Learning with subsets of the data

Aidan Boland & Nial Friel

Centre for Data Analytics

Insight

Introduction

In the current information age, there is huge interest in 'big' data, there are many likelihoods which become intractable as the size of the data increase. Gibbs random fields (GRF's) are an example of such a case, a GRF is a graphical model used in a variety of areas such as image analysis and network analysis. GRF's suffer from the curse of dimensionality, only trivially small cases can be dealt with using standard techniques.

Gibbs random field

Gibbs random fields (GRF's) are widely used to model complex dependency structure jointly in graphical models. The likelihood of a set of realised data y given a vector of parameters θ is defined as

$$f(y|\theta) = \frac{q(y|\theta)}{Z_\theta} = \frac{\exp(\theta^T s(y))}{\sum_{y \in Y} \exp(\theta^T s(y))}, \quad (1)$$

where $s(y)$ is a vector of statistics which are sufficient for the likelihood. The constant of proportionality Z_θ is a summation over all possible realisation of the GRF. Clearly, Z_θ is intractable for all but trivially small situations.

Example: Ising model

An example of a GRF is the Ising model. The Ising model is defined on a rectangular lattice or grid. It is used to model the spatial distribution of binary variables, taking values -1 and 1. The joint density of the Ising model can be written as

$$f(y|\theta) = \frac{1}{Z_\theta} \exp\left\{\theta \sum_{j=1}^{N} \sum_{i \sim j} y_i y_j\right\}$$

The normalising constant Z_θ is rarely available analytically since this relies on taking the summation over all different possible realisations of the lattice. For a lattice with N nodes this equates to $2^{\frac{N(N-1)}{2}}$ different possible lattice formations. The figure below shows an example of a 40×400 Ising lattice, each black square represents 1 and each white square represents -1.

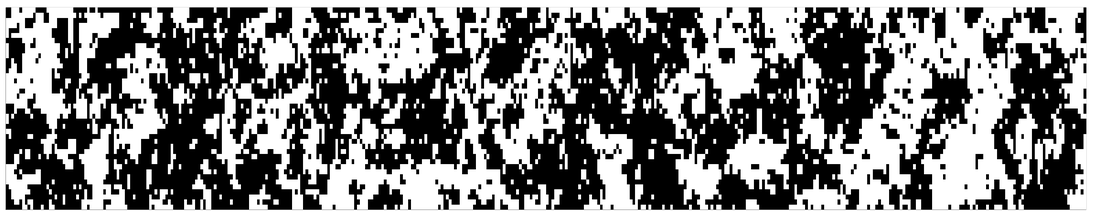

Figure: Ising example

Inference for GRF's

The distribution of interest is the posterior distribution of θ

$$\pi(\theta|y) \propto f(y|\theta)\pi(\theta)$$

Inference on θ is usually carried out using Markov chain Monte Carlo (MCMC), this works by creating a chain of values $(\theta_1, \ldots \theta_n)$, where the stationary distribution of this chain is that of our posterior distribution.

Due to the intractability of the likelihood it is difficult to sample values from the posterior distribution. Many techniques such as the exchange algorithm of Moller [4] or ABC [3] rely on sampling from the likelihood, however as the size of the data increases, sampling from the likelihood becomes very time consuming.

Composite likelihood

The composite likelihood is an approach to overcome the intractability. The idea is to use a product of full conditionals as an estimate to the true likelihood

$$f(y|\theta) \approx \prod_i f(y_{A_i}|\theta, y_{A_{-i}})$$

Where y_{A_i} is the ith group or block of observations from the data. Splitting the data into smaller blocks makes inference more computationally reasonable.

Contrastive Divergence

Asuncion [2] showed how the composite likelihood can be used with a technique known as contrastive divergence (CD) to get an estimate of the gradient of the posterior $\nabla\pi(\theta|y)$. Using CD an estimate of the Maximum a posteriori (MAP) can be found efficiently. We want to be able to get an estimate of the posterior distribution not just the MAP. Applying the work of Asuncion to Welling and Teh's Stochastic gradient fisher scoring algorithm, we obtain an estimate of the posterior distribution.

Stochastic Langevin

We create a chain of θ values using the following update.

$$\theta_{n+1} = \theta_n + \frac{\Sigma}{2}\widehat{\nabla} \log \pi(\theta_n|y) + C\eta \quad \eta \sim N(0, \Sigma).$$

At each step, the gradient $\widehat{\nabla}$ is updated using the methods discussed above. Alquier et al. [1] provide theoretical proofs of the convergence of noisy MCMC chains. We can apply these proofs to our algorithm.

Results

Ising Study

A grid of size 1000×1000 was simulated. An exchange algorithm [4] was run for 24 hours to get a 'ground' truth of the true posterior distribution. The stochastic Langevin algorithm was then run using 3 different block sizes, in each case 100 blocks were used at each step to estimate $\widehat{\nabla}$. The table and graph below show the results.

	Mean	SD	Time (Minutes)
Exchange	0.4001	0.00044	1440
Blocksize			
8×8	0.4005	0.00292	15
16×16	0.3999	0.00163	45
32×32	0.4001	0.00112	165

Table: Means and standard deviations.

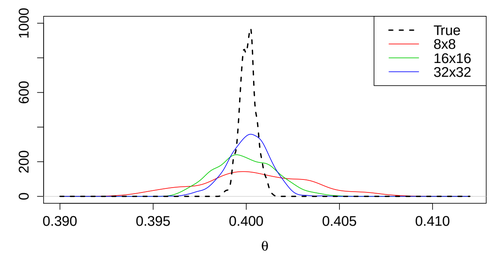

Figure: Results varying block sizes

We see how all 3 different block sizes give good estimates of the MAP, however they all under estimate the variance. The algorithm improves as the block size increases which is expected. This is also evident in the graph of the densities of the posterior estimates.

Future Work

The results show that our algorithm gives an efficient estimate of the MAP, it does not estimate the variance of the posterior well. Further work is needed to improve the algorithm, we need to be able to estimate this variance well. The choice of which blocks we use for the composite likelihood may be the key to this.

References

- P. Alquier, N. Friel, R. Everitt, and A. Boland. Noisy monte carlo: Convergence of markov chains with approximate transition kernels.
- A. U. Asuncion, Q. Liu, A. T. Ihler, and P. Smyth. Learning with blocks: Composite likelihood and contrastive divergence.
- J.-M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder. Approximate bayesian computational methods.
- J. Moller, A. Pettit, R. Reeves, and K. Bertheksen. An efficient markov chain monte carlo method for distributions with intractable normalising constants.

Funding

SimSci is Funded under the programme for Research in Third-level Institutions and co-funded under the European Regional Development fund

Investing In Your Future

The Insight Centre for Data Analytics is supported by Science Foundation Ireland under Grant Number SFI/12/RC/2289

DCU · NUI Galway OÉ Gaillimh · UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh · UCD Dublin · sfi Science Foundation Ireland