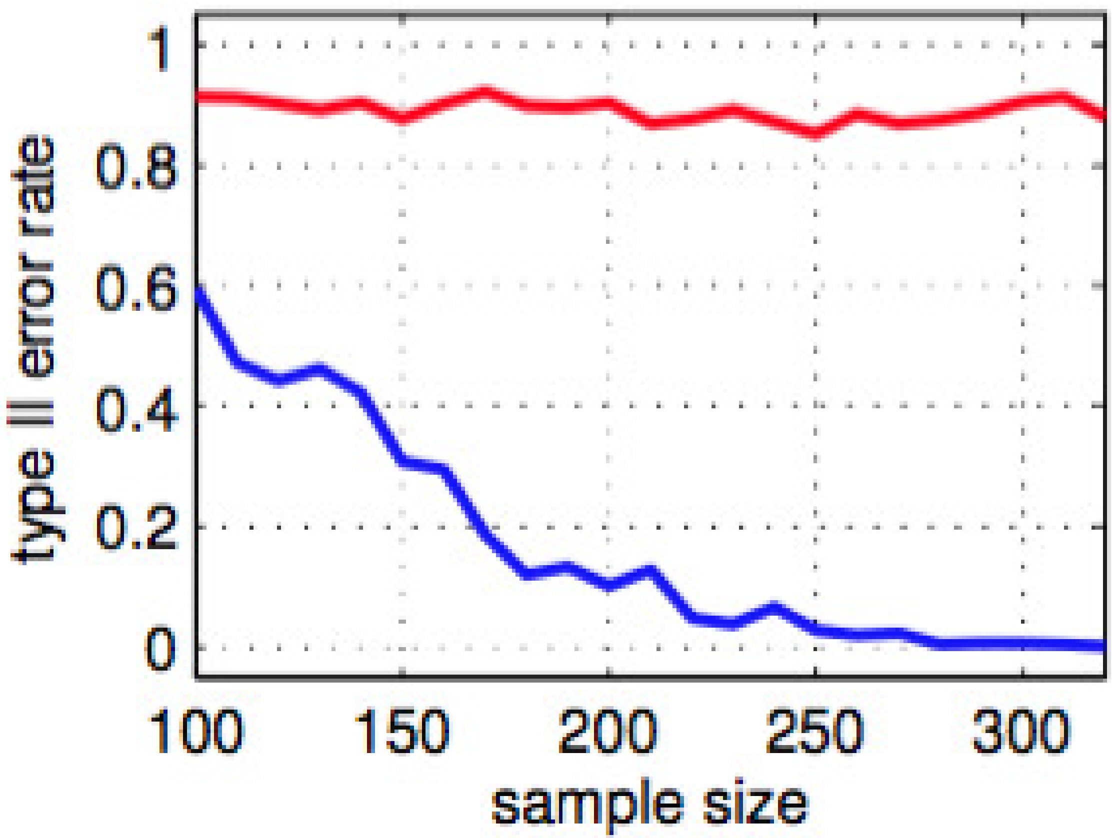
$$
V_n^* := \frac{1}{n} \sum_{i,j=1}^{m} W_i^* W_j^* \mathcal{K}((x_i, y_i), (x_j, y_j)).
$$

Time series experiments

Two time series, common variance (market volatility model) [Bauwens et al., 2006]

$$X_t = \epsilon_{1,t}\sigma_t^2, \quad Y_t = \epsilon_{2,t}\sigma_t^2, \quad \sigma_t^2 = 1 + 0.45(X_{t-1}^2 + Y_{t-1}^2)$$

$$\epsilon_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,1), \quad i \in \{1,2\}.$$



Outline

Testing for differences in marginal distributions of random processes (MMD):

- Markov chain convergence diagnostics
- Change point detection

Testing for independence between random processes (HSIC)

- Dependency structure in financial markets
- Brain region activation

Co-authors

- **From UCL:**
  - Kacper Chwialkowski
  - Paul Rubenstein
  - Heiko Strathmann
- **External:**
  - Dino Sejdinovic, Oxford
  - Bharath Sriperumbudur, Penn State

Selected references

Characteristic kernels and mean embeddings:

- Smola, A., Gretton, A., Song, L., Schoelkopf, B. (2007). A hilbert space embedding for distributions. ALT.
- Sriperumbudur, B., Gretton, A., Fukumizu, K., Schoelkopf, B., Lanckriet, G. (2010). Hilbert space embeddings and metrics on probability measures. JMLR.

Two-sample, independence, conditional independence tests:

- Gretton, A., Fukumizu, K., Teo, C., Song, L., Schoelkopf, B., Smola, A. (2008). A kernel statistical test of independence. NIPS
- Fukumizu, K., Gretton, A., Sun, X., Schoelkopf, B. (2008). Kernel measures of conditional dependence.
- Gretton, A., Borgwardt, K., Rasch, M., Schoelkopf, B., Smola, A. (2012). A kernel two- sample test. JMLR
- Chwialkowski, K., and Gretton, A. (2014). A Kernel Independence Test for Random Processes. ICML.
- Chwialkowski, K., Sejdinovic, D., and Gretton, A. (2014). A Wild Bootstrap for Degenerate Kernel Tests. NIPS.

Energy distance, relation to kernel distances

- Sejdinovic, D., Sriperumbudur, B., Gretton, A.,, Fukumizu, K., (2013). Equivalence of distance-based and rkhs-based statistics in hypothesis testing. Annals of Statistics.

Three way interaction

- Sejdinovic, D., Gretton, A., and Bergsma, W. (2013). A Kernel Test for Three-Variable Interactions. NIPS.
- Rubenstein, P., Chwialkowski, K., and Gretton, A. (2016). A Kernel Test for Three-Variable Interactions with Random Processes. UAI.

References

M. Arcones and E. Giné. On the bootstrap of u and v statistics. *The Annals of Statistics*, 20(2):655–674, 1992.

L. Bauwens, S. Laurent, and J.V.K. Rombouts. Multivariate GARCH models: a survey. *J. Appl. Econ.*, 21(1):79–109, 2006.

Andrey Feuerverger. A consistent test for bivariate dependence. *International Statistical Review*, 61(3):419–433, 1993.

A. Leucht and M. H. Neumann. Dependent wild bootstrap for degenerate U- and V-statistics. *Journal of Multivariate Analysis*, 117:257–280, 2013.

G. Székely and M. Rizzo. Brownian distance covariance. *Annals of Applied Statistics*, 4(3):1233–1303, 2009.

G. Székely, M. Rizzo, and N. Bakirov. Measuring and testing dependence by correlation of distances. *Ann. Stat.*, 35(6):2769–2794, 2007.