Suppose we have a group of N subjects. We randomly select L subjects and form a single group of subjects. We repeat this process K times to get K groups of L subjects each of which is subjected to a group ICA analysis. Given the number of subjects N, how should we choose L and K?

First, we discuss the choice of L. If $L = N$ then each of the K groups will contain the same N subjects. Hence there is no diversity in the K groups. We would like to control the amount of diversity in the K groups of L subjects. Consider any 2 subjects X and Y. The probability $P_{XY}(L)$ that both X and Y appear in a set of L randomly chosen subjects from N subjects is given by:

$$P_{XY}(L) = \frac{\binom{N-2}{L-2}}{\binom{N}{L}} \tag{1}$$

The expected number of times that X and Y appear together in sets of L subjects out of K independently drawn sets is:

$$E_{XY}(L) = K\, P_{XY}(L) \tag{2}$$

Ideally, we would like $E_{XY}(L)$ to be only a small fraction of K. Hence we impose the restriction:

$$E_{XY}(L) = K\, P_{XY}(L) \leq \alpha_{max}\, K \tag{3}$$

where α_{max} is a user defined constant such as $\alpha_{max} = 0.05$. This implies the chosen value of L must satisfy:

$$P_{XY}(L) \leq \alpha_{max} \tag{4}$$

In practice, we choose the largest value of L that satisfies this inequality. If $N = 23$ and $\alpha_{max} = 0.05$ then the largest value of L such that $P_{XY}(L) \leq \alpha_{max}$ is $L = 5$.

The number of group ICA runs K should be as large as possible. From our experiments on real fMRI data we can roughly say that values of $K > 50$ give equivalent results.

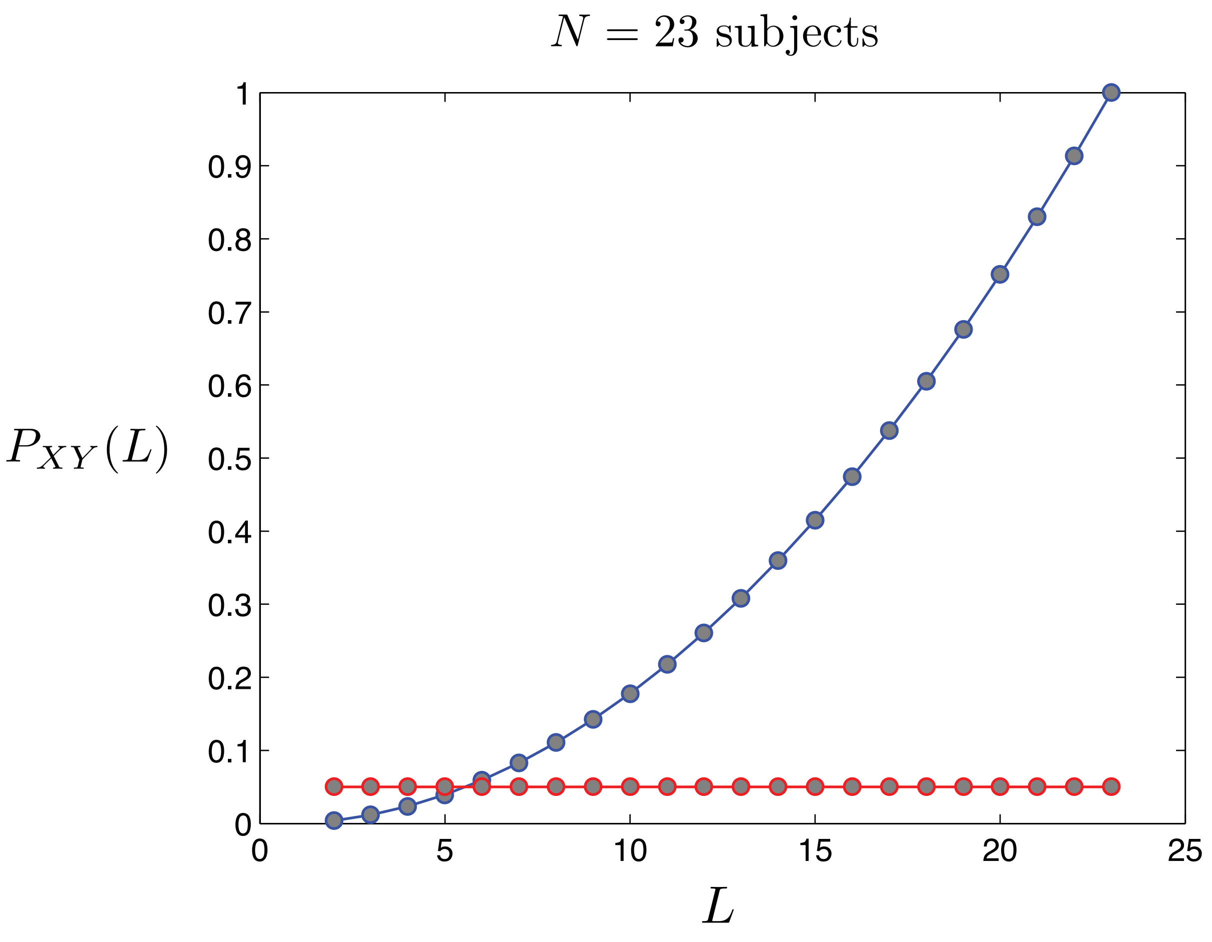

Figure shows a plot of $P_{XY}(L)$ vs L for $N = 23$ in blue. The red line shows the $\alpha_{max} = 0.05$ cutoff. The largest value of L for which $P_{XY}(L) \leq 0.05$ is $L = 5$.