

Bayesian Data Analysis : Chapter5

Hierarchical models

Andrew Gelman, John B.Carlin, Hal S.Stern,
David B.Dunson, Aki Vehtari, and Donald B.Rubin

山崎 遼也

情報学科 数理工学コース 4 回

2017/4/5～

Table Contents

1 5.1 Constructing a parameterized prior distribution

2 5.2 Exchangeability and setting up hierarchical models

Key Words



多くの統計の応用が、問題の構造によってある方向に関係または結びついているとして見れるような多パラメータを含み、これらのパラメータの結合確率モデルがそれらの依存を反映すべきであるということを示唆している。例えば、病院 j の患者が生存確率 θ_j であるとする心臓治療の有効性の研究において、病院のサンプルを表す θ_j の推定値がそれぞれに関連していると期待することは妥当であろう。 θ_j が共通の母分布のサンプルとみなされる事前分布を使用すると、これが自然な方法で達成されることがわかります。このような応用の重要な特徴は、 θ_j の値自体は観測されないものの、 θ_j の集団分布の側面を推定するために、 j でインデックス付けされたグループ内の i でインデックス付けされたユニットを有する観測データ y_{ij} を使用できることである。特定のパラメータに対して条件付きでモデル化された観測可能な結果に関し、このような問題を階層的にモデル化するのは当然である。それ自体は、超パラメータとして知られるさらなるパラメータに関して確率的な仕様が与えられます。このような階層的思考は、マルチパラメータの問題を理解するのに役立ち、また計算戦略を開発する上で重要な役割を果たす。

おそらく、実際にはもっと重要なのは、単純な非階層モデルは通常、階層データには不適切であるということである。多くのパラメータを使用すると、既存のデータにうまく適合するモデルを作成するという意味でそのようなデータを**オーバーフィット (overfit)**する傾向がありますが、新しいデータの予測は劣るものになります。対照的に、階層モデルは、データにうまくフィットするのに十分なパラメータを有する一方、母集団分布を使用してパラメータへの依存を構造化することにより、過適合の問題を回避することができる。この章の例で示すように、データ点よりも多くのパラメータで階層モデルを適合させることはしばしば賢明です。

5.1 節では、階層的な原則を用いて事前分布を構築するが、階層的な構造については正式な確率モデルを適合させないという問題を考慮する。最初に、過去のデータを使用して事前分布を作成する単一の実験の分析を検討してから、一連の実験のパラメータのためのもっともらしい事前分布を検討します。5.1 節の処理は、完全なベイズではありません。なぜなら、説明を簡単にするために、母集団分布のパラメータ (超パラメータ) について、完全な結合事後分布ではなく、点推定量を使用するためです。

5.2 節では, 完全なベイズ解析の文脈において, 階層的事前分布を構築する方法について議論する. 5.3-5.4 節は, 解析的手法と数値的手法を組み合わせる共役族の階層モデルを用いた一般的な計算方法を提示する. 階層ベイズモデルの重要な実用的および概念的な利点をすぐに調べるために, 最も一般的な計算方法の詳細を 3 部に引き継ぐ. この章では, 次の 2 つの拡張例を続けます. 同じ試験問題に関する別の研究の結果を組み合わせるために, 教育試験実験のための階層モデルと, 同じ研究問題に関する別の研究の結果を組み合わせるために, 医学研究で使用されるメタ解析の方法のベイズ取り扱い. 少数のグループからのデータにフィットする階層的モデルにとって重要となる, 弱情報事前分布を議論することで結論づけます.

5.1 パラメータ化事前分布の構築

階層データの文脈で1つの実験を解析する

階層的モデルの説明を始めるにあたり, 小さい実験のデータと同様の過去 (または過去の) 実験から構築された事前分布を用いてパラメータ θ を推定する問題を考察する. 数学的には, 現在および過去の実験を共通の母集団からの無作為標本とみなします.

例. ラット群における腫瘍のリスクの推定

可能性のある臨床応用のための薬物の評価において, 研究は齧歯類に対して日常的に行われている. 統計学の文献から得られた特定の研究では, 即時の目的は, ゼロ用量の薬物 (対照群) を受けるタイプ「F344」のメス実験ラットの集団における腫瘍の確率を推定することであると仮定する. データは, 14 匹のラットのうち 4 匹が子宮内膜間質ポリープ (腫瘍の一種) を発症したことを示している. θ の与えられた腫瘍数に対する二項モデルを仮定することは当然である. 便宜上, 共役族 $\theta \rightarrow \text{Beta}(\alpha, \beta)$ から θ の事前分布を選択する.

固定事前分布での解析

過去のデータから, タイプ F344 の女性ラットラットの群の腫瘍確率 θ が, 既知の平均および標準偏差を有する近似ベータ分布に従うことを知ったと仮定しよう. 腫瘍確率 θ は, ラットの違いおよび実験条件の違いにより変化する. ベータ分布 (付録 A 参照) の平均および分散の式を参照すると, 平均および標準偏差の所定の値に対応する α, β の値を見つけることができました. 次に, θ の $\text{Beta}(\alpha, \beta)$ 事前分布を仮定すると, θ の $\text{Beta}(\alpha + 4, \beta + 10)$ 事後分布が得られる.

履歴データを使用した母集団分布の近似推定量

典型的に, 潜在的な腫瘍のリスクの平均と標準偏差は利用可能ではない. むしろ, 過去のデータは, 類似のラットの群についての以前の実験で利用可能である. ラット腫瘍の例では, 過去のデータは, 実際に, 70 群のラットにおける腫瘍発生率の観察のセットであった (表 5.1). j 番目の歴史的実験では, 腫瘍を有するラットの数 y_j とし, ラットの総数を n_j とする. 我々は, y_j を独立した二項データとしてモデル化し, サンプルサイズ n_j と研究固有の手段 θ_j を与える. 過去の実験で, パラメータ (α, β) を用いたベータ事前分布が θ_j の母集団分布をよく表していると仮定すると, 図 5.1 のように階層モデルを模式的に表示できます.

70 個の値 $\frac{y_j}{n_j}$ の観測されたサンプル平均および標準偏差は, 0.136 および 0.103 である. 母集団分布の平均と標準偏差をこれらの値に設定すると, α と β を解くことができます (付録 A の 583 ページの (A.3) を参照). (α, β) の推定結果は (1.4, 8.6) である. これは, 指定された完全確率モデルに基づいていないため, ベイズ計算ではありません. 5.3 節のこの例では, (α, β) を推定するために, より完全なベイズ手法を提示する. 推定値 (1.4, 8.6) は, 単に母集団分布のパラメータを推定するという考え方を探る出発点に過ぎない.

現在の実験のための事前分布としての歴史的母集団分布の簡単な推定値を使用すると、 θ_{71} のベータ (5.4, 18.6) 事後分布が得られる：事後平均は 0.223 であり、標準偏差は 0.083 である。以前の情報は、経験の重さが現在の実験における腫瘍の数が異常に多いことを示すので、粗比 $4/14 = 0.286$ より実質的に低い事後平均をもたらした。

これらの分析では、現在の腫瘍リスク, θ_{71} , および 70 個の歴史的腫瘍リスク $\theta_1, \dots, \theta_{70}$ は、過去の実験がすべて実験室 A で行われたが、現在のデータが実験室 B で集められたことが知られている場合、または時間があれば無効となる仮定である、傾向は関連性があった。実際には、現行のデータと過去のデータとの間の相違を簡単に説明することはできますが、歴史的な分散を膨らませることで、ベータモデルでは、履歴分散を膨張させることは、 α, β を一定に保ちながら $(\alpha + \beta)$ を減少させることを意味する。腫瘍リスクの時間傾向などの他の体系的な相違は、より広範なモデルに組み込むことができます。

70 回の歴史的実験を用いて θ_{71} の事前分布を作成した後、最初の 70 回の実験 $\theta_1, \dots, \theta_{70}$ における腫瘍確率のベイズ推論を得るために、この同じ事前分布を使用することも望ましいかもしれない。既存のデータから事前分布を直接推定するアプローチには、いくつかの論理的かつ実際的な問題があります。

- ・ 最初の 70 回の実験について推定された事前分布を使用したい場合、データは 2 回使用されます。まずすべての結果を合わせて事前分布を推定し、各実験の結果を用いて θ を推定する。これは私たちの精度を過大評価するように思われる。

- ・ α, β の点推定値は任意であると思われ、 α, β の任意の点推定値を使用すると必然的にいくつかの事後不確かさは無視されます。

- ・ 私たちは反対の点を作ることもできます。 α, β を「推定する」のは意味がありますか？それらは「事前」流通の一部です。ベイズ推論の論理に基づいて、データが収集される前にそれらを知っていなければなりませんか？

結合情報の論理

これらの問題にもかかわらず、すべてのデータから母集団分布を推定し、71 個の値 θ_j を別々に推定するよりも、各 θ_j を推定するのに役立つ。それぞれ 20 匹のラットのうち 2 つの観察された腫瘍を用いた実験に対応する、 θ_{26}, θ_{27} の 2 つのパラメータの推論についての次の思考実験を考える。 θ_{26}, θ_{27} の両方の事前分布が 0.15 を中心とするとする。正確にデータ分析を完了した後に $\theta_{26} = 0.1$ と教えられた仮定する。これは θ_{27} の推定値に影響するはずでず。実際には、2 つのパラメータのデータが同一で、前回の分布から予想された値よりも 0.1 の仮定値が低いため、以前に信じられていたよりも θ_{27} が低いと考えるかもしれません。従って、 θ_{26}, θ_{27} は事後分布に依存すべきであり、それらは別々に解析すべきではない。

確率モデルをパラメータと実験のセット全体に置き、すべてのモデルパラメータの結合分布についてベイズ解析を実行することにより、データを使用して事前パラメータを推定し、前述の欠点をすべて排除するという利点があります。完全なベイズ解析については 5.3 節で説明する。**経験的ベイズ (empirical Bayes)** と呼ばれることもある、事前パラメータを推定するためのデータを用いた分析は、完全な階層ベイズ解析の近似値と見れる。ここで議論し、本書の残りの部分で使用する完全なベイズ手法は経験的ではないので、経験的ベイズという用語を避けることがある。

5.2 交換可能性と階層モデルの設定

前節の例から一般化し, 実験 $j = 1, \dots, J$ の集合を考える. ここで, 実験 j は尤度 $p(y_j|\theta_j)$ となる, データ (ベクトル) y_j とパラメータ (ベクトル) θ_j をもつ. (この章を通じて, 便利さから実験という単語を用いるが, この方法は非実験的データにも同様に適用することができる.) 異なる実験におけるパラメータのいくつかは重なり合っているもよい. 例えば, 各データベクトル y_j は, 平均 μ_j および共通分散 σ^2 を有する正規分布からの観測のサンプルであってもよく, その場合 $\theta_j = (\mu_j, \sigma^2)$ である. すべてのパラメータ θ に対する結合確率モデルを作成するために, 第1章で紹介した交換可能性という重要な概念を使用し, それ以来繰り返し使用しています.

交換可能性

θ_j を他と区別するためにデータ y 以外の情報が利用できず, パラメータの順序付けまたはグループ化ができない場合, それらの事前分布におけるパラメータ間の対称性を仮定しなければならない. この対称性は, 交換性によって確率論的に表される. $p(\theta_1, \dots, \theta_J)$ がインデックス $(1, \dots, J)$ の順列に対して不変であれば, パラメータ $(\theta_1, \dots, \theta_J)$ はその結合分布において交換可能である. 例えば, ラット腫瘍の問題では, おそらく θ_j の値に関係しないサンプルサイズ n_j 以外の 71 回の実験を区別する情報がないと仮定する. したがって, θ_j に対して交換可能なモデルを使用する.

すでに, 直接データのための独立同一分布のモデルを構築する際の交換可能性という概念に遭遇しました. 実際には, 無知は交換可能性を意味する. 一般的に, 問題について知らないほど, 自信を持って交換可能性の主張をすることができる. (これは, 統計分析に着手する前に, 問題の知識を制限する正当な理由を加えることを急がない!) さいころの目に類似して考える. 最初に 6 つの結果すべてに等しい確率を割り当てる必要がありますが, さいころの測定値を慎重に検討し, さいころの重さを慎重に検討すると, 最終的に不完全さに気付くことがあります.

交換可能な分布の最も簡単な式は、未知のパラメータベクトル ϕ で支配される事前 (または母) 分布からの独立なサンプルとして、パラメータ θ_j を持つ。これより、

$$p(\theta|\phi) = \prod_{j=1}^J p(\theta_j|\phi) \quad (5.1)$$

である。一般に、 ϕ が未知であるので、 θ の分布は ϕ の不確かさで平均をとる。

$$p(\theta) = \int \left(\prod_{j=1}^J p(\theta_j|\phi) \right) p(\phi) d\phi \quad (5.2)$$

独立同一分布の混合であるこの式は、通常、実際に交換可能性を確保するために必要なものです。

関連する理論的結果である, 1.2 節で言及した **de Finetti の定理**は, $J \rightarrow \infty$ の極限では, $(\theta_1, \dots, \theta_J)$ 上の適切に動作する交換可能な分布は, (5.2) と同様に, 独立同一分布のものである. J が有限であるとき, 定理は成立しない (演習 5.1, 5.2, および 5.4 を参照). 統計的には, 混合モデルは, 未知の超パラメータ ϕ によって決定される共通の **スーパー母集合 (super population)** から引き出されたパラメータ ϕ を特徴づける. すでにデータ y_1, \dots, y_n を, いくつかのパラメータベクトル θ を与えられた n 個の観測値が独立同一に分布する尤度の形で表す.

上記混合モデルの単純な反例として, 6 つの面のそれぞれに与えられたさいころの確率を考える. 確率 $\theta_1, \dots, \theta_6$ は交換可能ですが, 6 つのパラメータ θ_j は 1 に合計されるように制約されているため, 独立した同一の分布の混合でモデル化することはできません. それにもかかわらず, それらは交換可能にモデル化することができる.

