

SpatialStats: statistical models of tumour cell signalling

Kieran Campbell
kieran.campbell@dpag.ox.ac.uk
August 28, 2014

Contents

| | | |
|----------|-----------------------------|----------|
| 1 | Introduction | 1 |
| 1.1 | Signalling Model | 1 |
| 1.2 | LASSO regression | 1 |
| 1.2.1 | Significance testing | 2 |
| 1.2.2 | Multiple testing correction | 2 |
| 2 | Workflow | 2 |

1 Introduction

1.1 Signalling Model

This vignette introduces the workflow for elucidating spatial signalling pathways from spatially resolved proteomics data. The proposed model is of the form

$$y_i^{(p)} = \mu_p + \sum_{j \in NN(i)} w_{j|i} \sum_{q \in P} \beta_{qp} x_j^{(q)} + \epsilon_{ip} \quad (1)$$

where $y_i^{(p)}$ is the concentration of protein p in cell i , $NN(i)$ are the nearest neighbour cells of i , $w_{j|i}$ is the relative boundary size of cell j to cell i (that is, the proportion of cell i 's boundary made up of cell j)¹, P is the set of all proteins considered, β_{qp} is the effect of nearest neighbour protein q on protein p in the central cell, and $x_j^{(q)}$ is the concentration of protein q in cell j . The eventual quantities of real interest are the $\{\beta\}$, which give a measure of one protein's effect on the other in nearby cells and would eventually elucidate inter-cellular signalling pathways. The simple average can also be used, rather than the average weighted by the relative boundary size. The effect of this is that $w_{j|i} = \frac{1}{N_i}$ if cell i has N_i neighbours.

1.2 LASSO regression

Rather than ordinary least squares regression, *SpatialStats* uses LASSO regression - a form of regularization where the magnitude of each coefficient is penalised. This is motivated on three fronts: (1) it avoids overfitting given the large number of predictor variables, (2) it scales well to $p > n$ problems of a subset of cells is selected, and (3) it fits well with the assumption that the underlying interaction matrix is sparse, since we don't expect many channels to spatially interact with the others.

In LASSO regression the objective function becomes

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_i \left(y_i - \sum_j \beta_j x_{ij} \right)^2 + \lambda \sum_j |\beta_j| \quad (2)$$

¹Note that in general $w_{j|i} \neq w_{i|j}$

where the penalisation parameter λ is introduced, which controls the penalty the L^1 norm of the coefficients has on the minimisation. Several packages are available in R to do LASSO regression. Here we use two - *glmnet* and *lars* - depending on the type of significance testing used (see below).

1.2.1 Significance testing

1.2.2 Multiple testing correction

2 Workflow
