Appendix to 'Analysis of Conditional Randomisation and Permutation schemes with application to conditional independence testing'

Małgorzata Łazęcka*[1,2], Bartosz Kołodziejek†[2], and Jan Mielniczuk‡[1,2]

[1]Institute of Computer Science, Polish Academy of Sciences
[2]Faculty of Mathematics and Information Science, Warsaw University of Technology

In the appendix we first prove Theorem 1, Lemma 4, and Theorem 5, the equality $M = \tilde{M}$ used in the proof of Theorem 6 and then the two lemmas concerning properties of Kullback-Leibler projections. Finally we prove the ordering of covariance matrices in CP and CR scenarios discussed in Remark 3.
Below we give a proof of **Theorem 1**.

Proof of Theorem 1. We prove that $\mathbf{X}_n$ and $\mathbf{X}_n^*$ are exchangeable given $\mathbf{Z}_n = \mathbf{z}_n$. The proof that $\mathbf{X}_n, \mathbf{X}_{n,1}^*, \mathbf{X}_{n,2}^*, \ldots, \mathbf{X}_{n,B}^*$ are exchangeable is a straightforward extension as well as the proof of the fact that $(\mathbf{X}_n, \mathbf{Y}_n, \mathbf{Z}_n), (\mathbf{X}_{n,1}^*, \mathbf{Y}_n, \mathbf{Z}_n), (\mathbf{X}_{n,2}^*, \mathbf{Y}_n, \mathbf{Z}_n), \ldots, (\mathbf{X}_{n,B}^*, \mathbf{Y}_n, \mathbf{Z}_n)$ are exchangeable. We recall that the random variables $T_1, T_2, \ldots, T_s$ are exchangeable if their joint distribution is invariant under permutations of the components.

We denote by $\pi \in \Pi$ a permutation applied to $\mathbf{X}_n$ resulting in $\mathbf{X}_n^*$. That transformation consists of permutations on the layers $\mathbf{Z}_n = z$ denoted by π_z for $z \in \mathcal{Z}$ and we use a notation $i_z \in \{i : Z_i = z\}$ to denote the indices of subsequent observations on the layer $\mathbf{Z}_n = z$. Consider $P(\mathbf{X}_n = \mathbf{x}_n, \mathbf{X}_n^* = \mathbf{x}_n^* | \mathbf{Z}_n = \mathbf{z}_n, \Pi = \pi)$. Note that this probability equals $P(\mathbf{X}_n = \mathbf{x}_n | \mathbf{Z}_n = \mathbf{z}_n, \Pi = \pi)$ if $\mathbf{x}_n^*$ is an image of $\mathbf{x}_n$ under transformation π and 0 otherwise. Note that if $\mathbf{x}_n^*$ is an image of $\mathbf{x}_n$ then for all $z \in \mathcal{Z}$ and for all $i_z \in \{i : Z_i = z\}$

$$x_{i_z}^* = x_{\pi_z(i_z)}.$$

In case when $\pi(\mathbf{x}_n) = \mathbf{x}_n^*$ we have

$$P(\mathbf{X}_n = \mathbf{x}_n, \mathbf{X}_n^* = \mathbf{x}_n^* | \mathbf{Z}_n = \mathbf{z}_n, \Pi = \pi) = P(\mathbf{X}_n = \mathbf{x}_n | \mathbf{Z}_n = \mathbf{z}_n, \Pi = \pi) \tag{1}$$

and

$$\begin{aligned} P(\mathbf{X}_n = \mathbf{x}_n | \mathbf{Z}_n = \mathbf{z}_n, \Pi = \pi) &= P(\mathbf{X}_n = \mathbf{x}_n | \mathbf{Z}_n = \mathbf{z}_n) = \prod_z P(\forall_{i:Z_i=z} X_i = x_i | Z_i = z) \\ &= \prod_z \prod_{i_z} P(X_{i_z} = x_{i_z} | Z_{i_z} = z) = \prod_z \prod_{i_z} P(X_{i_z} = x_{i_z} | Z_{i_z} = z, \Pi = \pi) \\ &= \prod_z \prod_{i_z} P(X_{\pi_z(i_z)} = x_{i_z} | Z_{i_z} = z, \Pi = \pi) = P(\mathbf{X}_n = \mathbf{x}_n^* | \mathbf{Z}_n = \mathbf{z}_n, \Pi = \pi), \end{aligned}$$

where the first and the fourth equations follow from conditional independence of $\mathbf{X}_n$ and Π given $\mathbf{Z}_n$, and the second and the third use independence of $(X_i, Z_i)_{i=1}^n$. We also have that

$$P(\mathbf{X}_n = \mathbf{x}_n^*, \mathbf{X}_n^* = \mathbf{x}_n | \mathbf{Z}_n = \mathbf{z}_n, \Pi = \pi) = P(\mathbf{X}_n = \mathbf{x}_n^* | \mathbf{Z}_n = \mathbf{z}_n, \Pi = \pi),$$

where the above equation follows from analogous reasoning as in (1) applied to π^{-1}. When $\pi(\mathbf{x}_n) \neq \mathbf{x}_n^*$, then

$$P(\mathbf{X}_n = \mathbf{x}_n, \mathbf{X}_n^* = \mathbf{x}_n^* | \mathbf{Z}_n = \mathbf{z}_n, \Pi = \pi) = P(\mathbf{X}_n = \mathbf{x}_n^*, \mathbf{X}_n^* = \mathbf{x}_n | \mathbf{Z}_n = \mathbf{z}_n, \Pi = \pi) = 0.$$

Thus

$$P(\mathbf{X}_n = \mathbf{x}_n, \mathbf{X}_n^* = \mathbf{x}_n^* | \mathbf{Z}_n = \mathbf{z}_n, \Pi = \pi) = P(\mathbf{X}_n = \mathbf{x}_n^*, \mathbf{X}_n^* = \mathbf{x}_n | \mathbf{Z}_n = \mathbf{z}_n, \Pi = \pi).$$

*malgorzata.lazecka@ipipan.waw.pl
†bartosz.kolodziejek@pw.edu.pl
‡jan.mielniczuk@ipipan.waw.pl

and as the above equation holds for all $\pi \in \Pi$, we obtain

$$P(\mathbf{X}_n = \mathbf{x}_n, \mathbf{X}_n^* = \mathbf{x}_n^* | \mathbf{Z}_n = \mathbf{z}_n) = P(\mathbf{X}_n = \mathbf{x}_n^*, \mathbf{X}_n^* = \mathbf{x}_n | \mathbf{Z}_n = \mathbf{z}_n).$$

As we have proven the exchangeability of the sample and resampled samples given $\mathbf{Z}_n$, the test statistics based on them are also exchangeable given $\mathbf{Z}_n$. By averaging over $\mathbf{Z}_n$ the property also holds unconditionally.

For exchangeable random variables $T, T_1^*, T_2^*, \ldots, T_B^*$ and for $i \in \{1, \ldots, B, B+1\}$

$$P\left(1 + \sum_{b=1}^{B} \mathbb{I}(T \le T_b^*) = i\right) = \frac{1}{1+B}$$

as the rank of T among $T, T_1^*, T_2^*, \ldots, T_B^*$ is uniformly distributed on $\{1, \ldots, B+1\}$. Thus

$$P\left(1 + \sum_{b=1}^{B} \mathbb{I}(T \le T_b^*) \le i\right) = \frac{i}{1+B}$$

and from that we obtain

$$P\left(\frac{1 + \sum_{b=1}^{B} \mathbb{I}(T \le T_b^*)}{1+B} \le \frac{i}{1+B}\right) = \frac{i}{1+B}.$$

For any $\alpha \in \left[\frac{i}{B+1}, \frac{i+1}{B+1}\right)$ and $\alpha \le 1$ we thus have

$$P\left(\frac{1 + \sum_{b=1}^{B} \mathbb{I}(T \le T_b^*)}{1+B} \le \alpha\right) \le \alpha. \tag{2}$$

In the considered case of conditional independence the exchangeability of $T, T_1^*, T_2^*, \ldots, T_B^*$ holds given $\mathbf{Z}_n$=$\mathbf{z}_n$, thus the last inequality (2) holds given $\mathbf{Z}_n = \mathbf{z}_n$.It follows by averaging that (2) holds unconditionally. □

In order to prove **Lemma 4** we start with following simple lemma, which is crucial for our argument.

Lemma 1. *Assume that as* $r \to \infty$, $P(W_i^{(r)} \le t \mid W_1^{(r)}, \ldots, W_{i-1}^{(r)}) \xrightarrow{a.s.} P(Q_i \le t) =: F_i(t)$ *for all continuity points of* F_i, $i = 1, \ldots, d$. *Then* $(W_1^{(r)}, \ldots, W_d^{(r)}) \xrightarrow{d} (Q_1, \ldots, Q_d)$ *and* $(Q_i)_{i=1}^d$ *are independent.*

Proof. Assume that t_i is a continuity point of F_i. Then for $i = 1, \ldots, d$,

$$\begin{aligned} P(W_1^{(r)} \le t_1, \ldots, W_i^{(r)} \le t_i) &= P(W_1^{(r)} \le t_1, \ldots, W_{i-1}^{(r)} \le t_{i-1}) F_i(t_i) \\ &+ E\left[\mathbb{I}(W_1^{(r)} \le t_1, \ldots, W_{i-1}^{(r)} \le t_{i-1}) \left(P(W_i^{(r)} \le t_i \mid W_1^{(r)}, \ldots, W_{i-1}^{(r)}) - F_i(t_i)\right)\right]. \end{aligned}$$

By Lebesgue's dominated convergence theorem, the latter term converges to 0 as $r \to \infty$. Thus, by induction, the cumulative distribution function of $(W_1^{(r)}, \ldots, W_d^{(r)})$ converges to $F_1 \cdot \ldots \cdot F_d$ for all continuity points, which completes the proof. □

The above result generalizes to the case when all $W_i^{(r)}$ are multivariate.

Lemma 2. *Let* $m_r = (m_1^{(r)}, \ldots, m_d^{(r)})^\top \in \mathbb{N}^d$. *Suppose that* $W_r = (W_1^{(r)}, \ldots, W_d^{(r)})$ *has multivariate hypergeometric distribution* $\mathrm{Hyp}_d(n_r, m_r)$ *defined by*

$$P(W_r = (k_1, \ldots, k_d)) = \frac{\prod_{i=1}^d \binom{m_i^{(r)}}{k_i}}{\binom{|m_r|}{n_r}}, \qquad k_i \in \mathbb{N}, \quad k_i \le m_i^{(r)}, \quad \sum_{i=1}^{d} k_i = n_r.$$

Assume that as $r \to \infty$,

$$|m_r| \to \infty, \qquad n_r/|m_r| \to \alpha \in (0,1), \qquad m_r/|m_r| \to \beta = (\beta_1, \ldots, \beta_d) \in T_d.$$

Then

$$\frac{1}{\sqrt{|m_r|}}\left(W_r - \frac{n_r}{|m_r|} m_r^\top\right) \xrightarrow{d} N_d(0, \Sigma),$$

where Σ *is a* $(d-1)$*-rank matrix with elements* $\Sigma_{i,j} = \alpha(1-\alpha)\beta_i(\delta_{ij} - \beta_j)$.

The univariate case is proved in [1, Th. 2.1]. We could not find an appropriate reference for the general case. However, we refrain from giving a formal proof of the multivariate case, as it follows from the univiariate case in analogous way as Lemma 4 follows from Lemma 2 and we present a full argument below.

We now prove **Lemma 4**.

Proof. First, observe that (6) in Section 3 in the main text can be rewritten as

$$P(W_r = k) = \frac{\prod_{i=1}^{I} \binom{a_i^{(r)}}{k_{i1},\dots,k_{iJ}}}{\binom{n_r}{b_1^{(r)},\dots,b_J^{(r)}}},$$

where $\binom{a}{b_1,\dots,b_n} := \frac{a!}{\prod_{i=1}^n b_i!}$ whenever $a = |b|$. Denote by $W_i^{(r)}$ the ith row of matrix W_r, i.e. a random vector $(W_{ij}^{(r)})_{j=1}^J$, $i = 1,\dots,I$. Clearly, $W_1^{(r)} \sim \mathrm{Hyp}_J(a_1^{(r)}, b_r)$, where Hyp_J is defined in Lemma 2. Since $|b_r|= n_r$, by Lemma 2, we have

$$Z_1^{(r)} := \frac{1}{\sqrt{n_r}} \left(W_1^{(r)} - \frac{a_1^{(r)}}{n_r} b_r \right) \xrightarrow{d} Z_1 \sim N_d(0, \Sigma_1),$$

where $(\Sigma_1)_{i,j} = \alpha_1(1-\alpha_1)\beta_i(\delta_{ij} - \beta_j)$.

Now consider a conditional distribution of $W_i^{(r)}$ given $(W_k^{(r)})_{k<i}$, $i > 1$. We have

$$W_i^{(r)} \mid (W_k^{(r)})_{k<i} \sim \mathrm{Hyp}_J \left(a_i^{(r)}, b_r - \sum_{k=1}^{i-1} (W_k^{(r)})^\top \right).$$

Since $W_{ij}^{(r)}$ follows the hypergeometric distribution with parameters $\left(n_r, a_i^{(r)}, b_j^{(r)}\right)$ by the law of large numbers, we have

$$\frac{W_{ij}^{(r)}}{n_r} \xrightarrow{a.s.} \alpha_i \beta_j.$$

Observing that $m_r^{(i)} := |b_r - \sum_{k=1}^{i-1} W_k^{(r)}|= n_r - \sum_{k=1}^{i-1} a_k^{(r)}$, we have as $r \to \infty$,

$$\frac{a_i^{(r)}}{m_r^{(i)}} \to \frac{\alpha_i}{1 - \sum_{k=1}^{i-1} \alpha_k} \quad \text{and} \quad \frac{b_r^\top - \sum_{k=1}^{i-1} W_k^{(r)}}{m_r^{(i)}} \xrightarrow{a.s.} \beta.$$

We apply Lemma 2 conditionally on $(W_k^{(r)})_{k<i}$, to obtain for $i = 2,\dots,I$,

$$Z_i^{(r)} := \frac{1}{\sqrt{n_r - \sum_{k=1}^{i-1} a_k^{(r)}}} \left(W_i^{(r)} - \frac{a_i^{(r)}}{n_r - \sum_{k=1}^{i-1} a_k^{(r)}} \left(b_r^\top - \sum_{k=1}^{i-1} W_k^{(r)} \right) \right) \Big| \left(W_k^{(r)} \right)_{k<i} \xrightarrow{d} Z_i,$$

where $Z_i \sim N(0, \Sigma_i)$ with

$$(\Sigma_i)_{j,l} = \frac{\alpha_i}{1 - \sum_{k=1}^{i-1} \alpha_k} \left(1 - \frac{\alpha_i}{1 - \sum_{k=1}^{i-1} \alpha_k} \right) \beta_j (\delta_{jl} - \beta_l).$$

By Lemma 1, we have

$$(Z_1^{(r)}, \dots, Z_I^{(r)}) \xrightarrow{d} (Z_1, \dots, Z_I),$$

where $Z_1, \dots, Z_I$ are independent. By direct calculation, it is easy to see that

$$\frac{1}{\sqrt{n_r}} \left(W_i^{(r)} - \frac{1}{n_r} a_i^{(r)} b_r^\top \right) = \sum_{k=1}^{i} \gamma_{k,i}^{(r)} Z_k^{(r)},$$

where

$$\gamma_{k,i}^{(r)} = -\sqrt{\frac{n_r - \sum_{j=1}^{k-1} a_j^{(r)}}{n_r}} \frac{a_i^{(r)}}{n_r - \sum_{j=1}^{k} a_j^{(r)}} \quad \text{for } k < i \text{ and } \quad \gamma_{i,i}^{(r)} = \sqrt{\frac{n_r - \sum_{j=1}^{i-1} a_j^{(r)}}{n_r}}$$

We have $\lim_{r\to\infty} \gamma_{k,i}^{(r)} = \Gamma_{k,i}$, where

$$\Gamma_{k,i} = -\sqrt{1 - \sum_{j=1}^{k-1} \alpha_j} \frac{\alpha_i}{1 - \sum_{j=1}^{k} \alpha_j} \quad \text{for } k < i \text{ and } \quad \Gamma_{i,i} = \sqrt{1 - \sum_{j=1}^{i-1} \alpha_j}.$$

Thus,

$$\frac{1}{\sqrt{n_r}}\left(W_i^{(r)}-\frac{1}{n_r}a_i^{(r)}b_r^\top\right)_{i=1}^{I}\xrightarrow{d}\left(\sum_{k=1}^{i}\Gamma_{k,i}Z_k\right)_{i=1}^{I}=:Q\sim N(0,\Sigma),$$

where $\Sigma=(\Sigma_{i,j}^{k,l})$. $\Sigma_{i,j}^{k,l}$ denotes covariance of jth coordinate of ith consecutive subvector of the length J of Q with kth coordinate of the lth subvector. Thus

$$\Sigma_{i,j}^{k,l}=\operatorname{Cov}\left(\sum_{\ell=1}^{i}\Gamma_{\ell,i}Z_{\ell,j},\sum_{\ell=1}^{k}\Gamma_{\ell,k}Z_{\ell,l}\right).$$

Since no row is distinguished, in order to establish (7) in the main text it is enough to consider $i=1$ and $k\in\{1,2\}$. We have

$$\Sigma_{1,j}^{1,l}=\operatorname{Cov}(Z_{1,j},Z_{1,l})=(\Sigma_1)_{j,l}=\alpha_1(1-\alpha_1)\beta_j(\delta_{jl}-\beta_l)$$

and

$$\Sigma_{1,j}^{2,l}=\operatorname{Cov}\left(Z_{1,j},\sqrt{1-\alpha_1}Z_{2,l}-\frac{\alpha_2}{1-\alpha_1}Z_{1,l}\right)=-\frac{\alpha_2}{1-\alpha_1}(\Sigma_1)_{j,l}=-\alpha_1\alpha_2\beta_j\left(\delta_{jl}-\beta_l\right).$$

□

We prove now **Theorem 5**. The proof follows [2] and it is based on the multivariate Berry-Esseen theorem ([3]).

Proof of Theorem 5. Without loss of generality, we assume that $\mathcal{X}=\{1,2,\ldots,I\}$, $\mathcal{Y}=\{1,2,\ldots,J\}$ and $\mathcal{Z}=\{1,2,\ldots,K\}$ and let $M=I\cdot J\cdot K$. We define a function $k(\cdot)$, which assigns a triple $(x,y,z)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$ to each index $i=1,2,\ldots,M$, in the following way

$$k(i)=(x,y,z)\text{ and }i=x+I\cdot(y-1)+I\cdot J\cdot(z-1).$$

Thus, in the notation using the function k, we write e.g. a vector of all probabilities $(p(x,y,z))_{x,y,z}$ as $(p(k(i)))_{i=1}^M$. We let

$$\hat{p}^*(x,y,z)=\frac{n^*(x,y,z)}{n}=\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(X_i^*=x,Y_i=y,Z_i=z),$$

$p_{ci}=p(x|z)p(y|z)p(z)$ and we define $\hat{p}_{tci}$ (tci stands for *true conditional independence*) in the following way

$$\hat{p}_{tci}(x,y,z)=p(x|z)\frac{n(y,z)}{n(z)}\frac{n(z)}{n}=:p(x|z)\hat{p}(y|z)\hat{p}(z),$$

thus, since $\hat{p}^*$ follows the multinomial distribution with an observation (x,y,z) having a probability equal to $\hat{p}_{tci}(x,y,z)$, conditionally on the original sample we have that

$$\mathbb{E}^*\hat{p}^*(x,y,z)=p(x|z)\hat{p}(y|z)\hat{p}(z)$$

and

$$\left(\operatorname{Cov}^*\left((\hat{p}^*(x,y,z))_{x,y,z}\right)\right)_{x,y,z}^{x',y',z'}=\begin{cases}\frac{1}{n}\hat{p}_{tci}(x,y,z)(1-\hat{p}_{tci}(x,y,z)) & \text{if }(x,y,z)=(x',y',z')\\ -\frac{1}{n}\hat{p}_{tci}(x,y,z)\hat{p}_{tci}(x',y',z') & \text{if }(x,y,z)\neq(x',y',z')\end{cases}.$$

We define

$$\hat{\Sigma}_{x,y,z}^{x',y',z'}=n\left(\operatorname{Cov}^*\left((\hat{p}^*(x,y,z))_{x,y,z}\right)\right)_{x,y,z}^{x',y',z'}$$

and

$$Q_j^*:=\frac{1}{\sqrt{n}}\hat{\Sigma}_{-M}^{-1/2}\left(\mathbb{I}((X_j^*,Y_j,Z_j)=k(i))-\hat{p}_{tci}(k(i))\right)_{i=1}^{M-1},$$

$$W^*=\sum_{j=1}^{n}Q_j^*=\sqrt{n}\hat{\Sigma}_{-M}^{-1/2}\left(\hat{p}^*(k(i))-\hat{p}_{tci}(k(i))\right)_{i=1}^{M-1},$$

where $\hat{\Sigma}_{-M}=\operatorname{Cov}^*\left((\hat{p}^*(k(i)))_{i=1}^{M-1}\right)$. As $p(x,y,z)>0$ for all (x,y,z), the matrix $\hat{\Sigma}_{-M}$ is invertible, cf. e.g. [4]. One element of the vector $\hat{p}^*$ is omitted to ensure that the covariance matrix is invertible. As we have

$\sum_{x,y,z} \hat{p}^*(x,y,z) = 1$, the full dimension matrix $\hat{\Sigma}$ is singular. Then we apply multivariate Berry-Esseen theorem ([3])

$$\begin{aligned} &|P^*(W^* \in A) - P(Z \in A)| \\ &\leq K_d \sum_{j=1}^{n} \mathbb{E}^* \left\| \frac{1}{\sqrt{n}} \hat{\Sigma}_{-M}^{-1/2} \left(\mathbb{I}((X_j^*, Y_j, Z_j) = k(i)) - \hat{p}_{tci}(k(i)) \right)_{i=1}^{M-1} \right\|^3 \end{aligned} \tag{3}$$

and $d = M - 1$. We notice that as

$$\hat{p}_{tci} \to p_{ci} \text{ and } \hat{\Sigma}_{-M} \to \Sigma_{-M} \quad a.s.,$$

where Σ_{-M} denotes the matrix Σ without the last row and the last column, and for all $j = 1, 2, \ldots, M-1$

$$-1 \leq \mathbb{I}(X_j^* = x, Y_j = y, Z_j = z) - \hat{p}_{tci}(x,y,z) \leq 1,$$

we have that $\mathbb{E}^* \left\| \hat{\Sigma}_{-M}^{-1/2} \left(\mathbb{I}((X_j^*, Y_j, Z_j) = k(i)) - \hat{p}_{tci}(k(i)) \right)_{i=1}^{M-1} \right\|^3$ is bounded for almost all sequences. Thus in view of (3), conditionally, $W^* \to N(0, I)$ and as $\hat{\Sigma}_{-M}^{-1/2}$ converges to $\Sigma_{-M}^{-1/2}$ a.s., from Slutsky's theorem we have that

$$\sqrt{n} \left(\hat{p}^*(k(i)) - \hat{p}_{tci}(k(i)) \right)_{i=1}^{M} \xrightarrow{d} N(0, \Sigma_{-M}).$$

Now the conclusion follows by the continuous mapping theorem. □

We prove now the lemma which is used in the proof of **Theorem 6**.

Lemma 3. *Matrices $M = H_{CMI}\Sigma$ and $\tilde{M} = H_{CMI}\tilde{\Sigma}$ defined in the proof of Theorem 6 are equal, idempotent and their trace $tr(M) = tr(\tilde{M}) = (|\mathcal{X}|-1)(|\mathcal{Y}|-1)|\mathcal{Z}|$*

Proof. We show the result for $\tilde{M}$. The proof in the case of M is the same but more tedious (we skip the details). Matrix $M = H\Sigma = H_{CMI}(p_{ci})\Sigma$, where Σ is an asymptotic covariance matrix for CR scenario, has the following form

$$\begin{aligned} M_{x,y,z}^{x'',y'',z''} &= \mathbb{I}(x = x'', y = y'', z = z'') - \mathbb{I}(x = x'', z = z'')p(y''|z'') \\ &\quad - \mathbb{I}(y = y'', z = z'')p(x''|z'') + \mathbb{I}(z = z'')p(x''|z'')p(y''|z''). \end{aligned} \tag{4}$$

Multiplication of matrices H and Σ yields:

$$\begin{aligned} \tilde{M}_{x,y,z}^{x'',y'',z''} &= \sum_{x',y',z'} H_{x,y,z}^{x',y',z'} \tilde{\Sigma}_{x',y',z'}^{x'',y'',z''} = \sum_{x',y',z'} \Bigg(\underbrace{\frac{\mathbb{I}(x = x', y = y', z = z')}{p(x,y,z)}}_{a} - \underbrace{\frac{\mathbb{I}(x = x', z = z')}{p(x,z)}}_{b} \\ &- \underbrace{\frac{\mathbb{I}(y = y', z = z')}{p(y,z)}}_{c} + \underbrace{\frac{\mathbb{I}(z = z')}{p(z)}}_{d} \Bigg) \Big(- \underbrace{\mathbb{I}(y' = y'', z' = z'')p(x'|z')p(x''|z'')p(y',z')}_{e} \\ &+ \underbrace{\mathbb{I}(x' = x'', y' = y'', z' = z'')p(x'|z')p(y',z')}_{f} \Big) = - \underbrace{\mathbb{I}(y = y'', z = z'')p(x''|z'')}_{a \cdot e} \\ &+ \underbrace{\mathbb{I}(z = z'')p(x''|z'')p(y''|z'')}_{b \cdot e} + \underbrace{\mathbb{I}(y = y'', z = z'')p(x''|z'')}_{c \cdot e} - \underbrace{\mathbb{I}(z = z'')p(x''|z'')p(y''|z'')}_{d \cdot e} \\ &+ \underbrace{\mathbb{I}(x = x'', y = y'', z = z'')}_{a \cdot f} - \underbrace{\mathbb{I}(x = x'', z = z'')p(y''|z'')}_{b \cdot f} - \underbrace{\mathbb{I}(x = x'', z = z'')p(x''|z'')}_{c \cdot f} \\ &+ \underbrace{\mathbb{I}(z = z'')p(x''|z'')p(y''|z'')}_{d \cdot f} = \mathbb{I}(x = x'', y = y'', z = z'') - \mathbb{I}(x = x'', z = z'')p(y''|z'') \\ &- \mathbb{I}(y = y'', z = z'')p(x''|z'') + \mathbb{I}(z = z'')p(x''|z'')p(y''|z''). \end{aligned}$$

Below we present detailed calculations for the terms $c \cdot e$ and $d \cdot f$ (the calculations for other terms are analogous):

$$
\begin{aligned}
c\cdot e &= \sum_{x',y',z'} \mathbb{I}(y=y',z=z')\mathbb{I}(y'=y'',z'=z'')\frac{p(x'|z')p(x''|z'')p(y',z')}{p(y,z)} \\
&= \mathbb{I}(y=y'',z=z'')\sum_{x'}\frac{p(x'|z)p(x''|z'')p(y,z)}{p(y,z)} = \mathbb{I}(y=y'',z=z'')p(x''|z'')\sum_{x'}p(x'|z) \\
&= \mathbb{I}(y=y'',z=z'')p(x''|z''), \\
d\cdot f &= \sum_{x',y',z'} \mathbb{I}(z=z')\mathbb{I}(x'=x'',y'=y'',z'=z'')\frac{p(x'|z')p(y',z')}{p(z)} \\
&= \mathbb{I}(z=z'')\frac{p(x''|z'')p(y'',z'')}{p(z)} = \mathbb{I}(z=z'')p(x''|z'')p(y''|z'').
\end{aligned}
$$

We now show that $tr(\tilde{M}) = |\mathcal{X}|-1)(|\mathcal{Y}|-1)|\mathcal{Z}|$ and $\tilde{M}^2 = \tilde{M}$

$$
\begin{aligned}
\sum_{x,y,z}\tilde{M}^{x,y,z}_{x,y,z} &= \sum_{x,y,z}(1-p(y|z)-p(x|z)+p(x|z)p(y|z)) \\
&= |\mathcal{X}|\cdot|\mathcal{Y}|\cdot|\mathcal{Z}|-|\mathcal{X}|\cdot|\mathcal{Z}|-|\mathcal{Y}|\cdot|\mathcal{Z}|+|\mathcal{Z}|= (|\mathcal{X}|-1)(|\mathcal{Y}|-1)|\mathcal{Z}|
\end{aligned}
$$

We compute now $(\tilde{M}^2)^{x'',y'',z''}_{x,y,z}$. The first term in the first bracket is multiplied by the consecutive terms in the second bracket, then the second term in the first bracket and so on:

$$
\begin{aligned}
\sum_{x',y',z'}\tilde{M}^{x',y',z'}_{x,y,z}\tilde{M}^{x'',y'',z''}_{x',y',z'} &= \big(\mathbb{I}(x=x',y=y',z=z') - \mathbb{I}(x=x',z=z')p(y'|z') \\
&- \mathbb{I}(y=y',z=z')p(x'|z') + \mathbb{I}(z=z')p(x'|z')p(y'|z')\big)\cdot\big(\mathbb{I}(x'=x'',y'=y'',z'=z'') \\
&- \mathbb{I}(x'=x'',z'=z'')p(y''|z'') - \mathbb{I}(y'=y'',z'=z'')p(x''|z'') + \mathbb{I}(z'=z'')p(x''|z'')p(y''|z'')\big) \\
&= (\mathbb{I}(x=x'',y=y'',z=z'') - \mathbb{I}(x=x'',z=z'')p(y''|z'') - \mathbb{I}(y=y'',z=z'')p(x''|z'') \\
&+ \mathbb{I}(z=z'')p(x''|z'')p(y''|z'')) - (\mathbb{I}(x=x'',z=z'')p(y''|z'') - \mathbb{I}(x=x'',z=z'')p(y''|z'') \\
&- \mathbb{I}(z=z'')p(x''|z'')p(y''|z'') + \mathbb{I}(z=z'')p(x''|z'')p(y''|z'')) - (\mathbb{I}(y=y'',z=z'')p(x''|z'') \\
&- \mathbb{I}(z=z'')p(x''|z'')p(y''|z'') - \mathbb{I}(y=y'',z=z'')p(x''|z'') + \mathbb{I}(z=z'')p(x''|z'')p(y''|z'')) \\
&+ (\mathbb{I}(z=z'')p(x''|z'')p(y''|z'') - \mathbb{I}(z=z'')p(x''|z'')p(y''|z'') - \mathbb{I}(z=z'')p(x''|z'')p(y''|z'') \\
&+ \mathbb{I}(z=z'')p(x''|z'')p(y''|z'')) = \mathbb{I}(x=x'',y=y'',z=z'') - \mathbb{I}(x=x'',z=z'')p(y''|z'') \\
&- \mathbb{I}(y=y'',z=z'')p(x''|z'') + \mathbb{I}(z=z'')p(x''|z'')p(y''|z'') = M^{x'',y'',z''}_{x,y,z}.
\end{aligned}
$$

□

We prove now two lemmas which justify choice of null distributions in the numerical experiments.

Lemma 4. *Probability mass function $p_{ci}(x,y,z) = p(x|z)p(y|z)p(z)$ minimises $D_{KL}(p||q)$ over $q \in \mathcal{P}_{ci}$ defined as*

$$
\mathcal{P}_{ci} = \{q(x,y,z) : q(x,y,z) = q(x|z)q(y|z)q(z)\}.
$$

Proof. Indeed,

$$
\begin{aligned}
&D_{KL}(p||q) - D_{KL}(p||p_{ci}) \qquad (5)\\
&\quad = \sum_{x,y,z}p(x,y,z)\log\frac{p(x,y,z)}{q(x,y,z)} - \sum_{x,y,z}p(x,y,z)\log\frac{p(x,y,z)}{p(x|z)p(y|z)p(z)} \\
&\quad = \sum_{x,y,z}p(x,y,z)\log\frac{p(x|z)p(y|z)p(z)}{q(x|z)q(y|z)q(z)}.
\end{aligned}
$$

Next, by breaking the above expression into three sums, we obtain

$$
\sum_z p(z)\sum_x p(x|z)\log\frac{p(x|z)}{q(x|z)} + \sum_z p(z)\sum_y p(y|z)\log\frac{p(y|z)}{q(y|z)} + \sum_z p(z)\log\frac{p(z)}{q(z)}.
$$

The expression $\sum_x p(x|z)\log\frac{p(x|z)}{q(x|z)}$ is equal to Kullback-Leibler divergence of $p(x|z)$ and $q(x|z)$ for a fixed value of Z (similarly $\sum_y p(y|z)\log\frac{p(y|z)}{q(y|z)} = D_{KL}(p(\cdot|z)||q(\cdot|z))$ and $\sum_z p(z)\log\frac{p(z)}{q(z)} = D_{KL}(p||q)$). Thus (5) is non-negative and equal to 0 if and only if $q(x|z) = p(x|z)$, $q(y|z) = p(y|z)$ and $q(z) = p(z)$. □

Lemma 5. *Probability mass function p_{ci} minimises $D_{KL}(p_\lambda||q)$ over $q \in \mathcal{P}_{ci}$ such that*

$$\mathcal{P}_{ci} = \{q(x,y,z) : q(x,y,z) = q(x|z)q(y|z)q(z)\},$$

where $p_\lambda = \lambda p_{ci} + (1-\lambda)p$ and $\lambda \in [0,1]$

Proof. In view of Lemma 4 it is enough to show that $p_{\lambda,ci} = p_{ci}$ what, due to the form of p_{ci} will follow from $p_\lambda(x,z) = p(x,z)$ and $p_\lambda(y,z) = p(y,z)$.

We have that

$$p_\lambda(x,z) = \sum_y p_\lambda(x,y,z) = \sum_y \big(p_{ci}(x,y,z) + (1-\lambda)p(x,y,z)\big)$$
$$= \lambda p(x|z)\sum_y p(y|z)p(z) + (1-\lambda)p(x,z) = p(x,z).$$

Similarly, we have that $p_\lambda(y,z) = p(y,z)$. Thus $p_{\lambda,ci} = p_{ci}$. □

We prove now that the asymptotic covariance matrices in Conditional Permutation and Conditional Randomisation scenario are ordered (see **Remark 3** in the main text).

Lemma 6. *The covariance matrix for CR scenario dominates the covariance matrix for CP scenario:*

$$\tilde{\Sigma} \geq \Sigma$$

i.e. matrix $\tilde{\Sigma} - \Sigma$ is positive semi-definite.

Proof. We prove $\tilde{\Sigma} \geq \Sigma$. Define

$$(R)_{x,y,z}^{x',y',z'} = (\tilde{\Sigma} - \Sigma)_{x,y,z}^{x',y',z'} = \mathbb{I}(z=z')\big[\mathbb{I}(x=x')p(x|z)p(y,z)p(y',z)/p(z)$$
$$- p(x|z)p(x'|z)p(y,z)p(y',z)/p(z)\big].$$

We note that for any z the matrix $\tilde{R}(z)$ defined as

$$(\tilde{R}(z))_x^{x'} = r_x^{x'}(z) = \mathbb{I}(x=x')p(x|z) - p(x|z)p(x'|z)$$

is positive semi-definite. Now we define elements of matrix $\bar{R}(z) = (r_{x,y}^{x',y'}(z))_{x,y}^{x',y'}$ as

$$r_{x,y}^{x',y'}(z) = r_x^{x'}(z)p(y,z)p(y',z)$$

and we show that $\bar{R}(z) \geq 0$. Namely, for any non-zero vector $a = (a(x,y))_{x,y}$ it holds

$$a'\bar{R}(z)a = \sum_{x,y}\sum_{x',y'} a_{x,y} r_{x,y}^{x',y'}(z) a_{x',y'} = \sum_{x,y}\sum_{x',y'} a_{x,y} r_x^{x'}(z)p(y,z)p(y',z)a_{x',y'}$$
$$= \sum_{x,x'}\left(\sum_y a_{x,y}p(y,z)\right) r_x^{x'}(z) \left(\sum_{y'} a_{x',y'}p(y',z)\right) \geq 0,$$

where the last inequality follows as $\tilde{R}(z) \geq 0$. However,

$$(R)_{x,y,z}^{x',y',z'} = r_{x,y,z}^{x',y',z'} = r_{x,y}^{x',y'}\mathbb{I}(z=z')/p(z),$$

thus for any non-zero vector $a = (a(x,y,z))_{x,y,z}$ we have that

$$a'Ra = \sum_{x,y,z}\sum_{x',y',z'} a_{x,y,z} r_{x,y,z}^{x',y',z'} a_{x',y',z'} = \sum_{x,y,z}\sum_{x',y',z'} a_{x,y,z} r_{x,y}^{x',y'}(z)\mathbb{I}(z=z')/p(z)a_{x',y',z'}$$
$$= \sum_z \left(\sum_{x,y}\sum_{x',y'} a_{x,y,z} r_{x,y}^{x',y'}(z) a_{x',y',z}\right)/p(z) \geq 0.$$

□

References

[1] S. Lahiri and A. Chatterjee. A Berry-Esseen theorem for hypergeometric probabilities under minimal conditions. *Proc. Am. Math. Soc.*, 135(5):1535–1545, 2007.

[2] Kesar Singh. On the Asymptotic Accuracy of Efron's Bootstrap. *Ann. Stat.*, 9(6):1403–1433, nov 1981.

[3] V. Bentkus. A Lyapunov-type bound in $\mathbb{R}^d$. *Theory Probab. its Appl.*, 49:311–371, 2005.

[4] George A. F. Seber. *A Matrix Handbook for Statisticians.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Hoboken, New Jersey, 2008.