

統計的シミュレーションモデルをもとにした 意思決定支援システム

Statistics Model for Decision Making - A Bayesian Modelling

共同研究メンバー

○今泉忠*、久保田貴文*、増田浩通*（○代表、執筆者）

1. はじめに

近年「ビッグデータ」分析が盛んで、企業の戦略決定などに活用されている。例えば、Google Ngram Viewer を用いて“Big Data”という用語の出現傾向についてグラフを作成すると図1（2017年7月1日）が得られた。

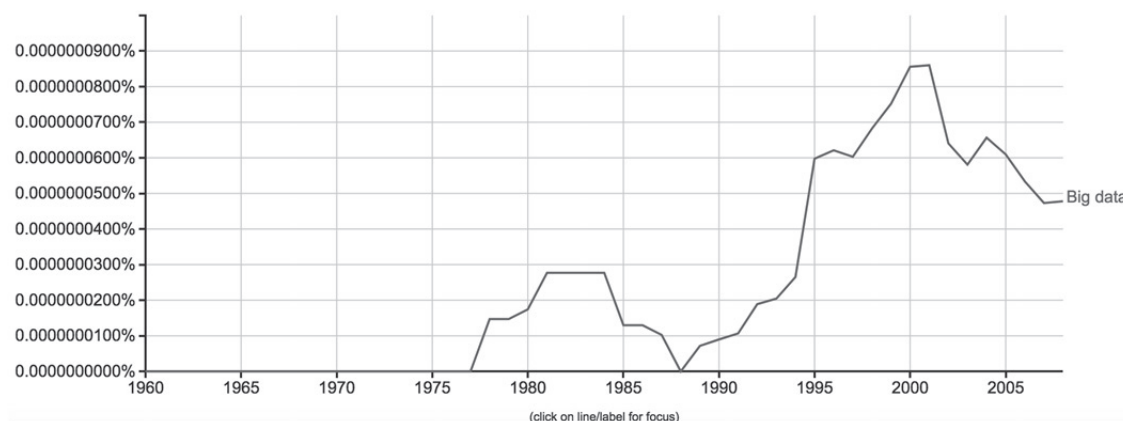


図1 “Big Data” の出現比率の推移（Google Ngram から）

図1では出現比率が2000年付近がピークになっており、その後は下降している。この原因としては様々想定されようが、データに基づいて意思決定を行うことが「当たり前」になってきていることを考えると、「ビッグデータ」が「データ」として当たり前のデータとして扱われるようになってきたとも考えられよう。

しかし、企業業績評価やコミュニティでの課題解決のように、複数のサブゴールが曖昧で相反しており、かつ、扱うデータがビッグデータである場合には、探索的な方法では意思決定支援ができない。そこで、全体-局所最適化を行えるシステムとしてベイズ統計モデルをもとに

* 多摩大学経営情報学部

したシミュレーションモデルとして構築することを提案する。解を1つのクリスピーな解ではなく、分布として表現することで、解の評価や比較を可能とする。

2. 分析システムの例

データに基づいて意思決定を支援する場合に、必ずしもデータ量の増加に比例して、データと取り出せる情報が比例的に増加するとは言えず、ややもすると分析結果がGIGO (Garbage In, Garbage Out) と表現されるような結果となる。このようになることを避けるためには、意思決定者を支援する分析者自身が、データの全体像を把握でき、また、分析の流れを他者に説明できるようになることが必要である。そのための支援システムとしては、統計分析のためのシステム R をもとにした Radiant や R Analyticflow などがある。

Radiant は Web ベースの対話型システムであり、分析者が分析パラメータを設定しながら分析を行える点が特徴である。R Analyticflow は分析の流れを GUI ベースで作成できるシステムであり、他者との分析フローの共有化が図れる点が特徴である。

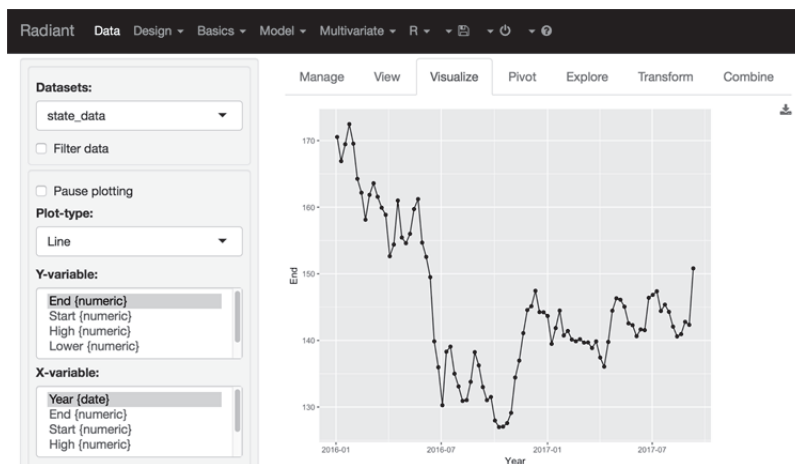


図2 Radiant の画面例

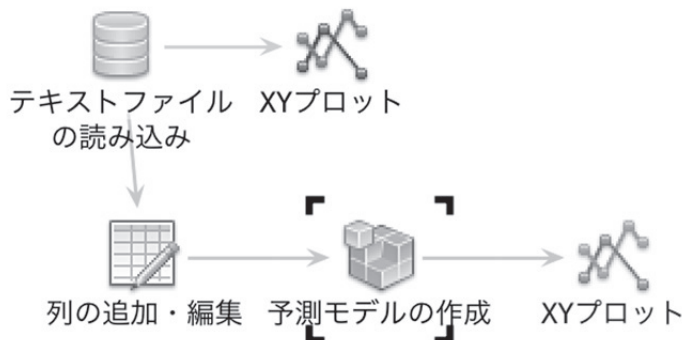


図3 R Analyticflow の例

3. シミュレーションモデルの構築

Radiant や R Analyticflow を用いることで適切な分析を行うことができる。しかし、為替変化などの将来の変化の傾向を、過去のデータを用いて予測する時系列データの分析では、上記のシステムでは不十分であると考えられる。それは、通常、データはある時点での標本の実現値であり、1つのデータであり、これを用いて平均値の変化を捉えることが難しいためである。例えば、2016年6月にはUKにおいて国民投票の結果、EU離脱が決定されて、さらに同年10月のメイ首相の演説で行われた。これに対応しポンドの為替レートでの変化が観られた。2016年1月から2017年8月までのGBP / JPY の為替レートの週ごと変化のデータを収集した。

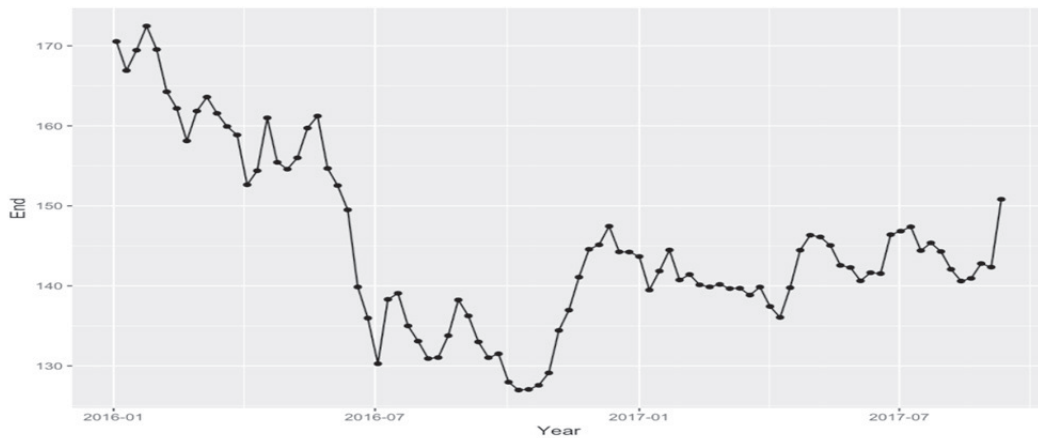


図4 GGP/JPY の週ごとの為替レートの推移（終値）

このデータは2016年6月に為替レートの変化があったことを示唆するが、変化の幅について情報がないために、その検証はできない。しかし、ここで以下のような構造モデルを想定してみる。ある時点 t での平均 μ_t とデータ Y_t について

$$\begin{aligned} \mu_t &\sim t(\mu_{t-1}, 1) \\ Y_t &\sim N(\mu_t, \sigma^2) \end{aligned}$$

とするのである。 $t(\mu_{t-1}, 1)$ は非心度 μ_{t-1} で自由度 1 の非心 t 分布である。これは、最も単純なモデルであるが、平均が分布をするというベイズモデルであり、誤差の分布が分散有限の正規分布でないとする事で、大きな変化が発生することをモデル化している。この平均の分布などの推定についてはMCMC (Markov Chain Monte Carlo) 法などのシミュレーション法を用いて行うことができる。為替レートデータに適用した場合に図5の結果を得た。

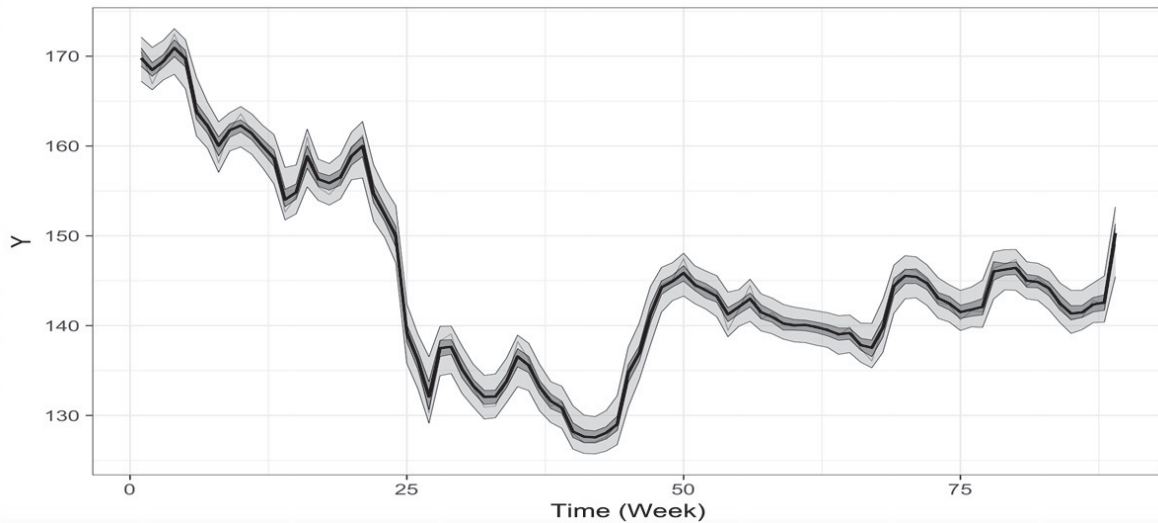


図5 MCMC サンプルの中央値と MCMC サンプル 95% 信用区間、MCMC サンプルの 50% 信用区間

この結果はシミュレーション結果であるので、再度分析を行えば解は変化するので、解の収束について確認するために複数回実施して確認した。この平均の分布から、国民投票時の為替レートの変化の大きさを評価できる。

Volume (容量)・Variety (種類)・Velocity (頻度・スピード)・Value (価値) の4Vで表現されるビッグデータに、ベイズモデルをどのように適用するかは残されているが、このようなフレームでの変化に関する評価が可能となることは意思決定の支援に関して大きな寄与となると考えられる。

参考文献

Vincent Nijs, Radiant: Business analytics using R and Shiny,

<https://www.rstudio.com/resources/videos/radiant-business-analytics-using-r-and-shiny/>

鈴木 了太, データ解析のための GUI R AnalyticFlow 3, 2015.12.5 <http://r.analyticflow.com/ja/>