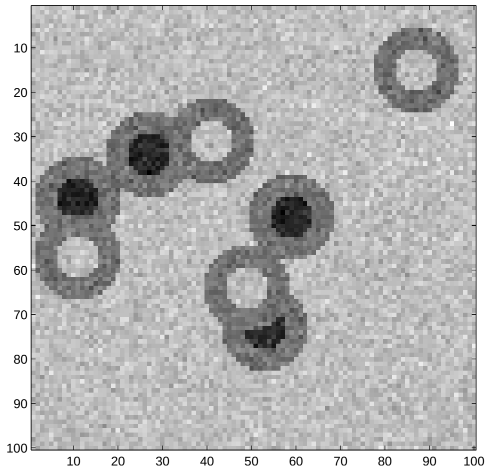
ELEC 446 : Computational Inference
Assignment 3

Due on Monday 19 September, or earlier. By email, please.

1. The following picture shows a (synthetic) noisy scene containing 'good' (black inside) and 'bad' (white inside) cells. Your task is to write an MCMC that automatically counts the number of good and bad cells, and displays the marginal posterior distribution over cell counts.



The image is a 100×100 pixel image, scaled between 0 (black) and 1 (white). You can download the array of numbers stored in the file `slide.mat` on the ELEC446 website. Also there is the code (`makefake.m`) that produced this image.

As you can see, the 'true' image is defined by an ordered set of (x, y) locations of cells (I have used pixel count as the units), with each location being either 'good' or 'bad'. You should model the *number* of cells as unknown, so random, with a geometric distribution with some sensible mean (you can investigate this hyperparameter later). The *location* of each cell can be modelled as uniformly distributed within the image.

Write down the prior distribution that this model implies.

By looking at the added noise term (in `makefake.m`), write down the likelihood function.

Now write an Metropolis-Hastings MCMC to explore the (un-normalized) posterior distribution. Your MCMC should utilize at least three *moves*:

(a) A birth-death move that creates or deletes cells.

(b) A swap move, that flips the label of a cell from 'good' to 'bad', and vice versa.

(c) A translate move, that randomly translates a cell within a window.

(d) Anything else you can think of that seems a good idea.

For each move, you will need to implement an appropriate accept-reject probability.

Show that the chain you simulate is irreducible, aperiodic, and in detailed balance with the target posterior distribution.

Plot the marginal posterior distribution over the number of good and bad cells.

Give a measure of the statistical efficiency of your MCMC algorithm.

2. (Challenge question – for the brave) Put a (hyper) prior distribution over the mean number of cells, and present results that integrate over this nuisance parameter. Does this make the counting more robust (simulate more data sets to test this in a frequentist sense.)

Colin Fox
Sep 2016