

8

Once we calculate Reproducibility (MC_1), Reproducibility (MC_2), ..., Reproducibility (MC_{n_C}), how can we calculate a cutoff to separate the high reproducibility ICs from the low reproducibility ICs?

unique integer associated with each component in each run

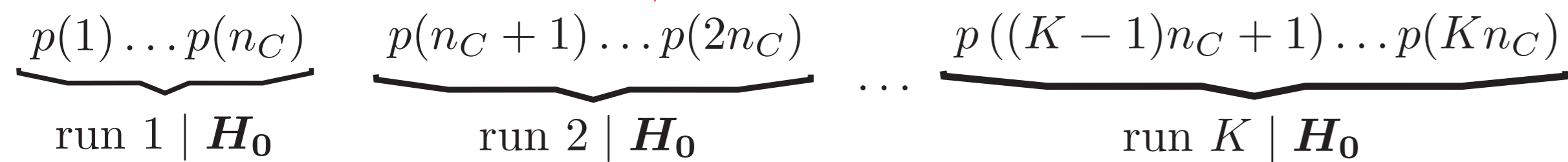


H_0 : Null hypothesis is that none of the ICs are reproducible. Hence we can randomly label IC i from run l as IC d from run s . This random labelling produces one realization of ICs under H_0 .

random labelling

$$p(i) \neq p(j) \text{ if } i \neq j$$

integers $1, 2 \dots Kn_C$ $\xrightarrow{\text{random permutation of integers}}$ permuted integers $p(1), p(2) \dots p(Kn_C)$



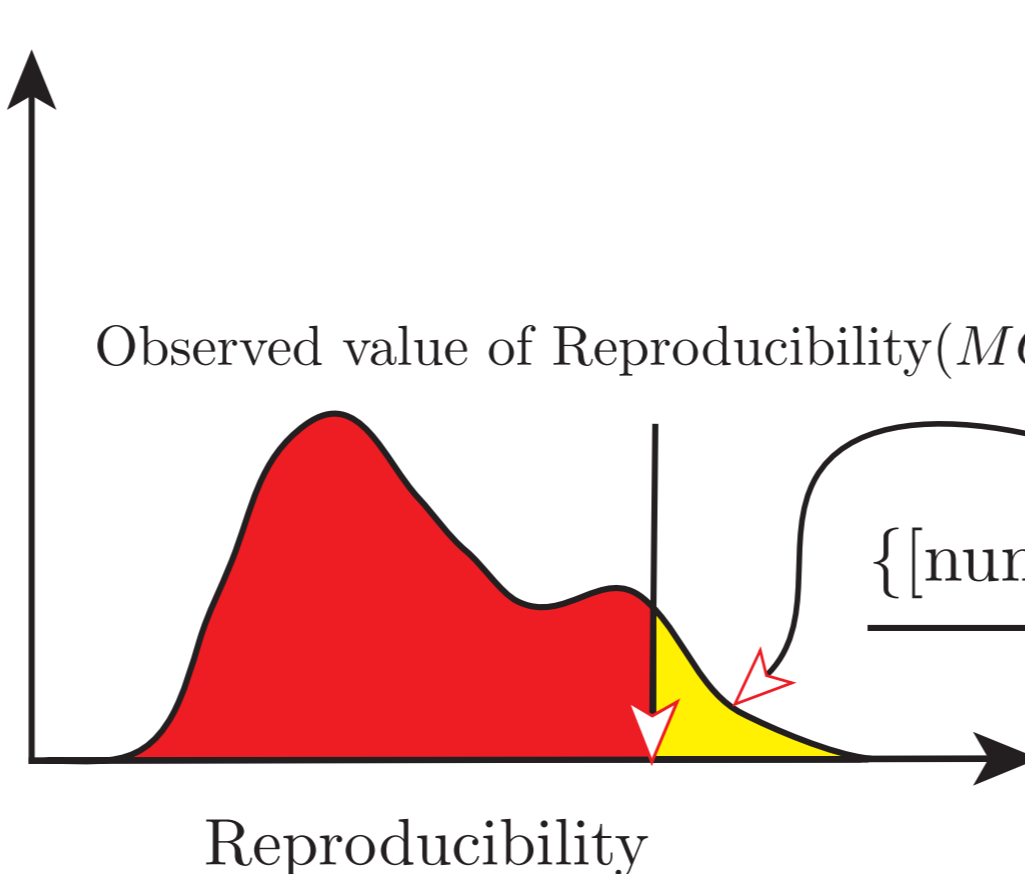
reproducibility calculation for 1 realization of ICs under H_0

- We repeat the reproducibility calculation for R realizations of ICs (e.g. $R = 100$) under H_0 using the random labelling process described above.
- This gives us a distribution of "Reproducibility" values under H_0 . Each realization of H_0 produces n_C "null" reproducibility values.
- Hence for R realizations we get $R \times n_C$ values that define the null distribution for testing the observed reproducibility values Reproducibility(MC_i).
- Denote the vector of this "null" reproducibility values by Reproducibility $_{Null}$

Calculate p -values for Reproducibility

13

Distribution of "Reproducibility" values under null hypothesis H_0



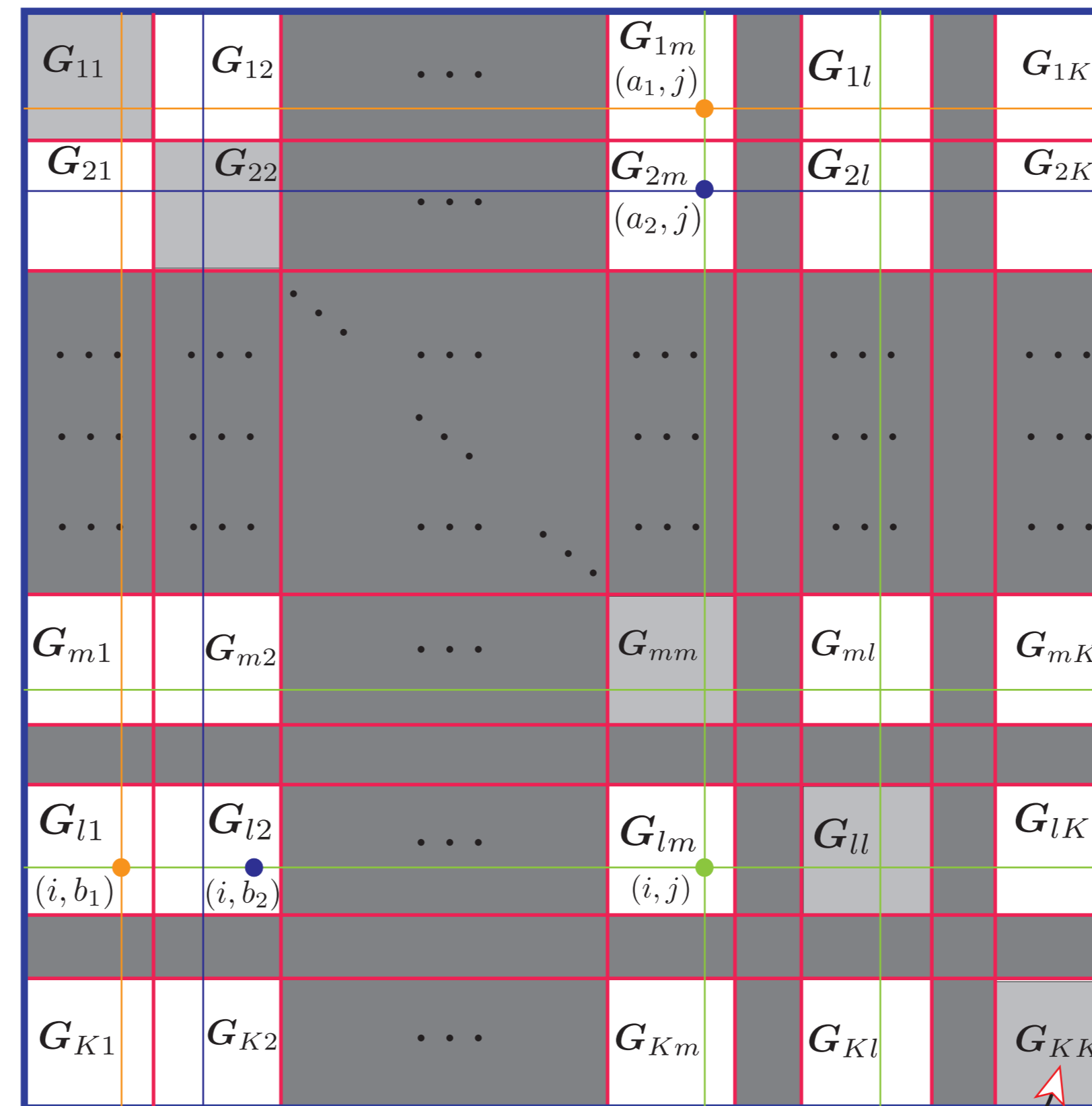
Reproducibility p -value for $MC_i =$

$$\frac{\{[\text{number of Reproducibility}_{Null} \geq \text{Reproducibility}(MC_i)] + 1\}}{(R n_C + 1)}$$

1

K = number of ICA runs
 n_C = number of extracted ICs in each run
 n = length of each IC
 $\mathbf{x}_j^{(m)}$ = $n \times 1$ vector of the j th IC from m th ICA run

cross-realization cross-correlation matrix (CRCM) \mathbf{G}



diagonal $n_C \times n_C$ block matrices are ignored in component matching

2

$$\mathbf{G}_{lm}(i, j) = |\text{corrcoef}(\mathbf{x}_i^{(l)}, \mathbf{x}_j^{(m)})|$$

3

- maximal element of \mathbf{G} occurs in \mathbf{G}_{lm} at position (i, j)
 - Hence component i from run l matches component j from run m
 - Denote this matched component by MC_1

4

- element (a_1, j) is the maximal element in the j th column of \mathbf{G}_{1m}
- element (i, b_1) is the maximal element in the i th row of \mathbf{G}_{l1}
 - In this case, $a_1 = b_1$.
 - Therefore we zero out the a_1 th row from $\mathbf{G}_{1r}, r = 1, \dots, K$
 - Similarly, we also zero out the b_1 th column from $\mathbf{G}_{r1}, r = 1, \dots, K$ (zeroed out rows and columns in this case are shown in Orange)

Add component a_1 from run 1 to MC_1

5

- element (a_2, j) is the maximal element in the j th column of \mathbf{G}_{2m}
- element (i, b_2) is the maximal element in the i th row of \mathbf{G}_{l2}
 - In this case, $a_2 \neq b_2$ and $\mathbf{G}_{2m}(a_2, j) > \mathbf{G}_{l2}(i, b_2)$
 - Therefore we zero out the a_2 th row from $\mathbf{G}_{2r}, r = 1, \dots, K$
 - Similarly, we also zero out the a_2 th column from $\mathbf{G}_{r2}, r = 1, \dots, K$ (zeroed out rows and columns in this case are shown in Blue)

Add component a_2 from run 2 to MC_1

6

After processing $\mathbf{G}_{rm}, r \neq l, m$ and $\mathbf{G}_{lr}, r \neq l, m$ we also

- Zero out the i th row of \mathbf{G}_{lr} and i th column of \mathbf{G}_{rl} for $r = 1, \dots, K$
- Zero out the j th column of \mathbf{G}_{rm} and the j th row of \mathbf{G}_{mr} for $r = 1, \dots, K$ (zeroed out rows and columns in this case are shown in Green)

12

$$\text{Reproducibility}(MC_1) = \left(\frac{2}{(K-1)K} \right) \sum_{a=1}^K \sum_{b=a+1}^K H_{MC_1}(a, b)$$

7

Suppose matched component 1, MC_1 consists of the matched ICs $\mathbf{x}_{i_1}^{(1)}, \mathbf{x}_{i_2}^{(2)}, \dots, \mathbf{x}_{i_K}^{(K)}$. Form the $K \times K$ cross-correlation matrix H_{MC_1} between the matched components in MC_1 . The (a, b) th element of this matrix is simply $H_{MC_1}(a, b) = |\text{corrcoef}(\mathbf{x}_{i_a}^{(a)}, \mathbf{x}_{i_b}^{(b)})|$. The normalized reproducibility of MC_1 is then defined as: