

Bayesian Data Analysis : Chapter3

Introduction to multiparameter 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

2017/4/2~3

Table Contents

1 3.1 Averaging over ‘ nuisance parameters ’

2 3.2 Normal data with a noninformative prior distribution

3 3.3 Normal data with a conjugate prior distribution

4 3.4 Multinomial model for categorical data

5 3.5 Multivariate normal model with known variance

6 3.6 Multivariate normal with unknown mean and variance

7 3.7 Example: analysis of a bioassay experiment

8 3.8 Summary of elementary modeling and computation

Key Words



事実上, 統計の実際の問題には, 複数の未知のまたは観察できない量が含まれています. ベイズ手法の単純な概念的枠組みが, 他の推論の方法よりも主要な利点を明らかにしているような問題に対処している. 問題にはいくつかの関心のあるパラメータが含まれることがあります, 結論はしばしば一度に 1 つ, またはほんのわずかのパラメータについて行われます. この場合, ベイズ分析の最終目的は, 関心のある特定のパラメータの周辺事後分布を得ることである. 原則として, この目標を達成するためのルートは明確です. まず, すべての未知数の結合事後分布を必要とし, 次にこの分布を, 望ましい周辺分布を得るために直ちに興味のない未知数に積分します. 同様に, シミュレーションを使用して, 結合事後分布からサンプルを抽出し, 対象のパラメータを見て, 他の未知数の値を無視します.

多くの問題では, 現実的なモデルを構築するためには必要ですが, 未知のパラメータの多くについて推論を行うことには興味がありません. このようなパラメータは, しばしば**迷惑パラメータ (nuisance parameter)**と呼ばれます. 典型的な例は, 測定問題におけるランダム誤差のスケールです. 本章では, 迷惑パラメータの一般的な扱いから始め, 3.2 節で未知の平均と分散をもつ正規分布をカバーする. 3.4, 3.5 節は, 多項式と多変量の正規分布の推論を示しています. これは, 離散と連続多変量データそれぞれの最も単純なモデルです. この章では, グリッド上の事後密度の数値計算を使用して, 非共役ロジスティック回帰モデルの解析を行います.

3.1 「迷惑パラメータ」の平均化

結合事後分布と周辺事後分布の考え方を数学的に表現するために, θ はベクトルの2つの部分, $\theta = (\theta_1, \theta_2)$ を持ち, さらに (少なくとも現時点では) θ_1 の推論であるため, θ_2 は迷惑パラメータとみなされる可能性があります. 例えば, 簡単な例では,

$$y|\mu, \sigma^2 \sim N(\mu, \sigma^2)$$

ここで, $\mu (= \theta_1), \sigma^2 (= \theta_2)$ は未知で, 一般に μ を中央にする関心量である. 観測データを与えられたもとでの興味のあるパラメータの条件付き分布を探す. この場合では, $p(\theta_1|y)$. これは, 結合事後分布

$$p(\theta_1, \theta_2|y) \propto p(y|\theta_1, \theta_2)p(\theta_1, \theta_2)$$

から, θ_2 での平均化

$$p(\theta_1|y) = \int p(\theta_1, \theta_2|y) d\theta_2$$

導かれる.

代わりに、結合分布は因数分解して、

$$p(\theta_1|y) = \int p(\theta_1|\theta_2, y)p(\theta_2|y)d\theta_2 \quad (3.1)$$

を得ることができる。これは、関心量の事後分布 $p(\theta_1|y)$ が迷惑パラメータ θ_2 が与えられた条件付き事後分布の混合である。

関心のある事後分布 $p(\theta_1|y)$ が迷惑パラメータ θ_2 が与えられた条件付き事後分布の混合である。ここで、 $p(\theta_2|y)$ は θ_2 の異なる取りうる値に対する重み付き関数である。重みは、 θ_2 の事後密度に依存し、したがって、データおよび事前モデルからの証拠の組み合わせに依存する。迷惑パラメータ θ_2 に対する平均化は一般的に解釈することができる。例えば、 θ_2 は、異なる可能性のあるサブモデルを表す離散成分を含むことができる。

積分 (3.1) を明示的に評価することはめったにありませんが、マルチパラメータモデルを使用して構築および計算するための重要な実践的戦略を示唆しています。事後分布は、周辺および事後シミュレーションによって計算することができます。まず、限界後置分布から最初に θ_2 を抽出した後、 θ_2 の抽出された値の下で条件付き事後分布から θ_1 を抽出する。このようにして、(3.1) で実施された統合は間接的に行われる。この形式の分析の標準的な例は、未知の平均と分散を持つ正規モデルによって提供されます。

3.2 無情報事前分布をもつ正規データ

サンプルから集団の平均を推定するプロトタイプの例として, 単変量正規分布 $N(\mu, \sigma^2)$ からの独立した観測値のベクトル y を考える. 多変量正規分布への一般化は 3.5 節にある. 我々は, 非公式な事前分布のもとでモデルを解析することから始め, これは解説の目的のための便利な仮定に過ぎず, 情報事前分布に容易に拡張されることを理解する.

無情報事前分布

2 章で, 位置パラメータとスケールパラメータの事前独立性を仮定し, μ, σ に対する現実的であいまいな密度は $(\mu, \log \sigma)$ 上で一様であるか, 同じであるが,

$$p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$$

であるということを見た.

結合分布 $p(\mu, \sigma^2|y)$

この従来の不適切な事前分布の下で結合事後分布は尤度関数に $1/\sigma^2$ をかけたものに比例する.

$$\begin{aligned} p(\mu, \sigma^2|y) &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right) \\ &= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2\right]\right) \\ &= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} [(n-1)\sigma^2 + n(\bar{y} - \mu)^2]\right) \end{aligned} \quad (3.2)$$

ここで,

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

は y_i のサンプル分散である. 十分統計量は \bar{y}, s^2 である.

条件付き事後分布 $p(\mu|\sigma^2, y)$

まず, 条件付き事後密度 $p(\mu|\sigma^2, y)$ を計算し, 次に境界後密度 $p(\sigma^2|y)$ を考察する. μ の事後分布を求めるために, σ^2 を与えて, 既知の分散と一様事前分布をもつ正規分布の平均について 2.5 節で得られた結果を単純に使用します.

$$\mu|\sigma^2, y \sim N(\bar{y}, \sigma^2/n) \quad (3.3)$$

周辺事後分布 $p(\sigma^2|y)$

$p(\sigma^2|y)$ を決定するために, 結合分布 83.2) を μ で平均化する必要がある.

$$p(\sigma^2|y) \propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y} - \mu)^2]\right) d\mu$$

この式を μ で積分するには, 積分 $\exp(-\frac{1}{2\sigma^2}n(\bar{y} - \mu)^2)$ を評価する必要があります. これは簡単な正規積分です. 従って,

$$\begin{aligned} p(\sigma^2|y) &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \sqrt{2\pi\sigma^2/n} \\ &\propto (\sigma^2)^{-(n+1)/2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \end{aligned} \quad (3.4)$$

これはスケーリングされた逆 χ^2 密度である.

$$\sigma^2|y \sim \text{Inv } \chi^2(n-1, s^2) \quad (3.5)$$

結合事後密度 (3.2) を条件事後分布と周辺事後分布の積として因数分解した. $p(\mu, \sigma^2 | y) = p(\mu | \sigma^2, y) p(\sigma^2 | y)$.

σ^2 に対する周辺事後分布はサンプリング定理の結果のアナロジーへの顕著な類似性を有する. σ^2 (と μ) の条件付きの, 近似的にスケールされた十分統計量, $\frac{(n-1)s^2}{\sigma^2}$ の分布は χ_{n-1}^2 分布である. 2.8 節でスケールパラメータの参照事前分布の偏差を考えるが, この結果は驚くべきことではない.

結合事後分布からのサンプリング

結合事後分布からサンプルを抽出するのは簡単である. まず (3.5) から σ^2 を抽出し, 次に (3.3) から μ を抽出する. また, 事後分布の解析結果は, 閉じた形で解くのに十分単純な少数多パラメータ問題の 1 つであるため, 解析結果を導出する.

μ の周辺事後分布の解析形式

母平均 μ は典型的に興味のある推定量であり、ベイズ分析の目的は μ の周辺事後分布である。これは、結合事後分布から σ^2 を積分することによって得られる。式 (3.1) は、 μ の事後分布が、分散 σ^2 のためのスケールされた逆 χ^2 分布上に混合された正規分布の混合物とみなされ得ることを示す。 σ^2 の上に結合事後密度を積分することによって、 μ についての周辺事後密度を導出することができる。

$$p(\mu|y) = \int_0^{\infty} p(\mu, \sigma^2|y) d\sigma^2$$

この積分は置換

$$z = \frac{A}{2\sigma^2}, \text{ ただし } A = (n-1)s^2 + n(\mu - \bar{y})^2$$

を使い、結果が非正規化ガンマ積分であると分かるので評価できる。

$$\begin{aligned} p(\mu|y) &\propto A^{-n/2} \int_0^{\infty} z^{(n-2)/2} \exp(-z) dz \\ &\propto [(n-1)s^2 + n(\mu - \bar{y})^2]^{-n/2} \propto \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2} \right]^{-n/2} \end{aligned}$$

これは $t_{n-1}(\bar{y}, s^2/n)$ 密度である. (付録 A を参照).

別の言い方をすると, $(\mu, \log \sigma^2)$ 上の非情報事前分布の下で, μ の事後分布が

$$\left. \frac{\mu - \bar{y}}{s/\sqrt{n}} \right| y \sim t_{n-1}$$

の形を有することを示した. ここで, t_{n-1} は, $n-1$ 自由度をの標準 t 密度 (位置 0, スケール 1) を示す. この周辺事後分布は, サンプル理論との別の興味深い比較を提供する. サンプル分布 $p(y|\mu, \sigma^2)$ の下で, 次の関係が成り立つ.

$$\left. \frac{\bar{y} - \mu}{s/\sqrt{n}} \right| \mu, \sigma^2 \sim t_{n-1}$$

ピボット量 $(y - \mu)/(s/\sqrt{n})$ のサンプル分布は, 迷惑パラメータ σ^2 に依存せず, その事後分布はデータに依存しない. 一般に, 推定量の中心的な量は, データと推定値の重要な関数であり, そのサンプル分布はすべてのパラメータとデータとは無関係です.

未来の観測値に対する事後予測分布

未来の観測値 \tilde{y} に対する事後予測分布は

$p(\tilde{y}|y) = \int \int p(\tilde{y}|\mu, \sigma^2, y)p(\mu, \sigma^2|y)d\mu d\sigma^2$ の混合として書ける. 積分の2つのうち1つ目は (μ, σ^2) が与えられた未来の観測値に対する正規分布で, y に全く依存しない. 事後予測分布から抽出するために, まず結合事後分布から μ, σ^2 を抽出し $\tilde{y} \sim N(\mu, \sigma^2)$ をシミュレートとする.

実際, \tilde{y} の事後予測分布は位置 \bar{y} , スケール $(1 + \frac{1}{n})^{1/2}s$, 自由度 $n-1$ に対する t 分布である. 解析形式は μ の事後分布の微分と同じ技術を用いることで得られる. 特に, 分布はパラメータ μ, σ^2 をその結合事後分布に従って積分することで得られる. 因数分解

$p(\tilde{y}|\sigma^2, y) = \int p(\tilde{y}|\mu, \sigma^2, y)p(\mu|\sigma^2, y)d\mu$ が $\mu|\sigma^2, y$ の分布と同様に, 変化したスケールファクタまで同じである $p(\tilde{y}|\sigma^2, y) = N(\tilde{y}|\bar{y}, (1 + \frac{1}{n})\sigma^2)$ を導くということに気付くことで, 簡単に結果を特定できる.

例. 光の速度の推定

Simon Newcomb は 1882 年に光の速度を測定する実験を行った.

Newcomb は, 光が 7442 メートルの距離を移動するのに必要な時間量を測定しました. Newcomb の 66 測定のヒストグラムを図 3.1 に示します. 異常に低い測定値が 2 つあり, おおよそ対称的に分布している測定値の集まりがあります. すべての 66 回の測定値が平均 μ と分散 σ^2 を持つ正規分布から独立していると仮定して, 正規モデルを (不適切に) 適用します. 主な実質的な目標は μ の事後推論である. 外側の測定値は通常モデルに適合しません. これらのデータに対する適合性の欠如を測定するためのベイズ法を 6.3 節で検討する. 66 回の測定の平均は $\bar{y} = 26.2$ であり, サンプルの標準偏差は $s = 10.8$ である. 非線形事前分布

$p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$ を仮定すると, μ の 95% 中央事後区間は, μ の $\bar{y} \pm 1.997s/\sqrt{66} = [23.6, 28.8]$ として μ の t_{65} 周辺事後分布から得られる.

事後区間は、シミュレーションによっても得ることができる。(3.5) と (3.3) で与えられる事後分布の因数分解の特徴づけに従って、まず $\sigma^2 \sim \text{Inv } \chi^2(65, s^2)$ のランダム値を χ^2_5 分布から無作為抽出した $65s^2$ として描く (付録 A を参照). そして、この σ^2 の値が与えられた、条件付事後分布 $N(26.2, \sigma^2/66)$ から μ を抽出する. (μ, σ^2) の 1000 個のシミュレートされた値に基づいて、 μ の事後中央値は 26.2 であり、解析的に計算された区間に近い μ に対する 95% 中央後区間が $[23.6, 28.9]$ であると推定される.

ちなみに、現在許容されている光速の値に基づいて、Newcomb の実験における μ の真の値は 33.0 であり、これは 95% の間隔をはずれている. これは、事後推論がデータを生成したモデルや実験と同じくらい良いことを強調します.

3.3 共役事前分布をもつ正規データ

共役事前分布族

より一般的なモデルへの最初のステップは、ちょうど考えられた無情報事前分布の代わりに、2パラメータ単変量正規サンプリグモデルに対する共役事前分布を仮定することである。(3.2) でみられる尤度の式とその後の議論は共役事前密度が積の形 $p(\sigma^2)p(\mu|\sigma^2)$ を持つことたなければならぬ事を示す。ここで、 σ^2 の周辺分布はスケーリングされた逆 χ^2 分布で、 σ^2 が与えられた μ の条件付き分布は (μ の周辺分布が t 分布となるように) 正規分布である。便利なパラメータ化は次の使用で与えられる。

$$\mu|\sigma^2 \sim N(\mu_0, \sigma^2/\kappa_0), \sigma^2 \sim \text{Inv } \chi^2(\nu_0, \sigma_0^2)$$

これは,

$$p(\mu, \sigma^2) \propto \sigma^{-1} (\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2} [\nu_0 \sigma^2 + \kappa_0 (\mu_0 - \mu)^2]\right) \quad (3.6)$$

これを $N\text{-Inv-}\chi^2(\mu_0, \sigma_0^2/\kappa_0; \nu_0, \sigma_0^2)$ の密度とする。その4つのパラメータは、 μ の位置およびスケール、ならびに σ^2 の自由度およびスケールとしてそれぞれ特定することができる。

$\mu|\sigma^2$ の条件付き分布における σ^2 の出現は μ と σ^2 が必然的にそれらの結合共役前の密度に依存することを意味する. 例えば, σ^2 が大きい場合, μ には分散が大きい事前分布が誘導される. 共役事前分布が主に便宜のために使用されることを考慮すると, この依存性は注目に値する. しかし, 反映すると, 平均の事前分散が観測 y のサンプリング分散である σ^2 に結びついていることがよくあります. このようにして, μ に関する事前の信念は, y の測定の尺度によって較正され, この尺度の κ_0 の事前測定と同等である.

結合分布 $p(\mu, \sigma^2 | y)$

事前密度 (3.6) に正規尤度をかければ, 事後分布は

$$\begin{aligned}
 p(\mu, \sigma^2 | y) &\propto \sigma^{-1} (\sigma^2)^{-(v_0/2+1)} \exp \left(-\frac{1}{2\sigma^2} [v_0 \sigma_0^2 + \kappa_0 (\mu - \mu_0)^2] \right) \\
 &\quad \times (\sigma^2)^{-n/2} \exp \left(-\frac{1}{2\sigma^2} [(n-1)s^2 + n(\bar{y} - \mu)^2] \right) \\
 &= \text{N-Inv-}\chi^2(\mu_n, \sigma_n^2 / \kappa_n; \nu_n, \sigma_n^2)
 \end{aligned} \tag{3.7}$$

ここで, いくつか計算すれば (演習 3.9 を参照),

$$\begin{aligned}
 \mu_n &= \frac{\kappa_0}{\kappa_0 + n} \mu_0 + \frac{n}{\kappa_0 + n} \bar{y}, \quad \kappa_n = \kappa_0 + n, \quad \nu_n = \nu_0 + n, \\
 \nu_n \sigma_n^2 &= \nu_0 \sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n} (\bar{y} - \mu_0)^2
 \end{aligned}$$

事後分布のパラメータは, 事前情報とデータに含まれる情報とを組み合わせる. 例えば μ_n は, 事前平均とサンプル平均の加重平均であり, 重みは 2 つの情報の相対精度によって決定されます. 事後自由度 ν_n は, 事前自由度に標本サイズを加えたものです. 事後二乗和 $\nu_n \sigma_n^2$ は, 事前二乗和, 標本二乗和, および標本平均と事前平均との差によって伝えられる追加の不確定性を組み合わせる.

条件付き事後分布 $p(\mu|\sigma^2, y)$

σ^2 が与えられた μ の条件付き事後分布は σ^2 を定数とし, 結合事後密度 (3.7) に比例する.

$$\begin{aligned}\mu|\sigma^2|y &\sim \text{N}(\mu_n, \sigma^2/\kappa_n) \\ &= \text{N}\left(\frac{\frac{\kappa_0}{\sigma^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}\right)\end{aligned}\tag{3.8}$$

これは σ^2 を固定した μ の 2.5 節での分析と必ず一致する.

周辺事後分布 $p(\sigma^2|y)$

σ^2 の周辺事後分布は, (3.7) より, スケーリングされた逆- χ^2 である.

$$\sigma^2|y \sim \text{Inv-}\chi^2(\nu_n, \sigma_n^2) \quad (3.9)$$

結合事後分布からのサンプリング

結合事後分布からサンプルするために, 前の節でみたように, まず周辺事後分布 (3.9) から σ^2 を抽出し, 正規条件付き事後分布 (3.8) から, σ^2 のシミュレートされた値を用いて, μ を抽出する.

μ の周辺事後分布の解析形式

前節で用いたのと同様の方法で, σ^2 に関する結合事後密度の積分は, μ の周辺事後密度が

$$\begin{aligned} p(\mu|y) &\propto \left(1 + \frac{\kappa_n(\mu - \mu_n)^2}{\nu_n \sigma_n^2} \right)^{-(\nu_n+1)/2} \\ &= t_{\nu_n}(\mu|\mu_n, \sigma_n/\kappa_n) \end{aligned}$$

であることを示している.

3.4 カテゴリカルデータに対する多項モデル

2章で強調された二項分布は2つ以上のありうる結果を許すように一般化できる. 多項サンプリング分布はそれぞれの観測値が k 個のありうる結果の1つであるような多項データを記述するのに用いられる. y がそれぞれの結果の観測値の数のカウントのベクトルならば,

$$p(y|\theta) \propto \prod_{j=1}^k \theta_j^{y_j}$$

ここで, 確率の和 $\sum_{j=1}^k \theta_j$ は1である. 分布は典型的に観測数 $\sum_{j=1}^k y_j = n$ を暗黙的に調整していると考えられている. 共役事前分布はディリクレ分布として知られる, ベータ分布の多変量一般化である.

$$p(\theta|\alpha) \propto \prod_{j=1}^k \theta_j^{\alpha_j - 1}$$

ここで, 分布は $\sum_{j=1}^k \theta_j = 1$ で, θ_j に制限される. 詳細については, 付録 A を参照. θ_j の結果として得られる事後分布は, パラメータ $\alpha_j + y_j$ を有するディリクレである.

事前分布は、 j 番目の結果カテゴリーの α_j 個の観測に関する $\sum_{j=1}^k \alpha_j$ 個の観測に対する尤度と数学的に同値である。二項分布の場合と同様で、いくつかのもっともらしい非情報 Dirichlet 事前分布が存在する。すべての j について $\alpha_j = 1$ と設定することにより、一様密度が得られる。この分布は $\sum_{j=1}^k \theta_j = 1$ を満たす任意のベクトル θ に等しい密度を割り当てる。すべての j に対して $\alpha_j = 0$ を設定すると、 $\log(\theta_j)$ の点で一様事前分布が不適切になります。得られた事後分布は、 k 個のカテゴリーのそれぞれに少なくとも 1 つの観測がある場合、 y の各成分が正であるように適切である。この章の最後の bibliographic note は、多項モデルの他の示唆されていない事前分布を指摘している。

例. 事前選挙投票

多項モデルの簡単な例として, 3 つの可能な回答を伴う調査質問のサンプルを検討する. 1988 年の 10 月の下旬, 調査は CBS News によりアメリカの 1447 人の大人に対して, 来たる大統領選挙の動向を調査するために執り行われた. 1447 人の中で, $y_1 = 727$ が George Bush を, $y_2 = 583$ が Michael Dukakis, $y_3 = 137$ がほかの候補者を支援または意見がないとした. 回答者の情報のほか何もないと仮定すると 1447 の観測は交換可能である. また, 簡単なランダム抽出 (すなわち, 1447 個を「ハットから引き出された」と名付ける) を仮定すると, データ (y_1, y_2, y_3) はパラメータ $(\theta_1, \theta_2, \theta_3)$, つまり母集団の調査の候補者の割合に対する多項分布に従う. 関心の推定値は $\theta_1 - \theta_2$ であり, これは 2 つの主要候補者に対する支持率の差である.

θ の無情報一様事前分布, $\alpha_1 = \alpha_2 = \alpha_3 = 1$ に関し, $(\theta_1, \theta_2, \theta_3)$ の事後分布は *Dirichlet*(728, 584, 138) である. 積分による $\theta_1 - \theta_2$ の事後分布を計算することができましたが, 後でディリクレ分布から 1000 点 $(\theta_1, \theta_2, \theta_3)$ を抽出し, それぞれの $\theta_1 - \theta_2$ を計算するほうが簡単である. その結果を図 3.2 に示す. 1000 回のシミュレーションのすべてが $\theta_1 > \theta_2$ であった. したがって, Bush がの Dukakis よりも多くの支持を得ている推定事後確率は 99.9%以上である.

実は, CBS の調査は, 独立したランダムサンプリングは使用されず, むしろ層別サンプリング計画の変形を使用しています. 8.3 節で (207 ページの表 8.2 を参照), サンプリングスキームの知識をいくつか使うことで, この調査の改善された分析を議論する.

複雑な問題 (例えば, 例えば, 多くの調査質問の結果の同時分析) で, 多項カテゴリの数, したがってパラメータが非常に大きくなり, モデル内で構造を追加することなく, 中程度のサイズのデータセットを有益に分析することは難しい. 正式には, 事前分布またはサンプリングモデルを通じて追加情報が分析に入力できます. 5 章で紹介する階層的モデリングのアイデアを用いて, 有益な事前分布を使用して複雑な問題の推論を改善することができます. あるいは, いくつかの調査質問を相互分類することによって生じる多項パラメータに構造を課すために, 対数線形モデルを使用することができます. 16.7 節では, 詳細と例を示します.

3.5 未知の分散に関する多変量正規モデル

ここでは, 多変量正規分布のパラメータに対するベイジアン推論の分布結果をやや正式に説明します. 多くの点で, これらの結果は単変量正規モデルに対して既に与えられているものと同等であるが, 適用された多くの統計的作業の中心的な活動である線形モデルの解析において重要な役割を果たす重要な新しい側面がある (5 章, 14,15) . この節は, この時点で, 将来の章の参考資料として参照することができます.

多変量正規尤度

議論されるべき基本モデルは、多変量正規分布を有する d 成分の観測可能なベクトル y に関するものであり、

$$y|\mu, \Sigma \sim N(\mu, \Sigma) \quad (3.10)$$

ここで、 μ は長さ d の (列) ベクトルで、対称正定値の Σ は $d \times d$ 分散行列である。単一観測値に対する尤度関数は

$$p(y|y_1, \dots, y_n | \mu, \Sigma) \propto |\Sigma|^{-1/2} \exp \left(-\frac{1}{2} (y - \mu)^T \Sigma^{-1} (y - \mu) \right)$$

で、 n 個の独立同一分布に従う観測値のサンプル y_1, \dots, y_n は

$$\begin{aligned} p(y_1, \dots, y_n | \mu, \Sigma) &\propto |\Sigma|^{-n/2} \exp \left(-\frac{1}{2} \sum_{i=1}^n (y_i - \mu)^T \Sigma^{-1} (y_i - \mu) \right) \\ &= |\Sigma|^{-n/2} \exp \left(-\frac{1}{2} \text{tr}(\Sigma^{-1} S_0) \right) \end{aligned} \quad (3.11)$$

ここで、 s_0 は μ に関する二乗和行の列で、

$$S_0 = \sum_i (y_i - \mu)(y_i - \mu)^T \quad (3.12)$$

共役分析

単変量正規モデルと同様に, 既知の Σ の場合を考慮して多変量正規モデルを解析する.

既知の Σ に関する μ に対する共役事前分布

対数尤度は μ の二次形式で, したがって μ に対する共役事前分布は, $\mu \sim N(\mu_0, \Lambda_0)$ としてパラメータ化される多変量正規分布である.

既知の Σ に関する μ に対する事後分布

μ に対する事後分布は

$$p(\mu|y, \Sigma) \propto \exp \left(-\frac{1}{2} \left((\mu - \mu_0)^T \Lambda^{-1} (\mu - \mu_0) + \sum_{i=1}^n (y_i - \mu)^T \Sigma^{-1} (y_i - \mu) \right) \right)$$

これは, μ の二次形式の指数である. 平方完成し定数項を引く (演習 3.13 を参照) と

$$\begin{aligned} p(\mu|y, \Sigma) &\propto \exp \left(-\frac{1}{2} (\mu - \mu_n)^T \Lambda_n^{-1} (\mu - \mu_n) \right) \\ &= N(\mu | \mu_n, \Lambda_n) \end{aligned}$$

ここで,

$$\mu_n = (\Lambda_0^{-1} + n\Sigma^{-1})^{-1}(\Lambda_0^{-1}\mu_0 + n\Sigma^{-1}\bar{y}), \Lambda_n^{-1} = (\Lambda_0^{-1} + n\Sigma^{-1})^{-1} \quad (3.13)$$

2.5 節の単変量正規モデルの結果と類似して, 事後平均はデータ平均と事前平均の加重平均であり, 加重はデータと以前の精度行列, $n\Sigma^{-1}, \Lambda_0^{-1}$ によって与えられ, 事後精度は, 事前精度とデータ精度の和である.

既知の Σ に関する μ の部分ベクトルの事後条件付き, 事後周辺分布

$\mu(1)$ は, 多変量正規分布 (付録 A を参照) の特性から, $\mu(1)$ の部分集合の周辺事後分布も多変量正規分布であり, 平均ベクトルは事後平均の適切な部分ベクトルに等しいベクトル μ_n および分散行列は, 適切な部分行列に等しい. また, 第 2 の部分集合 $\mu(2)$ の値を与えられた部分集合 $\mu(1)$ の条件付き事後分布は, 多変量正規分布である. 適切なサブベクトルと部分行列を示すために括弧内に上付き文字を書くと,

$$\mu^{(1)}|\mu^{(2)}, y \sim N\left(\mu_n^{(1)} + \beta^{1|2}(\mu^{(2)} - \mu_n^{(2)}), \Lambda^{1|2}\right) \quad (3.14)$$

ここで, 回帰係数 $\beta^{1|2}$ と条件付き分散行列 $\Lambda^{1|2}$ は

$$\eta^{1|2} = \lambda_n^{(12)} \left(\Lambda_n^{(22)}\right)^{-1}, \Lambda^{1|2} = \Lambda_n^{(11)} - \Lambda_n^{(12)} \left(\Lambda_n^{(22)}\right)^{-1} \Lambda_n^{(21)}$$

で定義される.

新しいデータに対する事後予測分布

新しい観測値に対する事後予測分布の解析形式 $\tilde{y} \sim N(\mu, \Sigma)$ を考える. 単変量正規分布と同様に, 結合分布 $p(\tilde{y}, \mu|y) = N(\tilde{y}|\mu, \Sigma)N(\mu|\mu_n, \Lambda_n)$ は (\tilde{y}, μ) の二次形式の指数である. 従って, (\tilde{y}, μ) は結合正規事後分布をもち, \tilde{y} の周辺事後分布は (多変量) 正規分布である. 依然として分散行列 Σ が既知であると仮定している. 単変量の場合と同様に, \tilde{y} の事後平均, 分散は, (2.7), (2.18) を使い,

$$E(\tilde{y}|y) = E(E(\tilde{y}|\mu, y)|y) = E(\mu|y) = \mu_n$$

$$\text{var}(\tilde{y}|y) = E(\text{var}(\tilde{y}|\mu, y)|y) + \text{var}(E(\tilde{y}|\mu, y)|y) = E(\Sigma|y) + \text{var}(\mu|y) = \Sigma + \Lambda_n$$

事後分布または事後予測分布からサンプリングするには, 指定平均および分散行列を用いて多変量正規分布からランダム抽出を生成する方法について付録 A を参照する.

μ に対する無情報事前分布

μ に関する無情報一様事前密度は, $|\Sigma_0^{-1}| \rightarrow 0$ の意味において事前精度がゼロになるので, 極限で得られる $p(\mu) \propto \text{const}$ である. 無限の極限で事前分散の限界 (事前精度ゼロ) において, 事前平均は無関係である. 事後密度は, 尤度 (3.11) に比例する. これは, $n \geq d$ の場合にのみ適切な事後分布であり, すなわち, 標本サイズが多変量正規分布の次元以上である場合に限られる. そうでなければ行列 S_0 はフルランクではない. $n \geq d$ の場合, 一様事前密度を仮定すると μ の事後分布は $\mu|\sigma, y \sim N(y, \sigma/n)$ である.

3.6 未知の平均と分散に関する多変量正規分布である 事前分布の共役逆-Wishart 分布族

未知の平均および分散を有する一変量正規分布の共役分布は、正規-逆- χ^2 分布 (3.6) であることを想起されたい。行列の事前分布を記述するために、スケーリングされた逆- χ^2 の多変量一般化である逆-Wishart 分布を用いることができる。 $(\mu_0, \Lambda_0/\kappa_0; \nu_0, \sigma_0)$ の観点から、 (μ, σ) の共役事前分布、正規-逆-Wishart は便宜的にパラメータ化される。

$$\Sigma \sim \text{Inv-Wishart}_{\nu_0}(\Lambda_0^{-1}), \mu|\Sigma \sim N(\mu_0, \Sigma/\kappa_0)$$

これは、結合事後分布

$$p(\mu, \Sigma) \propto |\Sigma|^{-((\nu_0+d)/2+1)} \exp\left(-\frac{1}{2}\text{tr}(\Lambda_0\Sigma^{-1}) - \frac{\kappa_0}{2}(\mu - \mu_0)^T\Sigma^{-1}(\mu - \mu_0)\right)$$

に対応する。パラメータ ν_0, Λ_0 は Σ の逆-Wishart 分布に対する自由度とスケール行列を表す。残るパラメータは、 Σ スケールでの事前平均 μ_0 と事前計量の数 κ_0 である。

事前密度に正規尤度をかければ、パラメータ

$$\mu_n = \frac{\kappa_0}{\kappa_0 + n} \mu_0 + \frac{n}{\kappa_0 + n} \bar{y}, \quad \kappa_n = \kappa_0 + n, \quad \nu_n = \nu_0 + n,$$

$$\Lambda_n = \Lambda_0 + S + \frac{\kappa_0 n}{\kappa_0 + n} (\bar{y} - \mu_0)(\bar{y} - \mu_0)^T$$

に関する同じ族の事前密度となる. ここで, S はサンプル平均についての二乗和行列である.

$$S = \sum_{i=1}^n (y_i - \bar{y})(y_i - \bar{y})^T$$

単変量正規分布からの他の結果は, 多変量の場合に容易に一般化する. μ の周辺事後分布は, 多変量 $t_{\nu_n-d+1}(\mu_n, \Lambda_n/(\kappa_n(\nu_n-d+1)))$ である. 新しい観測値 \tilde{y} の事後予測分布も, スケール行列の分子の $\kappa_n + 1$ の追加因子を持つ多変量 t です. 次の手順を使用して, (μ, Σ) の共役事後分布からのサンプルを容易に得ることができます. 最初に, $\Sigma|y \sim \text{Inv-Wishart}_{\nu_n}(\Lambda_n^{-1})$ を抽出し, $\mu|\Sigma, y \sim N(\mu_n, \Sigma/\kappa_n)$ を抽出する. 逆-Wishart および多変量正規分布からの抽出については, 付録 A を参照してください. 新しい観測の事後予測分布から抽出するには, 既に抽出された μ, Σ の値を与えて, $\tilde{y}|\mu, \Sigma, y \sim N(\mu, \Sigma)$ を抽出する.

異なる無情報事前分布

自由度 $d + 1$ の逆-Wishart 分布.

設定 $\Sigma \sim \text{Inv-Wishart}_{d+1}(I)$ の魅力的な特徴は, それぞれの相関関係は Σ わずかな事前分布しかない. (ただし, 相関行列が正定値であるという制約のため, 結合分布は一様ではない.)

自由度 $d - 1$ の逆-Wishart 分布.

別の提案された無情報事前分布は, **多変量 Jeffreys 事前密度**であり,

$$p(\mu, \Sigma) \propto |\Sigma|^{-(d+1)/2}$$

これは, 共役事前密度の $\kappa_0 \rightarrow 0, \nu_0 \rightarrow -1, |\Sigma_0| \rightarrow 0$ 極限である. 対応する事後分布は

$$\Sigma|y \sim \text{Inv-Wishart}_{n-1}(S^{-1}), \mu|\Sigma, y \sim N(\bar{y}, \Sigma/n)$$

として書ける. 事後分布が適切であると仮定して μ の周辺分布と \bar{y} の事後予測分布の結果は, 前の文脈に従う. 例えば, μ の周辺事後分布は, 多変量 $t_{n-d}(\bar{y}, S/(n(n-d)))$ である.

スケーリングされた逆-Wishart モデル

共分散行列をモデリングするときは、別々にモデル化できる一連のスケールパラメータを乗算することによって、逆-Wishart モデルを拡張するのに役立ちます。これにより、モデリングの柔軟性が得られ、分散パラメータを過度に制約することなく、相関の一樣または弱い事前分布を設定することができます。 Σ に対するスケーリングされた逆 Wishart モデルは

$$\Sigma = \text{Diag}(\xi) \Sigma_{\eta} \text{Diag}(\xi)$$

ここで、 η は逆-ウィシャート事前分布 (1 つの選択は $\text{Inv-Wishart}_{d+1}(I)$ であり、相関の周辺分布は一樣である) を与えられ、次に尺度パラメータ ξ は弱く有益な事前検体を与えられる。本節では、漸变的な勾配の階層的帰帰モデルに関連して 15.4 節でさらに議論する。

3.7 例: バイオアッセイ実験の分析

正規分布以外にも, 多パラメータサンプリングモデルでは事後分布の単純明示的計算が可能である. このようなモデルのデータ解析は, 本書の3部で説明した計算方法を使用して行うことができる. ここでは, 適用されたベイズ統計に関する文献から導かれたバイオアッセイ実験の非共役モデルの例を示します. このモデルは, 一般的な線形モデルの広範なクラスからの2つのパラメータの例であり, 16章でより完全に考慮する必要があります. 我々は, この2つのパラメータの例について十分に正確な推論を提供する点の2次元格子上で支持された離散分布によって事後分布を近似する, 特に単純なシミュレーション手法を使用する.

科学的な問題とデータ

薬物および他の化合物の開発においては、動物に急性毒性試験またはバイオアッセイ実験が一般的に行われている。そのような実験は、動物のバッチに様々な用量レベルの化合物を投与することによって進行する。動物の反応は、典型的には2つの結果。例えば、生存しているか死んでいる、腫瘍が存在していないか、または腫瘍がない。この種の実験は、

$$(x_i, n_i, y_i); i = 1, \dots, k$$

の形のデータを生じさせる。ここで、 x_i は n_i 動物に与えられた k 用量レベル (しばしば対数スケールで測定された) の i 番目を表し、そのうち y_i はその後陽性結果で応答する。そのような実験の実際のデータの例を表3.1に示します。20匹の動物を試験し、4つの用量レベルのそれぞれで5匹を試験した。

用量-反応関係のモデリング

これまで見てきたことを考えれば, 各グループ i 内の 5 匹の動物の結果を交換可能とモデル化しなければならず, データ点 y_i が二項分布していることを意味する等しい確率で独立とモデル化することは妥当である.

$$y_i | \theta_i \sim \text{Bin}(n_i, \theta_i)$$

ここで θ_i は投与量 x_i を与えられた動物の死亡確率である. (独立性と二項モデルが適切でない状況の例は, その死が伝染病によって引き起こされたかどうかである). この実験ではパラメータ $\theta_1, \dots, \theta_4$ が与えられた場合, 4 つの群の結果を互いに独立して扱うことも妥当である.

最も簡単な解析は, おそらく $p(\theta_1, \dots, \theta_4) \propto 1$ のような無情報密度を使用して, 4 つのパラメータ θ_i を前の分布で交換可能として扱い, この場合, パラメータ θ_i は独立したベータ事後分布を有する. しかし, θ_i パラメータの交換可能な事前モデルには重大な欠陥があります. 私たちは各グループ i の線量レベル x_i を知っており, 用量の関数として系統的に変化する死の確率を期待しています.

用量-反応関係の最も単純なモデル, つまり θ_i と x_i の関係は線形, $\theta_i = \alpha + \beta x_i$ である. 残念ながら, このモデルには, 低または高用量で x_i が $\pm\infty$ に近づくという欠陥があります (用量は対数スケールで測定されます). 一方, 確率である θ_i は 0 と 1 の間にあるように制約する必要があります. 標準的な解法は, ロジスティックのような θ の変化を用量-反応関係で使用する事です.

$$\text{logit}(\theta_i) = \alpha + \beta x_i \quad (3.15)$$

ここで, $\text{logit}(\theta_i) = \log(\theta_i/(1 - \theta_i))$ は (1.10) で定義された. これを **対数回帰 (logistic regression) モデル** と呼ばれる.

尤度

モデル (3.15) の下で, パラメータ α, β に関するグループ i に対する, サンプルング分布, または尤度を

$$p(y_i|\alpha, \beta, n_i, x_i) \propto [\text{logit}^{-1}(\alpha + \beta x_i)]^{y_i} [1 - \text{logit}^{-1}(\alpha + \beta x_i)]^{n_i - y_i} \quad (3.16)$$

として書ける. モデルは結合事後分布が

$$p(\alpha, \beta|y, n, x) \propto p(\alpha, \beta|n, x)p(y|\alpha, \beta, n, x) \quad (3.17)$$

$$\propto p(\alpha, \beta) \prod_{i=1}^k p(y_i|\alpha, \beta, n_i, x_i) \quad (3.18)$$

となるパラメータ α, β で特徴づけられる. この分析のためにサンプルサイズ n_i および用量レベル x_i を定数として考慮し, 後続の表記法における (n, x) の条件付けを省略する.

事前分布

2つのパラメータにおいて独立かつ局所的に均一な (α, β) の事前分布に基づく分析を提示する. すなわち, $p(\alpha, \beta) \propto 1$ である. 実際には, パラメータに関する事前の知識が実際でない場合, またはこの実験単独の簡単な分析を提示したい場合は, 事前に分布を均一にすることがあります. 無情報事前分布を用いた分析が十分に正確でない場合, 有益な事前分布を構築するために他の情報源 (例えば, 他のバイオアッセイ実験から) を用いることを検討することができる.

パラメータの大まかな推定

点 (α, β) のグリッドで結合事後分布 (3.16) を計算しますが, その前に (α, β) の大まかな推定値を取得してどこを見るかを知ることをお勧めします. おおまかな見積もりを得るために, 既存のソフトウェアを使用してロジスティック回帰を実行します. 表 3.1 の 4 つのデータ点について, (3.16) の (α, β) の最尤推定値を求める. 推定値は $(\alpha, \beta) = (0.8, 7.7)$ であり, α と β の標準誤差はそれぞれ 1.0 と 4.9 である.

結合事後密度の等高線図の取得

点 (α, β) のグリッドで事後密度を計算する準備が整った. いくつかの実験の後、範囲 $(\alpha, \beta) \in [-5, 10] \times [-10, 40]$ を使用して、事後分布のほぼすべての質量を捕捉する. 結果の等高線プロットは図 3.3a のようになります. 2次元プロットの最小輪郭レベルを 0.05 に設定するための一般的な正当性は、4.1 節の 85 ページに記載されている.

結合事後分布からのサンプリング

(α, β) の有効範囲をカバーする値のグリッドで正規化されていない事後密度を計算すると、分布をグリッド上のステップ関数として近似し、グリッド内の合計確率を 1 に設定することによって正規化することができる。1000 回の無作為抽出 (α^s, β^s) を以下の手順で事後分布から計算する。

1. 図 3.3a のグリッド上で計算された離散分布の β を数値的に合計することによって、 α の周辺事後分布を計算する。
2. $s = 1, \dots, 1000$
 - (a) 離散的に計算された $p(\alpha|y)$ から α を抽出する。これは、1.9 節で説明した逆 cdf メソッドの離散バージョンとして見ることができる。
 - (b) サンプリングされた α の値が与えられた場合、離散条件付き分布 $p(\beta|\alpha, y)$ から β^s を引き出す。
 - (c) サンプリングされた α と β のそれぞれに対して、サンプリンググリッドの間隔に等しい幅でゼロを中心とする一様ランダムジッタを加える。これにより、シミュレーションでは連続した分布が得られます。

図 3.3b の散布図に 1000 回の抽出 (α^s, β^s) が表示されます. 図 3.3a の尺度と同じ, プロットの尺度は, すべて 1000 回の抽出がグラフに収まるほど大きく設定されています.

この 2 次元グリッド近似を適用する際には, 多くの実際的な考慮があります. グリッドポイントの正確な位置とスケールを見つけるのが難しい場合があります. あまりにも小さな領域に定義されたグリッドは, グリッドの外側にある事後分布の重要な特徴を見逃す可能性があります. ポイント間の間隔が広い広いエリアに定義されたグリッドは, グリッドポイント間にある重要な特徴を見逃す可能性があります. 事後分布を計算する際にオーバーフローやアンダーフローの操作を避けることも重要です. 正規化されていない事後分布の対数を計算し, べき乗する前に最大値を減算することは, 通常は良い考えです. これは最大値 1 を持つ正規化されていない離散近似を作成し, 正規化することができます (グリッドの合計確率を 1 に設定することによって).

LD50 の事後分布

バイオアッセイ研究で共通の興味のあるパラメータは LD50(死亡確率が 50 %である用量レベル) であるロジスティックモデルでは, 50%生存割合は

$$\text{LD50: } E\left(\frac{y_i}{n_i}\right) = \text{logit}^{-1}(\alpha + \beta x_i) = 0.5$$

を意味する. そして, $\alpha + \beta x_i = \text{logit}(0.5) = 0$ で, LD50 は $x_i = -\alpha/\beta$ である. ベイズ手法における任意の要約の事後分布の計算は, 1.9 節の最後で説明するように, 簡単です. LD50 の事後分布をシミュレートするのは簡単です. 図 3.3b に描かれている (α, β) の 1000 回の抽出に対して $-\alpha/\beta$ を計算するだけです.

薬物が有益である場合の LD50 パラメータ化の困難

この例の文脈では、 $\beta \leq 0$ の場合、LD50 は無意味な概念であり、この場合、用量を増加させても死亡確率は増加しない。薬物が腫瘍率を低下させることができないことが確かであれば、 β の値を 0 未満にするようにパラメータ空間を制限する必要があります。しかし、ここで $\beta \leq 0$ の可能性を許容することがより理にかなっているようであり、LD50 はこの場合には解釈が難しいことに注意する。2つの結果を報告することにより、LD50 スケールの推論をまとめる。(1) $\beta > 0$ 、すなわち薬物が有害であるという事後確率、および(2) LD50 の事後分布は $\beta > 0$ である。1000 回のシミュレーション抽出のすべてが β の正の値を有するので、 $\beta > 0$ である事後確率はおおよそ 0.999 を超えると推定される。シミュレーションの抽出の LD50 は β の正の値で計算されます(この例では 1000 回抽出されます)。図 3.4 にヒストグラムが表示されます。この例は、周辺事後平均が必ずしもパラメータに関する推論の良い要約ではないことを示している。一般的に、LD50 の事後平均には関与していません。なぜなら、事後平均には用量反応関係が負の場合が含まれているからです。

3.8 基本的なモデリングと計算のまとめ

事後分布の計算を容易にするマルチパラメータモデルの欠如は、3つの主な理由のための主要な実用的ハンディキャップではない。第1に、パラメータが少ない場合、非共役多パラメータモデルにおける事後推論は、バイオアッセイの例に見られるように、単純なシミュレーション方法によって得ることができる。第2に、洗練されたモデルは、5章で述べるように、効果的な計算戦略が利用可能である(3部で一般的に議論するように)、階層的または条件的な方法で表されることが多い。最後に、4章で議論するように、事後分布に正規近似を適用することが多いため、通常モデルの共役構造は、実際には明示的に正規のサンプリングモデルへの適用を超えて重要な役割を果たすことができます。

バイオアッセイの例の我々の成功した分析は、単純なベイズ事後分布の計算のための以下の戦略を示唆している。以下は真に一般的なアプローチではありませんが、これまでに行ったことを要約して、逐次近似に基づく一般的な方法を紹介します。

1. θ のない任意の因子を無視して, モデルの尤度部分, $p(y|\theta)$ を書く.
2. 事後密度 $p(\theta|y) \propto p(\theta)p(y|\theta)$ を書く. 事前情報がよく形式化されている場合は, $p(\theta)$ に含める. さもないければ, 前の密度が後で追加の情報または構造を含むように変更され得るという理解をもって, 弱情報事前分布を使用するか, 一時的に $p(\theta) \propto \text{const}$ を設定する.
3. 出発点として使用するパラメータ θ のあらい推定値を作成し, 次のステップで計算との比較を行う.
4. 事後分布からシミュレーション $\theta^1, \dots, \theta^S$ を抽出する. 関心のある θ の関数の事後密度を計算するには, サンプルドローを使用します.
5. 任意の予測量, \tilde{y} が関心のある場合は, 抽出された θ^s の条件でサンプル分布から \tilde{y}^s を抽出することで $\tilde{y}^1, \dots, \tilde{y}^S$ を抽出する. 6 章では, θ, \tilde{y} の事後シミュレーションを使用して, データと実際の知識に対するモデルの適合性をチェックする方法を説明する.

非共役モデルの場合, 上記のステップ 4 は困難です. 3 部で議論するように, 複雑なモデルで事後シミュレーションを描くために様々な方法が開発されてきた. 時には, 高次元の問題は, 解析的シミュレーションと数値シミュレーションの方法を組み合わせることで解決できます. θ に 1, 2 個の成分しかない場合, 前のバイオアッセイの例で説明したように, グリッドでの計算によってシミュレーションを作成することができる.