A Review of Bayesian Statistical Methods Applied to Big Data
Farzana Jahan¹ and Kerrie Mengersen²

¹PhD Student, Science and Engineering Faculty, Queensland University of Technology
² Distinguished Professor of Statistics, Science and Engineering Faculty, Queensland University of Technology

Introduction
Sheer amount of information is now being generated, accumulated, collected or integrated in every fraction of second formulating the term “Big Data”. There has been no unique definition of Big Data, to date there is a large body of literature which attempt to define Big Data effectively. (1,2)

A Definition of Big Data
“Big Data includes the following aspects: “Volume”, “Velocity” and “Variety’, to describe the characteristics of information, “Technology” and “Analytical Methods”, to describe the requirements needed to make proper use of such information, and “Value”, to describe the transformation of information into insights that may create economic value for companies and the society.” (1)

There have been a wide range of research going on involving Big Data. We can broadly categorize the published literature on Big Data as follows:

Objective of the Study
• This study attempts to review the published studies that present Bayesian statistical models specifically for Big Data and discuss the reported and perceived benefits of these approaches.
• We aim to answer, whether focusing only on improving computational algorithms and infrastructure will be enough to face the challenges of Big Data.

Motivation of the Study
• Although, there have been many review papers on Big Data (1,3,4,5,6,7 etc.), to the best of our knowledge, to the best of out knowledge, the survey of Bayesian Statistical modelling applied to Big Data to address our question above remains unexplored.
• We attempt to make a potential contribution to the knowledgebase for the Statisticians to understand the usefulness and necessity to develop scalable Bayesian Statistical modelling for Big Data.

Bayesian Approaches in Big Data
The number of articles applying Bayesian approaches to Big Data is not very large (9,10,11,12)

Majority of the works conducted from the perspectives of Bayesian statistics were concerned with designing scalable algorithms to be able to analyse Big Data as noted from the reviews made. There is lack of research on Bayesian Modelling in Big Data situations.

Critical Reflection
• Though, we can see emphasis of theoretical developments regarding new methods or models to analyse Big Data in some articles (13,14,15), little work is concentrated on statistical or in particular Bayesian Statistical modelling to Big Data.
• The advantages of Bayesian Statistical modelling as: incorporation of multiple sources of information into the model via prior and flexibility in model estimation with high dimensional parameter space etc. (16) which are still under explored in Big Data perspectives.

Conclusion
• We are living in the era of Big Data and continuous research works are in progress to make the best use of the available data. Our review identified the application or development of Bayesian statistical modelling of Big Data.
• This review identified the need for further research to identify the advantages and/or drawbacks of Bayesian Statistical modelling in Big Data situations. The area is yet to be explored to be able to answer whether existing models are enough to face the Big Data challenges with scaling algorithms or there is need for new models to be developed for improved decision making with Big Data.
• As literature suggests, careful modelling and sound theoretical platform is needed along with the computational advances (13,14,15), we may conclude that there should be more works concentrating on Bayesian Statistical modelling in Big Data perspectives in order to reveal the usefulness of Bayesian inference and also to identify the challenges to address while developing or applying Bayesian Statistical models for Big Data.

Key References