

VERSION: 2012-04-22 18:42

# データ解析のための統計モデリング入門

一般化線形モデル・階層ベイズモデル・MCMC

久保拓弥

2012年4月22日

箱を開いて見ればその内容は確定されるという  
ような単純な事柄に関する真偽の議論  
の場合を除き，我々が追求する真理は，現  
在の知識に依存するという意味で相対的な，  
対象の一つの近似を与えるモデルによって  
表現されるようなものに過ぎない

赤池弘次 (1927 – 2009) [2]

# まえがき

この本では統計モデリング—「モデルを作って観測データにあてはめて現象を理解する」方法について説明します。内容は入門的なものばかりですが、初学者むけの普通の統計学教科書とはずいぶんと異なった方向のものです。

## この本のねらいと想定している読者

この本の読者として想定しているのは、いわゆる理系・文系の区別とは関係なく、「数理モデルで現象を表現・説明する」基礎訓練を受けていない人たちです。数理モデリングや解析プログラミングに関する講義・演習などをほとんど、あるいはまったく経験していない\*1 にもかかわらず、複雑なデータ解析に取りくまねばならない研究者が増えています。このような乖離をいくらかでもうめるように書いてみました。

読者は「確率・統計」も含む高校数学の範囲を理解しているという想定で説明します。高校数学であつかえる範囲をこえた内容については、各章の末尾で基礎的な統計学書籍を紹介しますので、そちらを参照してください。

この本では、章ごとに異なる例題を解決していく過程で、統計モデルの基本となる考えかたを説明します。あつかう内容は、応用範囲のひろい統計モデルのひとつである、一般化線形モデル (generalized linear model, GLM) の基礎と

---

\*1 たとえば私の専門分野である環境科学や生物学は、いわゆる理系なのでしょうが、現象と対応づけられる数理モデルを研究の道具として使える人は多くありません。私自身も体系だったトレーニングはほとんど受けていません。

発展だけに限定します。本の前半では GLM 入門的な内容を、後半では実際のデータ解析に使えるようにベイズ統計モデル化する方法を説明します。

私は生態学\*<sup>2</sup> のデータを解析する機会が多いので、各章の例題はそのような分野の大学院生あいてに説明するものを使用しました。とはいえ、これらの例題は本物のデータではなく、現実をかなり単純化した架空データ\*<sup>3</sup> を使った、たいへん簡単なものばかりです。つまり、この本を読みすすめるときには、生態学の専門知識はまったく必要ではありません。

## 使用する統計ソフトウェアとサポート web サイト

現実のデータ解析は複雑な構造をもつデータを処理していく作業であり、統計ソフトウェアによる支援が必要不可欠となります。この本の例題のデータ構造は単純なものばかりですが、データ解析の作業例を示すために、統計ソフトウェア R を使った例題データの図示・要約・統計モデルあてはめの過程を示します。この R 以外のソフトウェアとしては、またこの本の後半に登場する階層ベイズモデルのパラメーター推定のために、WinBUGS も使用します\*<sup>4</sup>。

データ解析が必要とされる多くの学術分野において、R は急速に普及しつつある統計ソフトウェア\*<sup>5</sup> です。R は誰でも無料で入手できるだけでなく、“free software” でもあり、ソースコードが完全に公開されています。つまり、いろいろな処理の過程でどのような計算をしているのか確認できるので、R は学術研究には適したソフトウェアであると言えます。

これに対して、WinBUGS は R のような free software ではありません。ただし誰でも無料で入手できるソフトウェア (freeware) であり、それなりに性能が良いので、とりあえずこの本の例題の解決に使用します。この点については、本文で WinBUGS が登場するときに、もう少し説明します。

---

\*<sup>2</sup> おもに個体以上のスケールの現象をあつかう生物学。動物・植物などの個体数の変動や空間分布、環境に対する個体の応答などを研究します。

\*<sup>3</sup> 実際の生物や観測データを例題としてとりあげると、どうしても統計モデル説明の本筋から外れた詳細について解説したり検討する必要があつて煩雑になるのでさげました。

\*<sup>4</sup> なお使用した R の version は 2.14.1, WinBUGS は 1.4.3 です。version に依存して、結果の出力が微妙に異なる場合があるかもしれません。

\*<sup>5</sup> データ解析環境とも呼ばれます。

数理モデルのあつかいに慣れている解析者であれば，考えかたの概要と数式で記述されたモデルを示せば，それにもとづいて自分自身でソフトウェアを実装できるでしょう．しかしこの本は，まさにそういったトレーニングを受けていない読者を想定していますので，必要に応じて R と WinBUGS のコードを示しながら具体的に説明を進めます．

各章の例題の架空データと作図・解析用の R と WinBUGS のコードはこの本のサポート web サイト <http://goo.gl/Ufq2> からダウンロードできます．この本のまちがいや補足説明なども上記サイトに掲載し，更新していきます．

## (第 2 章以降の説明の流れ)

