

第 12 章 ベイズ推定

～主観も活かした確率分布の逐次推定～

名探偵シャーロック・ホームズは、見たこと聞いたことのあらゆる情報を総動員して推理し、難問を解決する。我々も、情報を無駄にしないよう見る目をしっかり持って、チャンスの確率を目いっぱい上げるように努めたい。そのような時、ベイズ推定は、思考回路のエンジンとして我々に力を与えてくれるであろう。主観も取り入れて柔軟な推定ができるベイズ統計は、AI 時代の有力な推理規範としてその重要度が高まっている。その応用は、大規模複雑系に代表されるゲーム理論（数理科学）、社会経済、素粒子物理学など多岐にわたるが、本章では、本書の主題である確率分布の母数（パラメータ）推定に絞り、得られたデータ毎に逐次更新して精度を高めてゆくベイズ推定の基本部分について、その概略を述べる。

12.1 ベイズの定理

ベイズ推定は、数学を趣味とするイギリスの長老派の牧師トーマス・ベイズが 1740 年代に発見した以下の定理に端を発する。二つの事象 A と B があり、図 12.1 のような関係であるとする。その共通部分に $A \cap B$ をもつ。

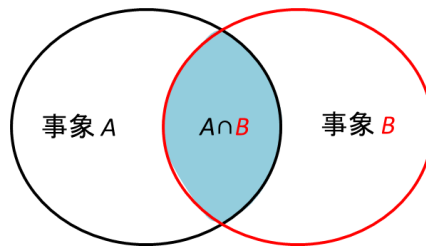


図 12.1 事象 A, B の関係

事象 A, B が起きる確率を $p(A), p(B)$ とし、その両方が同時に起きる確率を $p(A, B)$ とする。さらに、事象 A の中での事象 B の生起確率を $p(B|A)$ 、事象 B の中での事象 A の生起確率を $p(A|B)$ とするとき、これらは条件付確率と呼ばれる。これらの確率は以下の式で結ばれる。

$$p(A, B) = \begin{cases} p(B|A) p(A) \\ p(A|B) p(B) \end{cases} \quad (12.1)$$

上式の右辺同士を等号で結んで、以下のように整理できる。

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} \quad (12.2)$$

この式が**ベイズの公式**、あるいは**ベイズの定理**と呼ばれるものである。

12.2 ベイズ推定

(12.2)式に対して、事象 A に「原因」を与えるもの、事象 B に「結果」を与えるものを紐付けることによって、推定式としての効力を発揮するものになる。原因にも結果にも複数の事象があり、 A_i 、 B_j で表す。結果が B_j であったとき、その原因が A_i である確率 $p(A_i|B_j)$ は、基本式(12.2)より、次式で表される。

$$\begin{aligned} p(A_i|B_j) &= \frac{p(B_j|A_i)p(A_i)}{p(B_j)} \\ &= \frac{p(B_j|A_i)p(A_i)}{\sum_i p(B_j|A_i)p(A_i)} \end{aligned} \quad (12.3)$$

この式は、 $p(A_i)$ を事象 A_i の**事前確率**とみなし、それぞれに対する条件付確率 $p(B_j|A_i)$ がわかれば、 B_j という結果をもたらした後の A_i の確率、すなわち、**事後確率** $p(A_i|B_j)$ が推定できるという式である。この式が、**ベイズ推定式**である。新しい結果（情報）が入るたびに事後確率が更新され、推定精度が上がってゆくという性質を持つ。式中の条件付確率 $p(B_j|A_i)$ は**尤度**とも呼ばれる。大事な式なので、その構造を図 12.2 にまとめて示す。左辺を見たら、右辺がすらすら書けるように慣れておくと良い。

$$p(A_i|B_j) = \frac{p(B_j|A_i)p(A_i)}{\sum_i p(B_j|A_i)p(A_i)}$$

図 12.2 ベイズ推定の基本式

大変有用と期待される統計的推定法であるが、正統な統計学者達（R.A.フィッシャーやJ.ネイマンがその代表）からは、事前確率の定め方にあいまい性や恣意性が入る余地があるので、それが嫌われて日陰に追いやられてきた長い歴史がある。（このあたりの歴史的事情については[1]に詳しい）。一方で、第2次世界大戦中、英国のアラン・チューリングはベイズ推定によりナチス・

ドイツ軍の暗号エニグマを解いたと言われている。限られた情報からでも不完全な情報からでも（すなわち事前確率があいまいでも）、新たな情報が入るたびに、事後確率の再計算を続けてゆけば、十分満足できる判断（すなわち、最善でなくてもよい）を迅速に導き出せるのがベイズ推定である。カーナビが常に最適な道を教えてくれるとは限らないが、言うことを聞いていれば、ちゃんと目的地にたどり着く。それでよいのである。最善に近い答えを迅速に導き出してくれる推定法は、完璧な答えを時間を掛けて導き出す推定法よりも実用と言う意味で優れていることが多い。ベイズ推定はそのような性格を持つ推定法である。

ここまでの理解で応用できる例として、付録に紹介する**モンティ・ホール問題**を挙げたい。

12.3 確率分布母数のベイズ推定

12.3.1 確率分布と尤度関数

確率分布で与えられる**母集団**があり、そこから、無作為に取り出した複数の**標本値**から母集団の確率分布の**パラメータ値（母数）**を推定したい、という問題を考える。例えば、正規分布ならば、

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (12.4)$$

に対して、標本値（データ） $x_1, x_2, \dots, x_N$ からパラメータ μ （平均値）と σ^2 （分散）を推定したいという問題である。既に学んでいるように、この種の王道は**最尤推定**であり、「そのような結果（標本値）が得られたのは、そのような結果を生み出す確率がもっと高い確率分布であったから」という考えに従う。確率分布を表す式の変数とパラメータの関係を逆転させ、パラメータを変数と見立てて、最大確率を与えるパラメータ値を求めるのである。(12.4)式の場合には

$$L(\mu, \sigma^2) = f(x_1|\mu, \sigma^2) \times f(x_2|\mu, \sigma^2) \times \dots \times f(x_N|\mu, \sigma^2) = \prod_{n=1}^N f(x_n|\mu, \sigma^2) \quad (12.5)$$

を求め、この値を最大にする μ と σ^2 を見出すことになる。 L は**尤度関数**と呼ばれる。正規分布の場合は、その値（**最尤推定値**： $\hat{\mu}, \hat{\sigma}^2$ ）は解析的に解けて、 $\hat{\mu}$ は標本平均、 $\hat{\sigma}^2$ は標本分散になる（ $\hat{\sigma}^2$ が分散推定で広く用いられる不偏分散でなく、標本分散になる理由は、9章8頁のティータム欄を見てほしい）。

母集団が正規分布の場合は解析的に解けるが、このようなケースは限られていて、一般的には数値計算になる。以下、一般の場合で整理する。

母集団の確率分布（確率密度関数）： $f(x|\theta)$

母数（パラメータ）： $\theta \equiv (\theta_1 \ \theta_2 \ \dots \ \theta_M)$ （ M : パラメータ数）

標本値（データ）： $\mathbf{x} \equiv (x_1 \ x_2 \ \dots \ x_N)$ （ N : データ数）

尤度関数 : $L(\boldsymbol{\theta} | \mathbf{x}) = \prod_{n=1}^N f(x_n | \boldsymbol{\theta})$

最尤推定値 : 尤度関数を最大にする $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$

12.3.2 事前確率分布と事後確率分布

ここからがベイズ推定の本丸。ベイズ推定によって確率分布のパラメータ値を推定してゆく方法について述べる。

ベイズ推定の基本式(12.3)では、事前確率から事後確率を求める構造になっているが、これを
事前確率 → 母数 (パラメータ) の **事前確率分布** : $\pi(\boldsymbol{\theta})$

事後確率 → 母数の **事後確率分布** : $\pi(\boldsymbol{\theta} | \mathbf{x})$

に置き換える。定めたい本来の確率分布 (母集団の確率分布) ではなく、母集団のパラメータに対する確率分布と言うところに大きな違いがある。これをベイズ推定の式で表すと次式である。

$$\pi(\boldsymbol{\theta} | \mathbf{x}) = \frac{L(\mathbf{x} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta})}{P(\mathbf{x})} \quad \left(P(\mathbf{x}) = \int_{\boldsymbol{\theta}} L(\mathbf{x} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \right) \quad (12.6)$$

上式右辺の分母の P は、母数の数だけの多重積分になるが得られた結果は定数であり、**周辺尤度**あるいは**ベイズ因子**と呼ばれる。左辺の事後確率分布は、確率分布の性質より、パラメータの全範囲で積分すれば1になる。周辺尤度 P はそうなるような比例定数を担う意味を持ち、**正規化ファクター**とも呼ばれる。(12.6)式は母数に対する確率分布のベイズ推定において大事な式なので、その構造を図12.3にまとめて示す。

$$\pi(\boldsymbol{\theta} | \mathbf{x}) = \frac{L(\mathbf{x} | \boldsymbol{\theta}) \times \pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} L(\mathbf{x} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}}$$

図 12.3 確率分布の母数 (パラメータ) の値をベイズ推定するときの基本式

実際には、母集団のパラメータは定数であって、分布するようなものではない。故に、パラメータの分布を表す π は、母数の代表値 (最尤推定値、あるいは平均値など) を求めたいためだけに使われる便宜的なものである。故に、その分布形に拘る必要は無く、うまく代表値が得られるようなものなら何でもよいということになる。しかし実際にパラメータの分布形を決める際には、その分布を規定するパラメータ、すなわち、母数 (パラメータ) の確率分布を決めるパラメ

ータが求められる。正規分布を仮定して追い込みたいと思う場合には、その平均値と分散である。ここで言うパラメータは、母集団の分布のパラメータとは何の関係もないもので、勝手に決めているものと言うことになる。故に、このパラメータ（＝パラメータのパラメータ）は母集団のパラメータと区別して、**ハイパーパラメータ**と呼ばれる。

ベイズ推定の基本式(12.6)から次の性質が読み取れる

- ① 事後確率分布は事前確率分布と尤度関数の積に比例する
- ② 事前確率分布が一様分布の場合（＝事前情報が何もないときに採用）、事後確率分布は尤度関数に比例する（＝分布形状が同じ）
- ③ 周辺尤度は事後確率分布が確率分布としての必要条件を満たすための正規化ファクタである

では、実際に、具体的な母集団の確率分布について、次節で調べてみよう。

12.4 自然共役事前分布と母数代表値の推定

12.4.1 自然共役事前分布とは

(12.6)式において、事前確率分布と事後確率分布の分布形は、パラメータの値を追い込む目的のためだけにあるので、同じであっても異なっても問題ない。しかし、数式で解析したい場合には、事前と事後の確率分布が同じであると見通しが良い。このように、事前と事後の確率分布が同じ分布形で表される分布は**自然共役事前分布**（略して共役事前分布）と呼ばれる。

自然共役事前分布であるためには、前節最後にまとめた①の性質「事後確率分布は事前確率分布と尤度関数の積に比例」に縛られるので、うまく共役事前分布が見つけれられるかどうかは母集団の確率分布の形による。実はそのようなものは、非常に少なく、表 12.1 にその例を示す。

表 12.1 自然共役事前分布の例

母集団の確率分布 (尤度関数)	自然共役事前分布 (事後確率分布も同じ)
ベルヌイ分布 2項分布 ポアソン分布 正規分布(分散既知) 正規分布(分散未知)	ベータ分布 ベータ分布 ガンマ分布 正規分布 逆ガンマ分布

12.4.2 母数の点推定と確信区間

(12.6)式によって、母数の事後確率分布が得られたとしよう。前項でも述べたように、事後確率分布の分布形そのものには意味が無く、母数そのものの値が知りたいのである。それには、次の二つの量が代表的である。値を一つの点で代表して求めるので**点推定**である。

事後期待値 (expected a posteriori, EAP)：事後分布の期待値を推定値とする

$$\theta_{i,EAP} = E[\theta_i | \mathbf{x}] = \int_{\theta_i} \theta_i \iint_{\substack{\theta=\theta_1 \sim \theta_M \\ \theta \neq \theta_i}} \cdots \int \pi(\boldsymbol{\theta} | \mathbf{x}) d\theta_M \cdots d\theta_1 \quad (12.7)$$

事後確率最大値 (maximum a posteriori, MAP) : 事後分布の最大値を推定値とする

$$\hat{\boldsymbol{\theta}}_{MAP} = \max_{\boldsymbol{\theta}} \pi(\boldsymbol{\theta} | \mathbf{x}) \quad (12.8)$$

事後確率分布が凸型の形状であれば、MAP は分布の最頻値（モード）になる。

伝統的な統計的推定（頻度論）では、母数の最尤推定値は尤度関数の最大値を与えるものに定めるが、ベイズ推定の事後確率最大値（MAP）は尤度関数と事前確率分布の積の分布の最大値である。事前確率分布を考慮するか否かの違いであり、これがベイズ推定の特徴にある。事前情報がない場合は、事前確率分布を一様分布に採ることが一般的であり、その場合は MAP と最尤推定値は同じになる。

従来の統計的手法では、最尤推定値を点推定する方法に加え、信頼区間幅で推定する **区間推定** 法がある。確率水準 p を 90%, 95%, 99% などに採り、その範囲の中にある推定値を排除しないと言う推定法である（本書の 10 章参照）。式(12.6)によるベイズ推定においても、母数の事後確率分布が得られるので、累積確率 50% を中心とする $p\%$ の幅を **確信区間** として、EAP や MAP の確信性を求めることができる。用語の確信区間 (credible interval) は信頼区間 (confidence interval) と意味合いは近いのであるが、定義や考え方に違いがあり、区別して呼ばれている。

12.4.3 母集団の確率分布が2項分布の場合

表 12.1 に示したように、事前確率分布と事後確率分布が同じ分布形を使うことができる母集団分布（あるいは尤度関数）は限られるのであるが、それができる二つのケースを本項と次項で述べる。本項では、母集団の分布が2項分布の場合について説明する。同表より、2項分布に対応する自然共役事前分布はベータ分布である。2項分布はゲームの勝敗から勝率を求めたいようなときに現れる。ベータ分布は確率変数が 0~1 の範囲にある様々な形状を表すのに便利な分布である。最初に、この二つの分布を簡単に説明する。

$$\text{2項分布: } f(x | \theta) = {}_N C_x \theta^x (1-\theta)^{N-x}, \quad x=0, 1, \dots, N \quad (12.9)$$

$${}_N C_x = \frac{N!}{x! (N-x)!}$$

確率変数 $X=x$ は 0 から N までの整数で、勝率を θ とするとき、 N 回の試合において x 回勝つ確率を求めるときなどに現れる分布である。その期待値と分散は、

$$E[X] = N\theta$$

$$V[X] = N\theta(1-\theta)$$

となる。

$$\text{ベータ分布: } f(x|p, q) = \frac{1}{B(p, q)} x^{p-1} (1-x)^{q-1}, \quad 0 \leq x \leq 1 \quad (12.10)$$

$$\begin{aligned} \text{ベータ関数 } B(p, q) &= \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \quad (\Gamma: \text{ガンマ関数}; p, q \text{ が正の実数のとき}) \\ &= \frac{(p-1)!(q-1)!}{(p+q-1)!} \quad (p, q \text{ が正の整数のとき}) \end{aligned}$$

$$E[X] = \frac{p}{p+q}$$

$$V[X] = \frac{pq}{(p+q)^2(p+q+1)}$$

$$\text{最頻値 (モード)} \quad \frac{p-1}{p+q-2} \quad (p, q > 1)$$

ベータ分布は変数の範囲が0～1の範囲で、パラメータ p と q の値の設定により、様々な分布の形状（一様、凸型、右凸、左凸、凹型など）を実現できることに特徴がある。図 12.4 はその一例を示している。物理的な意味を問われることがない事前確率分布を表すには便利な分布である。

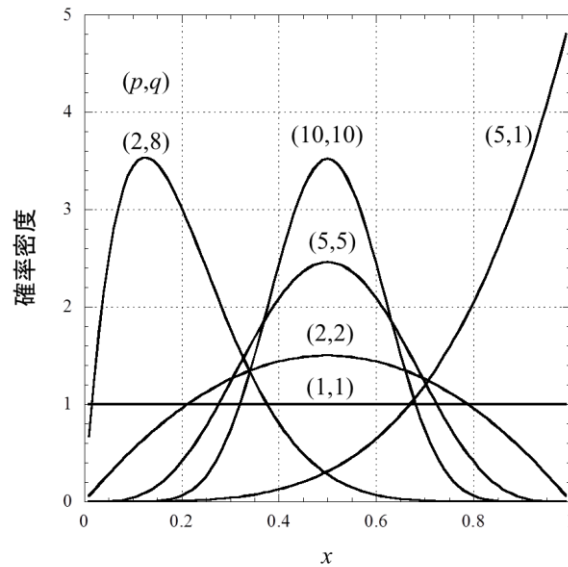


図 12.4 ベータ分布の形状

確率分布のカーネルと事後確率分布

確率分布のカーネルに着目する。**カーネル**（核）とは、確率分布の内の変数が関数になっている部分（＝定数係数を除いた部分）を言い、確率分布の形状を定める本質的な部分を言う（正規分布であれば $\exp(\cdots)$ の部分）。本項で扱う尤度関数が2項分布のときの、事後確率分布のカーネル部分は、尤度関数（2項分布）に対して確率変数： $X=k$ 、事前確率分布（ベータ分布）に対

して $X=\theta$ とすると、

$$\begin{aligned}\pi(\theta|k) &\propto \theta^k (1-\theta)^{N-k} \times \theta^{p-1} (1-\theta)^{q-1} \\ &= \theta^{k+p-1} (1-\theta)^{N-k+q-1}\end{aligned}\quad (12.11)$$

となり、事後確率分布は、ハイパーパラメータが $p \rightarrow k+p$ 、 $q \rightarrow N-k+q$ となって、ベータ分布は事後確率分布に対しても維持され、自然共役事前分布であることが確認できる。これにより、事後確率は、新たに得られた情報を D として、次式で定められる。

$$\pi(\theta|D) = \frac{1}{B(p', q')} \theta^{p'-1} (1-\theta)^{q'-1} \quad (p' = k+p, q' = N-k+q) \quad (12.12)$$

N, k に関する新たな情報が得られた場合には、(12.12)式の事後確率分布を次の段階の事前確率分布にして、ベイズ推定により順次更新して行けばよい。

以上は式だけで説明したので、具体的なイメージが湧きにくいと思う。以下ではその一例を示す。

具体例

太郎と二郎は将棋が好きで、時々二人で将棋を指している。三郎は、以前に、その対戦を 10 回くらい見たことがあり、7 割方太郎が勝っていた印象を持っている。最近の結果を太郎に聞いたところ 4 勝 3 敗だったと言う。では、太郎の実力（勝率）はどのくらいだろう。

設定の前半部分をわざと曖昧にしているのは、それでもベイズ推定なら答えを出せると言いたためであり、そのようなあいまい設定が、伝統的な統計学者からベイズ推定が嫌われた理由でもある。母数（この場合は勝率）の事前確率分布をベータ分布で表すときのパラメータ（この場合はハイパーパラメータ）の設定には、丁寧な説明が必要であるが、ここでは、以下のようなものと受け入れてほしい。

データ数（試合数）： $p+q-2$ 、勝ち数： $p-1$ 、負け数： $q-1$ （ p : 勝ち数+1、 q : 負け数+1）

ゆえに。この場合の事前確率分布のハイパーパラメータは、あいまいな情報と知りつつも、事前の知識を 7 勝 3 敗だったとみて、 $p=8, q=4$ としてみよう。そうすると事後分布では、 $p'=8+4=12$ 、 $q'=4+3=7$ となる。勝率 θ の期待値（EAP）は $p'/(p'+q')=12/19 \approx 0.63$ 、事後確率最大値（MAP）は $(p'-1)/(p'+q'-2)=11/17 \approx 0.65$ となる。このようにして新たな情報に基づく勝率推定ができる。事前確率分布の情報があいまいであっても、新たなデータが加わるたびに、事後確率は本来の母数の値に近づいてくるのがベイズ推定の特徴である。また、あいまい情報とした事前確率部分が、確定情報であった場合も母数推定は同じ結果になる（これなら伝統的統計学者からの文句も出ない）。さらに、全く事前情報がない場合は、事前確率分布は $p=q=1$ のベータ分布、すなわち、一様分布になり、尤度関数そのものが事後確率分布になる。

12.4.4 母集団の確率分布が正規分布で母分散が既知の場合

自然共役事前分布が存在するもう一つの例、尤度関数が正規分布（平均値が未知で分散が既知の場合）の場合についてみてゆきたい。正規分布は、基本中の基本の分布であるが、母数（パラメータ）が二つあるので結構複雑である。故に簡易な例として、母分散が既知の場合を示す。母平均も母分散も未知の場合には、分散推定のために表 12.1 に示した逆ガンマ分布が自然共役事前分布になる。その場合の求め方も、以下で示す分散が既知の場合と同じ考え方で進めればよいのであるが、式展開が煩雑になる（正規分布と逆ガンマ分布の積の分布が自然共役事前分布になる）。ゆえに、分散が未知の場合については[2]（の 6.5 節及び付録 F）に譲り、ここでは、母集団が正規分布で、かつ、その母分散が既知の場合についての母平均推定の求め方を示す。（平均が未知で分散が既知という設定は現実的ではないが、求め方を学ぶ例題だと思ってほしい）

問題設定

- 1) 母集団の確率変数は正規分布し、平均 μ 、分散 σ^2 である
- 2) 母分散 σ^2 は既知である
- 3) 事前情報として（＝それまでの測定、あるいは主観情報によって）、母集団の平均は μ_0 、平均 μ_0 の不確かさを表す分散が b_0 に絞り込まれている。（この場合の分散は母集団の分散（母分散）ではなく、平均値 μ を推定する際のばらつきの大きさ（ハイパーパラメータ）であり、母分散 σ^2 と区別するために、別の文字 b を使っている）
- 4) N 個の標本を抽出し、標本平均 $\bar{x}$ （標本分散 S^2 ）が得られたとする。

母平均 μ の自然共役事前分布（正規分布）のカーネル $e^{-(\mu-\mu_0)^2/b_0}$

尤度関数

$$L(\mu, \sigma^2 | \mathbf{x}) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^N e^{-\frac{1}{2\sigma^2} \sum_{n=1}^N (x_n - \mu)^2} = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^N e^{-\frac{N(\mu - \bar{x})^2 + NS^2}{2\sigma^2}}$$

尤度関数のカーネル $e^{-\frac{N(\mu - \bar{x})^2}{2\sigma^2}}$

事後確率分布（正規分布）のカーネル $e^{-\frac{1}{2} \left(\frac{N(\mu - \bar{x})^2}{\sigma^2} + \frac{(\mu - \mu_0)^2}{b_0} \right)}$

上式カーネルの指数部の（ ）の部分を整理すると

$$\begin{aligned} & \frac{N(\mu - \bar{x})^2}{\sigma^2} + \frac{(\mu - \mu_0)^2}{b_0} \\ &= \frac{(\sigma^2 + Nb_0)\mu^2 - 2(\sigma^2\mu_0 + Nb_0\bar{x})\mu + \sigma^2\mu_0^2 + Nb_0\bar{x}^2}{b_0\sigma^2} \end{aligned}$$

$$\begin{aligned}
&= \frac{\left(\mu - \frac{\sigma^2 \mu_0 + Nb_0 \bar{x}}{\sigma^2 + Nb_0} \right)^2 + \frac{\sigma^2 \mu_0^2 + Nb_0 \bar{x}^2}{\sigma^2 + Nb_0} - \left(\frac{\sigma^2 \mu_0 + Nb_0 \bar{x}}{\sigma^2 + Nb_0} \right)^2}{\frac{b_0 \sigma^2}{\sigma^2 + Nb_0}} \\
&= \frac{\left(\mu - \frac{\sigma^2 \mu_0 + Nb_0 \bar{x}}{\sigma^2 + Nb_0} \right)^2}{\frac{b_0 \sigma^2}{\sigma^2 + Nb_0}} + \text{定数}
\end{aligned}$$

であるので、母平均 μ の事後確率分布は

$$\begin{aligned}
\pi(\mu | D) &= \frac{1}{\sqrt{2\pi b_1}} e^{-\frac{(\mu - \mu_1)^2}{2b_1}} \quad (12.13) \\
\mu_1 &= \frac{\sigma^2 \mu_0 + Nb_0 \bar{x}}{\sigma^2 + Nb_0}, \quad b_1 = \frac{b_0 \sigma^2}{\sigma^2 + Nb_0}
\end{aligned}$$

となり、新たな情報によって (μ_0, b_0) が (μ_1, b_1) に更新される。

12.5 多重積分の数値解法：ベイズ推定を強力なものにする MCMC 法

自然共役事前分布を用いて数式的に解くのは、表 12.1 に可能な組み合わせを挙げているように、それができるケースが極めて限られている。ベイズ推定の思想から言えば、母数の確率分布が逐次追い込めればよいのであって、事前確率分布と事後確率分布が同じ分布形である必要は全くないのであるが、解析的に求めようとすれば、そのような縛りが出る。12.4 節で挙げた 2 項分布や正規分布の母数推定では、事前に有する主観的な感覚に新たに得られた情報を加えて考えれば、ある程度の見込みがつくものであり、ベイズ推定が有用な手段であるということの実感が薄いのではないだろうか。

ベイズ推定が真に有効となる土俵は、母数（パラメータ）の数が多く、分布の形も複雑なものに対して、その都度得られる情報をもとに、母数の推定値を逐次更新したいような場合にある。この場合でも基本式(12.6)に従って更新して行けばよいのであるが、パラメータの数が増えてくると積分計算が困難になる。ベイズ推定において多重積分が現れるのは、次の 2 点においてである。

- 1) (12.6)式右辺分母の周辺尤度計算
- 2) 事後確率分布から期待値を求める演算 (式(12.7))

数値演算によって多重積分をまともにやろうとすると、積分計算に寄与が少ない分布の端の方を含めて無駄な演算が多く、多重度が高い場合、計算時間を著しく増やすことになる（スーパーコンピュータでも手に負えなくなる場合が出てくる）。そこで、乱数を用いて、分布の高確率

部分を効率的にサンプリングして少ない演算数で、積分を収束させる方法がある。一言で言えば、母集団の確率分布に近い乱数を生成する方法である。

その考えに立って編み出された乱数利用積分法が、モンテカルロシミュレーションにマルコフ連鎖を採り入れた Markov Chain Monte Carlo 法 (MCMC 法) である。MCMC 法についての基本的な考え方と数値積分アルゴリズムについては、それ自体が独立した学問分野になっているので、その専門書 (例えば、[1], [3]) で学んでほしい。

もちろん、MCMC がなんでもできる打ち出の小槌ではない。利用者自身がその基礎を身に着け、独自の工夫を重ねて利用する結果として目的とする果実を手に入れることができる。数値解析手段と言うと無味乾燥なイメージを持つと思うが、その応用の先に広がる豊かな世界は[3]の第7章を読めば感動をもって味わうことができると思う。

12.6 蛇足

ここではベイズ推定を取り上げたが、言いたかったことは次の3点である。

- ① 物事の判断において、得られた情報を無駄にしないこと (例えそれが曖昧なものであっても)
- ② 新たな情報によって、事前確率が事後確率に変わる環境適応性があること
- ③ 一旦定式化できれば、基本式に従って機械的に確率を導き出せるようにアルゴリズム化されていること

本章は文献[1]~[3]を読んで学んだことを筆者流にまとめている。これまで、土俵としてきた電波伝搬や無線システムの研究に用いた経験がないので、この分野での具体的な応用例を示すことができなかったが、それは、次世代を担う皆さんに託したい。AI が情報技術分野を席卷する時代には、ベイズ推定が有力な武器になるだろうと強く確信する。

付録 モンティ・ホール問題をベイズ推定する

(1) モンティ・ホール問題

アメリカのテレビ番組に、モンティ・ホールという人が司会するゲームショーがあったと言う[4]。ルールはこうである。解答者の目の前に三つのドア (ここでは便宜上①、②、③とする) がある (図 12.5)。このドアの部屋のどれか一つに宝物が入っていて、残りは外れ。解答者が当てたらその宝物がもらえる。解答者は、どれか一つの部屋を選ぶように言われる。選んだ後、答えを知っている司会者が「君にチャンス을あげよう」といって、選んでいない二つの部屋のうち外れの部屋のドアを一つ開けてくれる。この状態で、「もう一度よく考えて選んでください」と言う。仮に解答者が最初に①を選んだときは、外れである②か③のどちらかを開けてくれるのである。さあ、どうする。これがゲームの全てである。解答者を悩ますゲームの性格から、モンティ・ホール・ジレンマとも呼ばれている。

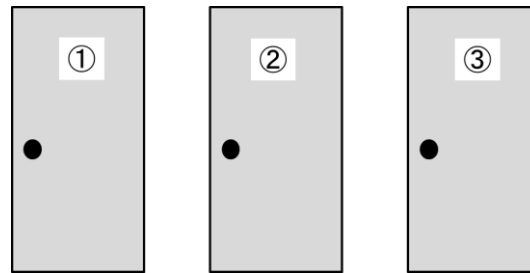


図 12.5 モンティ・ホール問題 (三つの部屋の一つに宝物が入っている。開けるチャンスは一回)

何も情報がないときは当たる確率は $1/3$ 。①を選んだとき、②を開けてくれたとしたら、選ぶべきドアは①か③の二つに絞られた。チャンスの確率を司会者が $1/3$ から $1/2$ にしてくれたように感じる。解答者は、頭の中でコイン投げをして、どちらかもう一度決めようと思う。でも、いや待てよ、確率が $1/2$ ならコイン投げしなくても、最初に決めたドアを開ければよいだけではないかと気が付き、そのまま①を開けた。この判断は正しいのだろうか。

この問題は、気軽に楽しめる数学ゲームとして、いろいろなバリエーションがあり (例えば、3 囚人問題[5]など)、モンティ・ホール問題もその一つである。正解は、もう一つのドア (②が開けられれば③に) に持ち替える。それにより、宝物を得る確率が $2/3$ になるのである。組み合わせを総ざらいし、その一つ一つについて当たり・外れを調べれば、そういうことになる (その確認は読者にお任せ)。 $1/2$ ではなかったと言うところに意外性があり、頭の体操として楽しめるのである。

最も直感的な理解には以下の説明がよいであろう。最初に①を選んだとき、その当たり確率は $1/3$ 。残りの②と③あわせて考えれば、そちらで当たる確率は $2/3$ 。②と③のうち、一方にはないよと教えてくれたのであるから、当たり確率 $2/3$ がそのまま残りのドア (②が開けられたとしたら③) に移ったと考えれば良い。この問題は、普通の確率計算で解け、ベイズ推定でなければ解けないと言うわけではない。そのことを知った上で、次節のベイズ推定を見てほしい。解法の特徴がよく見えるのである。

以下、雑談。筆者はモンティ・ホール・ジレンマの話の数学者エルデシュの伝記物語[6]の中で知った。本には次のような話が紹介されている。世界一の IQ (228) とギネスに認められたマリリン・フォス・セイブアントというマダムが週刊誌に人気コラム (マリリンにお任せ) を連載していた。そこに、読者からこのモンティ・ホール問題にどうすれば良いかとの質問が来た。マリリンは、「持ち替えなさい、そうすれば確率は2倍に上がる」と答え、「ドアが百万個あると想像してみなさい。司会者が残り 999,999 個のドアのうち、一つを残してみんな開けてくれたら、あなたは一瞬にして考えを変えるでしょう」。さらに、「司会者がドアを開けてはずれを見せたばかりのところへ UFO に乗った宇宙人が現れ、その何も知らない宇宙人にどちらかを選べと言えば $1/2$ でしょう。あなたはそれ以上のものを見ているのではないですか」と。この記事を見た世の数学者達との論争が面白いのであるが、それは、本でのお楽しみということで。(この話題は[4]にも)。

(2) ベイズ推定する

原因となる事象 A と事後の結果を定める事象 B の対応付けがうまくでき、かつ原因と結果の条件付確率である尤度が定められれば、求める確率は機械的に求まるのがベイズ推定の特徴である。

解答者が最初にドア①を指定したところからはじめよう。それぞれの事象は以下のように分類できる。

原因となる事象

A_1 : ドア①の部屋に宝物がある

A_2 : ドア②の部屋に宝物がある

A_3 : ドア③の部屋に宝物がある

解答者が①を選んだ後に行われる動作

B_2 : 司会者がドア②を開ける

B_3 : 司会者がドア③を開ける

そうすると、知りたい事後確率は

持ち替えない場合

$p(A_1|B_2)$: 司会者がドア②を開けたときにドア①の部屋に宝物がある確率

$p(A_1|B_3)$: 司会者がドア③を開けたときにドア①の部屋に宝物がある確率

持ち替える場合

$p(A_2|B_3)$: 司会者がドア③を開けたときにドア②の部屋に宝物がある確率

$p(A_3|B_2)$: 司会者がドア②を開けたときにドア③の部屋に宝物がある確率

事前確率は $p(A_1)=p(A_2)=p(A_3)=1/3$ である (こう決めつけて良いかどうかはわからないが、ここではそうであるとしよう)。条件付確率 (尤度) $p(B_j|A_i)$ ($i=1,2,3; j=2,3$) は、司会者に特別な意図がない限り表 12.2 になろう。(別の可能性もあることを (3) 項で述べる)。これらの条件により、上記 4 つの事後確率は式(12.3)により機械的に計算でき、それぞれ次の値になる。

表 12.2 条件付確率 (尤度) $p(B_j|A_i)$ の値

	B_2	B_3
A_1	1/2	1/2
A_2	0	1
A_3	1	0

持ち替えない場合

$$p(A_1|B_2) = \frac{p(B_2|A_1)p(A_1)}{\sum_i p(B_2|A_i)p(A_i)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3}} = \frac{1}{3}$$

$$p(A_1|B_3) = \frac{p(B_3|A_1)p(A_1)}{\sum_i p(B_3|A_i)p(A_i)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} = \frac{1}{3}$$

持ち替える場合

$$p(A_2|B_3) = \frac{p(B_3|A_2)p(A_2)}{\sum_i p(B_3|A_i)p(A_i)} = \frac{1 \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} = \frac{2}{3}$$

$$p(A_3|B_2) = \frac{p(B_2|A_3)p(A_3)}{\sum_i p(B_2|A_i)p(A_i)} = \frac{1 \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3}} = \frac{2}{3}$$

これより、持ち替えない場合はチャンスの確率 1/3 のままであるが、持ち替えることによって 2/3 に上げることができるのである。複雑な計算をしているように見えるかもしれないが、事後確率の式（上式の左辺）を適切に表すことが出来さえすれば、右辺は形の決まった式を手順通りに計算してゆけばよく、頭を悩ますところがない。アルゴリズムに従った機械的な計算だと思えばよく、それがベイズ推定の特徴である。

(3) さらにもう一押し

トムはこのクイズ番組が好きで、毎回欠かさず見ている。そしてあることに気がついた。司会者は、(解答者が①を指定した後) ②か③を開けるのに関して、②が外れであればいつも②を、②に宝物がある時だけ③を選ぶようにしているということ。司会者にとっての判断の面倒をなくす手抜きアルゴリズムである。これによって、何がどう変わるのだろうか。

ルールが変わったわけではないので、基本推定式は(12.3)式で同じである。違いは、表 12.2 に示した条件付確率（尤度）が表 12.3 に変わったことである。

表 12.3 条件付確率 (尤度) $p(B_i|A_i)$ の値 (司会者に選択の癖がある場合)

	B_2	B_3
A_1	1	0
A_2	0	1
A_3	1	0

最初に①を指定してから後のことを、表 A2 を使って前節と同じ計算をしてみよう。

持ち替えない場合

$$p(A_1|B_2) = \frac{p(B_2|A_1)p(A_1)}{\sum_i p(B_2|A_i)p(A_i)} = \frac{1 \times \frac{1}{3}}{1 \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3}} = \frac{1}{2}$$

$$p(A_1|B_3) = \frac{p(B_3|A_1)p(A_1)}{\sum_i p(B_3|A_i)p(A_i)} = \frac{0 \times \frac{1}{3}}{0 \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} = 0$$

持ち替える場合

$$p(A_2|B_3) = \frac{p(B_3|A_2)p(A_2)}{\sum_i p(B_3|A_i)p(A_i)} = \frac{1 \times \frac{1}{3}}{0 \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} = 1$$

$$p(A_3|B_2) = \frac{p(B_2|A_3)p(A_3)}{\sum_i p(B_2|A_i)p(A_i)} = \frac{1 \times \frac{1}{3}}{1 \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3}} = \frac{1}{2}$$

ドアを持ち替えた場合、事後確率はドア②が開けられたとき $p(A_3|B_2)=1/2$ 、ドア③だと $p(A_2|B_3)=1$ となった。やった、平均 $3/4$ になった、と喜ぶのは早合点。なぜなら、 $p(A_1)=p(A_2)=p(A_3)=1/3$ が変わらず、そんなことになるはずはないからである。注意深く見れば、司会者によってドア②が開けられる確率は $2/3$ (表 A2 より)、③は $1/3$ であるので、その比率で加重平均すると、

$$\text{当たり確率 (ドア替えない)} = p(A_1|B_2)p(B_2) + p(A_1|B_3)p(B_3) = \frac{1}{2} \times \frac{2}{3} + 0 \times \frac{1}{3} = \frac{1}{3}$$

$$\text{当たり確率 (ドア替える)} = p(A_2|B_3)p(B_3) + p(A_3|B_2)p(B_2) = 1 \times \frac{1}{3} + \frac{1}{2} \times \frac{2}{3} = \frac{2}{3}$$

となる。結局、トムが手にした新たな情報では当たり確率を変えるとこまでは行かなかった。司会者が③を開けた瞬間に大喜びできる、というのが新たな情報の価値であり、残念ながらトムが気づいたことの恩恵は小さかった（＝結果に影響しない）のである。

最近、解答者にも知恵がついて、持ち替える方ばかり選ぶようになってきた。それを、トムがさらに注意深く見ていると、大部分の人は最初に①を選ぶのだが、結果は意外に半々であることがわかってきた。トムがクイズの解答者になった時、どうすればよいであろう。（司会者の癖は変わってないとして）

参考文献

- [1] 豊田秀樹（編著）、基礎からのベイズ統計学、朝倉書店、2012.
- [2] 涌井良幸、涌井貞美、身につくベイズ統計学、技術評論社、2016.
- [3] 花田政範、松浦壮、ゼロからできる MCMC：マルコフ連鎖モンテカルロ法の実践的入門、講談社、2020.
- [4] ウィキペディア、[モンティ・ホール問題](#)
- [5] ウィキペディア、[3 囚人問題](#)
- [6] ポール・ホフマン（平石律子訳）、放浪の天才数学者エルデシュ、草思社、2000.

目次のページにもどる