Normalised HSIC and Lancaster are aymptotically degenerate V-statistics

Paul Rubenstein

June 2015

1 Introduction/why we care about this

In order to use the Leucht wild bootstrap, we have to show that the test statistic under consideration satisfies certain conditions. One of these conditions is that the statistic is a V-statistic whose core is a degenerate kernel. We instead show that our test statistics asymptotically approach such V-statistics.

This document shows the above for HSIC and the Lancaster test statistic under their respective null hypotheses - for HSIC, that the distribution generating the observations factorises as $P_{XY} = P_X P_Y$; for Lancaster, that there is *some* factorisation, which due to symmetry reduces down to considering the two cases $P_{XYZ} = P_{XY} P_Z$ and $P_{XYZ} = P_X P_Y P_Z$.

2 Layout of this document

In the first half of this document, we prove the above for HSIC. Some parts of this proof are then reused in the longer and more algebraically involved second half in which we prove the result for Lancaster. More detailed list:

- Prove that one can re-centre HSIC gram matrices wrt any point

- Prove that with population centering and the null hypothesis, the terms that we don't care about go to 0.

- Give Lancaster in terms of gram matrices

- Prove that you can recentre Lancaster wrt any point

- Rewrite Lancaster in terms of inner products of covariances and tensor products of means

- Derive asymptotic behaviour of covariances under null

- Using this behaviour, show that most terms in expansion of Lancaster as inner products go to zero.

- Show that the terms which do not go to zero form a type of HSIC

- Apply HSIC result.

- ??????

- Acquisition by google   **Kc: groupies would be nicer**

3   HSIC

3.1   Notation

[Dear reader: throughout this document I will try to stick to the convention that tilde means empirical centering and bar means population centering. However, because I am adapting this from previous versions in which I used different notation, there may be some mistakes. In particular, any $\bar{\mu}$ in this document is probably supposed to be $\tilde{\mu}$]

Let k, l be kernels with associated feature maps ϕ and ψ. Given observations $X_i, Y_i, i = 1, 2, \ldots n$ let K and L be the gram matrices such that $K_{ij} = \langle \phi(X_i), \phi(X_j) \rangle$ and $L_{ij} = \langle \psi(Y_i), \psi(Y_j) \rangle$.

Let $\tilde{\mu}_X = \frac{1}{n} \sum_{i=1}^{n} \phi(X_i)$ and let $\tilde{K}_{ij} = \langle \phi(X_i) - \tilde{\mu}_X, \phi(X_j) - \tilde{\mu}_X \rangle$, and similar for $\tilde{\mu}_Y$ and $\tilde{L}$. These are the *empirically* centred gram matrices.

Let $\mu_X = \mathbb{E}_X \phi(X)$ and let $\bar{K}_{ij} = \langle \phi(X_i) - \mu_X, \phi(X_j) - \mu_X \rangle$, and similar for μ_Y and $\bar{L}$. These are the *population* centred gram matrices.

Kc: Define operators K_{++}, K_{+} and $L \circ K$ for matrices

3.2   The test statistic for HSIC

The biased estimator for HSIC that we are interested in is the V-statistic

$$HSIC_b = \frac{1}{n^2}(K \circ L)_{++} - \frac{2}{n^3}(KL)_{++} + \frac{1}{n^4}K_{++}L_{++}$$

By the paper 'a kernel statistical test of independence', this can also be written as

$$HSIC_b = \frac{1}{n^2}(\tilde{K} \circ \tilde{L})_{++}$$

and therefore

$$\frac{1}{n^2}(\tilde{K} \circ \tilde{L})_{++} = \frac{1}{n^2}(K \circ L)_{++} - \frac{2}{n^3}(KL)_{++} + \frac{1}{n^4}K_{++}L_{++}$$

Remark. *Our ultimate aim is to show that $nHSIC_b$ is asymptotically a normalised degenerate V-statistic. We will show that*

$$\frac{1}{n}(\tilde{K} \circ \tilde{L})_{++} = \frac{1}{n}(\bar{K} \circ \bar{L})_{++} - \frac{2}{n^2}(\bar{K}\bar{L})_{++} + \frac{1}{n^3}\bar{K}_{++}\bar{L}_{++}$$

and then show that the last two terms go to zero as $n \longrightarrow \infty$ under the null hypothesis and that the first term of the right hand side is a degenerate normalised V-statistic

<u>Claim:</u>

$$\frac{1}{n^2}(\tilde{K} \circ \tilde{L})_{++} = \frac{1}{n^2}(K' \circ L')_{++} - \frac{2}{n^3}(K'L')_{++} + \frac{1}{n^4}K'_{++}L'_{++}$$

where K' and L' are K and L centred with respect to *any* point in feature space.

In particular,

$$\frac{1}{n^2}(\tilde{K} \circ \tilde{L})_{++} = \frac{1}{n^2}(\bar{K} \circ \bar{L})_{++} - \frac{2}{n^3}(\bar{K}\bar{L})_{++} + \frac{1}{n^4}\bar{K}_{++}\bar{L}_{++}$$

<u>Proof:</u> Given feature maps ϕ and ψ and observations $X_i, Y_i, i = 1, \ldots, n$ define

$$T(\phi, \psi, \{X_i\}, \{Y_i\}) = \frac{1}{n^2}\sum_{ij}\left\langle \phi(X_i) - \frac{1}{n}\sum_k \phi(X_k), \phi(X_j) - \frac{1}{n}\sum_k \phi(X_k)\right\rangle$$

$$\times \left\langle \psi(Y_i) - \frac{1}{n}\sum_k \psi(Y_k), \psi(Y_j) - \frac{1}{n}\sum_k \psi(Y_k)\right\rangle$$

Note that

$$T(\phi, \psi, \{X_i\}, \{Y_i\}) = \frac{1}{n^2}(\tilde{K} \circ \tilde{L})_{++}$$

By the paper, we also have that

$$T(\phi, \psi, \{X_i\}, \{Y_i\}) = \frac{1}{n^2}(K \circ L)_{++} - \frac{2}{n^3}(KL)_{++} + \frac{1}{n^4}K_{++}L_{++}$$

where K and L are the gram matrices of the observations with respect to the feature maps ϕ and ψ.

Now define, for constants W and Z, the new feature maps $\phi'(X) = \phi(X) - W$ and $\psi'(Y) = \psi(Y) - Z$. Let K' and L' be the gram matrices with respect to the feature maps ϕ' and ψ'. Then Kc: this can be shorten - since $\phi'(X_i) - \frac{1}{n}\sum_k \phi'(X_k) = \phi(X_i) - \frac{1}{n}\sum_k \phi(X_k)$, then the claim

$$T(\phi', \psi', \{X_i\}, \{Y_i\}) = \frac{1}{n^2} \sum_{ij} \left\langle \phi'(X_i) - \frac{1}{n} \sum_k \phi'(X_k), \phi'(X_j) - \frac{1}{n} \sum_k \phi'(X_k) \right\rangle$$

$$\times \left\langle \psi'(Y_i) - \frac{1}{n} \sum_k \psi'(Y_k), \psi'(Y_j) - \frac{1}{n} \sum_k \psi'(Y_k) \right\rangle$$

$$= \frac{1}{n^2} \sum_{ij} \left\langle \phi(X_i) - W - \frac{1}{n} \sum_k (\phi(X_k) - W), \phi(X_j) - W - \frac{1}{n} \sum_k (\phi(X_k) - W) \right\rangle$$

$$\times \left\langle \psi(Y_i) - Z - \frac{1}{n} \sum_k (\psi(Y_k) - Z), \psi(Y_j) - Z - \frac{1}{n} \sum_k (\psi(Y_k) - Z) \right\rangle$$

$$= \frac{1}{n^2} \sum_{ij} \left\langle \phi(X_i) - \frac{1}{n} \sum_k \phi(X_k), \phi(X_j) - \frac{1}{n} \sum_k \phi(X_k) \right\rangle$$

$$\times \left\langle \psi(Y_i) - \frac{1}{n} \sum_k \psi(Y_k), \psi(Y_j) - \frac{1}{n} \sum_k \psi(Y_k) \right\rangle$$

$$= T(\phi, \psi, \{X_i\}, \{Y_i\})$$

By the paper again, we have that

$$T(\phi', \psi', \{X_i\}, \{Y_i\}) = \frac{1}{n^2} (K' \circ L')_{++} - \frac{2}{n^3} (K'L')_{++} + \frac{1}{n^4} K'_{++} L'_{++}$$

Hence letting $Z = \mathbb{E}_X \phi(X)$ and $W = \mathbb{E}_Y \psi(Y)$, we see that

$$\frac{1}{n^2} (\tilde{K} \circ \tilde{L})_{++} = \frac{1}{n^2} (\bar{K} \circ \bar{L})_{++} - \frac{2}{n^3} (\bar{K}\bar{L})_{++} + \frac{1}{n^4} \bar{K}_{++} \bar{L}_{++}$$

∎

Multiplying the above equation through by n yields the equation in the remark before the preceeding claim. We now have three things that remain to be shown, assuming the null hypothesis that $P_{XY} = P_X P_Y$.

<u>**Claim:**</u> **[1]**

$$\frac{1}{n^3} \bar{K}_{++} \bar{L}_{++} \longrightarrow 0$$

<u>**Claim:**</u> **[2]**

$$\frac{1}{n^2} (\bar{K}\bar{L})_{++} \longrightarrow 0$$

<u>**Claim:**</u> **[3]**

$$\frac{1}{n} (\bar{K} \circ \bar{L})_{++}$$

is a degenerate normalised V-statistic

4

Remark. *In the following we will make use of the fact that* $\|\tilde{\mu} - \mu\| = O(\frac{1}{\sqrt{n}})$
(note to self - find reference for this)

Proof: [1]

$$\frac{1}{n^3}\bar{K}_{++}\bar{L}_{++} = \frac{1}{n^3}\sum_{i,j}\Big\langle \phi(X_i) - \mu_X, \phi(X_j) - \mu_X\Big\rangle\sum_{k,l}\Big\langle \psi(Y_k) - \mu_Y, \psi(Y_l) - \mu_Y\Big\rangle$$

$$= n\Big\langle \frac{1}{n}\sum_i(\phi(X_i) - \mu_X), \frac{1}{n}\sum_j(\phi(X_j) - \mu_X)\Big\rangle\Big\langle \frac{1}{n}\sum_k(\psi(Y_k) - \mu_Y), \frac{1}{n}\sum_l(\psi(Y_l) - \mu_Y)\Big\rangle$$

$$= n\Big\langle \bar{\mu}_X - \mu_X, \tilde{\mu}_X - \mu_X\Big\rangle\Big\langle \tilde{\mu}_Y - \mu_Y, \tilde{\mu}_Y - \mu_Y\Big\rangle$$

$$= n\|\tilde{\mu}_X - \mu_X\|^2\|\tilde{\mu}_Y - \mu_Y\|^2$$

$$= nO(\frac{1}{n})O(\frac{1}{n})$$

$$= O(\frac{1}{n}) \longrightarrow 0$$

∎

Proof: [2] Define $\tilde{C}_{XY}$ to be the uncentred covariance function *(is this the right name to give it?)*. That is,

$$\tilde{C}_{XY} = \frac{1}{n}\sum_i \phi(X_i) \otimes \psi(Y_i)$$

Observe that

$$\|\tilde{C}_{XY} - \tilde{\mu}_X \otimes \tilde{\mu}_Y\|^2 = \langle \tilde{C}_{XY} - \tilde{\mu}_X \otimes \tilde{\mu}_Y, \tilde{C}_{XY} - \tilde{\mu}_X \otimes \tilde{\mu}_Y \rangle$$

$$= \langle \tilde{C}_{XY}, \tilde{C}_{XY} \rangle - 2\langle \tilde{C}_{XY}, \tilde{\mu}_X \otimes \tilde{\mu}_Y \rangle + \langle \tilde{\mu}_X \otimes \tilde{\mu}_Y, \tilde{\mu}_X \otimes \tilde{\mu}_Y \rangle$$

$$= \frac{1}{n^2} \sum_{ij} \langle \phi(X_i) \otimes \psi(Y_i), \phi(X_j) \otimes \psi(Y_j) \rangle$$

$$- \frac{2}{n^3} \sum_{ijk} \langle \phi(X_i) \otimes \psi(Y_i), \phi(X_j) \otimes \psi(Y_k) \rangle$$

$$+ \frac{1}{n^4} \sum_{ijkl} \langle \phi(X_i) \otimes \psi(Y_j), \phi(X_k) \otimes \psi(Y_l) \rangle$$

$$= \frac{1}{n^2} \sum_{ij} \langle \phi(X_i), \phi(X_j) \rangle \langle \psi(Y_i), \psi(Y_j) \rangle$$

$$- \frac{2}{n^3} \sum_{ijk} \langle \phi(X_i), \phi(X_j) \rangle \langle \psi(Y_i), \psi(Y_k) \rangle$$

$$+ \frac{1}{n^4} \sum_{ijkl} \langle \phi(X_i), \phi(X_j) \rangle \langle \psi(Y_k), \psi(Y_l) \rangle$$

$$= \frac{1}{n^2} (K \circ L)_{++} - \frac{2}{n^3} (KL)_{++} + \frac{1}{n^4} K_{++} L_{++}$$

$$= HSIC_b[P_{XY}, P_X P_Y]$$

and therefore under the null hypothesis we have that

$$n\|\tilde{C}_{XY} - \tilde{\mu}_X \otimes \tilde{\mu}_Y\|^2$$

converges to some random variable (by 'a kernel statistical test of independence'). Therefore

$$|n\langle \tilde{C}_{XY}, \tilde{\mu}_X \otimes \tilde{\mu}_Y \rangle - n\langle \tilde{\mu}_X \otimes \tilde{\mu}_Y, \tilde{\mu}_X \otimes \tilde{\mu}_Y \rangle| = |n\langle \tilde{C}_{XY} - \tilde{\mu}_X \otimes \tilde{\mu}_Y, \tilde{\mu}_X \otimes \tilde{\mu}_Y \rangle|$$

$$\leqslant \sqrt{n}\|\tilde{C}_{XY} - \tilde{\mu}_X \otimes \tilde{\mu}_Y\| \sqrt{n}\|\tilde{\mu}_X \otimes \tilde{\mu}_Y\|$$

$$= O(1)\sqrt{n}O(\frac{1}{n}) = O(\frac{1}{\sqrt{n}})$$

Now, since

$$n\langle \tilde{\mu}_X \otimes \tilde{\mu}_Y, \tilde{\mu}_X \otimes \tilde{\mu}_Y \rangle = n\|\tilde{\mu}_X \otimes \tilde{\mu}_Y\|^2$$

$$= O(\frac{1}{n})$$

This implies that

$$n\langle \tilde{C}_{XY}, \tilde{\mu}_X \otimes \tilde{\mu}_Y \rangle = O(\frac{1}{\sqrt{n}})$$

Note that we could repeat all of the above analysis using the feature maps $\bar{\phi}$ and $\bar{\psi}$ defined by

$$\bar{\phi}(X) = \phi(X) - \mu_X$$

and

$$\bar{\psi}(Y) = \psi(X) - \mu_Y$$

The *uncentred* covariance function for $\bar{\phi}$ and $\bar{\psi}$ is the empirical *population centred* covariance function for ϕ and ψ:

$$\hat{C}_{XY} = \frac{1}{n}\sum_i (\phi(X) - \mu_X) \otimes (\psi(Y) - \mu_y)$$

and the sample means for $\bar{\phi}$ and $\bar{\psi}$ are the population centred sample means for ϕ and ψ:

$$\frac{1}{n}\sum_i \bar{\phi}(X) = \frac{1}{n}\sum_i (\phi(X) - \mu_X) = \tilde{\mu}_X - \mu_X$$

$$\frac{1}{n}\sum_i \bar{\psi}(Y) = \frac{1}{n}\sum_i (\psi(Y) - \mu_Y) = \tilde{\mu}_Y - \mu_Y$$

Thus, by the above analysis we have that

$$n\langle \hat{C}_{XY}, (\tilde{\mu}_X - \mu_X) \otimes (\tilde{\mu}_Y - \mu_Y)\rangle = O(\frac{1}{\sqrt{n}})$$

It therefore follows that $\frac{1}{n^2}(\bar{K}\bar{L})_{++} \longrightarrow 0$ since

$$
\begin{aligned}
\frac{1}{n^2}(\bar{K}\bar{L})_{++} &= \frac{1}{n^2}\sum_{ijk}\langle\phi(X_i) - \mu_X, \phi(X_j) - \mu_X\rangle\langle\psi(Y_j) - \mu_Y, \psi(Y_k) - \mu_Y\rangle \\
&= \frac{1}{n^2}\sum_{ijk}\langle(\phi(X_j) - \mu_X)\otimes(\psi(Y_j) - \mu_Y), (\phi(X_i) - \mu_X)\otimes(\psi(Y_k) - \mu_Y)\rangle \\
&= n\langle\frac{1}{n}\sum_j[(\phi(X_j) - \mu_X)\otimes(\psi(Y_j) - \mu_Y)], [\frac{1}{n}\sum_i(\phi(X_i) - \mu_X)]\otimes[\frac{1}{n}\sum_k(\psi(Y_k) - \mu_Y)]\rangle \\
&= n\langle\tilde{C}_{XY}, (\tilde{\mu}_X - \mu_X)\otimes(\tilde{\mu}_Y - \mu_Y)\rangle \\
&= O(\frac{1}{\sqrt{n}})
\end{aligned}
$$

Kc: Can't you use clt for $\sqrt{n}\tilde{C}_{XY}$, $\sqrt{n}(\tilde{\mu}_Y - \mu_Y)$, $\sqrt{n}(\tilde{\mu}_X - \mu_X)$? You would have

$$\frac{1}{\sqrt{n}}\langle\sqrt{n}\tilde{C}_{XY}, \sqrt{n}(\tilde{\mu}_X - \mu_X)\otimes\sqrt{n}(\tilde{\mu}_Y - \mu_Y)\rangle.$$

∎

Proof: [3]

$$\frac{1}{n}(\bar{K} \circ \bar{L})_{++} = \frac{1}{n}\sum_{ij}\langle\phi(X_i) - \mu_X, \phi(X_j) - \mu_X\rangle\langle\psi(Y_i) - \mu_Y, \psi(Y_j) - \mu_Y\rangle$$

Letting $S = (X, Y)$, observe that this is a normalised V-statistic with core

$$h(S_i, S_j) = \langle\phi(X_i) - \mu_X, \phi(X_j) - \mu_X\rangle\langle\psi(Y_i) - \mu_Y, \psi(Y_j) - \mu_Y\rangle$$

Under the null hypothesis the joint factorises as $P_{XY} = P_X P_Y$. The expecation operator $\mathbb{E}_{XY}$ therefore factorises as

$$\mathbb{E}_{XY} = \mathbb{E}_X \mathbb{E}_Y$$

Thus

$$\begin{aligned}
\mathbb{E}_{S_j}h(s_i, S_j) &= \mathbb{E}_{X_j Y_j}\langle\phi(x_i) - \mu_X, \phi(X_j) - \mu_X\rangle\langle\psi(y_i) - \mu_Y, \psi(Y_j) - \mu_Y\rangle \\
&= \langle\phi(x_i) - \mu_X, \mathbb{E}_{X_j}\phi(X_j) - \mu_X\rangle\langle\psi(y_i) - \mu_Y, \mathbb{E}_{Y_j}\psi(Y_j) - \mu_Y\rangle \\
&= \langle\phi(x_i) - \mu_X, 0\rangle\langle\psi(y_i) - \mu_Y, 0\rangle = 0
\end{aligned}$$

∎

4 Lancaster

The Lancaster test statistic we use is

$$\|\Delta_L \hat{P}\|^2 = \frac{1}{n^2}(\tilde{K} \circ \tilde{L} \circ \tilde{M})_{++}$$

An alternative expression for this can be calculated given the following definition

$$\|\Delta_L \hat{P}\|^2 = \|\hat{P}_{XYZ} - \hat{P}_{XY}\hat{P}_Z - \hat{P}_{YZ}\hat{P}_X - \hat{P}_{XZ}\hat{P}_Y + 2\hat{P}_X\hat{P}_Y\hat{P}_Z\|^2$$

which, using the table of inner products of products of marginal distributions given in the Gretton Lancaster paper gives the following expression in terms of gram matrices *(note to self - omit proof of this for the meantime but the working out for this should go in an appendix or something)*

$$\|\Delta_L \hat{P}\|^2 = \frac{1}{n^2}(K \circ L \circ M)_{++} \qquad -\frac{2}{n^3}((K \circ L)M)_{++} - \frac{2}{n^3}((K \circ M)L)_{++}$$

$$-\frac{2}{n^3}((M \circ L)K)_{++} \qquad +\frac{1}{n^4}(K \circ L)_{++}M_{++} + \frac{1}{n^4}(K \circ M)_{++}L_{++}$$

$$+\frac{1}{n^4}(L \circ M)_{++}K_{++} \qquad +\frac{2}{n^4}(MKL)_{++} + \frac{2}{n^4}(KLM)_{++}$$

$$+\frac{2}{n^4}(KML)_{++} \qquad +\frac{4}{n^4}tr(K_+ \circ L_+ \circ M_+) - \frac{4}{n^5}(KL)_{++}M_{++}$$

$$-\frac{4}{n^5}(KM)_{++}L_{++} \qquad -\frac{4}{n^5}(LM)_{++}K_{++} + \frac{4}{n^6}K_{++}L_{++}M_{++}$$

<u>**Claim:**</u>

$$\frac{1}{n^2}\big(\tilde{K} \circ \tilde{L} \circ \tilde{M}\big)_{++} = \frac{1}{n^2}(\bar{K} \circ \bar{L} \circ \bar{M})_{++} \qquad -\frac{2}{n^3}((\bar{K} \circ \bar{L})\bar{M})_{++} - \frac{2}{n^3}((\bar{K} \circ \bar{M})\bar{L})_{++}$$

$$-\frac{2}{n^3}((\bar{M} \circ \bar{L})\bar{K})_{++} \qquad +\frac{1}{n^4}(\bar{K} \circ \bar{L})_{++}\bar{M}_{++} + \frac{1}{n^4}(\bar{K} \circ \bar{M})_{++}\bar{L}_{++}$$

$$+\frac{1}{n^4}(\bar{L} \circ \bar{M})_{++}\bar{K}_{++} \qquad +\frac{2}{n^4}(\bar{M}\bar{K}\bar{L})_{++} + \frac{2}{n^4}(\bar{K}\bar{L}\bar{M})_{++}$$

$$+\frac{2}{n^4}(\bar{K}\bar{M}\bar{L})_{++} \qquad +\frac{4}{n^4}tr(\bar{K}_+ \circ \bar{L}_+ \circ \bar{M}_+) - \frac{4}{n^5}(\bar{K}\bar{L})_{++}\bar{M}_{++}$$

$$-\frac{4}{n^5}(\bar{K}\bar{M})_{++}\bar{L}_{++} \qquad -\frac{4}{n^5}(\bar{L}\bar{M})_{++}\bar{K}_{++} + \frac{4}{n^6}\bar{K}_{++}\bar{L}_{++}\bar{M}_{++}$$

<u>**Proof:**</u> **Inspired by the observation that**

$$\|\Delta_L \hat{P}\|^2 = \frac{1}{n^2}(\tilde{K} \circ \tilde{L} \circ \tilde{M})_{++}$$

$$= \frac{1}{n^2}\sum_{i,j}\Big\langle \phi(X_i) - \frac{1}{n}\sum_k \phi(X_k), \phi(X_j) - \frac{1}{n}\sum_k \phi(X_k)\Big\rangle$$

$$\times \Big\langle \psi(Y_i) - \frac{1}{n}\sum_k \psi(Y_k), \psi(Y_j) - \frac{1}{n}\sum_k \psi(Y_k)\Big\rangle$$

$$\times \Big\langle \omega(Z_i) - \frac{1}{n}\sum_k \omega(Z_k), \omega(Z_j) - \frac{1}{n}\sum_k \omega(Z_k)\Big\rangle$$

let us define the following operator. Given feature maps ϕ, ψ and ω, and data $\mathcal{D} = \{X_1, \ldots, X_n, Y_1, \ldots, Y_n, Z_1, \ldots, Z_n\}$ define

$$T(\phi, \psi, \omega, \mathcal{D}) = \frac{1}{n^2}\sum_{i,j}\Big\langle \phi(X_i) - \frac{1}{n}\sum_k \phi(X_k), \phi(X_j) - \frac{1}{n}\sum_k \phi(X_k)\Big\rangle$$

$$\times \Big\langle \psi(Y_i) - \frac{1}{n}\sum_k \psi(Y_k), \psi(Y_j) - \frac{1}{n}\sum_k \psi(Y_k)\Big\rangle$$

$$\times \Big\langle \omega(Z_i) - \frac{1}{n}\sum_k \omega(Z_k), \omega(Z_j) - \frac{1}{n}\sum_k \omega(Z_k)\Big\rangle$$

Denoting by K, L and M the gram matrices of the data with respect to ϕ, ψ and ω, we observe the following two equalities:

$$T(\phi, \psi, \omega, \mathcal{D}) = \|\Delta_L \hat{P}\|^2_{\phi \otimes \psi \otimes \omega}$$

And therefore

$$
\begin{aligned}
T(\phi, \psi, \omega, \mathcal{D}) = {}& \frac{1}{n^2}(K \circ L \circ M)_{++} && -\frac{2}{n^3}((K \circ L)M)_{++} - \frac{2}{n^3}((K \circ M)L)_{++} \\
& -\frac{2}{n^3}((M \circ L)K)_{++} && +\frac{1}{n^4}(K \circ L)_{++}M_{++} + \frac{1}{n^4}(K \circ M)_{++}L_{++} \\
& +\frac{1}{n^4}(L \circ M)_{++}K_{++} && +\frac{2}{n^4}(MKL)_{++} + \frac{2}{n^4}(KLM)_{++} \\
& +\frac{2}{n^4}(KML)_{++} && +\frac{4}{n^4}tr(K_+ \circ L_+ \circ M_+) - \frac{4}{n^5}(KL)_{++}M_{++} \\
& -\frac{4}{n^5}(KM)_{++}L_{++} && -\frac{4}{n^5}(LM)_{++}K_{++} + \frac{4}{n^6}K_{++}L_{++}M_{++}
\end{aligned}
$$

Now, define the following new feature maps: $\phi'(X) = \phi(X) - a$, $\psi'(Y) = \psi(Y) - b$ and $\omega'(Z) = \omega(Z) - c$ where a, b and c are constants in the relevant feature space. Denoting by K', L' and M' the gram matrices with respect to the new feature maps, we observe firstly by the equality immediately above that

$$
\begin{aligned}
T(\phi', \psi', \omega', \mathcal{D}) = {}& \frac{1}{n^2}(K' \circ L' \circ M')_{++} && -\frac{2}{n^3}((K' \circ L')M')_{++} - \frac{2}{n^3}((K' \circ M')L')_{++} \\
& -\frac{2}{n^3}((M' \circ L')K')_{++} && +\frac{1}{n^4}(K' \circ L')_{++}M'_{++} + \frac{1}{n^4}(K' \circ M')_{++}L'_{++} \\
& +\frac{1}{n^4}(L' \circ M')_{++}K'_{++} && +\frac{2}{n^4}(M'K'L')_{++} + \frac{2}{n^4}(K'L'M')_{++} \\
& +\frac{2}{n^4}(K'M'L')_{++} && +\frac{4}{n^4}tr(K'_+ \circ L'_+ \circ M'_+) - \frac{4}{n^5}(K'L')_{++}M'_{++} \\
& -\frac{4}{n^5}(K'M')_{++}L'_{++} && -\frac{4}{n^5}(L'M')_{++}K'_{++} + \frac{4}{n^6}K'_{++}L'_{++}M'_{++}
\end{aligned}
$$

Next, by the definition of T

$$T(\phi', \psi', \omega', \mathcal{D}) = \frac{1}{n^2} \sum_{i,j} \Big\langle \phi'(X_i) - \frac{1}{n}\sum_k \phi'(X_k), \phi'(X_j) - \frac{1}{n}\sum_k \phi'(X_k)\Big\rangle$$

$$\times \Big\langle \psi'(Y_i) - \frac{1}{n}\sum_k \psi'(Y_k), \psi'(Y_j) - \frac{1}{n}\sum_k \psi'(Y_k)\Big\rangle$$

$$\times \Big\langle \omega'(Z_i) - \frac{1}{n}\sum_k \omega'(Z_k), \omega'(Z_j) - \frac{1}{n}\sum_k \omega'(Z_k)\Big\rangle$$

$$= \frac{1}{n^2} \sum_{i,j} \Big\langle \phi(X_i) - a - \frac{1}{n}\sum_k [\phi(X_k) - a], \phi(X_j) - a - \frac{1}{n}\sum_k [\phi(X_k) - a]\Big\rangle$$

$$\times \Big\langle \psi(Y_i) - b - \frac{1}{n}\sum_k [\psi(Y_k) - b], \psi(Y_j) - b - \frac{1}{n}\sum_k [\psi(Y_k) - b]\Big\rangle$$

$$\times \Big\langle \omega(Z_i) - c - \frac{1}{n}\sum_k [\omega(Z_k) - c], \omega(Z_j) - c - \frac{1}{n}\sum_k [\omega(Z_k) - c]\Big\rangle$$

$$= \frac{1}{n^2} \sum_{i,j} \Big\langle \phi(X_i) - \frac{1}{n}\sum_k \phi(X_k), \phi(X_j) - \frac{1}{n}\sum_k \phi(X_k)\Big\rangle$$

$$\times \Big\langle \psi(Y_i) - \frac{1}{n}\sum_k \psi(Y_k), \psi(Y_j) - \frac{1}{n}\sum_k \psi(Y_k)\Big\rangle$$

$$\times \Big\langle \omega(Z_i) - \frac{1}{n}\sum_k \omega(Z_k), \omega(Z_j) - \frac{1}{n}\sum_k \omega(Z_k)\Big\rangle$$

$$= T(\phi, \psi, \omega, \mathcal{D})$$

$$= \|\Delta_L \hat{P}\|^2_{\phi \otimes \psi \otimes \omega}$$

And thus the Lancaster statistic is invariant to translating the origin in feature space. Hence setting $a = \mu_X$, $b = \mu_Y$ and $c = \mu_Y$ we see that we can write

$$\|\Delta_L \hat{P}\|^2_{\phi \otimes \psi \otimes \omega} = \frac{1}{n^2}(\bar{K} \circ \bar{L} \circ \bar{M})_{++} \quad -\frac{2}{n^3}((\bar{K} \circ \bar{L})\bar{M})_{++} - \frac{2}{n^3}((\bar{K} \circ \bar{M})\bar{L})_{++}$$

$$-\frac{2}{n^3}((\bar{M} \circ \bar{L})\bar{K})_{++} \quad +\frac{1}{n^4}(\bar{K} \circ \bar{L})_{++}\bar{M}_{++} + \frac{1}{n^4}(\bar{K} \circ \bar{M})_{++}\bar{L}_{++}$$

$$+\frac{1}{n^4}(\bar{L} \circ \bar{M})_{++}\bar{K}_{++} \quad +\frac{2}{n^4}(\bar{M}\bar{K}\bar{L})_{++} + \frac{2}{n^4}(\bar{K}\bar{L}\bar{M})_{++}$$

$$+\frac{2}{n^4}(\bar{K}\bar{M}\bar{L})_{++} \quad +\frac{4}{n^4}tr(\bar{K}_+ \circ \bar{L}_+ \circ \bar{M}_+) - \frac{4}{n^5}(\bar{K}\bar{L})_{++}\bar{M}_{++}$$

$$-\frac{4}{n^5}(\bar{K}\bar{M})_{++}\bar{L}_{++} \quad -\frac{4}{n^5}(\bar{L}\bar{M})_{++}\bar{K}_{++} + \frac{4}{n^6}\bar{K}_{++}\bar{L}_{++}\bar{M}_{++}$$

∎

Remark. *From here onwards, we will assume WLOG that the population means μ_X, μ_Y and μ_Z are all equal to 0. This can be done precisely because of the*

translation invariance - rather than working with the original feature maps ϕ, ψ and ω we can use the population centred maps defined by

$$\phi'(X) = \phi(X) - \mu_X$$

$$\psi'(Y) = \psi(Y) - \mu_Y$$

$$\omega'(Z) = \omega(Z) - \mu_Z$$

and then drop the primes (') for notational convenience. The consequences of this are that the means are 0 and the (uncentred emprical) covariances are actually population centred. That is,

$$\tilde{C}_{(\cdot)} = \hat{C}_{(\cdot)}$$

where $(\cdot)$ represents any subset of the variables.

Let us now consider the normalised Lancaster statistic using its representation in terms of population centred gram matrices.

$$
\begin{aligned}
n\|\Delta_L \hat{P}\|^2 = {} & \frac{1}{n}(\bar{K} \circ \bar{L} \circ \bar{M})_{++} && -\frac{2}{n^2}((\bar{K} \circ \bar{L})\bar{M})_{++} - \frac{2}{n^2}((\bar{K} \circ \bar{M})\bar{L})_{++} \\
& -\frac{2}{n^2}((\bar{M} \circ \bar{L})\bar{K})_{++} && +\frac{1}{n^3}(\bar{K} \circ \bar{L})_{++}\bar{M}_{++} + \frac{1}{n^3}(\bar{K} \circ \bar{M})_{++}\bar{L}_{++} \\
& +\frac{1}{n^3}(\bar{L} \circ \bar{M})_{++}\bar{K}_{++} && +\frac{2}{n^3}(\bar{M}\bar{K}\bar{L})_{++} + \frac{2}{n^3}(\bar{K}\bar{L}\bar{M})_{++} \\
& +\frac{2}{n^3}(\bar{K}\bar{M}\bar{L})_{++} && +\frac{4}{n^3}tr(\bar{K}_+ \circ \bar{L}_+ \circ \bar{M}_+) - \frac{4}{n^4}(\bar{K}\bar{L})_{++}\bar{M}_{++} \\
& -\frac{4}{n^4}(\bar{K}\bar{M})_{++}\bar{L}_{++} && -\frac{4}{n^4}(\bar{L}\bar{M})_{++}\bar{K}_{++} + \frac{4}{n^5}\bar{K}_{++}\bar{L}_{++}\bar{M}_{++}
\end{aligned}
$$

Up to symmetries $K \leftrightarrow L \leftrightarrow M$, there are 7 different terms here. Each of them can be written as an inner product between tensor products of covariance operators and means. Assume WLOG that the the means of the feature maps are 0. In this case, the covariance matrices are centred and $\bar{\mu} \longrightarrow 0$

The 7 terms and their form as inner products are:

(note to self: this isn't hard to show at all but it does take a lot of space. Omit proofs of this here, but include in an appendix)

$$\frac{1}{n}(A \circ B \circ C)_{++} = n\langle \tilde{C}_{abc}, \tilde{C}_{abc}\rangle$$

$$\frac{1}{n^2}((A \circ B)C)_{++} = n\langle \tilde{C}_{ab} \otimes \tilde{\mu}_c, \tilde{C}_{ab} \otimes \tilde{\mu}_c\rangle$$

$$\frac{1}{n^3}((A \circ B)_{++}C)_{++} = n\langle \tilde{C}_{ab} \otimes \tilde{\mu}_c, \tilde{\mu}_a \otimes \tilde{\mu}_b \otimes \tilde{\mu}_c\rangle$$

$$\frac{1}{n^3}((ABC)_{++} = n\langle \tilde{C}_{ab} \otimes \tilde{\mu}_c, \tilde{\mu}_a \otimes \tilde{C}_{bc}\rangle$$

$$\frac{1}{n^3}tr(A_+ \circ B_+ \circ C_+) = n\langle \tilde{C}_{abc}, \tilde{\mu}_a \otimes \tilde{\mu}_b \otimes \tilde{\mu}_c\rangle$$

$$\frac{1}{n^3}(AB)_{++}C_{++} = n\langle \tilde{C}_{ab} \otimes \tilde{\mu}_c, \tilde{\mu}_a \otimes \tilde{\mu}_b \otimes \tilde{\mu}_c\rangle$$

$$\frac{1}{n^3}A_{++}B_{++}C_{++} = n\langle \tilde{\mu}_a \otimes \tilde{\mu}_b \otimes \tilde{\mu}_c, \tilde{\mu}_a \otimes \tilde{\mu}_b \otimes \tilde{\mu}_c\rangle$$

Hence we can write the normalised lancaster statistic as

$$
\begin{aligned}
n\|\Delta_L \hat{P}\|^2 = {} & n\langle \tilde{C}_{XYZ}, \tilde{C}_{XYZ}\rangle \\
& - 2n\langle \tilde{C}_{XYZ}, \tilde{C}_{XY} \otimes \tilde{\mu}_Z\rangle \\
& - 2n\langle \tilde{C}_{XZY}, \tilde{C}_{XZ} \otimes \tilde{\mu}_Y\rangle \\
& - 2n\langle \tilde{C}_{YZX}, \tilde{C}_{YZ} \otimes \tilde{\mu}_X\rangle \\
& + n\langle \tilde{C}_{XY} \otimes \tilde{\mu}_Z, \tilde{C}_{XY} \otimes \tilde{\mu}_Z\rangle \\
& + n\langle \tilde{C}_{XZ} \otimes \tilde{\mu}_Y, \tilde{C}_{XZ} \otimes \tilde{\mu}_Y\rangle \\
& + n\langle \tilde{C}_{YZ} \otimes \tilde{\mu}_X, \tilde{C}_{YZ} \otimes \tilde{\mu}_X\rangle \\
& + 2n\langle \tilde{\mu}_Z \otimes \tilde{C}_{XY}, \tilde{C}_{ZX} \otimes \tilde{\mu}_Y\rangle \\
& + 2n\langle \tilde{\mu}_X \otimes \tilde{C}_{YZ}, \tilde{C}_{XY} \otimes \tilde{\mu}_Z\rangle \\
& + 2n\langle \tilde{\mu}_X \otimes \tilde{C}_{ZY}, \tilde{C}_{XZ} \otimes \tilde{\mu}_Y\rangle \\
& + 4n\langle \tilde{C}_{XYZ}, \tilde{\mu}_X \otimes \tilde{\mu}_Y \otimes \tilde{\mu}_Z\rangle \\
& - 4n\langle \tilde{C}_{XY} \otimes \tilde{\mu}_Z, \tilde{\mu}_X \otimes \tilde{\mu}_Y \otimes \tilde{\mu}_Z\rangle \\
& - 4n\langle \tilde{C}_{XZ} \otimes \tilde{\mu}_Y, \tilde{\mu}_X \otimes \tilde{\mu}_Z \otimes \tilde{\mu}_Y\rangle \\
& - 4n\langle \tilde{C}_{YZ} \otimes \tilde{\mu}_X, \tilde{\mu}_Y \otimes \tilde{\mu}_Z \otimes \tilde{\mu}_X\rangle \\
& + 4n\langle \tilde{\mu}_X \otimes \tilde{\mu}_Y \otimes \tilde{\mu}_Z, \tilde{\mu}_X \otimes \tilde{\mu}_Y \otimes \tilde{\mu}_Z\rangle
\end{aligned}
$$

We have two cases to consider under the null hypothesis due to symmetry. First, $P_{XYZ} = P_{XY}P_Z$. Second, $P_{XYZ} = P_X P_Y P_Z$. Let us consider the first case first.

If $P_{XYZ} = P_{XY}P_Z$, then, considering $X \otimes Y$ to be a single variable with kernel $k \otimes l$ and expanding the following expression in terms of gram matrices, we observe that

$$\|\tilde{C}_{XYZ} - \tilde{C}_{XY} \otimes \tilde{\mu}_Z\|^2 = HSIC_b(P_{XYZ}, P_{XY}P_Z)$$

and therefore, under the null,

$$\|\tilde{C}_{XYZ} - \tilde{C}_{XY} \otimes \tilde{\mu}_Z\|^2 \longrightarrow 0$$

and

$$n\|\tilde{C}_{XYZ} - \tilde{C}_{XY} \otimes \tilde{\mu}_Z\|^2 \longrightarrow randomvariable$$

By marginalising $P_{XYZ} = P_{XY}P_Z$ with respect to **X** or **Y**, we also obtain $P_{XZ} = P_X P_Z$ and $P_{YZ} = P_Y P_Z$. Thus, similarly

$$\|\tilde{C}_{XZ} - \tilde{\mu}_X \otimes \tilde{\mu}_Z\|^2 \longrightarrow 0$$
$$n\|\tilde{C}_{XZ} - \tilde{\mu}_X \otimes \tilde{\mu}_Z\|^2 \longrightarrow rv$$
$$\|\tilde{C}_{YZ} - \tilde{\mu}_Y \otimes \tilde{\mu}_Z\|^2 \longrightarrow 0$$
$$n\|\tilde{C}_{YZ} - \tilde{\mu}_Y \otimes \tilde{\mu}_Z\|^2 \longrightarrow rv$$

Now since $\|\tilde{\mu}_X\| = O(\frac{1}{\sqrt{n}})$ and similar for Y and Z, it is the case that $\tilde{\mu}_X \otimes \tilde{\mu}_Z, \tilde{\mu}_X \otimes \tilde{\mu}_Y, \tilde{\mu}_X \otimes \tilde{\mu}_Z = O(\frac{1}{n})$ and also $\tilde{C}_{XZ}, \tilde{C}_{YZ} = O(\frac{1}{\sqrt{n}})$ by the above limits (I think!). In general, since $\tilde{C}_{XY}$ converges to a fixed non-zero quantity, $\|\tilde{C}_{XY} \otimes \tilde{\mu}_Z\| = O(\frac{1}{\sqrt{n}})$, and so $\|\tilde{C}_{XYZ}\| = O(\frac{1}{\sqrt{n}})$

We now go back to the expansion of the normalised lancaster statistic and tackle each term from bottom to top.

4.1 Asymptotic behaviour of each term of normalised lancaster

Note that here we use big-O notation. **O(1) means bounded for large n.**

$$
\begin{aligned}
4n\langle \tilde{\mu}_X \otimes \tilde{\mu}_Y \otimes \tilde{\mu}_Z, \tilde{\mu}_X \otimes \tilde{\mu}_Y \otimes \tilde{\mu}_Z \rangle &= 4n\langle \tilde{\mu}_X, \tilde{\mu}_X \rangle\langle \tilde{\mu}_Y, \tilde{\mu}_Y \rangle\langle \tilde{\mu}_Z, \tilde{\mu}_Z \rangle \\
&= 4n\|\tilde{\mu}_X\|^2\|\tilde{\mu}_Y\|^2\|\tilde{\mu}_Z\|^2 \\
&= O(\frac{1}{n^2})
\end{aligned}
$$

$$|4n\langle \tilde{C}_{YZ} \otimes \tilde{\mu}_X, \tilde{\mu}_Y \otimes \tilde{\mu}_Z \otimes \tilde{\mu}_X \rangle| = |4n\langle \tilde{C}_{YZ}, \tilde{\mu}_Y \otimes \tilde{\mu}_Z \rangle\langle \tilde{\mu}_X, \tilde{\mu}_X \rangle|$$
$$\leqslant 4n\|\tilde{C}_{YZ}\|\|\tilde{\mu}_Y \otimes \tilde{\mu}_Z\|\|\tilde{\mu}_X\|^2$$
$$= 4nO(\frac{1}{n})O(\frac{1}{n})O(\frac{1}{n})$$
$$= O(\frac{1}{n^2})$$

$$|4n\langle \tilde{C}_{XZ} \otimes \tilde{\mu}_Y, \tilde{\mu}_X \otimes \tilde{\mu}_Z \otimes \tilde{\mu}_Y \rangle| = |4n\langle \tilde{C}_{XZ}, \tilde{\mu}_X \otimes \tilde{\mu}_Z \rangle\langle \tilde{\mu}_Y, \tilde{\mu}_Y \rangle|$$
$$\leqslant 4n\|\tilde{C}_{XZ}\|\|\tilde{\mu}_X \otimes \tilde{\mu}_Z\|\|\tilde{\mu}_Y\|^2$$
$$= 4nO(\frac{1}{n})O(\frac{1}{n})O(\frac{1}{n})$$
$$= O(\frac{1}{n^2})$$

$$|4n\langle \tilde{C}_{XY} \otimes \tilde{\mu}_Z, \tilde{\mu}_X \otimes \tilde{\mu}_Y \otimes \tilde{\mu}_Z \rangle| = |4n\langle \tilde{C}_{XY}, \tilde{\mu}_X \otimes \tilde{\mu}_Y \rangle\langle \tilde{\mu}_Z, \tilde{\mu}_Z \rangle|$$
$$\leqslant 4n\|\tilde{C}_{XY}\|\|\tilde{\mu}_X \otimes \tilde{\mu}_Y\|\|\tilde{\mu}_Z\|^2$$
$$= 4nO(1)O(\frac{1}{n})O(\frac{1}{n})$$
$$= O(\frac{1}{n})$$

$$|4n\langle \tilde{C}_{XYZ}, \tilde{\mu}_X \otimes \tilde{\mu}_Y \otimes \tilde{\mu}_Z \rangle| \leqslant 4n\|\tilde{C}_{XYZ}\|\|\tilde{\mu}_X\|\|\tilde{\mu}_Y\|\|\tilde{\mu}_Z\|$$
$$= 4nO(\frac{1}{\sqrt{n}})O(\frac{1}{\sqrt{n}})O(\frac{1}{\sqrt{n}})O(\frac{1}{\sqrt{n}})$$
$$= O(\frac{1}{n})$$

$$|2n\langle \tilde{\mu}_X \otimes \tilde{C}_{ZY}, \tilde{C}_{XZ} \otimes \tilde{\mu}_Y \rangle| \leqslant 2n\|\tilde{\mu}_X\|\|\tilde{C}_{ZY}\|\|\tilde{C}_{XZ}\|\|\tilde{\mu}_Y\|$$
$$= 2nO(\frac{1}{\sqrt{n}})O(\frac{1}{\sqrt{n}})O(\frac{1}{\sqrt{n}})O(\frac{1}{\sqrt{n}})$$
$$= O(\frac{1}{n})$$

$$|2n\langle \tilde{\mu}_X \otimes \tilde{C}_{YZ}, \tilde{C}_{XY} \otimes \tilde{\mu}_Z \rangle| \leqslant 2n\|\tilde{\mu}_X\|\|\tilde{C}_{YZ}\|\|\tilde{C}_{XY}\|\|\tilde{\mu}_Z\|$$
$$= 2nO(\frac{1}{\sqrt{n}})O(\frac{1}{\sqrt{n}})O(1)O(\frac{1}{\sqrt{n}})$$
$$= O(\frac{1}{n^{3/2}})$$

$$|2n\langle\tilde{\mu}_Z\otimes\tilde{C}_{XY},\tilde{C}_{ZX}\otimes\tilde{\mu}_Y\rangle|\leqslant 2n\|\tilde{\mu}_Z\|\|\tilde{C}_{XY}\|\|\tilde{C}_{ZX}\|\|\tilde{\mu}_Y\|$$
$$=2nO(\frac{1}{\sqrt{n}})O(1)O(\frac{1}{\sqrt{n}})O(\frac{1}{\sqrt{n}})$$
$$=O(\frac{1}{n^{3/2}})$$

$$|n\langle\tilde{C}_{YZ}\otimes\tilde{\mu}_X,\tilde{C}_{YZ}\otimes\tilde{\mu}_X\rangle|=n\|\tilde{C}_{YZ}\|^2\|\tilde{\mu}_X\|^2$$
$$=nO(\frac{1}{n})O(\frac{1}{n})$$
$$=O(\frac{1}{n})$$

$$|n\langle\tilde{C}_{XZ}\otimes\tilde{\mu}_Y,\tilde{C}_{XZ}\otimes\tilde{\mu}_Y\rangle|=n\|\tilde{C}_{XZ}\|^2\|\tilde{\mu}_Y\|^2$$
$$=nO(\frac{1}{n})O(\frac{1}{n})$$
$$=O(\frac{1}{n})$$

$$|n\langle\tilde{C}_{XY}\otimes\tilde{\mu}_Z,\tilde{C}_{XY}\otimes\tilde{\mu}_Z\rangle|=n\|\tilde{C}_{XY}\|^2\|\tilde{\mu}_Z\|^2$$
$$\longrightarrow\|C_{XY}\|^2\times rv$$

$$|2n\langle\tilde{C}_{YZX},\tilde{C}_{YZ}\otimes\tilde{\mu}_X\rangle|\leqslant 2n\|\tilde{C}_{YZX}\|\|\tilde{C}_{YZ}\|\|\tilde{\mu}_X\|$$
$$=2nO(\frac{1}{\sqrt{n}})O(\frac{1}{\sqrt{n}})O(\frac{1}{\sqrt{n}})$$
$$=O(\frac{1}{\sqrt{n}})$$

$$|2n\langle\tilde{C}_{XZY},\tilde{C}_{XZ}\otimes\tilde{\mu}_Y\rangle|\leqslant 2n\|\tilde{C}_{XZY}\|\|\tilde{C}_{XZ}\|\|\tilde{\mu}_Y\|$$
$$=2nO(\frac{1}{\sqrt{n}})O(\frac{1}{\sqrt{n}})O(\frac{1}{\sqrt{n}})$$
$$=O(\frac{1}{\sqrt{n}})$$

Finally, observe that

$$|2n\langle \tilde{C}_{XYZ}, \tilde{C}_{XY} \otimes \tilde{\mu}_Z\rangle| \leqslant 2n\|\tilde{C}_{XYZ}\|\|\tilde{C}_{XY}\|\|\tilde{\mu}_Z\|$$
$$= 2nO(\frac{1}{\sqrt{n}})O(1)O(\frac{1}{\sqrt{n}})$$
$$= O(1)$$

and hence Cauchy-Schwarz does not suffice to bound this last quantity, nor the quantity three above it.

To summarise, we are left with the following asymptotic equality.

$$n\|\Delta_L \hat{P}\|^2 \longrightarrow n\langle \tilde{C}_{XYZ}, \tilde{C}_{XYZ}\rangle - 2n\langle \tilde{C}_{XYZ}, \tilde{C}_{XY} \otimes \tilde{\mu}_Z\rangle + n\langle \tilde{C}_{XY} \otimes \tilde{\mu}_Z, \tilde{C}_{XY} \otimes \tilde{\mu}_Z\rangle$$

Remark. *To reiterate our objective - we need to show that $n\|\Delta_L\hat{P}\|^2$ tends to some V-statistic with a core that is a degenerate kernel. In the following, we will massage the right-hand side of the equation above and show that* it *tends the desired type of quantity. Since $a \longrightarrow b \longrightarrow c$ implies that $a \longrightarrow c$, this suffices.*

Observe that

$$n\langle \tilde{C}_{XYZ}, \tilde{C}_{XYZ}\rangle - 2n\langle \tilde{C}_{XYZ}, \tilde{C}_{XY} \otimes \tilde{\mu}_Z\rangle + n\langle \tilde{C}_{XY} \otimes \tilde{\mu}_Z, \tilde{C}_{XY} \otimes \tilde{\mu}_Z\rangle$$

$$= n\langle \frac{1}{n}\sum_i \phi(X_i) \otimes \psi(Y_i) \otimes \omega(Z_i), \frac{1}{n}\sum_j \phi(X_j) \otimes \psi(Y_j) \otimes \omega(Z_j)\rangle$$

$$- 2n\langle \frac{1}{n}\sum_i \phi(X_i) \otimes \psi(Y_i) \otimes \omega(Z_i), \frac{1}{n^2}\sum_{jk} \phi(X_j) \otimes \psi(Y_j) \otimes \omega(Z_k)\rangle$$

$$+ n\langle \frac{1}{n^2}\sum_{ij} \phi(X_i) \otimes \psi(Y_i) \otimes \omega(Z_j), \frac{1}{n^2}\sum_{kl} \phi(X_k) \otimes \psi(Y_k) \otimes \omega(Z_l)\rangle$$

$$= \frac{1}{n}\sum_{ij} \langle\phi(X_i), \phi(X_j)\rangle\langle\psi(Y_i), \psi(Y_j)\rangle\langle\omega(Z_i), \omega(Z_j)\rangle$$

$$- \frac{2}{n^2}\sum_{ijk} \langle\phi(X_i), \phi(X_j)\rangle\langle\psi(Y_i), \psi(Y_j)\rangle\langle\omega(Z_j), \omega(Z_k)\rangle$$

$$+ \frac{1}{n^3}\sum_{ijkl} \langle\phi(X_i), \phi(X_j)\rangle\langle\psi(Y_i), \psi(Y_j)\rangle\langle\omega(Z_k), \omega(Z_l)\rangle$$

$$= \frac{1}{n}((\bar{K} \circ \bar{L}) \circ \bar{M})_{++} - \frac{2}{n^2}((\bar{K} \circ \bar{L})\bar{M})_{++} + \frac{1}{n^3}(\bar{K} \circ \bar{L})_{++}\bar{M}_{++}$$

$$= nHSIC_b[P_{XYZ}, P_{XY}P_Z]$$

Where we are considering $\bar{K} \circ \bar{L}$ to be the gram matrix for the random variable (X, Y) with kernel $k \otimes l$. Observe that the expectation of this random variable is C_{XY} (this is the covariance function obtained

from using the (separately) population centred variables X and Y). Since

$$
\begin{aligned}
(\bar{K} \circ \bar{L})_{ij} &= \langle \phi(X_i), \phi(X_j) \rangle \langle \psi(Y_i), \psi(Y_j) \rangle \\
&= \langle \phi(X_i) \otimes \psi(Y_i), \phi(X_j) \otimes \psi(Y_j) \rangle
\end{aligned}
$$

we have that the population centred gram matrix $\overline{\bar{K} \circ \bar{L}}$ satisfies

$$
(\overline{\bar{K} \circ \bar{L}})_{ij} = \langle \phi(X_i) \otimes \psi(Y_i) - C_{XY}, \phi(X_j) \otimes \psi(Y_j) - C_{XY} \rangle
$$

Now, since HSIC is invariant to recentring the gram matrices, we have that

$$
n\|\Delta_L \hat{P}\|^2 \longrightarrow \frac{1}{n}((\overline{\bar{K} \circ \bar{L}}) \circ \bar{M})_{++} - \frac{2}{n^2}((\overline{\bar{K} \circ \bar{L}})\bar{M})_{++} + \frac{1}{n^3}(\overline{\bar{K} \circ \bar{L}})_{++}\bar{M}_{++}
$$

Alert! I realised that in the following, I have assumed that $\|\tilde{C}_{XY} - C_{XY}\|^2 = O(\frac{1}{n})$ which is analogous to $\|\tilde{\mu}_X - \mu_X\|^2 = O(\frac{1}{n})$. Then the analysis used for HSIC in the previous part of this document holds exactly again here

Extra note: I've realised that I'm a bit confused about the time series assumptions. Are we assuming that each of X, Y and Z are (separately) stationary processes with the desired mixing properties, or that the tuple (X, Y, Z) is a stationary process with the desired mixing properties? If the latter, then is it the case that the marginal processes (if that is the right term to use), eg (X, Y) and Z, have the desired properties? If so then we are all good.

For the same reasons as is the case for HSIC, we thus have that

$$
\begin{aligned}
n\|\Delta_L \hat{P}\|^2 &\longrightarrow \frac{1}{n}((\overline{\bar{K} \circ \bar{L}}) \circ \bar{M})_{++} \\
&= \frac{1}{n} \sum_{ij} \langle \phi(X_i) \otimes \psi(Y_i) - C_{XY}, \phi(X_j) \otimes \psi(Y_j) - C_{XY} \rangle \langle \omega(Z_i), \omega(Z_j) \rangle
\end{aligned}
$$

This is a normalised V-statistic with core h being degenerate. Let $S_i = (X_i, Y_i, Z_i)$. Then

$$
h(S_i, S_j) = \langle \phi(X_i) \otimes \psi(Y_i) - C_{XY}, \phi(X_j) \otimes \psi(Y_j) - C_{XY} \rangle \langle \omega(Z_i), \omega(Z_j) \rangle
$$

Since $P_{XYZ} = P_{XY}P_Z$, we have that $\mathbb{E}_{XYZ} = \mathbb{E}_{XY}\mathbb{E}_Z$ and thus

$$
\begin{aligned}
\mathbb{E}_{S_j} h(s_i, S_j) &= \mathbb{E}_{X_j Y_j Z_j} \langle \phi(x_i) \otimes \psi(y_i) - C_{XY}, \phi(X_j) \otimes \psi(Y_j) - C_{XY} \rangle\langle \omega(z_i), \omega(Z_j) \rangle \\
&= \mathbb{E}_{X_j Y_j} \mathbb{E}_{Z_j} \langle \phi(x_i) \otimes \psi(y_i) - C_{XY}, \phi(X_j) \otimes \psi(Y_j) - C_{XY} \rangle\langle \omega(z_i), \omega(Z_j) \rangle \\
&= \langle \phi(x_i) \otimes \psi(y_i) - C_{XY}, \mathbb{E}_{X_j Y_j}[\phi(X_j) \otimes \psi(Y_j) - C_{XY}]\rangle\langle \omega(z_i), \mathbb{E}_{Z_j} \omega(Z_j) \rangle \\
&= \langle \phi(x_i) \otimes \psi(y_i) - C_{XY}, 0 \rangle\langle \omega(z_i), 0 \rangle \\
&= 0
\end{aligned}
$$

Observe that if moreover $P_{XYZ} = P_X P_Y P_Z$ then all of the above analysis holds but additionally we have that $C_{XY} = 0$ and so

$$
n\|\Delta_L \hat{P}\|^2 \longrightarrow \frac{1}{n}(\bar{K} \circ \bar{L} \circ \bar{M})_{++}
$$

Further notes

Note that the test statistic in each case is actually the same thing as

$$
\frac{1}{n} \sum_{ij} \langle \phi(X_i) \otimes \psi(Y_i) \otimes \omega(Z_i) - C_{XYZ}, \phi(X_j) \otimes \psi(Y_j) \otimes \omega(Z_j) - C_{XYZ} \rangle
$$

5   Appendix

5.1   Decomposition of lancaster in terms of gram matrices

5.2   Decomposition of lancaster in terms of inner products of covariances and tensors of means