

第 1 章

ベイズ推測法

最尤法では観測値の持つパラメータやモデルに関する情報を尤度関数を通して引き出す。また、最尤法を利用すれば、パラメータの推定からモデル選択まで一貫したフレームワークの下で統計的推測を行うことができる。最尤法の性質の多くは、大標本下での漸近的な挙動により保証されている。もちろん、これらの性質は極端に標本数が大きくなくとも実際上は問題なく利用できるが、現実的にはパラメータの数に比して少ない情報量の観測値しか得られない場合もあるだろう。

例えば、ある生物個体群から 10 個体をランダムにサンプリングし雌雄それぞれの数を観察したところ、すべて雌であったとする。雌雄で生息域が異なるような場合は別として、集団からランダムにサンプリングができていると考えられる場合に、このような観測値に遭遇すれば「想定していることと合致しない」と考えるであろう。この例の場合、母集団中のオスの割合を θ とおくと、 θ は 0.5 に近い値をとるという事前の知識があるからである。逆にこのような「想定していること」あるいは「事前の知識」を確率的に表現し推測に利用する方法をベイズ法 (Bayesian method) とよぶ。

ベイズ法の元々の由来は上記のように事前情報を推測に取り入れることにあったが、ベイズ法はその特徴を生かして様々な方向へと応用されてきている。例えば、Stein 推定 (Stein 1956) に代表される縮小推定では、類似のパラメータを多く含むモデルにおいて、ベイズ的推定法が最尤法を優越することが証明された。このアイデアは “borrowing strength” というコンセプトとしてランダム効果モデルなど多くのモデルで利用されてきた。また、ベイズ法は潜在変数を含む階層構造をもつモデルとも大変相性がよい。実は一昔前までは、ベイズ法の理論的な良さが計算上の困難さから実践で利用できないこともあった。しかし、最近ではマルコフ連鎖モンテカルロ法に代表されるような計算アルゴリズムの進化に加え、WinBUGS などのソフトウェア環境の発展もあり、その障壁はかなり低くなってきている。そのため、最尤法における最適化における困難を避け、計算の利便性から敢えてベイズ法を利用することもしばしばある。

さてこの節では、ベイズ法の基本的な考え方のみを 1 パラメータモデルを通して述べる。2 つ以上のパラメータを含むモデルを含め、その他の拡張や応用例については、後述の関連する章にて紹介する。

1.1 事前分布と事後分布

例題 2.XX

複数の分集団で構成されるメタ個体群がある。このメタ個体群中の一つに分集団から 10 個体のメスをサンプリングし、妊娠しているかどうかを観測したところ妊娠個体は 1 個体のみであった。このメタ個体群全体における過去の観測結果から、妊娠個体数の割合 (妊娠率) の事前情報として期待値 0.3、標準偏差 0.05 の確率分布が想定できるとする。この事前情報と観測値の情報を基に、この分集団の現在の妊娠率をベイズ法により推定せよ。

一つの大きな個体群全体が、多くの小さな分集団によって構成されているとき、これをメタ個体群とよぶ。分集団どうしで個体の交流がランダムに生じる場合が最も典型的なメタ個体群であるが、必ずしもその形に限らない。いま、このメタ個体群全体の妊娠率を θ とおき、妊娠率は分集団間で共通とする。このとき、サンプリングの対象となった分集団から観測値 ($Y = y$ とおく) は 2 項分布 $Bin(N, \theta)$ に従う (先の例では $N = 10$ である)。また $Y = y$ の確率関数は

$$f(y|\theta) = \binom{N}{y} \theta^y (1 - \theta)^{N-y} \quad (1.1)$$

と表せる。ベイズ法の枠組みでは、この観測値の確率分布を、パラメータ θ が与えられたときの条件付分布とみなす。

次に、事前情報の利用について考える。事前情報を確率分布で表現するとき、これを事前分布 (prior distribution) とよぶ。事前分布の表現方法は一意的ではないが、 θ が割合を表すパラメータで定義域が $0 < \theta < 1$ であるから、ここでは自然な確率分布としてベータ分布 $Beta(a, b)$ を仮定する。この事前分布の確率密度関数を

$$\pi(\theta) = \frac{1}{B(a, b)} \theta^{a-1} (1 - \theta)^{b-1} \quad (1.2)$$

と表すことにする。ベータ分布の期待値と分散の定義より、この場合の事前分布のパラメータは $a = 24.9, b = 58.1$ となる。

次に、この事前分布を観測値に基づいて更新してみよう。いまパラメータに関する事前情報を確率分布の形で表現した。このような事前の知識の確率的表現は、実は既に条件付確率の項の課題問題 2.1 でも述べている。すなわち、由来の分からない個体が 2 つの湖のいずれから来た確率をそれぞれ $1/2$ とし、この事前確率をその個体の遺伝子の観測に基づき条件付確率を通して更新した。このようにして、情報の追加により確率的な主張を更新する際、条件付確率および条件付分布が自然に利用できる。この例でも、パラメータの事

前分布を $Y = y$ を与えたときの θ の条件付分布として次のように更新する.

$$\begin{aligned}
 \pi(\theta|y) &= \frac{f(y|\theta)\pi(\theta)}{f(y)} = \frac{f(y|\theta)\pi(\theta)}{\int_0^1 f(y|\theta)\pi(\theta)d\theta} \\
 &= \frac{\binom{N}{y}\theta^y(1-\theta)^{N-y}\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}}{\int_0^1 \binom{N}{y}\theta^y(1-\theta)^{N-y}\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}d\theta} \quad (1.3) \\
 &= \frac{1}{B(y+a, N+y-b)}\theta^{y+a-1}(1-\theta)^{N-y+b-1}
 \end{aligned}$$

このようにして更新されるパラメータに対する情報を事後分布 (posterior distribution) とよぶ. 図 1 では標本数の増大に応じて事前分布が更新されていく様子を示している. 標本数が小さい時には事前分布の影響が大きく, 標本数が増大すればその影響も自然と薄れ, 観測値の情報が支配していく点がベイズ推測の一つの良い性質である.

ここで述べた 2 項分布の例では, 事前分布および事後分布ともにベータ分布となる. このように, 事前分布の分布型が事後分布でも保存されるとき, この事前分布のことを共役事前分布 (conjugate prior distribution) とよぶ. この他, ポアソン分布の期待値パラメータに対してガンマ分布を仮定した場合や, 正規分布の期待値パラメータに正規分布を仮定する場合も共役事前分布となる. 共役事前分布を仮定すると, 事後分布が解析的に求まるという利点があるが, 数値計算により事後分布が求められる現在の状況では, 必ずしも事前分布を共役事前分布に制約する必要はない.

共役事前分布の他に, 無情報事前分布 (noninformative prior) とよばれる事前分布がある. 無情報事前分布は, パラメータに関する情報が事前には無いがベイズ流に確率分布を通してパラメータの推測結果を表現したい場合や, 形式的にベイズの枠組みを用いて推定精度を上げたい場合, そして小標本の場合の極端な推測結果を避けたい場合など, いくつかの場面で仮定される. 利用される理由は一意的ではなく, また定義の仕方も議論のあるところである. 例えば, この 2 項分布の例においても, 自然に考えれば $Beta(1, 1) = U(0, 1)$ が無情報事前分布となるが, Jeffreys の定義した無情報事前分布は, Fisher 情報量を用いて $\pi(\theta) = \sqrt{I(\theta)}$ で定義するため, $Beta(1/2, 1/2)$ が無情報事前分布となる. この Jeffreys の無情報事前分布はパラメータを変換しても無情報事前分布の定義が維持され, パラメータの変換に対する不変性を有しているが, 複数のパラメータを含むモデルでは必ずしもこの性質が成り立たないことが知られている. なお, ここではこの問題には深入りせず, 議論はベイズの専門書に譲る (例えば, Gelman et al. 2004).

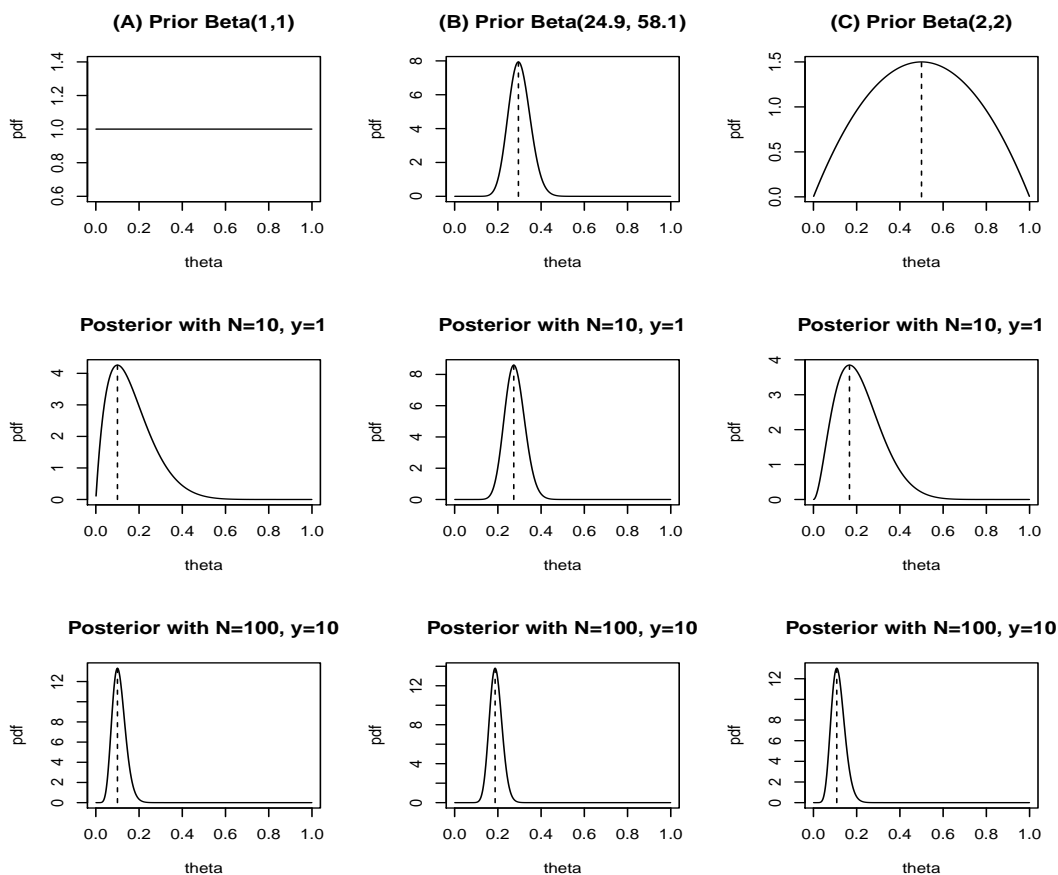


図 1.1: 事前分布から事後分布への更新の様子. (A) 事前分布が $\text{Beta}(1,1)$ のとき事後分布の最大値は最尤推定値と同じ. (B) 事前分布が $\text{Beta}(24.9, 58.1)$ のとき事前の情報強く更新が慎重に行われている. (C) 事前分布 $\text{Beta}(2,2)$ では事前分布の情報 $\text{Beta}(24.9, 58.1)$ と比較して強くないため事前分布の仮定が即座に薄められる. 作図の例は”Ch2.bayes.binom.r”を参照のこと.

1.2 事後分布を用いたパラメータの推定

1.2.1 点推定

前節で述べたように、ベイズ法ではパラメータに対する事前の知識や情報を「事前分布」として表し、この事前分布を観測値に基づいて「事後分布」へと更新することがその基礎となる. ところで、最尤法では尤度関数を最大化することでパラメータを推定し、プロファイル尤度や種々の漸近的性質を通して信頼領域他を構成した. ベイズ推測法では事後分布が同様の役割を果たす.

事後分布の最大値を**事後モード** (posterior mode) とよぶ. 図 1 の事後分布における破線は事後モードの値を示している. 一様分布型の無情報事前分布を仮定するとき、事後分

布の最大値は最尤推定値と同じであるが，他の事前分布を仮定する場合には事前分布の形状に応じて事後モードの更新の様子が変わる．事後モードはベイズ的なパラメータの推定方法の一つであり，数値的な最適化を伴う場合には威力を発揮するが，最尤法の場合とは異なりパラメータの変換に不変ではない点に注意が必要である．なお，2項分布の例では事後モードは $\hat{\theta}_{mode} = (y + a - 1)/(N + a + b - 2)$ で求まる．

事後モードと並んで利用される点推定の方法として事後平均 (posterior mean) や事後メディアン (posterior median) が挙げられる．事後平均はその名の通り，事後分布 $\pi(\theta|y)$ に関して θ の条件付期待値を求める．すなわち，2項分布の θ の場合には，

$$\hat{\theta}_{mean} = E[\theta|Y = y] = \int_0^1 \theta \pi(\theta|y) d\theta = \frac{y + a}{N + a + b} \quad (1.4)$$

で求まる．2項分布とベータ分布のように共役な関係にあるとき，事後分布は解析的に美しい性質をもつ．すなわち，事後平均は事前平均と最尤推定値 $\hat{\theta}_{ML} = y/N$ の重み付き平均

$$E[\theta|Y = y] = \frac{a + b}{N + a + b} E[\theta] + \frac{N}{N + a + b} \hat{\theta}_{ML} \quad (1.5)$$

で表せ， N の値が小さいとき事前平均の重みが大きくなり，逆に N が大きくなるにしたがって最尤推定値が優勢となる．なお，事後平均の精度は下記の事後分散で評価される．

$$V[\theta|Y = y] = \int_0^1 (\theta - E[\theta|Y = y])^2 \pi(\theta|y) d\theta = \frac{1}{N + a + b + 1} \cdot \frac{y + a}{N + a + b} \cdot \frac{N - y + b}{N + a + b} \quad (1.6)$$

この他，事後メディアンは事後分布をシミュレートする場合に利用しやすく，とくに事後分布の形に歪みがある場合や，裾の重い事後分布の場合に効果的である．

例題 2.XXX の例における推定結果は， $\hat{\theta}_{mode} = 0.274$, $\hat{\theta}_{mean} = 0.278$ ($SE = 0.0465$)，そして $\hat{\theta}_{med} = 0.277$ となり，ほぼ等しいことが分かる．なお，最尤推定値は $\hat{\theta}_{ML} = 0.1$ ($SE = 0.095$) である．

ここで忘れてはならないのは，ベイズ法の利点は事後平均や事後モードでパラメータを推定することだけでなく，事後分布全体を通して確率的にパラメータを推測することである．次節ではその一例として，ベイズ信用区間について述べる．

1.2.2 ベイズ信用区間

最尤法では，最尤推定値の付近で尤度プロファイルがシャープであれば，それだけ信頼区間が狭くなる．ベイズの事後分布においても同じ事が当然いえる．例えば図 1 において，いずれの事前分布の場合にも，標本数 $N = 10$ と $N = 100$ とでは事後分布の形状が異なる．もちろん標本数を大きくすることにより，確率的に事後モードの周りに集中した事後分布となる．

ところで，最尤法に代表される頻度主義的な統計推測とベイズ法の大きな違いは，パラメータをある値とみなすか，あるいは確率変数とみなすかという点である．例えば頻度主

義における 95% 信頼区間は、同様の実験を 100 回繰り返した場合に 95 回は真のパラメータ値を含むという意味であり、決して真のパラメータが提示された信頼区間に含まれる確率が 95% であるという主張ではない。一方で、ベイズ法ではパラメータは確率変数であり、確率的なステートメントが可能であるため、95% の確率で真のパラメータを含む区間を定義することができる。このように、データ解析における実際上の解釈は別として、ベイズ法の信頼区間は厳密には頻度主義的な信頼区間とは異なる意味をもつ。そこでベイズ法では信用区間 (credible interval) とよび区別する。

信用区間の計算方法には大きく 2 種類あり、一つは $100\% - 95\% = 5\%$ の確率を分布の裾の右と左に 2.5% ずつ当配分する方法、もう一つは最も狭い区間となるように確率的に高いパラメータ値の区間を構成する方法である。後者はすなわち

$$\int_{\{\theta|\pi(\theta|y)>c\}} \pi(\theta|y)d\theta = 0.95 \quad (1.7)$$

となるような定数 c を求め、 $\{\theta|\pi(\theta|y) > c\}$ により区間 (あるいは領域) を求める方法である。これにより、区間中のすべての値は区間外のどの値よりも高い事後確率密度となり、かつ密度の累積 (積分値) が 95% となる。このような信用区間を特に最高事後密度区間 (highest posterior density interval) とよぶ。

先の 2 項分布の例では、事後分布の形状がモードの付近で対称に近く、等確率に配分する方法では $95\%CI = (0.193, 0.373)$ 、最高事後密度区間法では $95\%CI = (0.190, 0.369)$ となり、比較的近い値をとることがわかる。

```

1 # 95% Credible Interval (equal tail)
2 CI_1 <- function(a, b, N, y, level){
3   percentile <- c((1-level)/2, 1-(1-level)/2)
4   CI <- qbeta(percentile, y+a, N-y+b)
5   CI
6 }
7
8 CI_1.res <- CI_1(a=24.9, b=58.1, N=10, y=1, level=0.95); CI_1.res
9
10 # 95% Credible Interval (grid HPDI)
11 CI_2 <- function(a, b, N, y, level, split=1000){
12   theta <- seq(0,1,length.out=split+1)[-c(1,(split+1))]
13   PDF <- dbeta(theta, y+a, N-y+b)/split
14   clist <- seq(0.00001,max(PDF),length.out=100)
15   SS <- numeric(100)
16   for(i in 1:100){ SS[i] <- sum(PDF[PDF > clist[i]]) }
17   Num <- which.min(abs(SS-level))
18   INT <- theta[PDF > clist[Num]]
19   CI <- c(min(INT), max(INT))
20   CI
21 }
22
23 CI_2.res <- CI_2(a=24.9, b=58.1, N=10, y=1, level=0.95); CI_2.res
24

```

```

25 # Posterir and Credible Interval
26 a<-24.9; b<-58.1; N<-10; y<-1
27 theta <- seq(0.001,0.999,0.001)
28 win.graph()
29 plot(theta, dbeta(theta, y+a, N-y+b), type="l", ylab="pdf")
30 theta.mode<-(a+y-1)/(N+a+b-2)
31 theta.mean<-(a+y)/(N+a+b)
32 lines(c(theta.mode, theta.mode),c(0,dbeta(theta.mode, y+a, N-y+b)), lty=2)
33 points(theta.mean, 0, pch=4, col="red", cex=2)
34 lines(CI_2.res, c(0,0), lwd=2, col="blue")

```

(Ch2.bayes.interval.r)

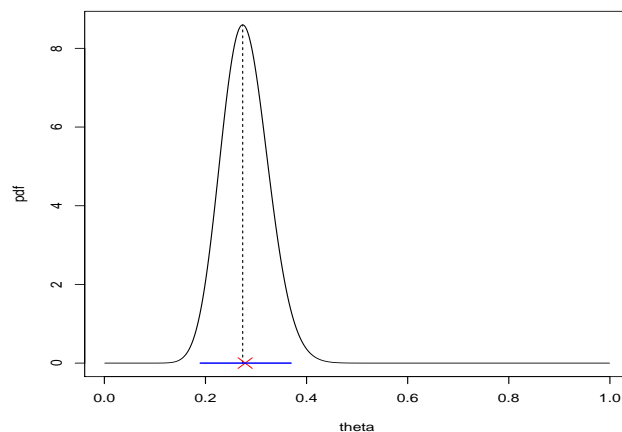


図 1.2: 事前分布 $\text{Beta}(24.9, 58.1)$, 標本数 $N = 10$, 観測オスの数 $y = 1$ の場合の 95% 高事後密度区間 (青線). 赤×印は事後平均の値.

1.2.3 仮説の検証およびモデルの妥当性評価

基礎的な仮説検定の問題では $H_0: \theta = \theta_0$ のような単純帰無仮説に対して, $H_1: \theta \neq \theta_0$ のような複合対立仮説を設定して検定を行う. ところが, パラメータを確率変数と捉えるベイズ法では単純仮説は扱いにくい. なぜなら, 通常パラメータ値は連続変数として扱うが, 連続分布において 1 点の値の確率はその定義上 0 とならざる得ないからである. そこでベイズ法では, θ_0 の周りに幅を持たせて便宜上帰無仮説を区間として扱うという方法も考えられている. しかし, 区間の幅の決め方が恣意的となる欠点がある. 別の方法として, $H_0: \theta = \theta_0$ となるモデルに対する事前確率を ω , $H_1: \theta \neq \theta_0$ であるモデルに対する事前確率を $1 - \omega$ とし, さらに H_1 の下ではパラメータに対して別途事前分布を想定するという方法もある. ただ, ベイズ流の単純仮説の検定には Lindley パラドックス (e.g. Berger 1985) に代表される問題点もある. 単純仮説の検定が目的であれば, 信用区間を求める方が自然であろう.

それでは複合仮説の検定はどうであろうか. いま, 帰無仮説 $H_0 : \theta < \theta_0$ に対して, 対立仮説 $H_1 : \theta \geq \theta_0$ をとる. 事後分布が求まっていれば, これらの仮説の事後確率を積分して求めることにより, それぞれの仮説の妥当性を評価できる. 再び頻度主義的な仮説検定との対比を行うと, 頻度主義では有意水準を定め, これを基準に採択域と棄却域を設ける. ここでも, 有意水準は「帰無仮説が誤って棄却される確率」ではなく, 同様の検定を行った際に帰無仮説が棄却される頻度を表している. そして検出力も同様に解釈される. 一方でベイズ法の場合には, 直接的にそれぞれの仮説の妥当性を確率的に評価することになる. したがって, そこには有意水準や検出力のように言わば非対称な概念は存在しない.

さて, いま例題 XXX の問題において, 2つの仮説 $H_0 : \theta < 0.3$ と $H_1 : \theta \geq 0.3$ の妥当性を評価してみよう. 事後確率 $P(\theta < 0.3|y)$ は `pbeta(0.3, 1+24.9, 10-1+58.1)` より 0.687 と計算できる. これにより仮説 H_0 の方が妥当性が高く, その確率的評価も可能となる. ところで, 事前分布として一様分布型の無情報事前分布を仮定すると, それぞれの確率は最尤推定の結果に大きく引きずられ, H_0 を支持する確率が 0.972 と高くなる. このように, 仮説に対する確率は事前確率の設定に依存する. ただし, 観測値の情報量が多いときにも, ベイズ推測は事前分布に敏感ではなくなり, この場合にも基本的には最尤法から得られる結果と類似するであろう.

例 1.1 [ベイズモデル選択] 例題 XXX のメタ個体群モデルにおいて, 別の分集団からも同様のサンプリングを行ったところ, メス 30 個体のうち 6 個体が妊娠をしていたとする. 分集団間で妊娠率に差があるかどうかをモデルの事後確率を通して検討したい. いま, 確率モデルとして

$$\begin{aligned} \text{Model } M = 1 \text{ のとき} \quad & Y_1 | \theta \sim \text{Bin}(N_1, \theta) \\ & Y_2 | \theta \sim \text{Bin}(N_2, \theta) \\ & \theta \sim \text{Beta}(a, b) \end{aligned} \tag{1.8}$$

$$\begin{aligned} \text{Model } M = 2 \text{ のとき} \quad & Y_1 | \theta_1 \sim \text{Bin}(N_1, \theta_1) \\ & Y_2 | \theta_2 \sim \text{Bin}(N_2, \theta_2) \\ & \theta_1, \theta_2 \sim (\text{iid}) \text{Beta}(a, b) \end{aligned}$$

を考える. またモデルに対する事前確率として

$$P(M = 1) = P(M = 2) = 0.5 \tag{1.9}$$

とする. このとき, それぞれのモデル 1 の事後確率は

$$\begin{aligned} P(M = 1 | Y_1 = y_1, Y_2 = y_2) &= \frac{P(M = 1, Y_1 = y_1, Y_2 = y_2)}{P(Y_1 = y_1, Y_2 = y_2)} \\ &= \frac{f(y_1, y_2 | M = 1) P(M = 1)}{f(y_1, y_2 | M = 1) P(M = 1) + f(y_1, y_2 | M = 2) P(M = 2)} \\ &= \frac{f(y_1, y_2 | M = 1)}{f(y_1, y_2 | M = 1) + f(y_1, y_2 | M = 2)} \end{aligned} \tag{1.10}$$

で定義される。ここで、 $f(y_1, y_2 | M = 1)$ および $f(y_1, y_2 | M = 2)$ はそれぞれモデル 1 とモデル 2 が正しいと仮定したときの観測値 $Y_1 = y_1, Y_2 = y_2$ の周辺分布であり、実質的にはこれらを計算すればモデルの事後確率が計算できる。

$$\begin{aligned}
 f(y_1, y_2 | M = 1) &= \int_0^1 f(y_1 | \theta, M = 1) f(y_2 | \theta, M = 1) \pi(\theta | M = 1) d\theta \\
 &= \frac{1}{B(a, b)} \binom{N_1}{y_1} \binom{N_2}{y_2} \int_0^1 \theta^{y_1 + y_2 + a - 1} (1 - \theta)^{N_1 + N_2 - y_1 - y_2 + b - 1} \\
 &= \binom{N_1}{y_1} \binom{N_2}{y_2} \frac{B(y_1 + y_2 + a, N_1 + N_2 - y_1 - y_2 + b)}{B(a, b)}
 \end{aligned} \tag{1.11}$$

となる。細かい計算は省略するが、同様にしてモデル 2 ($M = 2$) が正しいと仮定するとき、

$$\begin{aligned}
 f(y_1, y_2 | M = 2) &= \int_0^1 \int_0^1 f(y_1 | \theta_1, M = 2) f(y_2 | \theta_2, M = 2) \pi(\theta_1 | M = 2) \pi(\theta_2 | M = 2) d\theta_1 d\theta_2 \\
 &= \binom{N_1}{y_1} \binom{N_2}{y_2} \frac{B(y_1 + a, N_1 - y_1 + b) B(y_2 + a, N_2 - y_2 + b)}{B(a, b)^2}
 \end{aligned} \tag{1.12}$$

となり、したがってモデル $M = 1$ の事後確率が式 (10) により計算できる。実際に例題の値で計算すると、妊娠率が同じとするモデル 1 の事後確率は 0.184 と高くないことがわかる。

```

1 # Comparison of models using Bayes factor
2 N1 <- 10; y1 <- 1
3 N2 <- 30; y2 <- 9
4 a <- 24.9; b <- 58.1
5
6 # Marginal probability given Model 1
7 Marginal_M1 <- beta(y1+y2+a, N1+N2-y1-y2+b)/beta(a, b)
8 Marginal_M1 <- choose(N1, y1)*choose(N2, y2)*Marginal_M1
9 Marginal_M1
10
11 # Marginal probability given Model 2
12 Marginal_M2 <- beta(y1+a, N1-y1+b)*beta(y2+a, N2-y2+b)/beta(a, b)^2
13 Marginal_M2 <- choose(N1, y1)*choose(N2, y2)*Marginal_M2
14 Marginal_M2
15
16 # Posterior probability of Model 1 and 2
17 Posterior_M1 <- Marginal_M1/(Marginal_M1+Marginal_M2)
18 Posterior_M2 <- 1- Posterior_M1
19
20 # Bayes factor
21 BF12 <- Posterior_M1/Posterior_M2
22
23 # Results

```

```

24 Posterior_M1
25 Posterior_M2
26 BF12

```

(Ch2.bayes.modelcomp.r)

最尤法の枠組みではモデル選択で AIC を利用することが多い。AIC ではモデルの複雑さに対するペナルティとして推定するパラメータの数が課される。ベイズ法におけるモデルの事後確率では、表面上はパラメータ数が現れないが、これらは積分する変数として反映されていると考えればよい。ところで、ベイズ法でもモデル選択の手法がいくつか提案されている。そのうちのひとつがベイズ因子 (Bayes factor) とよばれる方法である。

先の例のように異なる 2 つのモデルがあるとき、ベイズ因子は一般に

$$BF = \frac{\text{モデルの事後確率の比}}{\text{モデルの事前確率の比}} \quad (1.13)$$

で定義される。これはモデルの事前確率の比が、観測値によってどの程度更新されるかを測る尺度である。先の妊娠率の例では、モデルの事前確率が等しいから、ベイズ因子は下記のように事後確率の比として計算される。

$$BF_{12} = \frac{P(M=1|y_1, y_2)}{P(M=1|y_1, y_2)} \bigg/ \frac{P(M=1)}{P(M=1)} = \frac{f(y_1, y_2|M=1)}{f(y_1, y_2|M=2)} = 0.226 \quad (1.14)$$

このようなベイズ因子を用いることと、モデルの事後確率を求めることに本質的には差は無く、むしろモデルの事後確率を計算するメリットの方が大きいと思われる。例えば、2 つの異なるモデル $M=1, 2$ が同じパラメータ θ を含んでいるとする。このモデルを与えたときのパラメータの事後分布がそれぞれ $\pi(\theta|y, M=1)$, $\pi(\theta|y, M=2)$ で与えられ、さらにモデルの事後分布 $P(M=1|y)$, $P(M=2|y)$ も同時に求まるとする。このとき、モデルに依存しない θ の事後分布は

$$\pi(\theta|y) = \pi(\theta|y, M=1)P(M=1|y) + \pi(\theta|y, M=2)P(M=2|y) \quad (1.15)$$

として定義される。このような考えに基づく事後分布の構成をモデルアヴェレージング (model averaging) とよぶ (e.g. Hoeting et al. 1999)。モデルの不確実性を考慮した推測法の枠組みは、野生生物の計量的なリスク評価においても実践的なアプローチとなる。なおこの方法の適用例は XXX 章で述べる。

1.3 潜在変数モデル

例題 2.XX ではメタ個体群内の妊娠率を一定と仮定した。しかしながら、妊娠率は分集団の環境や密度の違いによって不均一性を持つかもしれない。いま、各分集団の妊娠率がランダムにある種の確率分布に従うと仮定し、メタ個体群全体として妊娠率の期待および分集団間での妊娠率の確率的変動の程度を記述したいとする。そこで、メタ個体群から K

個の分集団をランダムに選び、さらにそれらの分集団からそれぞれ N_i 個体を選び妊娠の有無を観測する。分集団によっては十分な標本数を確保できないこともあるだろう。メタ個体群全体の情報を利用して、各分集団の妊娠率の推定精度を向上できるであろうか。

この検証のために次のようなモデルを考えよう。

$$\begin{aligned} Y_i | \theta_i &\sim \text{Bin}(N_i, \theta_i) \quad (i = 1, 2, \dots, K) \\ \theta_1, \theta_2, \dots, \theta_K &\sim (\text{iid}) \text{Beta}(a, b) \end{aligned} \quad (1.16)$$

1.8 節で述べたように Y_i の周辺分布はベータ 2 項分布

$$f(y_i | a, b) = \binom{N_i}{y_i} \frac{B(y_i + a - 1, N_i - y_i + b - 1)}{B(a, b)} \quad (1.17)$$

にしたがう。よって、パラメータ a, b の尤度関数は以下の式で与えられる。

$$L(a, b) = \prod_{i=1}^K f(y_i | a, b) = \prod_{i=1}^K \binom{N_i}{y_i} \frac{B(y_i + a - 1, N_i - y_i + b - 1)}{B(a, b)} \quad (1.18)$$

最尤推定値 (\hat{a}, \hat{b}) が求まれば、観測値を与えたときの θ_i ($i = 1, 2, \dots, K$) の条件付分布 (事後分布) がベータ分布 $\text{Beta}(y_i + \hat{a}, N_i - y_i + \hat{b})$ であることを利用し推定することができる。例えば事後平均を用いる場合、

$$\hat{\theta}_i = \frac{y_i + \hat{a}}{N_i + \hat{a} + \hat{b}} \quad (1.19)$$

で推定できる。ここで、 (\hat{a}, \hat{b}) は i 番目の分集団だけでなく他の分集団の妊娠率の観測情報を含んでいる。小さな標本数しか得られない分集団でも、このような “borrowing strength” のコンセプトで推定精度は向上する。

なお、この例では θ_i に対して定義された確率分布は、事前情報の利用というよりも、分集団間の妊娠率の変動を記述するために仮定された。このように、事前分布とは異なった意味でパラメータに確率分布を導入するとき、このパラメータを θ_i を潜在変数 (latent variable) とよぶ。潜在変数モデルの概念は、本来のベイズ推定とは異なるが、形式的にはベイズ推定における事前分布および事前分布の表現と違いがないため、ここでは敢えて事前分布および事後分布とよぶことにする。また、未知であった事前分布のパラメータの推定値を事後分布に代入した。このような手続きによる推定方法を一般に経験的ベイズ法 (empirical Bayes method) という。(理論的側面: Morris らの論文引用?) 経験的ベイズ法の利点は様々な文献で議論されており、しばしば最尤推定法よりも安定した推定結果を示すことが知られている。

このように、生物資源解析の分野では、とくにベイズ法にこだわらず、潜在変数を用いて階層的な確率モデルを構築し最尤法によりパラメータの推測を行うことがある。ランダム効果モデルや混合効果モデルのその典型的な例であり、いくつかの適用例については XXX 章で紹介する。

```

1  # Function for simulation
2  Sim.EB <- function(a,b,K,N,Nit){
3
4  theta.true <- array(0, c(Nit, K))
5  theta.estEB <- array(0, c(Nit, K))
6  theta.estML <- array(0, c(Nit, K))
7  a.est <- numeric(Nit)
8  b.est <- numeric(Nit)
9
10 for(i in 1:Nit){
11   theta.true[i,] <- rbeta(K,a,b)
12   y <- rbinom(K, N, theta.true)
13
14   NLL.BB <- function(par){
15     a <- exp(par[1])
16     b <- exp(par[2])
17     LL <- lgamma(y+a)+lgamma(N-y+b)-lgamma(N+a+b)+lgamma(a+b)-lgamma(a)-lgamma(b)
18     NLL <- -1.0*sum(LL)
19     NLL
20   }
21
22   res <- optim(c(log(a), log(b)), NLL.BB, method="BFGS")
23   a.est[i] <- exp(res$par[1])
24   b.est[i] <- exp(res$par[2])
25   theta.estEB[i,] <- (y+a.est[i])/(N+a.est[i]+b.est[i])
26   theta.estML[i,] <- y/N
27 }
28
29 Out <- as.list(NULL)
30 Out$theta.true <- as.vector(theta.true)
31 Out$theta.estEB <- as.vector(theta.estEB)
32 Out$theta.estML <- as.vector(theta.estML)
33 Out$a.est <- a.est
34 Out$b.est <- b.est
35 Out
36
37 }
38
39 # Do simulation and summary
40 res.sim <- Sim.EB(a=30, b=70, K=10, N=rep(20,10), Nit=100)
41 par(mfrow=c(2,2))
42 hist(res.sim$a.est, xlab="a", main="Estimate of a")
43 hist(res.sim$b.est, xlab="b", main="Estimate of b")
44 plot(res.sim$theta.true, res.sim$theta.estEB, xlim=c(0,1), ylim=c(0,1),
45   xlab="True theta", ylab="EB estimate of theta", main="Empirical Bayes")
46 abline(0,1)
47 plot(res.sim$theta.true, res.sim$theta.estML, xlim=c(0,1), ylim=c(0,1),
48   xlab="True theta", ylab="ML estimate of theta", main="ML estimation")
49 abline(0,1)

```

(Ch2.EB.r)

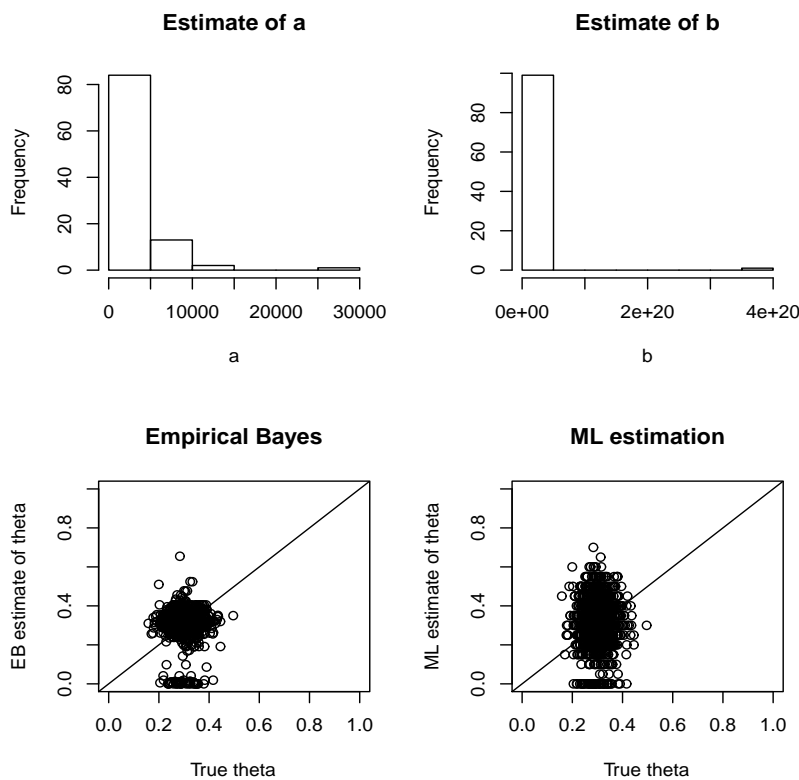


図 1.3: 潜在変数モデルにおける経験ベイズ推定量と最尤推定量の比較 (過小評価のところを直すこと)

1.4 ベイズ法で利用する計算アルゴリズムについての補足

これまで基本的に 1 変数のパラメータを含むモデルに対してベイズ法を述べてきた。1 変数モデルでも共役事前分布以外を仮定すると事後分布が解析的に求まらないことがあるが、実際に計算する上で困ることはないだろう。例えば、パラメータ θ の定義域 (a, b) でパラメータ値を $a < \theta_1 < \theta_2 < \dots < \theta_k < b$ のようにグリッド状に区切るとする。このように離散近似した各点において、尤度 $f(y|\theta)$ と事前分布 $\pi(\theta)$ の値を計算できれば、各点における事後分布は

$$\pi(\theta_i|y) = \frac{f(y|\theta_i)\pi(\theta_i)}{\sum_{j=1}^k f(y|\theta_j)\pi(\theta_j)} \quad (1.20)$$

で近似的に計算できるからである。このような計算方法をグリッド法 (grid method) という。グリッド法を用いれば事後平均も以下のように求められる。

$$E[\theta|y] = \sum_{i=1}^k \theta_i \pi(\theta_i|y) \quad (1.21)$$

このグリッド法は2変数くらいまでなら利用できるが、それ以上のパラメータを伴う場合には大変効率が悪い。なぜならば、パラメータの組み合わせの数が増えるだけでなく、事前確率や尤度値の小さなパラメータも含めて網羅的に計算するため、次元の呪いの影響をまともに受けることになるからである。このような問題を回避するために、数々の数計算アルゴリズムが編み出されてきた。その代表的な方法がマルコフ連鎖モンテカルロ法 (Markov chain Monte Carlo method, MCMC) である (e.g. Gilks et al. 1996, Gamerman 1997, Gelman et al. 2004, Robert and Casella 2004, Givens and Hoeting 2012)。この方法では、マルコフ連鎖として定義される条件付分布を利用して、事後分布に従うパラメータの乱数系列を効率的に生成する。

マルコフ連鎖モンテカルロ法は総称として使われることが多いが、状況によって利用するアルゴリズムが異なる。最も基礎となるのが Metropolis-Hasting アルゴリズムであり、1ステップ前のパラメータ値を所与とした任意の提案分布から次の候補となる乱数を生成し、これを採択するか棄却するかを事後分布の比 (厳密には尤度と事前分布の積の比) で判断する。提案分布は任意であるが、1ステップ前の値から大きく変動する場合には採択率が低くなり効率が悪く、また変動が小さすぎると採択率は高くなるがパラメータ空間内を自由に動けずやはり効率の悪いアルゴリズムとなる。複数のパラメータを含むモデルで、提案分布を厳密な条件付分布とすると、採択率が必ず1となる。このようなアルゴリズムをギブスサンプリング (Gibbs sampling) という。このアルゴリズムは有名な WinBUGS というソフトウェアに実装されており、R2WinBUGS というライブラリーで R から実行および結果の取り纏めが可能となる。

ベイズ推測で利用される計算アルゴリズムはマルコフ連鎖モンテカルロ法だけではない。代表的な代替手段は、重点サンプリング (importance sampling) や **sampling importance-resampling** 法 (SIR) である。SIR 法はパラメータ数が多いときにはしばしば計算の効率が悪いが、比較的簡便にベイズ法による計算を実装することができる。なお、SIR 法を用いた解析例は第7章で解説する。

ベイズ法とは異なり潜在変数モデルでは、事後分布をシミュレートする代わりに、EM アルゴリズムやその拡張型、そして Laplace 近似等を駆使して数値計算を行うことが多い。例えば Laplace 近似を用いる場合には、ADMB-RE という C 言語の実行ファイルを簡単に生成するソフトを用いるのが便利である。Gibbs sampling で WinBUGS を R から呼び出すのと同様に、ADMB-RE の実行ファイルを R から呼び出し計算できる。これについても、第XXX章の解析例で紹介する。

なお、ベイズ法はパラメータ推定やモデルの妥当性評価だけでなく、将来の観測値を予測することとも大変相性が良い。そのような理由で予測を必然的に伴う時系列解析の分野では、ベイズ法が盛んに利用されてきた。この話題は本書の範囲を超えるため触れないが、同シリーズの姜 (2010) を参照されたい。