

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

2017/4/5～11

Table Contents

- 1** 5.1 Constructing a parameterized prior distribution
- 2** 5.2 Exchangeability and setting up hierarchical models
- 3** 5.3 Fully Bayesian analysis of conjugate hierarchical models
- 4** 5.4 Estimating exchangeable parameters from a normal model
- 5** 5.5 Example: parallel experiments in eight schools
- 6** 5.6 Hierarchical modeling applied to a meta-analysis
- 7** 5.7 Weakly informative priors for hierarchical variance parameters

Key Words

- 母集団分布
- 超パラメータ
- オーバーフィット
- メタ解析
- 経験ベイズ
- de Finetti の定理
- 超母集団
- 超事前分布
- 経験ロジット
- 較正, 誤較正
- 推論的事前分布
- 逆-gamma(ϵ, ϵ) 事前分布
- 半-Cauchy 事前分布

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なせるような多パラメータを含む。これらはパラメータの結合確率モデルがその依存性を反映すべきであるということを示唆している。例えば、心臓治療の有効性の研究において、病院 j の患者が生存確率 θ_j であるとする、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**母集団分布 (population distribution)** のサンプルとみなすような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 の母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 (hyperparameters)** 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當であるということである。少ないパラメータに関して、一般的に大きなデータセットを正確に適合することはできない。しかし、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分布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. 同じテスト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, 医療研究で使われる **メタ解析 (metaanalysis)** の方法のベイジアン処理. 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る.

5.1 パラメータ化事前分布の構築

過去のデータの文脈で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, 小さい実験からのデータを使いパラメータ θ を推定する問題と, 同様の以前 (または過去の) 実験から構築された事前分布について考える. 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ランダムサンプルとみなす.

例. ラット群における腫瘍のリスクの推定

可能な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, 目先の目的は, 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 θ を推定することであるとする. 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, 共役分布族 $\theta \sim \text{Beta}(\alpha, \beta)$ 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, 既知の平均と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する. 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る. 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.

過去のデータを使用した母集団分布の近似推定

典型的に、腫瘍のリスクの潜在的な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集合であった (表 5.1). j 番目の過去の実験では、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。 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と標準偏差は、0.136, 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) の推定は (1.4, 8.6) である。これは、特定の完全な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。

現在の実験についての事前分布としての過去の母集団分布の簡単な推定値を使用すると, θ_{71} に対する $\text{Beta}(5.4, 18.6)$ 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. 事前情報は, 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, 粗い比 $4/14 = 0.286$ 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 個の過去の腫瘍リスク $\theta_1, \dots, \theta_{70}$ 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, $\frac{\alpha}{\beta}$ 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 回の過去の実験を用いて 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 回の実験についての推定に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。これは精度を過大評価するように思われる。
- ・ α, β の点推定量は任意であると思われ、 α, β の任意の点推定量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。
- ・ 反対の点を作ることもできます。 α, β を推定することに意味がありますか？ それらは事前分布の一部です。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おくべきか？

情報結合の論理

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} の両方の事前分布は 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 $\theta_{26} = 0.1$ と教えられたとする。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、事前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、本書の残りで使用する完全なベイズ手法は経験的ではないことを誤って示唆するので、経験的ベイズという用語を避ける。

まとめ 5.1

- ・ 階層モデル

…別々の実験の背景にあると仮定される共通の母集団分布について考えるもの. 複雑なモデルに適合できるほど十分多くのパラメータを持ち, 母集団分布を仮定することによって過適合を抑えられる.

- ・ 完全なベイズ解析

…確率モデルをパラメータと実験の集合全体に置き, 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するというもの.

- ・ 経験的ベイズ

…事前パラメータを推定するためのデータの一部を用いる分析. 完全な階層ベイズ解析の近似値と見ることができる.

- ・ 経験ベイズの問題点

1. 事前パラメータ推定に使ったデータを 2 回使うと精度が過大評価になる.
2. 点推定値を使用すると必然的にいくつかの事後不確実性は無視される.
3. ベイズ推論の一部であるのにデータを得る前に知る必要があるのか?

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 および共通の分散 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, 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、交換可能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 以外に 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、直接データに対し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。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できる。(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ぐということではない!) さいころの目のアナロジーを考える。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るが、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る。それによって、他のものよりも 1 つの結果が起こりやすく、6 つの結果の間の対称性がなくなる可能性がある。

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 (または母集団) 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成り立たせ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, 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**超母集団 (supper 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.

例. 交換可能性とサンプリング

以下の思考実験は、ランダムサンプリングからの推論における交換可能性の役割を示している. 簡単化のために, θ ではなく y のレベルで交換可能な非階層的な例を使用します.

米国から 8 州を選択し, 1981 年に各州の人口 1000 人あたりの離婚率を記録した. これらを y_1, \dots, y_8 とする. 8 つの州の離婚率 y_8 についてあなたは何について言うことができるだろうか?

8 つの州のいずれかを他の州と区別するための情報がないので, それらを交換可能とモデル化する必要がある. 8 つの y_j に対してベータ分布, ロジット正規, または $[0, 1]$ の範囲に限定された他の事前分布を使用することができます. 米国の離婚統計に精通していない限り, (y_1, \dots, y_8) の分布はかなり曖昧なはずです.

これらの 8 つから 7 つの州を無作為に抽出し, 離婚率を教える. それぞれ 1000 人あたりの離婚数 (年あたり) は, 5.8, 6.6, 7.8, 5.6, 7.0, 7.1, 5.4 である. 主にデータに基づいて, 残りの値 y_8 の合理的な事後 (予測) 分布は, おそらく 6.5 付近に集中し, 質量の大部分は 5.0~8.0 の間にあると考えらる. 添え字を変更しても, 結合分布は変更されない.

残りの値を他の y_j とすると、事後推定値は同じになる。 y_j は交換可能であるが、8 番目の観察されていない状態での離婚率はおそらく観察された率と類似していると仮定しているので、それらは独立していない。

最初は、8 つの州が山間部の州であるという事前情報を与えたとする。アリゾナ州、コロラド州、アイダホ州、モンタナ州、ネバダ州、ニューメキシコ州、ユタ州、およびワイオミング州であるが、ランダムな順序で選択されている。どの観測された率がどの州に対応しているかはまだ分からない。今、7 つのデータ点が観測される前に、8 つの離婚率は引き続き交換可能にモデル化されるべきです。しかし、8 つの値の事前分布 (すなわち、データを見る前に) は変わるはずです。モルモン教徒が多いユタ州の離婚率がはるかに低く、自由離婚法を持つネバダ州は残りの 6 州より離婚率がはるかに高いと仮定するのが合理的である。おそらく、分布内の異常値の期待値を前提とすると、事前分布には幅の広いテールがあるはずです。この余分な情報 (8 つの州の名前) を見ると、7 つの観測値を見ると、それらの数値が近似していることに気付くと、8 番目の州がネバダ州またはユタ州であることが合理的な推測に見えるかもしれません。したがって、その値は、観察された 7 つの値よりもはるかに低いか、または非常に高いと予想される。

これは、2つのもっともらしいシナリオを説明するために、二峰性または三峰性の事後分布につながる可能性がある。ただし、どの状態がどのインデックス番号に対応しているかを示す情報がないため、8つの値 y_j の事前分布は交換可能です。(演習 5.6 を参照)

最後に、サンプリングされていない状態 (y_8 に対応する) はネバダだったと伝えます。今度は、7つの観測値を見る前に、 y_8 を他の7つの数字と区別する情報があるので、交換可能な事前分布を8つの離婚率のセットに割り当てることはできません。いったん y_1, \dots, y_7 が観測された場合、 y_8 の合理的な事後分布は、その最大の観測された率、すなわち

$p(y_8 > \max(y_1, \dots, y_7) | y_1, \dots, y_7)$ は大きくなければならない。

ちなみに、1981年のネバダの離婚率は、人口1000人当たり13.9だった。

ユニットで追加情報が利用可能な場合の交換可能性

多くの場合、観測は完全には交換可能ではないが、部分的または条件付きで交換可能である。

- ・ 観測値をグループ化できる場合、階層モデルを作成することができます。各モデルにはサブモデルがありますが、グループ特性は不明です。グループ特性が交換可能であると仮定すると、グループ特性に同一の事前分布を使用できる。

- ・ y_i が交換可能ではなく (y_i, x_i) 交換可能であるように y_i が追加情報 x_i を有する場合、 $y_i|x_i$ について (y_i, x_i) または条件付きモデルの同一モデルを作成することができる。

ラット腫瘍の例では、実験条件で追加の知識が得られなかったので、 y_j は交換可能であった。特定のバッチの実験が異なるラボで行われたことが分かっているならば、部分的な交換可能性を仮定し、2つのレベルの階層モデルを使用して各ラボやラボ間のバリエーションをモデル化することができます。

離婚の例では, $j = 1, \dots, 8$ に対する, ただどの添え字がどの状態に対応していない, 去年の状態 j の離婚率 x_j を知っていれば, 確かに y_j の 8 つの値を区別することができるだろうが, 結合事前分布 $p(x_j, y_j)$ は各状態について同じであろう. 去年の同じ離婚率 x_j を有する州では, グループ分けを使用して部分的な交換可能性を仮定することができ, あるいは x_j の可能な値が多い場合 (条件付き交換可能性を仮定すると), 条件付き交換可能性を仮定し x_j を回帰モデルにおいて共変数として使う.

一般に, 共変数に関する交換可能性をモデル化する通常の方法は条件付き独立性を通じてである. $x = (x_1, \dots, x_J)$ に関して,
$$p(\theta_1, \dots, \theta_J | x_1, \dots, x_J) = \int [\prod_{j=1}^J p(\theta_j | \phi, x_j)] p(\phi | x) d\phi.$$
このようにして, 交換可能なモデルは, 普遍的な適用が可能となります. なぜなら, 異なる単位を区別するために利用できる情報は, x および y 変数にエンコードする必要があるからです.

ラット腫瘍の例では, サンプルサイズ n_j が異なる実験を区別するために利用可能な唯一の情報であることに既に留意した. 腫瘍率をモデル化するために n_j が有用な変数ではないように思われるかもしれないが, 興味があれば, J 対 $(n, y)_j$ の交換可能なモデルを作成することができる. 自然な最初のステップは, モデル化することができ, 明らかな関係を見るために $\frac{y_j}{n_j}$ 対ニュージャージー州をプロットすることであろう. 例えば, 研究者の中には, より希少な事象を正確に疑っているため, おそらくいくつかの研究 j はより大きなサンプルサイズ n_j を有していた. すなわち, より小さい θ_j , したがって $\frac{y_j}{n_j}$ のより小さい期待値. 実際, ここには示されていない $\frac{y_j}{n_j}$ 対 n_j のプロットは, 2 つの変数間に明白な関係を示さない.

交換可能なモデルへの反対

事実上あらゆる統計的適用において, ユニットが実際に異なるという理由で交換可能性に反対するのは当然のことです. 例えば, 71 回のラット腫瘍実験は, 異なる時間, 異なるラット, およびおそらく異なる実験室で行われた. ただし, そのような情報は交換可能性を無効にしません. 実験が異なるということは, θ_j が異なることを意味するが, 共通の分布から引き出されたかのように考えることができる. 実際, それらを区別するための情報がないため, 論理的な選択肢はありませんが, θ_j を交換可能とモデル化する必要があります. 無知をモデル化するための交換可能性に反対することは, 一般的な回帰モデルに反対する, あるいは個々のラベルのない散布図に点を表示することに反対する, 共通の母集団からのサンプルの独立かつ同一分布のモデルに反することよりも合理的ではない. 回帰と同様に, 有効な関心事は交換可能性に関するものではなく, 関連する知識を説明変数として可能な限り符号化することに関するものです.

階層モデルの完全なベイズ取り扱い

推論の問題に戻り, 大事なこれらのモデルの階層部分は ϕ が未知で, 従って事前分布 $p(\phi)$ を有するということである. 適当なベイズ事後分布はベクトル (ϕ, θ) である. 結合事前分布は

$$p(\phi, \theta) = p(\phi)p(\theta|\phi)$$

で, 結合事後分布は

$$p(\phi, \theta|y) \propto p(\phi, \theta)p(y|\phi, \theta) = p(\phi, \theta)p(y|\theta) \quad (5.3)$$

ただし, データ分布 $p(y|\phi, \theta)$ は θ のみに依存するので, 後者の簡略化が成り立っている. 超パラメータ ϕ は θ を通してのみ y に影響している. 前に, 非現実的であるが, ϕ は既知であると仮定した. いま, モデルの ϕ に不確実性を含んでいる.

超事前分布

(ϕ, θ) の結合確率分布を作成するためには, 事前分布を ϕ に割り当てなければならない. ϕ についてはほとんど知られていない場合, 事前分布を分散することができますが, 不適切な事前密度を使用して事後分布が適切であることを確認する際には注意が必要です. ほとんどの実際の問題では, 少なくとも超パラメータを有限領域に制限するために, 実質的な超事前分布を割り当てない場合, ϕ のパラメータに関する十分な実質的知識が必要です. 非階層的モデルの場合と同様に, ϕ に関する単純で比較的非情動的な事前分布から出発し, 事後分布の変動があまりにも大きければ, より多くの事前情報を加えることがしばしば実用的である.

ラット腫瘍の例では, 超パラメータは (α, β) であり, θ のベータ分布を決定する. 次の節で, その例の続きで適切な**超事前分布 (hyperprior distribution)** を構築するための 1 つのアプローチを示します.

事後予測分布

階層的モデルは、超パラメータ ϕ , およびパラメータ θ の両方によって特徴付けられる。データアナリストにとって興味深い2つの事後予測分布があります。(1) 既存の θ_j に対応する将来の観測値 \tilde{y} の分布, または (2) 同じ超母集団から引き出される将来の θ_j に対応する観測値 \tilde{y} の分布. 将来の θ_j を $\tilde{\theta}$ とラベル付けする. 6章で議論するように, 両方の種類の複製を使用してモデルの妥当性を評価することができます. ラット腫瘍の例では, 将来の観察は, (1) 既存の実験からの追加のラット, または (2) 将来の実験の結果であり得る. 前者の場合, 事後予測抽出は, 既存の実験での θ_j の事後抽出に基づいている. 後者の場合, 母集団分布から新しい実験のために $\tilde{\theta}$ を抽出し, ϕ の事後抽出を与えてから, シミュレートされた $\tilde{\theta}$ を与えて \tilde{y} を抽出する必要があります.

まとめ 5.2

- ・ 交換可能性

… $p(\theta_1, \dots, \theta_J)$ がインデックス $(1, \dots, J)$ の順列に対して不変であること. 独立同一分布のモデルを構築するために必要. 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できる.

- ・ **de Finetti の定理**

…確率変数の無限系列 $\theta_1, \theta_2, \dots$ は交換可能であるとする. このとき, この中の任意の N 個の観測値の結合確率は, ある確率変数 ϕ を用いて, 独立同分布に得られたかのように表現することが可能.

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

- ・ 交換可能, 独立同一分布

…交換可能であっても独立同一分布としてモデル化できないパターンがある.

- ・ **超事前分布**

…事前分布のパラメータである超パラメータに対する事前分布のこと.

5.3 共役階層モデルの完全なベイズ解析

階層型モデルの推論戦略は、3.8 節で提示された多パラメータ問題への一般的なアプローチに従いますが、実際には階層型モデルに多数のパラメータが存在するため、実際には困難です。特に、 (θ, ϕ) の共役事後分布から、等高線をプロットしたり、シミュレーションの散布図を表示することはできません。しかし、以前と同様のシミュレーションベースのアプローチに従うことができます。

この節では、母集団分布 $p(\theta|\phi)$ が尤度 $p(y|\theta)$ と共役であるラット-腫瘍の例のベータ-二項モデルについて、結合事後分布 $p(\theta, \phi|y)$ からのシミュレーションを得るための解析的方法と数値的方法とを組み合わせたアプローチを提示する。実際に発生する多くの非共役階層モデルでは、本書の3部に示すより高度な計算方法が必要です。しかし、より複雑な問題であっても、共役分布を用いた手法は、より正確な計算のための近似推定値および開始点を得るために有用である。

条件付き分布と周辺分布の解析的導出

最初に分析的に以下の3つのステップを実行します.

- 1, 超事前分布 $p(\phi)$, 母集団分布 $p(\theta|\phi)$, 尤度 $p(y|\theta)$ の積として, 非正規化形式で結合事後密度 $p(\theta, \phi|y)$ を書く.
- 2, 超パラメータ ϕ が与えられたときの θ の条件付き事後密度を解析的に決定する. 固定観測 y については, これは ϕ の関数 $p(\theta|\phi, y)$ である.
- 3, ベイズパラダイムを用いて ϕ を推定する. すなわち, その周辺事後分布 $p(\phi|y)$ を得る.

第1のステップは直ちに, 第2のステップは共役モデルにとって容易である. なぜなら, ϕ の条件付き分布は, θ の母集団分布が独立して同一分布のモデル (5.1) であるため, 条件付き事後密度は, 成分 θ_j の共役事後密度の積である.

第3のステップは、 θ についての結合事後分布を積分することによって強引に実行することができる。

$$p(\phi|y) = \int p(\theta, \phi|y) d\theta \quad (5.4)$$

しかしながら、正規分布を含む多くの標準モデルでは、 ϕ の周辺事後分布は、条件付き確率式

$$p(\phi|y) = \frac{p(\theta, \phi|y)}{p(\theta|\phi, y)} \quad (5.5)$$

を用いて代数的に計算することができる。この式は、分子がちょうど結合事後分布 (5.3) であり、分母が ϕ が既知であれば θ の事後分布であるので、この式は有用である。いくつかの標準共役モデルについて (5.5) を使用することの難しさは、固定 y に対して θ と ϕ の両方の関数として見なされる分母 $p(\theta|\phi, y)$ は、 y だけでなく ϕ に依存する正規化係数である。ベイズの定理における比例定数に注意する必要があります。特に、階層モデルを使用する場合は、実際に一定であることを確認する必要があります。演習 5.11 には、積分 (5.4) に閉形式解がない非共役モデルの例があるので、(5.5) は助けにならない。

事後分布からの抽出シミュレーション

次の戦略は、この章で考察されているような単純な階層モデルの場合、結合事後分布、 $p(\theta, \phi|y)$ からの抽出をシミュレーションするのに役に立つ。

1, 周辺事後分布 $p(\phi|y)$ から、超パラメータ ϕ のベクトルを抽出する。もし、 ϕ が低次元ならば、3 章で議論された手法が使える。高次元の ϕ ならば、3 部で議論されるようなより込み入った手法が必要となる。

2, 抽出された ϕ の値が与えられた、条件付き事後分布 $p(\theta|\phi, y)$ からパラメータベクトル θ を抽出する。この章で考察した例では、因数分解 $p(\theta|\phi, y) = \prod_j p(\theta_j|\phi, y)$ が成り立つので、成分 θ_j は一度に 1 つずつ独立して抽出することができる。

3, 必要に応じて、抽出された θ を与えた事後予測分布から予測値 \tilde{y} を抽出する。問題に応じて、前節の最後で説明したように、 ϕ を与えた新しい値 $\tilde{\theta}$ を抽出する必要があります。

通常、上記のステップは、 L 個の抽出集合を得るために L 回実行される。 θ, \tilde{y} の事後結合シミュレーションから、関心のある任意の推定量または予測量の事後分布を計算することができる。

ラット腫瘍モデルへの応用

ここでは, 5.1 節で説明したラット腫瘍実験の完全なベイズ分析を実行します. 再び, 実験 $j = 1, \dots, J, J = 71$ からのデータは, 独立した二項分布に従うと仮定される.

$$y_j \sim \text{Bin}(n_j, \theta_j)$$

ただし, ラットの数 n_j は既知である. パラメータ θ_j はベータ分布からの独立なサンプルとして仮定される.

$$\theta_j \sim \text{Beta}(\alpha, \beta)$$

で, 未知の超パラメータについての無知を反映するために, 無情報超事前分布を割り当てる. いつものように, 無情報という言葉は, モデルのこの部分に対する態度を示しており, この特定の分布に特別な特性があることを暗示するものではありません. 超事前分布が推論のために重要であることが判明した場合, これを報告し, 可能であれば, よりよい情報事前分布を構築するために使用できるさらなる実質的な知識を求めるべきである. 超パラメータ (α, β) の不適切な事前分布を割り当てる場合, 事後分布が適切であることを確認する必要があります. 事後密度の積分可能性を検査するまで, この特定の分析の比較的恣意的で重要でない無情報超事前分布の選択を延期する.

結合, 条件付き, 周辺事後分布

事後分布の解析形式の決定のためにまず 3 つのステップを行う. すべてのパラメータの結合事後分布は

$$\begin{aligned}
 p(\theta, \alpha, \beta | y) &\propto p(\alpha, \beta) p(\theta | \alpha, \beta) p(y | \theta, \alpha, \beta) \\
 &\propto p(\alpha, \beta) \prod_{j=1}^J \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \theta_j^{\alpha-1} (1 - \theta_j)^{\beta-1} \prod_{j=1}^J \theta_j^{y_j} (1 - \theta_j)^{n_j - y_j}
 \end{aligned} \tag{5.6}$$

となる. (α, β) が与えられたとき, θ の要素は $\theta_j^A (1 - \theta_j)^B$ の形である独立な事後密度, すなわち, ベータ密度をもち, その結合密度は,

$$p(\theta | \alpha, \beta, y) = \prod_{j=1}^J \frac{\Gamma(\alpha + \beta + n_j)}{\Gamma(\alpha + y_j) \Gamma(\beta + n_j - y_j)} \theta_j^{\alpha + y_j - 1} (1 - \theta_j)^{\beta + n_j - y_j - 1} \tag{5.7}$$

となる. 条件付き確率式 (5.5) に (5.6) と (5.7) を代入することにより, (α, β) の周辺事後分布を求めることができる.

$$p(\alpha, \beta | y) \propto p(\alpha, \beta) \prod_{j=1}^J \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} \frac{\Gamma(\alpha + y_j) \Gamma(\beta + n_j - y_j)}{\Gamma(\alpha + \beta + n_j)} \tag{5.8}$$

式 (5.8) の積は分析的に単純化することはできませんが, ガンマ関数を計算するために標準ルーチンを使用して (α, β) の指定された値に対して計算するのは簡単です.

標準的なパラメータ化を選択し, 無情報超事前分布を設定する

ラットの母集団でラット腫瘍の分布についてすぐに利用可能な情報がなければ, (α, β) に対して比較的拡散した超事前分布を求める. 超事前分布を割り当てる前に, $\text{logit}(\frac{\alpha}{\alpha+\beta}) = \log(\frac{\alpha}{\beta})$ と $\log(\alpha + \beta)$ に関して再パラメータ化する. これは, θ のベータ母集団分布における平均値とサンプルサイズの対数のロジットである. 事前平均値とサンプルサイズに独立した超事前分布を割り当てることは合理的なようであり, ロジスティック変換と対数変換を使用してそれぞれを $(-\infty, \infty)$ スケールに置きます. 残念なことに, これらの新たに変換されたパラメータ上の一様事前密度は, 極限 $(\alpha + \beta) \rightarrow \infty$ で無限の積分を伴う不適切な事後密度を生じ、この特定の事前密度はここでは使用できない.

合理的に大量のデータを含むこのような問題では, 尤度によって支配され, 適切な事後分布をもたらす無情報超事前密度を設定することが可能である.

拡散した超事前密度の 1 つの妥当な選択は、 $(\frac{\alpha}{\alpha+\beta}, (\alpha + \beta)^{-1/2})$ 上で一様であり、適切なヤコビアンで乗算すると、元のスケールで以下の密度が得られ、

$$p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2} \quad (5.9)$$

で自然な変換スケールで、

$$p\left(\log\left(\frac{\alpha}{\beta}\right), \log(\alpha + \beta)\right) \propto \alpha\beta(\alpha + \beta)^{-5/2} \quad (5.10)$$

である。この事前密度については演習 5.9 を参照。

適切な超事前分布を使用する場合、事後密度の積分可能性をチェックする数学的な努力を避けることができます。もう 1 つのアプローチは、 $p(\frac{\alpha}{\alpha+\beta}, \alpha + \beta) \propto 1$ または $p(\alpha, \beta) \propto 1$ などの平坦な超事前密度を暫定的に使用し、その後、事後密度から輪郭およびシミュレーションを計算する。(詳しくは以下) 結果は、事後等高線が無限に向かって移動することをはっきりと示し、事後密度がその極限内で積分できないことを示す。積分可能な事後密度を得るために、事前分布を変更しなければならない。

ちなみに, $[-10^{10}, 10^{10}] \times [-10^{10}, 10^{10}]$ のような曖昧で有限の範囲で $(\log(\frac{\alpha}{\beta}), \log(\alpha + \beta))$ の事前分布を一様にするのは, この問題の許容可能な解法は, この場合, ほぼすべての事後質量が, 無限大に近い α, β の範囲にあり, これは分散がゼロの $\text{Beta}(\alpha, \beta)$ 分布に相当し, すべてが θ_j パラメータは, 事後分布において本質的に等しくなる. 尤度が積分可能でない場合, 一様事前密度に遠く離れた有限カットオフを設定しても, 必ずしもその問題が解消されるわけではありません.

超パラメータの周辺事後密度の計算

データとパラメータの完全確率モデルを確立したので, 超パラメータの周辺事後分布を計算する. 図 5.2 は, $(\log(\frac{\alpha}{\beta}), \log(\alpha + \beta))$ の値のグリッド上の正規化されていない周辺事後密度の等高線プロットを示す. プロットを作成するには, 濃度 $p(\log(\frac{\alpha}{\beta}), \log(\alpha + \beta) | y)$ を得るためにヤコビアンを掛けて密度関数 (5.8) の対数を事前密度 (5.9). 以前の点推定値 $(-1.8, 2.3)$ を中心とする範囲 $(\log(\frac{\alpha}{\beta}), \log(\alpha + \beta)) \in [-2.5, -1] \times [1.5, 3]$ (すなわち, $(\alpha, \beta) = (1.4, 8.6)$) であり, 各パラメータにおいて 4 の係数をカバーする. 次に, 計算上のオーバーフローを避けるために, グリッド上の各点からログ密度の最大値を引き, 累乗し, 正規化されていない周辺事後密度の値を生成する.

等高線プロットの最も明白な特徴は, (1) モードが点推定値 (我々が予想する) から遠くはないこと, (2) 周辺後方分布の重要な部分がグラフの範囲外にあることです.

このとき $p(\log(\frac{\alpha}{\beta}), \log(\alpha + \beta)) \in [-2.3, -1.3] \times [1, 5]$ の範囲で

$p(\log(\frac{\alpha}{\beta}), \log(\alpha + \beta)|y)$ を再計算する. 結果のグリッドは, 図 5.3a に示すように, 本質的にすべての周辺事後分布を表示します. 図 5.3b は, 数値的に計算された事後分布からの 1000 回のランダム描画を表示しています. グラフは, この変換の下での超パラメータの周辺事後分布が, おおよそ $(-1.75, 2.8)$ のモードに関して対称であることを示している. これは, $(\alpha, \beta) = (2.4, 14.0)$ の近似値に相当し, 先に得られた粗い推定値とは多少異なる.

(α, β) の有効範囲をカバーするグリッドの相対事後密度を計算した後, グリッド上のステップ関数として分布を近似し, グリッド内の全確率を 1 に設定することによって正規化する.

次に, $(\log(\frac{\alpha}{\beta}), \log(\alpha + \beta))$ のグリッドに基づいて事後モーメントを計算することができる. 例えば,

$$E(\alpha|y) \text{ is estimated by } \sum_{\log(\frac{\alpha}{\beta}), \log(\alpha+\beta)} \alpha \cdot p(\log(\frac{\alpha}{\beta}), \log(\alpha + \beta)|y)$$

図 5.3 のグリッドから, $E(\alpha|y) = 2.4$ と $R(\beta|y) = 14.3$ を計算します. 事後分布は $(\log(\frac{\alpha}{\beta}, \log(\alpha + \beta))$ のスケール上でほぼ対称であるため, 上記の図 5.3a のモードに基づく推定値に近い. グリッドを平均化することのより重要な結果は, $(\alpha + \beta)$ の事後不確実性を説明することであり, これは点推定では捕捉されない.

パラメータと超パラメータの結合事後分布からサンプリングする

次のように, $(\alpha, \beta, \theta_1, \dots, \theta_J)$ の結合事後分布から 1000 個のランダムサンプルを抽出する.

1, 3.8 節のバイオアッセイの例に対して, 図 3.3b の $(\alpha + \beta)$ を抽出するのに使用された同じ離散グリッドサンプリング手順を使用して, 図 5.3 に表示された事後分布から $(\log(\frac{\alpha}{\beta}, \log(\alpha + \beta))$ を 1000 回の抽出をシミュレートする.

2, $l = 1, \dots, 1000$ に対して,

(a) $(\log(\frac{\alpha}{\beta}, \log(\alpha + \beta))$ の l 番目の抽出をスケール (α, β) に変換して, それらの周辺事後分布から超パラメータを抽出する.

(b) 各 $j = 1, \dots, J$ に対して, θ_j をその条件付き事後分布 $\theta_j | \alpha, \beta, y \sim \text{Beta}(\alpha + y_j, \beta + n_j - y_j)$ からサンプリングする.

結果を表示する

図 5.4 は, 事後中央値と θ_j の 95%区間をシミュレーションで計算したものです. 速度 θ_j は, それらの標本点推定値 $\frac{y_i}{n_j}$ から母集団分布に近似平均 0.14 で縮小される. より少ない観測値での実験は, より収縮し, より高い事後分散を有する. 結果は, かなり多くの実験のために, この例では意味をなさない, 超パラメータのポイント推定に基づいて得られるものとは表面的に類似している. しかし, 主要な相違点が残っており, 特に, 超パラメータの事後不確定性を反映して, 完全なベイズ解析では事後変動がより高い.

まとめ 5.3

- ・条件付き分布と周辺分布の解析的導出

1. 超事前分布 $p(\phi)$, 母集団分布 $p(\theta|\phi)$, 尤度 $p(y|\theta)$ の積として, 非正規化結合事後密度 $p(\theta, \phi|y)$ を導く.
2. 超パラメータ ϕ が与えられたときの θ の条件付き事後密度を解析的に決定する. 固定観測 y については, これは ϕ の関数 $p(\theta|\phi, y)$ である.
3. ベイズパラダイムを用い ϕ を推定し, 周辺事後分布 $p(\phi|y)$ を得る.

- ・事後分布からの抽出シミュレーション

1. 周辺事後分布 $p(\phi|y)$ から, 超パラメータ ϕ のベクトルを抽出する. もし, ϕ が低次元ならば, 3 章で議論された手法が使える, 高次元ならば, 3 部で議論されるようなより込み入った手法が必要となる.
2. 抽出された ϕ の値が与えられた, 条件付き事後分布 $p(\theta|\phi, y)$ からパラメータベクトル θ を抽出する. $p(\theta|\phi, y) = \prod_j p(\theta_j|\phi, y)$ (交換可能) ならば, 成分 θ_j は一度に 1 つずつ独立に抽出することができる.
3. 必要に応じて, 抽出された θ を与えた事後予測分布から予測値 \tilde{y} を抽出する. 問題に応じて, ϕ を与えた新しい値 $\tilde{\theta}$ を抽出する必要があります. 通常, 上記のステップは, L 個の抽出集合を得るために L 回実行される.

5.4 正規モデルからの交換可能なパラメータの推定

ここでは、観察されたデータが、各グループまたは実験ごとに異なる平均で正規分布に基づいた単純な階層モデルの完全な扱いを示しています。このモデルは、既知のデータ分散を有する一方向の通常のランダム効果モデルと呼ばれることもあり、広く適用されており、15章で一般的に取り扱う階層正規線形モデルの重要な特殊ケースです。この節では、5.3節の計算手法に従った一般的な処理を提示する。次の節では、詳細な例を示します。代数的な詳細に気をつけていない人は、動機付けの例を先取りして見たいかもしれません。

データ構造

それぞれ既知の誤差分散 σ^2 で n_j 個の独立で正規分布に従うデータ点 y_{ij} , すなわち,

$$y_{ij}|\theta_j \sim N(\theta_j, \sigma^2), \text{ for } i = 1, \dots, n_j, j = 1, \dots, J \quad (5.11)$$

からパラメータ θ_j を推定する, J 個の独立な実験を考える. 分散の解析から標準的な表記を用いることで, グループ j のサンプル平均を

$$\bar{y}_{\cdot j} = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$$

としてラベル付けする. ただし, サンプル分散

$$\sigma_j^2 = \sigma^2 / n_j$$

である. 各 θ_j に対する尤度を十分統計量 $\bar{y}_{\cdot j}$ を使うことによって,

$$\bar{y}_{\cdot j}|\theta_j \sim N(\theta_j, \sigma_j^2) \quad (5.12)$$

とかける.

各グループ j の平均に対して別々の分散 σ_j^2 を許容する柔軟性のために、後で有用となる表記法である。この章の残りの部分では、すべての式は既知の値 σ_j^2 に対して暗黙的に条件付きである。未知の分散を有する手段の集合を推定する問題は、11.6, 13.6 節に示されるいくつかの追加の計算方法を必要とする。まれに厳密には真実であるが、モデルのサンプリングレベルでの既知の分散の仮定は、しばしば適切な近似である。

この章で提供されるモデルの処理は、実験のデータポイント数以外の理由で分散が異なる状況にも適しています。実際、尤度 (5.12) はここで述べたよりもはるかに一般的な文脈で現れることがあります。例えば、グループサイズ n_j が十分に大きい場合には、平均 y である。 j は、データ y_{ij} がそうでないときでも、 θ_j を与えられて、ほぼ正規に分布する。実際の可能性が (5.12) でよく近似されている他の応用は、次の 2 つのセクションに表示されます。

実用的な配慮から事前分布を構築する

パラメータベクトル $\theta = (\theta_1, \dots, \theta_J)$ の事前分布を特定する問題を直ちに考えるのではなく、与えられたデータ (y_{ij}) に対してどんな種類の事後推定が妥当であろうと考えてみよう. 簡単な自然なアプローチは、実験 j の平均結果である θ_j を $\bar{y}_{.j}$ と推定することです. しかし、例えば、実験群あたり $n_j = 2$ の観察しかない $J = 20$ の実験があり、群は本質的に同一の条件下で同じラットの同じ系統から採取された 20 対の検定である場合はどうなるだろうか? グループごとの 2 つの観測では、正確な見積りができません. 20 群は同じラットの系統に由来するので、ここでは、各 θ_j をプールされた推定値

$$\bar{y}_{..} = \frac{\sum_{j=1}^J 1/\sigma_j^2 \bar{y}_{.j}}{\sum_{j=1}^J 1/\sigma_j^2} \quad (5.13)$$

で推定することを好むかもしれない.

どの推定値を使用するかを決定するために、古典的統計からの伝統的なアプローチは、分散分析を実行して、平均間の差異をテストすることです. J グループ平均が有意に変動すると思われる場合は、別々の標本平均を選択し、グループ平均間の分散がグループ内の個々の変動によって説明できるものより著しく大きくない場合は、 $\bar{y}_{..}$ を使用します.

分散表の理論的解析は、以下の通りである。ここで、 τ_2 は $\theta_1, \dots, \theta_J$ の分散である。わかりやすくするために、すべての j について $n_j = n$ および $\sigma_j^2 = \sigma^2/n$ となるバランスの取れた設計の分散分析を示します。

	df	SS	MS	E(MS σ^2, τ)
Between groups	$J - 1$	$\sum_i \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2$	$SS/(J - 1)$	$n\tau^2 + \sigma^2$
Within groups	$J(n - 1)$	$\sum_i \sum_j (y_{ij} - \bar{y}_{.j})^2$	$SS/(J(n - 1))$	σ^2
Total	$Jn - 1$	$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$	$SS/(Jn - 1)$	

分散の古典的なランダム効果分析では、表の二乗和 (SS) と二乗平均 (MS) 列を計算し、 τ を推定するために 'between' と 'within' 平均二乗を使用します。平均二乗内の比が 1 よりも著しく大きい場合、分散分析は、各 j について別個の推定値 $\theta_j = y_{\cdot j}$ を示唆する。平均二乗の比が「統計的に有意」でない場合、F 検定は、 $\tau = 0$ であるという仮説を棄却することはできず、プールすることは合理的である。 $\theta_j = y_{\cdot j}$ 階層的回帰モデルの文脈における 15.6 節のベジアン分散分析について議論する。

しかし、私たちは完全なプールと無しのどちらかを選択することを余儀なくされていません。代わりに、重み付けされた組み合わせを使用することができます。

$$\hat{\theta}_j = \lambda_j \bar{y}_{\cdot j} + (1 - \lambda_j) \bar{y}_{\cdot \cdot}$$

ここで、 λ_j は 0 と 1 の間にある。

どのような事前モデルがこれらの様々な事後推定を生成するのか？

1, J 個の値 θ_j が $(-\infty, \infty)$ の独立した一様事前密度を有する場合、プールされていない推定値 $\theta_j = \bar{y}_{\cdot j}$ は事後平均である。

2, プールされた推定値 $\theta = y_{\cdot \cdot}$ は、 J 個の値 θ_j が等しくなるように制限され、共通の θ 上の一様事前密度を有する場合の事後平均である。

3, 重み付けされた組合せは、 J 個の値 θ_j が独立しており、正規事前密度と同じに分布している場合、事後平均である。

これら 3 つのオプションはすべて θ_j で交換可能であり、オプション 1 と 2 はオプション 3 の特別な場合です。すべての j について $\lambda_j \equiv 1$, θ_j について無限の事前分散に対応するプーリングはなく、完全なプーリングは、すべての j について $\lambda_j \equiv 0$ に、 θ_j についてゼロの事前分散に対応する。

階層モデル

共役性 (より正確には, 部分的共役性) の便利さについて, パラメータ θ_j は超パラメータ (μ, τ) に関する正規分布から抽出されると仮定する.

$$p(\theta_1, \dots, \theta_J | \mu, \tau) = \prod_{j=1}^J N(\theta_j | \mu, \tau^2) \quad (5.14)$$

$$p(\theta_1, \dots, \theta_J) = \int \prod_{j=1}^J [N(\theta_j | \mu, \tau^2)] p(\mu, \tau) d(\mu, \tau)$$

すなわち, θ_j は (μ, τ) が与えられた条件付き独立である. 階層モデルはまた, ラット腫瘍について図 5.1 に示すように, 共有集団分布からのランダムサンプルとしての θ_j の解釈を可能にする.

τ が与えられた, μ に対する無情報一様超事前分布を割り当てる.

$$p(\mu\tau) = p(\mu|\tau)p(\tau) \propto p(\tau) \quad (5.15)$$

μ の均一な事前密度は, この問題に対して一般に妥当である. すべての J 個の実験からの結合されたデータは, 一般に μ について非常に有益であるため, 事前分布についてあいまいである可能性がある. τ の事前分布については, 解析の後半まで議論を延期するが, 関連する原則は既にラット腫瘍の例において議論されている. いつものように, 最初に超パラメータ上の条件式の答えを求め, その事前分布と事後分布を考慮する.

結合事後分布

観測可能な y_{ij} とサンプリングモデルと事前分布を組み合わせると, 十分統計量 \bar{y}_j で表すことができるすべてのパラメータと超パラメータの結合事後分布が得られる.

$$\begin{aligned} p(\theta, \mu, \tau | y) &\propto p(\mu, \tau) p(\theta | \mu, \tau) p(y | \theta) \\ &\propto p(\mu, \tau) \prod_{j=1}^J N(\theta_j | \mu, \tau^2) \prod_{j=1}^J N(\bar{y}_{\cdot j} | \theta_j, \sigma_j^2) \end{aligned} \quad (5.16)$$

この分析では, y とパラメータ σ_j のみに依存する因子を無視することができます.

超パラメータが与えられた正規平均の条件付き事後分布

一般的な階層構造と同様に, パラメータ θ_j は事前分布 (μ と τ が与えられている) において独立であり, 尤度 (5.11) の異なる因子に現れる. したがって, 条件付き事後分布 $p(\theta|\mu, \tau, y)$ は, J 個の成分に影響を及ぼす.

超パラメータに関しては, 正規事前分布を仮定すると, J 個の独立した未知の正規平均を単に持つことができるので, 各 θ_j に対して 2.5 節の方法を独立して用いることができる. θ_j に対する条件付き事後分布は独立で,

$$\theta_j | \mu, \tau, y \sim N(\hat{\theta}_j, V_j)$$

となる. ここで,

$$\hat{\theta}_j = \frac{1/\sigma^2 \bar{y}_{\cdot j} + 1/\tau^2 \mu}{1/\sigma_j^2 + 1/\tau^2}, \quad V_j = \frac{1}{1/\sigma_j^2 + 1/\tau^2} \quad (5.17)$$

である. 事後平均は事前集団平均と j 番目のグループのサンプル平均の精度重みの重み付き平均である. $\hat{\theta}_j, V_j$ に対するこの式はデータと μ, τ の関数である. μ, τ が与えられた θ_j の条件付き事後密度は適切である.

超パラメータの周辺事後分布

解は, 未知の μ, τ に依存するため, まだ不完全である. アプローチの次のステップは, 超パラメータの完全なベイズ取り扱いである. 5.3 節では, 結合事後密度 $p(\theta, \mu, \tau|y)$ から $p(\mu, \tau|y)$ を得るための2つのアプローチとしての積分または解析計算について述べる. 階層正規モデルの場合, 超パラメータに関するデータによって直接提供される情報を直接考慮することができる.

$$p(\mu, \tau|y) \propto p(\mu, \tau)p(y|\mu, \tau)$$

多くの問題について, この分解は役に立たない. なぜなら, 周辺尤度 $p(y|\mu, \tau)$ は, 一般に閉じた形で書くことができないからである. しかし, 正規分布の場合, 周辺尤度は特に単純な形をとっている. 群の θ での平均 $y_{\cdot j}$ の周辺分布は, 独立して (ただし, 同一の分布ではない) 正規であることを意味する.

$$\bar{y}_{\cdot j}|\mu, \tau \sim N(\mu, \sigma_j^2 + \tau^2)$$

こうして, 周辺事後密度を次のように書き表せる.

$$p(\mu, \tau|y) \propto p(\mu, \tau) \prod_{j=1}^J N(\bar{y}_{\cdot j}|\mu, \sigma_j^2 + \tau^2) \quad (5.18)$$

τ が与えられた μ の事後分布

(5.18) を用いて事後分布 $p(\mu, \tau|y)$ を 2 つの変数の関数とし直接計算し、ラット腫瘍の例のように進めることができた。しかし、正規モデルでは、 $p(\tau|y)$ の単変量数値計算を残し、 μ の積分でさらに単純化できる。事前密度 (5.15) と同様に、超パラメータの周辺事後密度を因数分解する。

$$p(\mu, \tau|y)p(\mu|\tau, y)p(\tau|y) \quad (5.19)$$

(5.19) の右辺の第 1 成分は、 τ がわかっていれば μ の事後分布にすぎない。既知の仮定された τ を用いた (5.18) の検査から、一様条件付き事前密度 $p(\mu|\tau)$ により、対数事後分布は μ で二次的であると分かる。従って、 $p(\mu|\tau, y)$ は正規でなければならない。この分散の平均および分散は、分散 $(\sigma_j^2 + \tau^2)$ を有する μ の J 個の独立した推定値としてグループ平均 \bar{y}_j を考慮すると直ちに得られる。データを一様事前密度 $p(\mu|\tau)$ と組み合わせ、

$$\mu|\tau, y \sim N(\hat{\mu}, V_\mu)$$

が得られる。ここで、 $\hat{\mu}$ は \bar{y}_j の値の精度の重み付き平均で、 V_μ^{-1} は総精度

$$\hat{\mu} = \frac{\sum_{j=1}^J 1/(\sigma_j^2 + \tau^2) \bar{y}_j}{\sum_{j=1}^J 1/(\sigma_j^2 + \tau^2)}, \quad V_\mu^{-1} = \sum_{j=1}^J \frac{1}{\sigma_j^2 + \tau^2} \quad (5.20)$$

である。結果は τ が与えられた、 μ に対する適切な事後密度である。

τ の事後分布

(5.19) と (5.18) と (5.20) の分子と分母の置換から, τ の事後分布を解析的に求めることができる.

$$p(\tau|y) = \frac{p(\mu, \tau|y)}{p(\mu|\tau, y)} \propto \frac{p(\tau) \prod_{j=1}^J N(\bar{y}_{\cdot j} | \mu, \sigma_j^2 + \tau^2)}{N(\mu | \hat{\mu}, V_\mu)}$$

この等式は, μ の任意の値に対して成り立たなければなりません (言い換えると, μ のすべての要素は, 式が簡略化されるときに取り消されなければなりません). 特に, μ を $\hat{\mu}$ に設定すると, 式の評価が簡単になります.

$$\begin{aligned} p(\tau|y) &\propto \frac{p(\tau) \prod_{j=1}^J N(\bar{y}_{\cdot j} | \hat{\mu}, \sigma_j^2 + \tau^2)}{N(\hat{\mu} | \hat{\mu}, V_\mu)} \\ &\propto p(\tau) V_\mu^{1/2} \prod_{j=1}^J (\sigma_j^2 + \tau^2)^{-1/2} \exp\left(-\frac{(\bar{y}_{\cdot j} - \hat{\mu})^2}{2(\sigma_j^2 + \tau^2)}\right) \end{aligned} \quad (5.21)$$

ただし, $\hat{\mu}, V_\mu$ は (5.20) で定義される. 両方の式は τ の関数であり, $p(\tau|y)$ が τ の複雑な関数であることを意味する.

τ の事前平均

分析を完了するには、 τ に事前分布を割り当てる必要があります。便宜上、 τ には拡散しない無情報事前密度を使用するため、結果的に得られる事後密度を調べて有限の積分を保証する必要があります。例示的な分析のために、一様事前分布 $p(\tau) \propto 1$ を用いる。 τ の均一な事前密度が適切な事後密度をもたらす、対照的に、分散成分 $p(\log \tau) \propto 1$ の一見合理的な無情報事前分布が τ の不適切な事後分布をもたらすのを数学的に示すことを演習として残す。あるいは、応用では、母集団分散 τ の「最良の推測」と上限を決定するための余分な労力を必要とせず、スケーリングされた逆- χ^2 分布族 (分散パラメータの自然選択) から妥当な事前分布を構築することができます。スケーリングされた逆- χ^2 密度の平均に「最良の推測」を、99 上限パーセンタイルに上限をマッチングさせる。最初の分析が無情報一様事前密度を用いて行われると、より現実的な事前分布を有する感度分析がしばしば望ましい。

計算

このモデルに対する, θ の事後分布の計算はシミュレーションを通して最も簡単に行われる. 上で使われる因数分解に従って,

$$p(\theta, \mu, \tau | y) = p(\tau | y) p(\mu | \tau, y) p(\theta | \mu, \tau, y)$$

第1ステップは, τ をシミュレートすることで, (5.21) から計算された $p(\tau | y)$ について, τ の一様に間隔を置いた値のグリッド上で, 逆 cdf 法を用いることで数値的に簡単に行うことができる (1.9 節を参照). 第2と第3のステップは, μ, θ をシミュレートすることで, θ_j を独立に得るために, 正規分布からサンプリングすることで簡単に行うことができる. (5.20) から μ を求めて (5.17) から θ_j を得られる.

事後予測分布

新しいデータ, 現在または新しいバッチの事後予測分布からのサンプリングは自然に, パラメータの事後分布から与えられた抽出である. 2つのシナリオ, (1) 平均 $\theta = (\theta_1, \dots, \theta_J)$ で, 現在のバッチの集合から新しいデータ \tilde{y} , (2) 平均 $\tilde{\theta} = (\tilde{\theta}_1, \dots, \tilde{\theta})$ で, \tilde{J} 個の将来のバッチからの将来のデータ \tilde{y} を考える. 後者の場合, 将来のバッチに対する \tilde{J} 個の個々のサンプルサイズ \tilde{n}_j をも特定しなければならない.

パラメータ θ の現在のバッチから新しデータ \tilde{y} の事後予測分布からの抽出を得るために, まず $p(\theta, \mu, \tau | y)$ から抽出し (5.11) から予測データ \tilde{y} を抽出する.

J 個の新しいグループに対する新しいデータ \tilde{y} の事後予測シミュレーションを得るために, 次の3つのステップを行う. まず, 事後分布から (μ, τ) を抽出する. 次に, 母集団分布 $p(\tilde{\theta}_j | \mu, \tau)$ から \tilde{J} 個の新しいパラメータ $\tilde{\theta} = (\tilde{\theta}_1, \dots, \tilde{\theta})$ を抽出する. これは, (式 (5.14)) 超パラメータが与えられた θ に対する母集団, または事前分布である. 最後に, データ分布 (5.11) から θ が与えられた \tilde{y} を抽出する.

超パラメータの自然な非ベイズ推定による難しさ

完全なベイズ手法のいくつかの利点を見るために、データから μ, τ の点推定に基づいて時々使用される近似法と比較します。以前に提示された分散分析から導き出された不偏点推定値は、

$$\hat{\mu} = \bar{y}_{..}, \quad \hat{\tau}^2 = (MS_B - MS_W)/n \quad (5.22)$$

である。 MS_B, MS_W は、分散分析からそれぞれ「間」と「内」の平均二乗です。この代替アプローチでは、 $\theta_1, \dots, \theta_J$ に対する推論は条件付き事後分布 $p(\theta | \hat{\mu}, \hat{\tau})$ に基づく。

前節のラット腫瘍の例で見たように、超パラメータの点推定値を置き換えることの主な問題は、それらの本当の不確実性を無視することです。結果として生じる θ の推論は、ベイズ事後要約として解釈することはできない。さらに、(5.22) の推定 $\hat{\tau}^2$ には、それが負になりうるという欠点がある！

分散成分に対する負の推定量の問題は, MS_W が MS_B を超える場合に $\hat{\tau}^2$ を 0 に設定することによって回避することができるが, これは新たな問題を生じさせる. $MS_W > MS_B$ が強すぎると思われる場合は常に $\tau^2 = 0$ を推定する. $MS_W > MS_B$ である場合, サンプルサイズは, τ^2 がゼロから区別されるには小さすぎるが, これは, $\tau^2 = 0$ であることを知っているとは同じではない. 後者の主張は暗黙のうちに点数の推定によってなされ, すべての群が θ_j が絶対同一であることを意味し, 次の節の例に示すように, 科学的には不可解な主張につながる. この特定の困難を回避するために (μ, τ) の点推定値を構築することは可能であるが, 不確実性を無視して, すべての点推定値に共通の問題を依然として有する.

まとめ 5.4

- ・ 階層モデルの例

…例: 観察されたデータが, 各グループまたは実験ごとに異なる平均で正規分布に基づいている. ただ, 各グループに別々に, 通常の正規モデルを適応し, その結果の重み付き平均をとるといった感じで簡単.

5.5 例: 8つの学校での並列実験

ベイズ分析が他の方法との重要な点で異なる結論を与えるという問題を有する階層的正規モデルを説明する。

特別なコーチングプログラムがテストの得点に及ぼす影響を分析するための教育試験サービスの調査が実施された。SAT-V(修学能力試験-口頭)のコーチングプログラムが8つの高等学校のそれぞれに及ぼす影響を評価するために、別々のランダム化実験が行われました。各試験の結果変数は、SAT-V(教育試験サービスによって管理され、大学が入学決定をするのを助けるために使用される標準化多項選択試験)の特別な管理に関するスコアであった。スコアは200と800との間で変化することができ、平均は約500、標準偏差は約100である。SAT試験は、特に試験の成績を向上させるための短期的な努力に影響されにくいように設計されています。その代わりに、それらは、長年の教育によって培われた知識と能力を反映するように設計されています。それにもかかわらず、本研究の8つの学校のそれぞれは、短期間のコーチングプログラムでSATスコアの向上が成功すると考えた。また、8つのプログラムのいずれかが他のどのプログラムよりも効果的である、もしくは、他のプログラムと類似した効果を持つと考えられる理由は何もなかった。

実験の結果を表 5.2 に要約する. 実験に参加したすべての生徒はすでに PSAT(予備的 SAT) を受験しており, 指導生と非指導生の PSAT-M(数学) と PSAT-V テストの差異が生じました. 特に, 各学校において, 推定されたコーチング効果およびその標準誤差は, 共分散調整の分析によって得られた (すなわち, PSAT-M および PSAT-V を対照変数として用いて, 処置群に対して SAT-V の線形回帰を行った) 完全にランダム化された実験に適しています. 各学校について別の回帰が推定された. 単純なサンプル平均ではないが (共分散の調整のため), y_j とラベルされた推定コーチング効果とそのサンプリング分散 σ_j^2 は, 前節の $y_{.j}$ と $\sigma_{.j}^2$ と同じ役割を果たす. 推定値 y_j は独立した実験によって得られ, 8 つの実験のすべてのサンプルサイズが各学校の 30 人を超える比較的大きいため, すべての実用目的で既知のサンプリング分散を有するほぼ正規のサンプリング分布になる (4.1 節のデータ削減の議論を想起する). ちなみに, SAT-V 上の 8 点の増加は, 約 1 問以上の試験問題の正解に相当する.

非階層モデルに基づく推論とその問題

階層ベイズモデルをフィッティングする前に, 2 つのより簡単な非階層的
手法 (8 つの実験の効果を個別に推定し, プーリングを完了する) を検討
し, この例に対してこれらどちらのアプローチも適切ではことを議論
する.

別々の推定量

表 5.2 を大まかに見ると, まずコーチングプログラムの中には適度な効
果 (18-28 ポイントの範囲) のものがあり, ほとんどが小さな効果 (0-12 ポ
イント) を有し, 2 つはマイナスの効果があることが示唆される. しかし,
これらの推定された効果の標準誤差に注目すると, 実験のいずれかを統
計的に区別することは困難であることがわかります. 例えば, 各実験を
別々に処理し, 単純な正規分析をそれぞれ適用すると, 95%の事後区間が
すべて実質的に重複する.

プールされた推定量

独立な分析に基づく事後区間の一般的な重なりは, すべての実験が同じ
量を推定している可能性があることを示唆している. すべての実験が同
じ効果を持ち, この共通の効果の独立した推定値を生成するという仮説
の下で, 表 5.2 のデータを, 既知の分散を有する 8 つの正規分布観測とし
て扱うことができる.

無情報事前分布に関し, $y_{.j}$ の代わりに y_j を用いて, 式 (5.13) で定義されているように, 学校における共通コーチング効果の事後平均は $y_{..}$ となる. このプールされた推定量は 7.7 で, 8 つの実験は独立なので事後分散は $(\sum_{j=1}^8 1/\sigma_j^2)^{-1} = 16.6$ である. 従って, 共通の効果は, 標準誤差が $\sqrt{16.6} = 4.1$ である 7.7 であり, 95%の事後区間 $[-0.5, 15.9]$ または約 $[8 \pm 8]$ を導くと推定される. この分析を支持して, すべての θ_j が同じ量を推定しているという仮説の古典的検定は, その自由度 (この場合, 7) よりも小さい χ^2 統計量をもたらす. $\sum_{j=1}^8 (y_j - \bar{y}_{..})^2 / \sigma_i = 4.6$. 言い換えると, (5.22) からの推定値 $\hat{\tau}^2$ は負である.

8 つの学校すべてでコーチングの効果が本当に同じであれば, 1 つの学校の観察された効果をちょうど偶然に 28 にすることは可能だろうか? この仮定が真であるならば, 8 つの研究にわたって期待される自然変動の感覚を得るために, 推定される取り扱いの効果は, 平均 8 点および標準偏差 (8 つの分散 σ_j^2 の平均の平方根) 13 点を有する正規分布からの 8 つの独立した抽出であると仮定する. 次に, 通常の順序統計量の期待値に基づいて, y_j の最大観測値は約 26 点, 他のは順に約 19, 14, 10, 6, 2, 3, -9 である点. これらの予想される効果サイズは, 表 5.2 の観察された効果サイズのセットと一致する. したがって, 学校 A は実際には 28 ポイントの効果があると信じることは賢明ではないようです.

別々の推定量とプールされた推定量の問題

2つの極端な態度、(それぞれの θ_j を別々に考慮する別々の分析) とプールされた推定値につながる代替的な視点 (単一の共通の効果)) に関する問題を見るためには、学校 A での効果 θ_1 を考慮する。学校 A の効果は 28.4 と推定され、別の分析では標準誤差が 14.9 であり、プールされた推定値は 7.7 であり、共通誤差モデルでは標準誤差が 4.1 である。8つの学校の別々の分析は、「A の真の効果が 28.4 を超える確率は $\frac{1}{2}$ 」であり、他の7つの学校の結果を考慮すると疑わしい。一方、プールされたモデルは、次を示す。有意でない χ^2 検定にもかかわらず、「A の真の効果が 7.7 未満である確率は $\frac{1}{2}$ 」であり、私たちの知識の不正確な要約であると思われる。また、表 5.2 のデータから正当化することは困難である「確率は A における真の効果は C で真の効果未満であるという確率は $\frac{1}{2}$ である」をも示す。前節の理論的議論のように、どちらの推定も完全に満足できるものではなく、すべての θ_j が等しいと仮定せずに 8つの実験からの情報を組み合わせた妥協を望む。階層的モデルの下でのベイズ分析は、それを正確に提供する。

階層モデルの下での事後シミュレーション

結果として, 5.4 節で示される正規モデルに基づき, $\theta_1, \dots, \theta_8$ の事後分布を計算する. (この問題でこのモデルを適用することの妥当性については, 6.5, 17.4 節で詳しく説明しています). 前節の終わりで論じたように, 事後分布 τ, μ, θ をその順序でシミュレートすることにより, ベイズモデルに対する事後分布から抽出する. サンプル標準偏差 σ_j は既知であり, 表 5.2 の値と等しいと仮定し, μ, τ について独立した一様事前密度を仮定する.

結果

(5.21) より周辺事後密度 $p(\tau|y)$ は図 5.5 にプロットされている. 0 に近い τ の値がもっともらしい. 0 がもっともありそうな値で, 10 以上の τ の値は, $\tau = 0$ の半分の確率しかなく, $\Pr(\tau > 25) \approx 0$ である. 他のモデルのパラメータの周辺分布と結合分布を扱う推論はシミュレーション値から得られる. 詳細はこの節での議論で示される. しかしながら, 正規階層モデルでは, τ が与えられた (また, μ で平均化された) 条件付き事後分布を考えることでよい取り扱いを得る.

条件付き事後平均 $E(\theta_j|\tau, y)$ (μ で平均化された) は, 図 5.6 で τ の関数として示されている. 縦軸は θ_j に対するスケールを表している. 図 5.6 と横軸では同じスケールを持つ図 5.5 を比較すると, τ のありそうな値の多くに対して, 推定された効果が比較的近いということが見て取れる. τ が大きくなるにつれ, つまり, 学校でより多様化していくことに対応化したが, 表 5.2 で推定値が実際の値により近くなるのだ.

図 5.7 の線は, 条件付き標準偏差 $\text{sd}(\theta_j|\tau, y)$ を τ の関数として表している. τ が増加するにつれ, 母分布は 8 つの影響に従いそれぞれが違っていき, 従って, 各 θ_j の事後不確実性は増し, $\tau \rightarrow \infty$ の極限で表 5.2 の標準偏差に近づいていく. (τ が与えられた, 成分 θ_j に対する事後平均と標準偏差は, μ で平均化し, 平均と分散の公式 (2.7), (2.8) を, 用いることで計算される. 演習 5.12 を参照.)

図 5.5-5.7 を見たところ, 一般的な k 結論は, ある学校で 28.4 点以上の効果が出るということはあまりないということを示している. τ のありそうな値に対して, すべての学校の推定量は実質的に 28 点よりも低い. 例えば, $\tau = 10$ でさえ, 学校 A での効果が 28 点より低いという確率は, $\Phi[(28 - 14.5)/9.1] = 93\%$ である. ここで, Φ は標準正規累積分布関数である. ほかの学校では効果が 28 点より低いという確率は 99.5%, 99.2%, 98.5%, 99.96%, 99.8%, 97%, 98% である.

実質的に重要なことであるが, τ の事後モードで条件付けをするならば, データの正確な要約は得られない. τ のような超パラメータの形式的な値 (例えば, 最尤推定量) に条件付けするという技術は, 実際にはしばしば使われるが, 超パラメータの事後分布によりもたらされる不確実性を無視している. $\tau = 0$ では, 推論はすべての実験が同じサイズの影響, 7.7 点と同じ標準偏差, 4.1 点を有しているというものである. 図 5.5-5.7 は確かに, この回答は 8 つの学校の推定量に引きずられ過ぎているということを示している. 問題はこの問題では特に明確である. なぜなら, τ の事後モードがパラメータ空間の境界上にあるからである. $(\theta_1, \dots, \theta_J, \mu, \tau)$ の結合事後推定は一般的にもっと悪い問題で苦しめられる.

議論

表 5.3 は 8 つの学校に対して 200 個のシミュレートされた影響を要約している. ある意味で, これらの結果はプールされた 95% 区間 $[8 \pm 8]$ と似ている. そこでは, 8 つのベイズ 95% 区間は大きく重なっており, 5 と 10 の間に中央値中心となっている. また, 表の結果が 8 つの独立な回答への方向でプールされた推定量と異なる. 95% ベイズ区間は一般的な区間のほとんど 2 倍で, 実質的に 16 点よりも大きい影響の確率, 特に学校 A の, そして, 負の影響の確率, 特に学校 C のを示唆している. もし, 事後区間においてより良い精度が求められるならば, より多くのシミュレーション抽出を行うかもしれない. ここでは, 小さいシミュレーションが多くの実際の目的で適切な推論を与えるということを示すために, 200 回の抽出しか行っていない.

表 5.3 で示されている 8 つの学校の影響の順位付けは, 本質的に 8 つの別々の推定より得られるものと同様である. しかしながら, 詳しくは違いがある. 例えば, 学校 A の影響が 28 点よりも大きいというベイズ確率は 10% よりも小さい. これは実質的に, 学校 A に対する個別の推定に基づく 50% の確率よりも小さい.

シミュレーションに基づく事後結果の示すように, 学校 A の影響の 200 回のシミュレーションは図 5.8a に示されている. パラメータ θ を推定することで, このモデルのより複雑な問題に答えやすくなる. 例えば, 8 つのコーチングプログラムの最も成功した影響である, 最大値 $\{\theta_j\}$ の事後分布はどんなものか? 図 5.8b はこの事後分布からの 200 個の値のヒストグラムで, 28.4 以上の抽出は 22 個しかないということを示している.

従って, $\Pr(\max\{\theta_j\} > 28.4) \approx \frac{22}{200}$ である. 図 5.8a は学校 A の影響の周辺事後分布を与え, 図 5.8b は最大影響の事後分布を与えるため, 後者の数値はより大きい値を有する. 別の例として, コーチングプログラムが学校 A が学校 C よりも効果的である事後確率である $\Pr(\theta_1 > \theta_3 | y)$ を, $\theta_1 > \theta_3$ である θ のシミュレーションドローの割合によって推定することができる. 結果は $\frac{141}{200} = 0.705$ となる.

要約すると, この例のベイズ解析は, 関心のある多くのパラメータについての直接的な推論を可能にするだけでなく, データに適応するのに十分柔軟な階層モデルを使用し, 超パラメータの不確かさと同様に部分プールを考慮する事後推論を提供する.

5.6 メタ解析に適応される階層モデリング

メタ解析 (Meta-analysis) はますます一般的に重要な要約のプロセスとなっていて、特定の分野における研究の知見を統合する. いくつかの並列データソースからの情報を結合する方法として、メタ解析は階層モデリングに密接に関連しています. この節では、階層的モデリングを医療におけるメタ解析に比較的簡単な応用を考える. 9.2 節の決定問題の文脈における別のメタ解析問題を考察する.

医療事例のデータは、表 5.4 の最初の 3 つの列に表示されています. これは β 阻害薬 (中枢神経系に影響を及ぼし、心筋を弛緩させる薬剤のこと) を投与するか受けないかに無作為に割り振られた心臓発作患者の 2 つの群からなる 22 の臨床試験において、心筋梗塞後の死亡率を要約するものである. 死亡率は 3%~21%であり、そのほとんどは β 阻害薬の使用による恩恵は控えめではあるが統計的に有意な利点を示している. メタ解析の目的は、研究中の取り扱いの有益な効果についての証拠の全体的な強さを示す研究の総合的な分析を提供することである. 正式なメタ解析に進む前に、どの研究が含まれているかを決定する厳格な基準を適用することが重要です. (これは 8 章で議論されているように、観測研究のためのデータ収集における無視可能性の懸念に関連する)

各研究のパラメータの定義

β 阻害薬の例では, メタ解析には, いくつかの 2×2 テーブルの形式のデータが含まれています. 臨床試験 j (メタ解析の対象となる系列中) が, それぞれで y_{0j} 人および y_{1j} 人の死亡を含む統制群 n_{0j} 人および実験群の n_{1j} 人の被験者の使用を伴う場合, 通常のサンプリングモデルは, それぞれ死亡確率 p_{0j} および p_{1j} を有する 2 つの独立した二項分布を含む. 関心のある推定量は, 確率の差 $p_{1j} - p_{0j}$, 確率またはリスク比 p_{1j}/p_{0j} , およびオッズ比 $\rho_j = \frac{p_{1j}}{1-p_{1j}} / \frac{p_{0j}}{1-p_{0j}}$ が含まれる. ある範囲の研究計画 (症例対照研究ならびに臨床試験およびコホート研究を含む) における解釈可能性を含む多くの理由, および事後分布が比較的小さな試料サイズであっても正常に近いという事実のために, オッズ比の (自然) 対数 (これを $\theta_j = \log \rho_j$ とラベル付けをする) に対する推論である.

尤度の正規近似

パラメータ θ_j に対する正規近似尤度で各実験 j の結果を要約すると, 比較的単純なベイズメタ回帰は前節の通常 of 理論の結果を用いることで可能である. これは, 点推定値と標準誤差を生成する多くの標準的な解析手法で可能です. これは, 正規平均と正規偏差に近似するものとみなすことができます. 1つのアプローチは**経験ロジット (empirical logits)**に基づくものである. 各研究 j に対して,

$$y_j = \log \left(\frac{y_{1j}}{n_{1j} - y_{1j}} \right) - \log \left(\frac{y_{0j}}{n_{0j} - y_{0j}} \right) \quad (5.23)$$

により, 近似サンプリング分散

$$\sigma_j^2 = \frac{1}{y_{1j}} + \frac{1}{n_{1j} - y_{1j}} + \frac{1}{y_{0j}} + \frac{1}{n_{0j} - y_{0j}} \quad (5.24)$$

で θ_j を推定できる.

表記法 y_j と σ_j^2 を階層正規モデルの以前の式と一致させるために使用します. 関与する標本分布の漸近正規性を改善するこれらの推定量の様々な改良がある (特に, 2×2 表の 4 つの計数のそれぞれに 0.5 などの小数を追加することが推奨される) が, 特定のサンプルサイズが適度に大きい場合, そのような詳細は私たちに関係しません.

推定された対数オッズ比 y_j およびそれらの推定標準誤差 σ_j^2 は, 表 5.4 の第 4 列および第 5 列として表示されます. 階層的ベイズ解析を使用して, 22 の研究からの情報を結合し, 各 θ_j の改善された推定値と, すべての研究に対する影響の平均および分散の推定値を得る.

メタ解析の推論の目的

メタ分析の議論は、分析の対象となる推定値について、特に主要な焦点が、結合すべき研究のいずれにおいても効果がないという帰無仮説を検証することにある場合には、ときには不正確である。焦点は意味のあるパラメータを推定することであり、この目的のためには、広い意味での研究が同等であるという大まかな前提を受け入れ、3つの可能性があるようです。第1の可能性は、すべての研究における個体を、共通の集団からの独立したサンプルとみなし、同じ結果尺度などを用いて、互いに同一の複製として研究を見ることである。第2の可能性は、研究が異なっており、いずれかの研究の結果が他の研究の結果に関する情報を提供しないということである。第3の、より一般的な可能性は、研究を交換可能であるとみなしているが、必ずしも同一または完全に無関係であることは必要としない。言い換えれば、研究の違いを許しているが、その差異が先験的に他の研究よりも1つの研究に有利な予測可能な効果を期待するものではない。この章で詳しく説明するように、この第3の可能性は、両極端の連続体を表しており、ベイズ解析の基礎となる交換可能なモデル(母集団分布を特徴付ける未知の超パラメータを持つ)です。

交換可能性は, 研究効果の結合分布の形式を規定していない. 以下では, 変化するパラメータの正規分布の都合の良い仮定を採用する. 実際には, 6章で説明した技術のいくつかを使用してこの仮定をチェックすることが重要です.

メタ分析の最初の潜在的な推定量や一般的な階層的な構造化問題は, 観察された研究と交換可能であるとみなすことができるすべての研究の全体的な平均効果を表すので, 効果の大きさの分布の平均である. 可能性のある他の推定値は, 観察された研究のいずれかにおける効果サイズと, 別の同等の (交換可能な) 観察されない研究における効果サイズである.

交換可能性が不適切な場合はどうなるか?

交換可能性を仮定するとき, より複雑なモデルの基礎を形成する重要な共変量がないと仮定し, この仮定 (おそらく誤っている) はメタ分析で広く採用されている. メタ分析の J 個の研究を区別するための他の情報 (データ (n, y) に加えて) が利用可能であり, 交換可能なモデルが不適切な場合はどうなりますか? この状況では, 観察されたデータと共変量において, 例えば 15 章のような階層的回帰モデルを使用して, モデルの枠組みを拡張して, 治療効果が共変量の関数としてどのように挙動するかを推定することができる. 本当の目的は, 一般的に, 母集団の既知の特性とリスクへの曝露に基づいて影響を予測できるように, 応答面を推定することです.

階層正規モデル

研究に特有の効果推定値の近似的な正規サンプリング分布に関連する正規母分布は, 前節の SAT コーチングの例と同じ形式の分析を可能にする. y_j は, 一般に, (5.23) から得られる j 番目の研究における効果 θ_j の点推定値を表すものとする. ただし, $j = 1, \dots, J$. 階層的正規モデルの第 1 段階では,

$$y_j | \theta_j, \sigma_j \sim N(\theta_j, \sigma_j^2)$$

を仮定する. ここで, σ_j は, (5.24) からの対応する推定標準誤差を表す. β 阻害薬の例のほぼすべての研究では, 各サンプル群のサイズが 50 人以上であるため, 各研究の 2 項分散が正確に推定されるため, 既知の分散の単純化はほとんど効果がありません. 階層の第 2 段階では, 未知の超パラメータである平均 μ および標準偏差 τ を有する交換可能な正規事前分布を再び使用する. 最後に, μ と τ には超事前分布が必要です. この問題については, 少数の研究 (たとえば 5 または 10) であっても, 結合されたデータが効果サイズの母集団分布の中心に関して比較的情報的になるため, μ に関する無情報または局所的に一樣な事前密度を仮定することは妥当である. SAT コーチングの例と同様に, 本質的に便宜上, τ の局所的に一樣な事前密度を仮定するが, 事前情報を含むように分析を変更することは容易である.

分析の結果とより簡単な手法との比較

メタ解析モデルの分析は、前節とまったく同じ方法論に従っています。第 1 に、図 5.5 と同様のプロット (ここには示されていない) は、0 に近い値は明らかに妥当であるが、0 はモードでの値よりも 25% 小さい事後密度しか持たず、 τ の周辺事後密度は比ゼロの値でピークを迎えるということを示している。ロジットスケールに関する 22 の研究の効果 θ_j の事後分位数は、表 5.4 の最後の列として表示される。

τ の事後分布は、データのサンプリング標準偏差に比べて小さい値の周りに集中している (表 5.5 の τ の事後中央値を表 5.4 の η_j の値と比較すると) 特に内部精度の低い研究 (例えば、研究 1, 6, および 18) のために、ベイズ推定値の収縮が明らかである。研究間の実質的な均質性の程度は、研究固有の推定値から互いに強みを借りているベイズ推定値に移行するときに得られる事後分散の大幅な減少にさらに反映される。 τ を固定する近似的なアプローチを用いると、完全なベイズのものに比べて小さすぎる標準偏差が得られる。

個々の影響に対してシミュレートされた事後密度のヒストグラム (示されていない) は, 全体平均の中心値から離れる歪みを示すが, 全体平均の分布はより大きい対称性を有する. 2 と 18 のような不正確な研究は, 7 と 14 のようなより正確なものより長いテールの事後分布を示す.

メタ解析では, 関心はしばしば全体的平均効果の推定値 μ に焦点を当てる. τ の事後密度に τ を与えた μ の条件付き事後平均と標準偏差のグラフ (ここには図示せず) を重ね合わせると, $E(\mu|\tau, y)$ の妥当な値の小さな範囲が約 -0.26 からちょうど -0.24 以上であるが, $sd(\mu|\tau, y)$ は, τ のもっともらしい値の範囲で 2 以上の係数で変化する. 後者の特徴は, その推定における不確実性を十分に考慮するために, τ を平均化することの重要性を示している. 事実, 条件付き事後標準偏差 $sd(\mu|\tau, y)$ は, $\tau = 0.13$ で 0.060 の値を有するが, τ について事後分布にわたって平均すると, $sd(\mu|y) = 0.071$ の値が見出される.

表 5.5 は, 仮想的な将来の研究において, 超パラメータ μ および τ および予測される効果, $\tilde{\theta}_j$ の事後推論の概要を示す. オッズ比スケール (換言すれば, 累乗) に変換すると, μ の約 95% の最高事後密度区間は $[-0.37, -0.11]$, または $[0.69, 0.90]$ である.

対照的に, 完全なプールから生じる 95%の事後区間, すなわち $\tau = 0$ であると仮定すると, かなり狭い $[0.70, 0.85]$. これらのデータの最初の公表された議論では, 後者は「非常に狭い不確かさの範囲」のように見える. 階層的ベイズ分析は, これは, すべての研究が同一であると主張する効果を有する不適切なモデルの使用によるものであることを示唆している.

数学的に言えば, 完全なプールは, パラメータ τ が正確にゼロであるという仮定を作るが, データ供給は, τ がゼロに近いかもしれないが, おそらくは 0.3 まで高くなる可能性があるという証拠を提供する. 関連する懸念事項は, 一般的に使用される分析が全体的な平均効果の推論に過度の重点を置く傾向があることである. 研究が行われていない特定の集団 (または以前に研究された集団であるがわずかに改変された集団) における可能性のある治療効果についての不確実性は, 新しい研究効果の推論によってより合理的に表され, 全体的な平均よりむしろ研究が行われている. この場合, 表 5.5 の「予測される効果」の行に示されているように, 不確実性はさらに大きくなる. 個々の患者の不確実性には, さらに別の変動要素が含まれる. 特に, β 阻害薬のデータでは, 新しい研究での真の効果, θ_j が正 (その研究における死の確率を増加させる治療に対応する) である事後確率がわずかに 10%を超えている.

まとめ 5.6

- ・ **メタ解析**

…ますます一般的に重要な要約のプロセスとなってきた、特定の分野における研究の知見を統合する. いくつかの並列データソースからの情報を結合する方法として、階層モデリングに密接に関連している.

- ・ **経験ロジット**

$$y_j = \log \left(\frac{y_{1j}}{n_{1j} - y_{1j}} \right) - \log \left(\frac{y_{0j}}{n_{0j} - y_{0j}} \right)$$

というような変換.

- ・ **メタ解析の目的**

…メタ分析の議論は、分析の対象となる推定値について、特に主要な焦点が、結合すべき研究のいずれにおいても効果がないという帰無仮説を検証することにある場合には、ときには不正確である. 焦点は意味のあるパラメータを推定することであり、この目的のためには、広い意味での研究が同等であるという大まかな前提を受け入れるべき.

・メタ解析の議論

1. すべての研究における個々を, 共通の母集団からの独立なサンプルとみなし, 同じ結果尺度などを用いて, 同一の複製とみなす場合.
2. 研究が異なっていて, ある研究結果がほかの研究結果に情報を与えない場合.
3. 研究を交換可能とみなすが, 必ずしも同一, 無関係であることを必要としない場合.

5.7 階層分散パラメータに対する弱情報事前分布

上記の分析における重要な要素は、尺度パラメータ τ の事前分布である。一様分布を使用しましたが、ベイズ文献ではさまざまな無情報事前分布が提案されています。無情報事前分布の選択は、特にグループ J の数が少ないか、グループレベルの変動 τ が小さい問題について、推論に大きな影響を及ぼすことが分かる。ここでは、通常モデルの文脈でオプションを説明しますが、より一般的には、グループレベルの分散の推論に適用されます。

事前分布の選択に関係する概念

不適切な事前分布の極限

不適切な事前密度が、必ずしもそうではないが、適切な事後分布につながる可能性がある。混乱を避けるために、適切な分布の特定の極限として不適切な分布を定義すると便利です。グループレベルの分散パラメータでは、一般的に考えられる2つの不適切な密度は、 τ についての $A \rightarrow \infty$ とした $U(0, A)$ と、 τ^2 についての $\epsilon \rightarrow 0$ とした逆-gamma(ϵ, ϵ) である。

ここで見るように、 $U(0, A)$ モデルは、グループ J の数が少なくとも3である限り、 $A \rightarrow \infty$ として極限の適切な事後密度をもたらす。そして、有限ではあるが十分に大きい A については、推論は A の選択に敏感ではない。

対照的に、逆-gamma(ϵ, ϵ) モデルは、適切な極限で事後分布を持たない。その結果、事後推論は敏感であり、単に 0.001 などの低い値に簡単に設定することはできません。

校正

事後推定は、バイアスの古典概念に対するベイズアナロジーである事後平均の**校正 (calibration)** の概念を用いて評価することができる. 任意のパラメータ θ に対して, $\hat{\theta} = E(\theta|y)$ とすると, $E(\theta|\hat{\theta}) - \hat{\theta}$ として事後平均の**誤校正 (miscalibration)** を定義することができる. 事前分布が真である場合, すなわち, $p(\theta)$ から最初に θ を抽出して, データを構築した場合, $p(y|\theta)$ から y を引くと, 事後平均が自動的に校正されます. つまり, $\hat{\theta}$ のすべての値に対して, 誤校正は 0 になります.

修正するには, 古典的バイアス解析では, 真の θ を条件とし, データに基づく推定値 $\hat{\theta}$ の分布を見る. ベイズ校正分析では, データ y (したがって推定値 $\hat{\theta}$) を条件付けし, これらのデータを生成した可能性のあるパラメータ θ の分布を見る.

不適切なモデルを考えると、 θ を非正規化密度から抽出することは不可能であるため、理論を拡張しなければならない。この文脈で較正を評価するためには、ベイズ推論で使われる**推論的事前分布 (inferential prior distribution)** とともに θ が抽出される真の事前分布を仮定することが必要である。

8つの学校の階層モデルでは、 $A \rightarrow \infty$ の範囲 $(0, A)$ の一様事前密度の極限として、 τ の不適切な一様密度を考えることができます。 A の任意の有限値に対して、不適切な一様な密度は、正の較正の推論につながります。つまり、(平均して) τ を過大評価します。

2段階でこの較正を実証する。まず、 τ に対する真かつ事前の事前分布の両方が $(0, A)$ 上で一様であると仮定する。その後、較正はほとんどゼロになります。今、 $U(0, A)$ で真の事前分布を維持し、推論事前分布を $U(0, \infty)$ にする。これは真の θ を変更することなく、すべてのデータ y について必然的に θ を増加させる (今度は範囲 $[A, \infty)$ の θ の値を平均している) ので、較正の平均値は正になる。

階層的分散パラメータの無情報, 弱情報事前分布のクラス

一般的な結果

本質的に暫定的な前提として, 無情報事前分布や弱情報事前分布を見てみましょう. モデルがフィットされた後でその分布を見て, それが意味をなすかどうかを見なければなりません. 事後分布が意味をなさない場合, これは, モデルに含まれていない追加の事前知識が利用可能であり, これは使用された事前分布の前提と矛盾することを意味する. この外部の知識とより一貫して事前分布に戻って変更することが適切です.

一様事前分布

まず, 分布が定義されているスケールについて明示しなければならないことを認識しながら, 一様事前分布を考慮する. 分散パラメータをモデル化するための様々な選択肢が提案されている. $\log \tau$ の一様事前分布は, 正でなければならないパラメータの対数で自然に働くように見えるが, 不適切な事後分布をもたらす. 代替案は, コンパクト集合 (例えば, A のある大きな値に対して $[-A, A]$ の範囲内) で事前分布を定義することであろうが, 事後分布は事前サポートの下境界 $-A$ に大きく依存するだろう.

この問題は, ((5.16) の θ と μ に積分した後の) 周辺尤度 $p(y|\tau)$ が $\tau \rightarrow 0$ として有限の非ゼロ値に近づくために生じる. したがって, $\log \tau$ の事前密度が一様であれば, 事後確率は極限 $\log \tau \rightarrow -\infty$ に積分される無限質量を持つことになる. 別の言い方をすると, 階層モデルでは, データはグループレベルの分散 0 を排除できないため, 事前分布はこの領域に無限大を置くことはできません.

もう 1 つの選択肢は, τ 上の一様事前分布であり, これは $\tau = 0$ 近くで有限積分を有し, 従って上記の問題を回避する. 一般に, 応用の仕事 (5.5 節に示されている) でこの無情報密度を使用した, $\tau \rightarrow \infty$ の無限大の事前質量を持つ正の値に向かってやや不適当な較正を行っている. $J = 1$ または 2 のグループでは, 事実上不適切な事後密度となり, 本質的に $\tau = \infty$ となり, プーリングを行わない. ある意味では, 1 つまたは 2 つのグループからのデータでどれくらいのプーリングを行うべきかを決定することは, データだけでは困難と思われるので, これは合理的な動作です. しかし, ベイズの観点からは, データが問題になることなく, 事前に決定が下されるのは厄介です. さらに, 4 または 5 などの小さい J については, 事後分布の重い裾の右端が τ の過大評価につながり, その結果, 個々の θ_j を推定するのに最適ではないプールが生じることを心配する.

弱情報条件付き共役事前分布の極限として,これらの不適切な一様事前密度を解釈することができる. $\log \tau$ 上の一様事前分布は, 0 自由度の逆- χ^2 密度の形をした $p(\tau) \propto \tau^{-1}$ または $p(\tau^2) \propto \tau^{-2}$ と等価で, 適切な逆- γ 事前分布の極限としてとれる.

τ の一様密度は, -1 自由度の逆- χ^2 密度である $p(\tau^2) \propto \tau^{-1}$ と等価である. この密度は適切な逆- χ^2 密度の極限として容易には見られませんが (正の自由度を持たなければならないため), τ 上の半分 t 族の限界として解釈できます. ここで, スケールは ∞ に近づく. (そして, ν は任意の値.)

ベイズの文献で時々提案されている別の無情報事前分布は, τ^2 については一様である. 上記のように高い値に向かうように見えるが, そうであるように, また適切な事後分布に対して, 適切であるためには $J \geq 4$ グループである必要がある.

逆-gamma(ϵ, ϵ) 事前分布

モデル (5.21) のパラメータ τ は, その周辺尤度がすべての J グループからのデータに複雑な方法で依存するため, 単純な共役事前分布族を持たない. しかし, 逆-ガンマ分布族は, モデル内の他のパラメータが与えられた条件付き共役です. すなわち, τ^2 が逆-ガンマ事前分布を有する場合, 条件付き事後分布 $p(\tau^2 | \theta, \mu, \gamma)$ も逆-ガンマである. τ^2 の逆-gamma(α, β) モデルは, スケール $s^2 = \frac{\beta}{\alpha}$ と自由度 $\nu = 2\alpha$ の逆- χ^2 分布として表すこともできます. 逆- χ^2 パラメータ化は, 適切な事前分布の様々な選択の根底にある情報の理解に役立ちます. 逆-gamma(ϵ, ϵ) 事前分布は, 条件付き結合体系内の非情報性の試みであり, ϵ は 1 または 0.01 または 0.001 などの低い値に設定されます. この事前分布の難しさは, $\epsilon \rightarrow 0$ の極限では不適切な事後密度を生じるため, ϵ は妥当な値に設定されなければならない. 残念ながら, τ の値が低いデータセットでは, 推論はこのモデルでは非常に敏感になり, 図 5.9 に示すように, 事前分布はほとんど情報に見えません.

半-Cauchy 事前分布

また、標準ケースとコーシーがエッジケースとして含まれている代替クラスとして、分布の t 族 (実際には、スケールパラメータ τ が正であるため、半- t 族) を考慮する。最初に、この問題の t モデルを考慮した。なぜなら、これは、再パラメータ化を使用して τ に対する条件付き共役事前分布として表すことができるからである。

しかしここでの目的のためには、半-Cauchy が便利な弱情報分布族であることを認識すれば十分です。分布はゼロに広いピークを有し、単一のスケールパラメータを有し、それをいくらかの大きな値に設定できることを示すために A とラベル付けする。極限 $A \rightarrow \infty$ では、これは τ 上の一様事前密度となる。 A は、たとえ尾部であっても (例えば、正規分布の半分とは異なり) 緩やかな傾きを有し、その領域で尤度が強い場合にはデータを支配することができるため、弱情報と考える事前分布を表す。少数のグループから推測される分散パラメータのための半-Cauchy モデルを考察する (推論は弱情報事前分布の選択に敏感である)。

8つの学校の例への応用

5.5 節の 8 つの学校の例で, いくつかの提案された無情報事前密度の特性を示す. ここで, パラメータ $\theta_1, \dots, \theta_8$ は 8 つの異なる学校におけるコーチングプログラムの相対的効果を表し, τ はこれらの効果の学校間標準偏差を表す. 効果は, 試験での得点として測定され, 平均は約 500 で 200 から 800 の点数がつく. したがって, 可能な最大の影響範囲は約 300 点であり, τ の現実的な上限は 100 点である.

8つの学校の問題に対する無情報事前分布

図 5.9 は, 無情報的であることが意図されている事前分布の 3 つの異なる選択に起因する 8 校モデルの事後分布を示す. 左端のヒストグラムは, 一様事前密度を有するモデルの τ の事後推論を示す. データは, グループ J の数がわずか 8 であることを考慮すると, より大きい値の可能性を反映して, $\tau = 20$ 以下の値域のサポートを示し, その後はわずかなテールがあります. すなわち, 右の尾部に有限質量の適切な後部密度を確保するために必要な $J = 3$ をあまり超えるものではない.

対照的に、図 5.9 の中央のヒストグラムは、 τ^2 の逆-gamma(1, 1) 事前分布を伴う結果を示す。この新しい事前分布は異なる推論を導く。特に、 τ の事後平均および事後中央値はより低く、 θ_j の収縮は、 τ 上の一様事前分布を有する以前に当てはめられたモデルよりも大きい。これを理解するには、事後分布がかなりの範囲内の事前分布をグラフ化することが役立ちます。グラフは、事前分布が $[0.5, 5]$ の範囲に集中していることを示しています。これは、尤度がこれに比べてフラットに近い狭い領域です (τ の事後シミュレーションの分布が事前分布、 $p(\tau)$ が非常に近くなるため分かる)。比較すると、左側のグラフでは、 τ の一様事前分布は、この問題の場合、事後推論を制約するようには見えないという意味で、無情報に近いように見える。

最後に、図 5.9 の右端のヒストグラムは、 τ^2 の逆-gamma(0.001, 0.001) 事前分布を持つ対応する結果を示します。この事前分布は、 τ の周辺尤度はゼロ付近で高いままであるのに、ゼロ付近でさらに急激にピークに達し、さらに後方推論を歪ませる。

この例では, $\log \tau$ の一様事前密度は考慮しない. これは, 図 5.9 の一番右のグラフのように, $\tau = 0$ でのスパイクを伴う不適切な事後密度をもたらす. また, τ^2 上の一様事前密度も考慮しない. また, 図 5.9 の一番左のグラフに類似した事後分布を生成する, わずかに高い右の尾部を有する.

この例は, 最も単純なアプローチ, すなわち τ の一様事前密度がうまくいくように見える喜ばしいケースです. 付録 C に詳述されているように, このモデルは R または Stan で直接プログラムすることも簡単です.

図 5.9 のヒストグラムと密度プロットの外観は, τ のスケールでプロットする選択肢によって決定的に影響される. 代わりに $\log \tau$ のスケールでプロットした場合, 逆-gamma(0.001, 0.001) の事前密度が一番平坦であるように見える. しかし, 逆-gamma(ϵ , ϵ) 事前分布は, 結果的な事後分布が ϵ の選択に対して非常に敏感なままであるので, この問題のためにはまったく無情報的ではない. 階層的モデル尤度は $\log \tau$ を極限 $\log \tau \rightarrow -\infty$ に制約しないので, ログスケール上では無情報事前分布は機能しない.

3つの学校の問題に対する弱情報事前分布

一様事前分布は8つの学校の分析ではうまくいくように見えるが、グループJの数がはるかに少ないと問題が発生する。その場合、データはグループレベルの分散に関する情報をほとんど提供せず、適切ではないか適切であるが非現実的に広い事後分布を導く。最初の3つの学校のデータだけを使って、8つの学校の例を再分析することによって証明します。

図5.10は、2つの異なる事前分布に基づく τ の推論を示しています。最初、 $J=8$ (図5.9を参照)とうまく機能するデフォルトの一様分布を続ける。残念なことに、図5.10の左のヒストグラムが示すように、3つの学校のデータセットの結果的な事後分布は、非常に長い右の尾を持ち、妥当性が高すぎる τ の値を含んでいます。この重い尾は、Jが非常に低い(Jがそれよりも低く、右の尾が無限大の場合)予想され、事後分布としてこれを使用すると、学校効果 θ_j の推定値をアンダープールする効果があります。

図 5.10 の右のヒストグラムは、スケールパラメータ $A = 25$ の半-Cauchy 事前分布から得られる τ の事後推論を示しています (この教育的なテストの例の文脈で、基礎となる θ_j の標準偏差に対して予想される値よりも少し高い値を選んだので、モデルは τ を弱く制限します)。グラフ上の線が示すように、この事前分布は、 $\tau < 50$ の妥当な範囲にわたって高く、この点を超えると徐々に下回っている。この事前分布は、この例ではうまくいくように見えますが、 τ の低さでの尤度の限界を反映していますが、非現実的な上部テールの多くを除去しています。

この半-Cauchy 事前分布は、8 校の問題でもうまくいくでしょう。そのデフォルトの一樣事前分布の前提が妥当な結果をもたらしたため、これは不必要でした。わずか 3 つの学校では、 τ に関する実際の事前知識を表現することではなく、事後分布を制約するような弱情報提供的な事前分布を、データによって許容される範囲で使用するという問題に直面したのである。

まとめ 5.7

- ・ 不適切な事前分布の極限
- …不適切な事前密度が、適切な事後分布につながる可能性がある。
- ・ 階層モデルのグループレベルの分散パラメータ τ
- …一般的に考えられる不適切な密度は
 1. τ についての $A \rightarrow \infty$ とした $U(0, A)$.
グループ数 ≥ 3 ならば, $A \rightarrow \infty$ としたとき適切な事後分布を持つ.
 2. τ^2 についての $\epsilon \rightarrow 0$ とした逆-gamma($-\epsilon, \epsilon$).
適切な事後分布を持たない.
- ・ 逆-gamma($-\epsilon, \epsilon$) 事前分布
- …無情報事前分布として、スケールパラメータの事前分布に用いられる。
- ・ 半-Cauchy 事前分布
- …t 分布族の特例. t 分布族は τ の条件付き共役事前分布となる.

- ・ 校正

…偏差の古典概念に対するベイズアナロジーである事後推論は事後平均の校正の概念を用いることで評価できる.

- ・ 誤校正

…任意のパラメータ θ に対して, $\hat{\theta} = E(\theta|y)$ とすると, $E(\theta|\hat{\theta}) - \hat{\theta}$ として事後平均の誤校正 (miscalibration) を定義する.