Modeling cell populations in high content screening using copulas
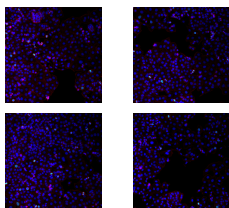
Edouard Pauwels, Mines Paristech, France

Introduction

Well

$\longrightarrow$

$\mathbf{X} = \{\mathbf{x_1}, \ldots, \mathbf{x_4}\}$
$\mathbf{x_i} \in \mathbb{R}^{+*} \times (0, 1)^4, \mathbf{i} = 1 \ldots 4$

Phenotypic classes
Outlier detection

Model

$P(\mathbf{X}|\Theta) = \sum_{\mathbf{Z}=1}^{K} P(\mathbf{Z}|\Theta) \prod_{\mathbf{j}=1}^{4} P(\mathbf{x_j}|\mathbf{Z}, \Theta)$

Class conditional distribution

- ▶ Gamma and beta univariate marginals
- ▶ Gaussian copula

$$P(x_1, \ldots, x_5 | R, \theta_1, \ldots, \theta_5) = c_R(F_1(x_1), \ldots, F_5(x_5)) \prod_{j=1}^{5} f_i(x_i)$$

Results

Comparison with a standard gaussian parametrization:

- ▶ Train and test likelihood
- ▶ Outlier detection
- ▶ Model distribution *v.s.* empirical distribution