BAYESIAN PROCESSING OF MICROARRAY IMAGES

Neil D. Lawrence†, Marta Milo†, Mahesan Niranjan†,
Penny Rashbass‡and Stephan Soullier‡

†Department of Computer Science,
Regent Court, 211 Portobello Road, Sheffield, S1 4DP, U.K.
{neil,marta,niranjan}@dcs.shef.ac.uk
‡Centre for Developmental Genetics,
University of Sheffield School of Medicine and Biomedical Science,
Firth Court, Western Bank, Sheffield, S10 2TN, U.K.
{P.Rashbass,S.Soullier}@sheffield.ac.uk

Abstract. Gene expression measurements quantify the level of mRNA produced from each gene. Two principal methods exist for producing slides for extracting these levels: photolithography and spotted arrays. One difficulty with the spotted array format is determining the size and location of the spots on the array. In this paper we present a Bayesian approach to processing images produced by these arrays that seeks posterior distributions over the size and positions of the spots. This enables us to estimate expression ratios *and* their variances. Exact inference for the model we specify is intractable; we develop an approximate inference technique which combines importance sampling with variational inference. Our technique has already been shown to be more consistent than both manual processing and another automated technique [6]. Here we present large-scale results for twenty-four microarray slides each representing 5760 genes and show the dramatic effects of incorporating variance in our downstream analysis. Software based on this algorithm is available for academic use.

INTRODUCTION

Microarrays allow the simultaneous measurement of thousands of gene expression levels [1]. In cDNA microarrays genetic material representing the sequences of individual genes are 'spotted' on a slide. mRNA from two biological samples is then 'reverse transcripted' into DNA which is 'tagged' with different fluorescent dyes. This DNA is then hybridised to the slide's spots and the slide is scanned to give two images, one for each biological sample, each of which varies in intensity according to the quantity of two dyes present (Figure 1). The intensity of each dye present at each spot is interpreted as indicating the corresponding gene's expression level

in the biological sample, or, more precisely, the ratios of the dyes are interpreted as indicating the ratio of the gene expression level. The two images associated with the different samples are normally visualised together in colour with one sample being placed on the red channel of the resulting image and the other sample placed on the green channel.

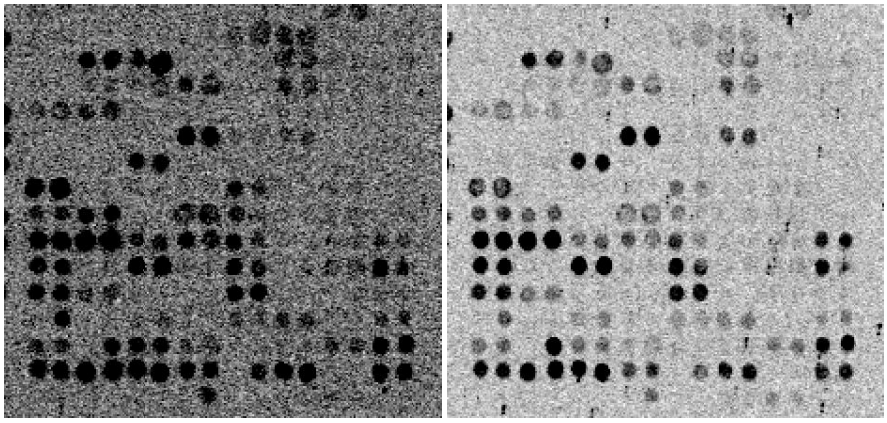

Figure 1: *Left and Right*: Two channels with intensities scaled by a factor of six from a microarray shown as 'negatives' for clearer printing. *Left*: The 'red' sample representing the response to the mutant's sample. *Right*: The 'green' channel representing the response to the wild-type sample. The images show one sub-grid of the slide which contains a total of forty-eight sub-grids. The position and shape of each of the spots must be identified. Note: (i) The higher background level for the red channel, (ii) the dust spot in the upper right of the green channel and (iii) a faint 'tide mark' on the right hand side of the green channel. (iv) Spots missing from the grid are either: genes that are not present in the sample or points where the cDNA laid down by the spotter has been damaged.

Extracting the Intensities. To determine the gene expression level some assessment of the dye intensity must be made at each spot. One common approach is to specify the x-radius and y-radius of an oval which separates each spot from the surrounding background. Pixels within the oval are then considered to represent the samples' gene expression levels whilst those outside the oval are termed as 'background'. ScanAlyze[1] is one widely utilised software package which allows the user to specify the parameters of the ovals and extracts the gene expression ratios given these ovals. Whilst an initial rough placement of grids of the ovals is often available, or simple to lay out, the process of refining the ovals' positions on an individual basis is repetitive and labourious. Here we present a Bayesian algorithm which not only refines these positions automatically, but also provides a variance on quantities of interest which may be used in further downstream analysis.

[1] ScanAlyze is available free for academic use and downloadable from `http://rana.lbl.gov/EisenSoftware.htm`.

Noise in the Measurements. The gene expression level extracted by the software is subject to 'noise' arising from several different sources. Techniques exist which aim to compensate for some of the sources of this noise, these include considering ratios of the intensities and normalising the image channels [11]. However, in practice, the only way forward is repeating experiments in an effort to quantify the noise. In this paper we are concerned with noise caused by misplacement of the oval. There are two important causes of this noise: firstly, manual processing of images containing thousands of spots leads to human error in placement of the ovals due to the tedious and lengthy[2] nature of the task. Secondly, even if ample time were given to a researcher, it is not necessarily clear which portions of the image should be assigned to the background and which we should assign to the foreground, *i.e.* there is uncertainty in the placement. The presence of uncertainty leads us to consider processing of the images in a Bayesian manner. It has already been shown how with a Bayesian approach more consistent results can be obtained [6], here we give details of the mathematical algorithm and demonstrate how accounting for uncertainty in the measurements modifies the downstream analysis.

THE BAYESIAN APPROACH

The Bayesian approach involves encapsulating the information we initially have about the ith oval's parameters, $\boldsymbol{\theta}_i$, in a prior distribution $p(\boldsymbol{\theta}_i)$. We then construct a model of the observed image, I, given the ith spot $p(I|\boldsymbol{\theta}_i)$ which is known as the likelihood. Bayesian inference involves utilising Bayes' rule to determine the posterior distribution,

$$p(\boldsymbol{\theta}_i|I) \propto \frac{p(I|\boldsymbol{\theta}_i)\, p(\boldsymbol{\theta}_i)}{Z_i}, \tag{1}$$

which represents our belief about the oval's parameters given the image. The difficulty in determining the posterior normally lies in evaluation of the constant of proportionality

$$Z_i = \int p(I|\boldsymbol{\theta}_i)\, p(\boldsymbol{\theta}_i)\, d\boldsymbol{\theta}_i. \tag{2}$$

The Prior. A common choice of prior distribution is that of the Gaussian distribution, a more general distribution however is that of the multi-variate Student-t,

$$p(\boldsymbol{\theta}_i|\mathbf{m}_i, \mathbf{S}) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)\left(1+\frac{1}{\nu}\left(\boldsymbol{\theta_i}-\mathbf{m}_i\right)^{\mathbf{T}}\mathbf{S}^{-1}\left(\boldsymbol{\theta_i}-\mathbf{m}_i\right)\right)^{-\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu}{2}\right)|\mathbf{S}|^{\frac{1}{2}}(\nu\pi)^{\frac{d}{2}}},$$

where d is the number of parameters in the parameter vector, $\mathbf{m}_i \in \Re^{d\times 1}$ is a vector of means, $\mathbf{S} \in \Re^{d\times d}$ is a matrix which is related to the scale of the parameters and ν is the degrees of freedom parameter. The Student-t becomes Gaussian with

[2]For our data accurate placement of the spots by human took about 4-6 hours for each slide. Our algorithm, written in a combination of C++ and MATLAB, runs unattended for each slide in about half an hour on a 1.2GHz Pentium III machine.

mean $\mathbf{m}_i$ and covariance $\mathbf{S}$ as $\nu \to \infty$, otherwise it represents a more heavy tailed distribution. As a further refinement we place a spherical Gaussian prior over the mean vector of the Student-t,

$$p\left(\mathbf{m}_i|\boldsymbol{\mu}_i, \beta\right) = \frac{\beta^{\frac{d}{2}}}{(2\pi)^{\frac{d}{2}}} \exp\left(-\frac{\beta}{2}\left(\mathbf{m}_i - \boldsymbol{\mu}_i\right)^{\mathrm{T}}\left(\mathbf{m}_i - \boldsymbol{\mu}_i\right)\right),$$

which is parameterised by a mean $\boldsymbol{\mu}_i$ and a precision parameter β. One reason for the popularity of a Gaussian prior is that if an appropriate likelihood function is selected the integral in (2) will be tractable. It is more difficult to develop an appropriate likelihood for the Student-t distribution. However, in this application, as we shall see in the next section, the likelihood function renders the integral intractable for both choices of prior. Consequently we develop, a novel approximating technique which alleviates the requirement of solving that integral directly. This technique is very general and has already been applied to visual tracking in [9].

The Likelihood For the definition of the likelihood, consider that the oval defines an area as being from the spot, or foreground, and an area as being from the background. Since there are many spots, and we do not wish our specification of the background to include all the other spots, we constrain the area of the background to be a box around the oval of fixed width and height, w_{box}. Thus each oval allocates the pixels to a set which belong to the background, $\mathcal{I}_{\text{back}}$, and a set which belongs to the foreground, $\mathcal{I}_{\text{fore}}$. For simplicity, we utilise a likelihood which assumes independence between the pixels,

$$p\left(I|\boldsymbol{\theta}_i\right) = \prod_{j \in \mathcal{I}_{\text{fore}}} p\left(I_j|F\right) \prod_{k \in \mathcal{I}_{\text{back}}} p\left(I_k|B\right),$$

where $p\left(I_j|F\right)$ is the distribution of an individual pixel's red and green channels given that it's from the foreground and $p\left(I_k|B\right)$ is the distribution of a pixel's channels given that it's from the background. Such likelihood models are commonly utilised in computer vision (see *e.g.* [3]). We further simplified the likelihood by assuming independence between the red and green channels, $p\left(I_j|F\right) = p\left(r_j|F\right)p\left(g_j|F\right)$ where r_j and g_j are, respectively the intensities of the red and green channels at pixel j. The background model was treated similarly. We built histograms, based on the initial rough grid localisation, which represented the intensities of the foreground and background pixels for both the red and green channels. Note that, despite our simplifying assumptions, the likelihood function we have defined is highly non-linear in the parameters $\boldsymbol{\theta}_i$, rendering the integral in (2) intractable. Furthermore, derivatives of the likelihood with respect to the parameters $\boldsymbol{\theta}_i$ may not be computed analytically, precluding techniques such as the Laplace approximation.

APPROXIMATE INFERENCE

Substituting the likelihood and prior distribution specified above into (2) renders evaluation of the marginal likelihood intractable. Instead we must look to approxi-

mations to make progress.

Variational Inference In variational inference [5] we choose to lower bound the log-likelihood by seeking a constrained form of the posterior distribution, specifically we assume it factorises across sub-sets of the state variables. Variational inference is facilitated if our distributions belong to the conjugate exponential family (see *e.g.* [2]), so our first step is to decompose the Student-t prior distribution into a multivariate Gaussian,

$$p\left(\boldsymbol{\theta}_i|\mathbf{m}_i,\mathbf{P}_i\right)=\mathcal{N}\left(\boldsymbol{\theta}_i|\mathbf{m}_i,\mathbf{P}_i^{-1}\right),$$

where

$$\mathcal{N}\left(\boldsymbol{\theta}_i|\mathbf{m}_i,\mathbf{P}_i^{-1}\right)=\frac{|\mathbf{P}_i|^{\frac{1}{2}}}{(2\pi)^{\frac{d}{2}}}\exp\left(-\frac{1}{2}\left(\boldsymbol{\theta}_i-\mathbf{m}_i\right)^{\mathrm{T}}\mathbf{P}_i^{-1}\left(\boldsymbol{\theta}_i-\mathbf{m}_i\right)\right),$$

whose precision matrix, $\mathbf{P}_i$, is sampled from a Wishart distribution

$$p\left(\mathbf{P}_i|\mathbf{S},\nu\right)=W\left(\mathbf{P}_i|\mathbf{S},\nu\right),$$

where

$$W\left(\mathbf{P}_i|\mathbf{S},\nu\right)=\frac{|\mathbf{P}_i|^{\frac{(\nu-d-1)}{2}}}{2^{\frac{\nu d}{2}}\pi^{\frac{d(d-1)}{4}}|\mathbf{S}|^{\frac{\nu}{2}}\prod_{i=1}^{d}\Gamma\left(\frac{\nu+1-i}{2}\right)}\exp\left(-\frac{1}{2}\mathrm{Tr}\left(\mathbf{S}^{-1}\mathbf{P}_i\right)\right).$$

This hierarchical formalism for our prior leaves us with three sets of variables, $\mathbf{m}_i$, $\mathbf{P}_i$ and $\boldsymbol{\theta}_i$, whose prior distributions all belong to the conjugate exponential family. In variational inference, we seek an approximation to the posterior which factorises across these variables

$$q\left(\mathbf{m}_i,\mathbf{P}_i,\boldsymbol{\theta}_i\right)=q\left(\mathbf{m}_i\right)q\left(\mathbf{P}_i\right)q\left(\boldsymbol{\theta}_i\right).$$

Minimising the variational lower bound on the log likelihood with respect to each factor of this distribution [10] gives

$$q\left(\mathbf{m}_i\right)\propto\exp\left\langle\ln p\left(\boldsymbol{\theta}_i|\mathbf{m}_i,\mathbf{P}_i\right)p\left(\mathbf{m}_i|\boldsymbol{\mu}_i,\beta\right)\right\rangle_{q(\mathbf{P}_i)q(\boldsymbol{\theta}_i)},$$

where $\langle\cdot\rangle_{p(\cdot)}$ is an expectation under the distribution $p\left(\cdot\right)$. Substituting in the relevant probability distributions gives

$$q\left(\mathbf{m}_i\right)=\mathcal{N}\left(\mathbf{m}_i|\bar{\mathbf{m}}_i,\Sigma_i\right),$$

where

$$\bar{\mathbf{m}}_i = \Sigma_i\left[\left\langle\mathbf{P}_i\right\rangle\left\langle\theta_i\right\rangle+\beta\bar{\mathbf{m}}_i\right], \tag{3}$$

$$\Sigma_i = \left[\left\langle\mathbf{P}_i\right\rangle+\beta\mathbf{I}\right]^{-1} \tag{4}$$

and moments of interest may be found as

$$\left\langle\mathbf{m}_i\right\rangle=\bar{\mathbf{m}}_i,\ \left\langle\mathbf{m}_i\mathbf{m}_i^{\mathrm{T}}\right\rangle=\Sigma_i+\bar{\mathbf{m}}_i\bar{\mathbf{m}}_i^{\mathrm{T}}. \tag{5}$$

A similar process leads to

$$q\left(\mathbf{P}_i\right) \propto \exp\left\langle \ln p\left(\boldsymbol{\theta}_i|\mathbf{m}_i,\mathbf{P}_i\right) p\left(\mathbf{P}_i|\nu,\mathbf{S}\right)\right\rangle_{q(\mathbf{m}_i)q(\boldsymbol{\theta}_i)},$$

giving

$$q\left(\mathbf{P}_i\right) = W\left(\mathbf{P}_i|\bar{\mathbf{S}}_i,\bar{\nu}\right)$$

where

$$\begin{aligned}\bar{\mathbf{S}}_i &= \left[\mathbf{S}^{-1} + \left\langle \boldsymbol{\theta}_i\boldsymbol{\theta}_i^{\mathrm{T}}\right\rangle - \left\langle \mathbf{m}_i\right\rangle\left\langle \boldsymbol{\theta}_i\right\rangle^{\mathrm{T}} - \left\langle \boldsymbol{\theta}_i\right\rangle\left\langle \mathbf{m}_i\right\rangle^{\mathrm{T}} + \left\langle \mathbf{m}_i\mathbf{m}_i\right\rangle^{\mathrm{T}}\right]^{-1}, \quad (6)\\ \bar{\nu}_i &= \nu + 1 \quad (7)\end{aligned}$$

and a moment of interest may be found as

$$\left\langle \mathbf{P}_i\right\rangle = \bar{\nu}\bar{\mathbf{S}}_i^{-1}. \quad (8)$$

Finally we may obtain

$$q\left(\boldsymbol{\theta}_i\right) \propto \exp\left\langle \ln p\left(I|\boldsymbol{\theta}_i\right) p\left(\boldsymbol{\theta}_i|\mathbf{m}_i,\mathbf{P}_i\right)\right\rangle_{q(\mathbf{P}_i)q(\mathbf{m}_i)} \quad (9)$$

unfortunately even with the variational approximation we may not compute the constant of proportionality associated with this distribution. We therefore turn to sampling methods to handle this distribution.

Variational Importance Sampler The update equations for the posterior distributions are dependent on the distribution in (9). Being unable to resolve them directly we turn to sampling methods for estimating them. A range of sampling methods are applicable, for simplicity we focus on importance sampling (see *e.g.* [7]). We may re-write (9) as

$$q\left(\boldsymbol{\theta}_i\right) = \frac{1}{Z'} p\left(I|\boldsymbol{\theta}_i\right) \mathcal{N}\left(\boldsymbol{\theta}_i|\left\langle \mathbf{m}_i\right\rangle, \left\langle \mathbf{P}_i\right\rangle^{-1}\right).$$

Expectations under this distribution may now be estimated by sampling vectors $\boldsymbol{\theta}_i^{(s)}$ from

$$R\left(\boldsymbol{\theta}_i\right) = \mathcal{N}\left(\boldsymbol{\theta}_i|\left\langle \mathbf{m}_i\right\rangle, \left\langle \mathbf{P}_i\right\rangle^{-1}\right),$$

the proposal distribution. The moments required for the previous section may then be estimated as

$$\begin{aligned}\left\langle \boldsymbol{\theta}_i\right\rangle &= \sum_{s=1}^{S} w_s \boldsymbol{\theta}_i^{(s)}, \quad (10)\\ \left\langle \boldsymbol{\theta}_i\boldsymbol{\theta}_i^{\mathrm{T}}\right\rangle &= \sum_{s=1}^{S} w_s \boldsymbol{\theta}_i^{(s)}\ \boldsymbol{\theta}_i^{(s)\,\mathrm{T}}, \quad (11)\end{aligned}$$

where the values

$$w_s = \frac{p\left(I|\boldsymbol{\theta}_i^{(s)}\right)}{\sum_{s=1}^{S} p\left(I|\boldsymbol{\theta}_i^{(s)}\right)} \quad (12)$$

are known as the 'importance weights'. A common problem with importance sampling is that small fraction of the importance weights dominate the sums in (10) and (11). This occurs when there is a large mis-match between the proposal distribution, $R(\boldsymbol{\theta}_i)$ and the distribution of interest $q(\boldsymbol{\theta}_i)$. The variational importance sampler alleviates this problem through an adaptive proposal distribution, R, which responds to the observed data. Iterating our algorithm improves the match between $R(\boldsymbol{\theta}_i)$ and $q(\boldsymbol{\theta}_i)$. The quality of the samples obtained is often summarised by the effective number of samples,

$$S_{\text{eff}} = \frac{1}{\sum_{s=1}^{S} w_s^2} \leq S.$$

Implementation Variational algorithms are often sensitive the order in which parameter updates occur. For this reason in Algorithm 1 we sketch out the order of the updates we used. These were found to be effective across a range of microarray slides. The algorithm, in common with most variational approaches, is sensitive to initialisations of the parameters. For our experiments we initialised the required parameters as $\beta = .2$, $\mathbf{S} = \mathbf{I}$, $\nu = 4$ and $\Sigma_i = \beta^{-1}\mathbf{I}$. The expectations of the precision matrix and the Student-t's mean were initialised as $\langle \mathbf{P}_i \rangle = (\nu + 1)\left[\mathbf{S}^{-1} + \beta^{-1}\mathbf{I}\right]$ and $\langle \mathbf{m}_i \rangle = \boldsymbol{\mu}_i$. The vector $\boldsymbol{\mu}_i$ was taken from rough grid layout parameters associated with the slide. The number of samples was set at $\mathcal{S} = 200$. Whilst convergence can be monitored for portions of the variational update, we wished to avoid the associated computational overload and thus restricted ourselves to the simple convergence criteria shown in Algorithm 1. Finally, for our experiments we set the maximum number of iterations as $K = 20$.

Algorithm 1 Variational importance sampler for Microarray images.

Require: A number of samples S, a number of iterations K, a variance of the mean, β, a scale matrix, $\mathbf{S}$, a degrees of freedom parameter, ν, and an an initial estimate of $\mathbf{m}_i$'s posterior covariance Σ_i.
for each spot, i, in image **do**
  repeat
    Obtain rough size and location of spot and place in $\boldsymbol{\mu}_i$.
    Sample S parameter vectors, $\boldsymbol{\theta}_i^{(s)}$, from $\mathcal{N}(\boldsymbol{\theta} | \langle \mathbf{m}_i \rangle, \langle \mathbf{P}_i \rangle)$.
    Evaluate the importance weights for each sample, w_s, according to (12) and compute S_{eff}.
    Estimate $\langle \boldsymbol{\theta}_i \rangle$ and $\left\langle \boldsymbol{\theta}_i \boldsymbol{\theta}_i^{\mathrm{T}} \right\rangle$ using eqns (10) and (11).
    for K iterations **do**
      Update $\bar{\mathbf{m}}_i$ and $\langle \mathbf{m}_i \rangle$ followed by $\left\langle \mathbf{m}_i \mathbf{m}_i^{\mathrm{T}} \right\rangle$ according to eqns (3) and (5).
      Update $\bar{\mathbf{S}}_i$ and $\bar{\nu}_i$ according to eqns (6) and (7)
      Update $\langle \mathbf{P}_i \rangle$and Σ_i according to the eqns (8) and (4).
    end for
  until $S_{\text{eff}} > \frac{S}{4}$.
end for

RESULTS

We considered twenty-four bespoke manufactured cDNA microarray slides from experiments on *aphakia* mice. The experiments involve a comparison of eye tissue extracted from wild-type and aphakia mice embryos for the presence or absence of 5,760 genes at three time points (10.5, 11.5 and 12.5 days gestation) with eight slides associated with each time point. For each gene, i, the measurement of interest is the $\log_2$ ratio between the red, r_i, and green, g_i, channels, $\gamma_i = \log_2 \frac{r_i}{g_i}$ each channel being associated with one of the two different biological samples. Note that this measure is extremely sensitive when the signal intensity, defined as $a_i = r_i \cdot g_i$, is low. It is of biological interest to examine genes for which the magnitude of γ_i is 'high' (normally above 1) as these genes are being expressed differently in the two samples and are therefore associated with the aphakia mutation. The Bayesian approach we have proposed allows us to compute estimates of $\langle \gamma_i \rangle_{q(\boldsymbol{\theta}_i)}$ and its variance, $\sigma_i^2 = \langle \gamma_i^2 \rangle_{q(\boldsymbol{\theta}_i)} - \langle \gamma_i \rangle^2_{q(\boldsymbol{\theta}_i)}$, using the importance weights that we determine. In Figure 2 a small portion of one of these images, chosen for the variety of spots types displayed, is shown together with the rough grid placement and the last set of samples from the proposal distribution (the importance weights associated with these samples are not visualised). Each gene at each time point is associated with a maximum of sixteen values for γ_i, as each of the eight slides contains the genes twice[3]. It is common, due to experimental variations, for the values to be missing in some slides. We are interested in an estimate of the actual $\log_2$ ratio of the expression levels, $\bar{\gamma}_i^{(t)}$, at each time point, $t = 10.5,\ 11.5,\ 12.5$. In standard approaches this may be estimated as the mean of all the γ_i values for each time-step.

Weighting by Variance. Our approach, as well as giving more consistent results than other automated approaches [6], allows us to weight the expectations by their variance to modify our estimate of $\bar{\gamma}_i^{(t)}$. This is achieved through a weighted mean which downweights observations with high σ_i^2.

To illustrate the effect of allowing for the variance we show, in Figure 3, plots of the temporal differences between the $\log_2$ ratios. We computed $\bar{\gamma}_i^{(t)}$ with and without considering variances. The plots are shown for 125 of the 5760 genes that were selected as being of interest for further biological investigation [8]. Inclusion of the variance has two important effects: firstly, we have been able to draw ellipses around the genes' positions representing one standard deviation of error. Secondly it is apparent that, when computing the weighted mean of the ratio, many of the genes make quite large movements on the plot. These movements are mostly diagonal or along the y-axis. They correspond to changes in the $\gamma_i^{(t)}$ values for $t = 11.5$ and 12.5. This reflects large uncertainty for these measurements. This uncertainty arises as many genes of interest will be switched off in the mutant strain for the latter time points. Genes which are switched off have low intensity a_i, which generally leads to greater uncertainty and is in turn reflected in higher σ_i^2. A larger range of values for σ_i^2 leads to a greater effect from using a weighted mean. The biological findings of

[3] In our slides the replicates are adjacent to each other, thus neighbouring dots have similar intensities in the images.

these experiments are presented in [8], details of the gene selection and hierarchical clustering will be made available at `http://www.dcs.shef.ac.uk/ml/`.

Figure 2: *Left*: Greyscale of one of the two channels of the microarray image, dashed lines superimposed on the image which represent the parameters from the 'grid'. The dashed circles represent the mean parameters of the ovals for each spot, μ_i. *Right*: Samples from the proposal distribution $R(\boldsymbol{\theta}_i)$. These samples have associated importance weights which are used to compute expectations of interest. Spot positions without samples have been detected as empty.

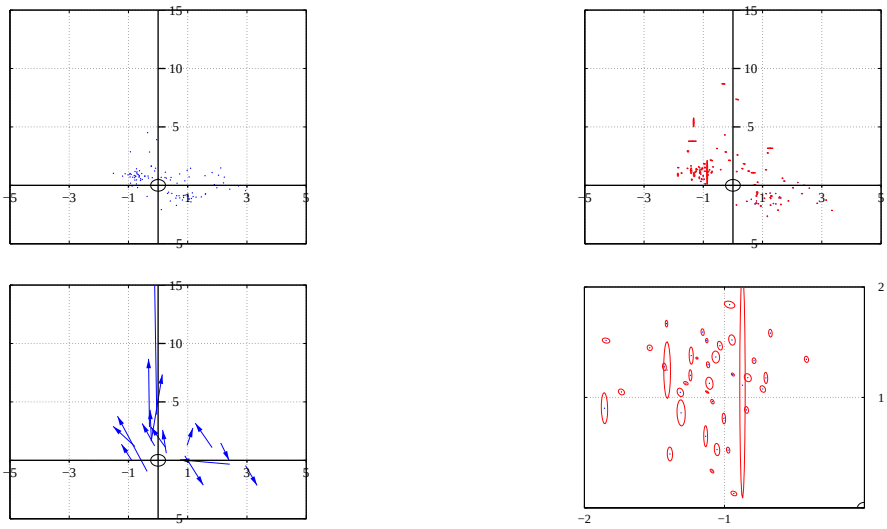

Figure 3: Plots of $\bar{\gamma}_i^{(11.5)} - \bar{\gamma}_i^{(10.5)}$ on the x-axis and $\bar{\gamma}_i^{(12.5)} - \bar{\gamma}_i^{(11.5)}$ on the y-axis. *Top Left* : values of $\bar{\gamma}_i$ calculated without taking account of uncertainties. *Top Right*: $\bar{\gamma}_i$ calculated with taking account of uncertainties. *Bottom Left*: Arrows indicating the direction of the the largest 15 moves. Movements on the y-axis are associated with changes in $\bar{\gamma}_i^{(12.5)}$, movements on the x-axis are associated with changes in $\bar{\gamma}^{(10.5)}$ and diagonal movements are associated with $\bar{\gamma}_i^{(11.5)}$. *Bottom Right*: Detail of plot from top right.

CONCLUSIONS

We have presented a Bayesian approach to microarray image processing which combines variational and sampling techniques. The importance sampler was used to fi-

nesse computational intractabilities induced by the likelihood. Variational inference enabled us to utilise a data-dependent proposal distribution for the importance sampler. Importance sampling on its own would have lead to a low effective number of samples, variational inference on its own is intractable. The changes we saw through incorporating the uncertainty reflected the fact that intensity for many genes of interest was low for $t = 11.5$ and 12.5. We are currently looking into propagating the uncertainty further through the downstream analysis, focussing on techniques which are popular in the biological community (such hierarchical clustering).

Software which implements our approach is downloadable from `http://www.dcs.shef.ac.uk/~neil/VIS`.

Acknowledgements PR is supported by a fellowship from the Lister Institute of Preventative Medicine and SS is a Fight for Sight research fellow.

References

[1] M. B. Eisen and P. O. Brown, "DNA Arrays for Analysis of Gene Expression," **Methods in Enzymology**, vol. 303, 1999.

[2] Z. Ghahramani and M. J. Beal, "Graphical Models and Variational Methods," in M. Opper and D. Saad (eds.), **Advanced Mean Field Methods — Theory and Practice**, MIT Press, 2001.

[3] M. Isard and J. MacCormick, "BraMBLe: A Bayesian Multiple-Blob Tracker," in **Proc. 8th Int. Conf. Computer Vision**, July 2001.

[4] M. I. Jordan (ed.), **Learning in Graphical Models**, vol. 89 of **Series D: Behavioural and Social Sciences**, Dordrecht, The Netherlands: Kluwer, 1998.

[5] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola and L. K. Saul, "An Introduction to Variational Methods for Graphical Models," in Jordan [4], pp. 105–162.

[6] N. D. Lawrence, M. Milo, M. Niranjan, P. Rashbass and S. Soullier, "Reducing the Variability in cDNA Microarray Image Processing by Bayesian Inference," Submitted to *Bioinformatics*.

[7] D. J. C. MacKay, "Introduction to Monte Carlo Methods," in Jordan [4], pp. 175–204.

[8] S. Soullier, M. Milo, J. Moss, N. D. Lawrence, D. Williams, L. Smith, V. van Heyningen, T. Freeman, A. Greenfield, M. Niranjan and P. Rashbass, "The Role of Pitx3 in Mouse Eye Development Dissected using Microarray Analysis," .

[9] J. Vermaak, N. Lawrence and P. Pérez, "Variational Inference for Visual Tracking," in **Computer Vision and Pattern Recognition**, IEEE Computer Society Press, 2003, To appear.

[10] S. Waterhouse, D. J. C. MacKay and T. Robinson, "Bayesian Methods for Mixtures of Experts," in D. S. Touretzky, M. C. Mozer and M. E. Hasselmo (eds.), **Advances in Neural Information Processing Systems**, Cambridge, MA: MIT Press, 1996, vol. 8, pp. 351–357.

[11] Y. H. Yang, M. J. Buckley, S. Dudoit and T. P. Speed, "Comparison of methods for image analysis on cDNA microarray data," Techn. report, **Department of Statistics, University of California, Berkeley**, 2000, Available from `http://www.stat.Berkeley.edu/~terry`.