Chapter 5. Bayesian Prediction 後半

5.4 DERIVED PARAMETERS AND OUT OF SAMPLE INFERENCE IN A DOSE-RESPONSE STUDY

5.4.1　5.4.2

2003 年 1 月に池で死亡したカナダガン(*Branta canadensis*) の肝臓と腎臓には高濃度の $NaVO_3$ が検出された。哺乳類では $NaVO_3$ の生体への影響が調べられているが鳥類では調べられていない。
カモに様々な濃度の $NaVO_3$ にさらす実験を行い、死亡と $NaVO_3$ の関係について調べた。

マガモ Mallard drakes (*Anas platyrhynchos*)
Treatment (32 + 4 個体):
10, 18, 34, 62, 113, 208, 382, 700 ppm の $NaVO_3$ にさらした（各 4 個体を 7 日間）。
Sham（4 個体）
15 個体で肝臓と腎臓の $NaVO_3$ 量を調べた。
全個体に死亡・生存の binary response データ、15 個体に $NaVO_3$ 量の bivariate continuous response データ

このような Dose-response study は毒性の検査では標準的な方法である。
動物個体 i は毒物への抵抗の限界値T_iを持っており、与えられるドース D がこれを越える（$D > T_i$）と「死亡」のような応答を示す。たとえばD_{50}は

$$Pr(T \leq D_{50}) = 0.50$$

となる。ドースレベルD_{50}は EC50 (effective concentration)または LD50 (lethal dose, 応答が死亡の時)と呼ばれる。

$$d_i = \log(D_i)$$

d^* をd_i の平均とすると

$$\text{logit}(p_i) = \alpha_R + \beta_R(d_i - d^*)$$

肝臓と腎臓の $NaVO_3$ 量（L_i および K_i）はドースレベルを介して関係していそう。
→ bivariate regression

$$\log(L_i) = \alpha_L + \beta_L(d_i - d^*) + \epsilon_i$$
$$\log(K_i) = \alpha_K + \beta_K(d_i - d^*) + \nu_i$$

(ϵ_i, ν_i)は平均ゼロ・分散行列Σの二変量正規分布（bivariate normal distribution）に従う。
Regression coefficient の事前分布には vague な正規分布、Σにも vague な事前分布（逆ウィシャート分布）。

5.4.3 Derived Parameters

Derived parameter は標準偏差σ_L、σ_K 、およびその相関 ρである。これらは全て Σ の関数である。

さらに、LD50 も derived parameter である。
$p(D) = 50\%$ を解いて、

$$\text{logit}(p(D)) = \alpha_R + \beta_R(\log(D) - d^*) = \text{logit}(0.50) = 0$$
$$D_{50} = \exp(d^* - \alpha_R/\beta_R)$$

5.4.4 Out of Sample Prediction

ガンがさらされた $NaVO_3$ のドースレベルは未知である。
実験下でカモが $L = 57.3, K = 226$ であり、かつ死亡したと仮定すると、どれくらいのドースレベルの $NaVO_3$ にさらされたと予測できるのか？

解決法は、仮想的なドースの事前分布D^{H}を使うこと、仮想データ$X^{\mathrm{H}} = (L^{\mathrm{H}}, K^{\mathrm{H}}, R^{\mathrm{H}})$ を事後予測分布を使って解析することである。（これは口で言うよりも“簡単”である）
$\boldsymbol{X}$を全ての観測データ、$\boldsymbol{\theta}$を未知のパラメータとする。
X^{H}が$\boldsymbol{X}$から conditionally independent (i.e., $\boldsymbol{\theta}$とD^{H}のもとで)であり、$\boldsymbol{\theta}$とD^{H}の事前分布が independent であるならば、

$$[D^{\mathrm{H}}, \boldsymbol{\theta} \mid \boldsymbol{X}, X^{\mathrm{H}}] \propto [X^{\mathrm{H}} \mid D^{\mathrm{H}}, \boldsymbol{\theta}][\boldsymbol{\theta} \mid \boldsymbol{X}][D^{\mathrm{H}}]$$

$\boldsymbol{\theta}$で両辺積分すると、$[D^{\mathrm{H}} \mid \boldsymbol{X}, X^{\mathrm{New}}]$という条件付き分布が得られる。
これは事後予測分布$[X^{\mathrm{H}} \mid D^{\mathrm{H}}, \boldsymbol{X}]$を事前分布$[D^{\mathrm{H}}]$で重みづけしたものに比例する。
この計算は Panel 5.3 に Panel 5.4 の BUGS コードを足すことで行うことができる(略)。

もしも仮想的なデータ$X^{\mathrm{H}} = (L^{\mathrm{H}}, K^{\mathrm{H}}, R^{\mathrm{H}})$を観測値であるかのように用いると、事後分布$[X^{\mathrm{H}} \mid D^{\mathrm{H}}, \boldsymbol{\theta}]$ が不適切な影響を受けるので、事後予測分布を用いる必要が生じる。

仮想的観測値にもとづく事後予測分布は平均 0.976、5^{th} percentile 0.889 となった。

→ガンの測定値（死亡個体の肝臓・腎臓の $NaVO_3$ 量）はカモの実験では LD97.6 に相当する。
90%の confidence で最低でも LD88.9 のドースレベルにさらされたと予測できる。

以下の理由でこの解析は satisfactory である。
データに拘束された近似はない。あいまいな漸近性（asymptotics）もない。
推論は分かりやすくに連結された（？articulated）前提条件に基づき、
全ての結論は事後分布(のサマリー)に基づくという包括的な原理に則している。

5.5 PRODUCT BINOMIAL REPRESENTATION OF THE CJS MODEL

ここでは派生パラメータを使ってベイズ統計の複雑な計算を単純化する方法を示す。

Cormack-Jolly-Seber model (CJS model)　詳細は 11 章
動物個体を捕獲して標識し、リリースするというサンプリングを *t* 回行うような状況を扱う。
CJS モデルはオープン個体群を記述する。出生・移入で生息地に入り、死亡・移出で生息地から出る。一時的な移出はないとする。死亡・移出せず、そこにとどまって生きていた場合を生存とする。生存・捕獲・リリースのそれぞれについてベルヌーイ試行として扱う。

このモデルの identifiable parameter は、

$$\theta = \{p_2, p_3, \dots, p_{t-1}; \phi_1, \phi_2, \dots, \phi_{t-2}, \lambda_{t-1}\}$$

時間 *i* から *i*+1 での生存率を ϕ_i （全標識個体で一定。個体標識個体と非標識個体の生存率は同じ。）
時間 *i* で捕獲された個体の再捕確率は p_i
時間 *i* でリリースされた総個体数はR_i （以前に標識されていたものとされていなかったものを両方含む。）
λ は、$\lambda_{t-1} = \phi_{t-1} p_t$

このときの尤度（Cormack 1964）は

$$\prod_{i=1}^{t-1} \phi_i^{v_i} p_{i+1}^{a_{i+1}} (1 - p_{i+1})^{v_i - a_i + 1} \chi_i^{c_i}$$

v_i は時間 i より前にリリースされ、のちに（時間 j ）再捕獲された個体数$(i \le j \le t)$
a_i は時間 i より前にリリースされ、のちに（時間 i ）再捕獲された個体数$(i \le j \le t)$
c_i は時間 i でリリースされ、その後$(\le t)$再捕獲されなかった個体数

χ_i は生存率と捕獲率のパラメータの関数

$$\chi_i = (1 - \phi_i) + \phi_i (1 - p_{i+1}) \chi_{i+1}$$

i =*t*-1, *t*-2,...,2,1

複雑な形の尤度を用いるとϕ_i と p_i の事後分布を計算することが困難になる→MCMC を使う。
full conditional distribution が分からないのでギブス・サンプリングはできない。→Metropolis-Hastings argorithm を使う。
たとえば、θ のコンポーネントに独立の一様な事前分布 U(0,1) を置くということができる。現在のパラメータ値の logit に $N(0, \sigma^2)$を足して、候補の値を生成する。この場合アルゴリズムの調整が必要。

以下に尤度の再パラメータ化と derived parameter の使用で計算効率が改善できることを示す。
θの十分統計量は

$$S = \{r_1, r_2, \dots, r_{t-1}; m_2, m_3, \dots, m_t\}$$

r_i は時間 *i* に標識してリリースされ、のちに（$\le t$）再捕獲された個体数
m_i は時間 *i* で標識された個体数
$T_2 = r_1$とする。

$$T_{i+1} = T_i - m_i + r_i$$

T_i は時間 i より前に標識・リリースされ、のちに（時間 j ）再捕獲された個体数($i \le j \le t$)
複雑な CJS 尤度を十分統計量と derived parameter で書き直す(Burnham 1991)。

$$\prod_{i=1}^{t-1} B(r_i; R_i, \lambda_i) \prod_{i=2}^{t-1} B(T_i; m_i, \tau_i)$$

λ_i と τ_i はϕ_i と p_i の関数(derived parameter)である。

この式の単純さは強調すべきである。
もし、λ_iやτ_i自体に興味があり、独立した事前分布をこれらにおけるなら、事後分布は共役事前分布を使い inspection によって得られる。
たとえばλ_i~Be(a, b)のとき、事後分布は

$$[\lambda_i|\text{Data}] \sim \text{Be}(a + r_i, b + R_i - r_i)$$

このことと、ベイズの枠組みでの derived parameter の扱いやすさは、CJS モデルの解析の 2 つの方法の基礎になる。

5.5.1 CJS with Uniform Prior on λ and τ

ϕ_i と p_i に事前分布を置くことを選択するのが、ふつうである。
これらは、このモデルで基本的で、もっとも関係（関心）のある量であり、対象にしている現象をもっとも自然に記述する。
(5.9)の derived parameter であるλ_i は時間 i でリリースされ、時間 i よりあとに再捕獲される確率であり、τ_i は時間 i より前にリリースされ、時間 i または、それより後に再捕獲される確率である。

これらはϕ_i （生存率）と p_i （捕獲率）に複雑に依存する。 λ_i とτ_i にどのような事前分布を置けばよいか分からないし、ϕ_i と p_i の事前分布の点で何を意味するのか不明瞭である。

しかし、以下のことが当てはまるケースは多い。
データが事前分布を超える: データが十分あれば、事前分布の選択は事後分布にはあまり影響しない。
そのため、λ_i とτ_iに独立した一様分布を置くと、(5.9)より、事後分布も独立になる。

$$[\lambda_i|\text{Data}] = \text{Be}(r_i + 1, R_i - r_i + 1) \quad (5.12)$$

$$[\tau_i|\text{Data}] = \text{Be}(T_i + 1, m_i - T_i + 1) \quad (5.13)$$

ソフトウェアでこれらの分布からサンプリングできる。
サンプリングされたλ_iとτ_iを(5.10)と(5.11)からϕ_i と p_i に変換する。

$$\phi_i = (\lambda_i/\lambda_{i+1})(1 - \tau_{i+1}(1 - \lambda_{i+1}))$$

i =1,2,...,t-2 (5.14)

$$p_i = (\tau_i\lambda_i)(1 - \tau_i(1 - \lambda_i))$$

i =1,2,...,t-2 (5.15)

(5.14)と(5.15)に、(5.12)と(5.13)からサンプルされた λ_iとτ_i を導入することで、（ψ（λ_iとτ_i） の事前分布が一様分布であるという条件での） ϕ_i と p_i の事後分布からサンプルを得ることができる。
変換されたパラメータであるψ に一様分布を置くことは、θ に一様分布を置くこととは異なり、一様でない（inducing な）分布を置くことを意味する。

これから示すように、rejection sampling を使って、このような induced prior を用いて得られた ϕ_i と p_iのサンプルを θ に一様分布を置いたときに得られるサンプルに変換することができる。

5.5.2 CJS with Uniform Prior on p and ϕ

rejection sampling (4.2.2 を参照)では、まず candidate の分布 $c(x)$ からサンプリングすることにより、target の分布 $t(x)$ からサンプルを得る。普通、$t(x)$は $c(x)$よりもサンプリングが難しい。Rejection sampling に必要な条件は$t(x) \leq Mc(x)$となるような M が存在することである($w(x) = t(x)/Mc(x) \leq 1$)。
Rejection sampling では$c(x)$から$X = x$を引き出し、成功確率が $w(x)$のベルヌーイ試行に基づいて、$t(x)$の値として accept または reject と判断する。

β の事後分布を $f_t(\beta|\mathrm{X}), f_c(\beta|\mathrm{X})$とし、それぞれ別々の事前分布$[\beta]_t$, $[\beta]_c$ に基づくとする。

$$\frac{f_t(\beta|\mathrm{X})}{f_c(\beta|\mathrm{X})} \propto \frac{[x|\beta][\beta]_t}{[x|\beta][\beta]_c} = \frac{[\beta]_t}{[\beta]_c}$$

事前分布$[\beta]_t$ が事前分布$[\beta]_c$の rejection sampling によってサンプリングできるとすれば、以下を用いて、

$$w(\beta) = \frac{[\beta]_t}{M[\beta]_c}$$

$f_t(\beta|\mathrm{X})$は同じ$w(x)$を用いて $f_c(\beta|X)$の rejection sampling によってサンプリングできる。
$w(\beta) \leq 1$となるように M が選ばれていることに注意。
M が、(ほとんどの$f_t(\beta|X)$の support において)$w(\beta) \leq 1$を保証するだけの大きさで良いとすれば、M の値は小さくなり、モデルのアクセプトの確率が高くなる。i. e. $f_t(\beta|X)$の rejection sampler の効率が改善される。

CJS モデルでθ (つまり ϕ_i と p_i)の事前分布を一様分布にすると、計算すべき事後分布は$[\theta]_t \equiv 1$ のときの$f_t(\theta|X)$である。Candidate は簡単にサンプリングできる$f_c(\theta|X)$で、これは ψの事前分布を一様分布にする事で induce される事前分布$[\theta]_c$(一様でない)に対応する。
以下 (Link and Barker 2008)は簡単に計算できる。

$$[\theta]_c \propto \frac{\lambda_{t-1}}{\lambda_1} \prod_{i=1}^{t-2} \phi_i$$

ガ(*Gonodontis bidentata*)の標識再捕獲データの例への応用
イングランドで 17 日間にわたって毎日捕獲・標識・リリースした。 データは 689 のオスからなる。
2 つの体色変異型の相対適応度の研究の一環として demographic パラメータが推定されている。
十分統計量は Table 5.3

(5.9)の Burnham による parameterization の興味深い点はτ_i とλ_i の全ての組み合わせが、意味のあるϕ_i と p_iの値に対応しているわけではないことである。つまり、τ と λ を ϕ_i と p_iに変換する際、$\phi_i > 1$の値が出る可能性がある。
頻度主義における類似の問題は、最尤推定値が非現実的な(許容できない)値をとるときに起こる。
このようなときは「尤度は問題になっている生物学的現象について何も知らない。」ことを考えるとよい。
数学的に問題なくても生物学的には非現実的なパラメータ値が出ることがある。

このような場合には情報のある事前分布を取ることに意味がある（パラメータに非現実的な値をとらせない）。
→可能なパラメータ値の範囲内で一様分布を取る。
もし、5.5.1 で述べられた方法を用いてτとλ に一様事前分布を置き、それをϕ とp に変換するなら、ϕ とp でおかしな値が出たら全て捨てる（reject）ことになる。ここでは、τとλ の単位区間（0 から 1）の一様事前分布を、許容できる範囲の一様分布に置き換えた。
一様な事前分布から得られた生存率ϕ_iの事後分布はかなり平ら（diffuse）である(Fig 5.5)。
かなりの値が 1.0 付近に集中している。

この例では 94.5%の値をリジェクトした。
Rejection sampling の利点は、事後分布から直接サンプリングすることによる計算効率が random walk Metropolis-Hastings よりも上昇することである。

この例では、(5.12)と(5.13)のβ 分布から τ とλ を生成し、2,001,676 セットの ϕとp を得た。これにより”ψ-uniform”な事後分布を考えることができる。rejection sampling を行うと、302,424 セットのサンプリングにまで減らすことができた（”θ-uniform”な事後分布）。
θ-uniform な事後分布から直接サンプリングするよりも計算時間は短い（1 222 秒と 10 000 秒）。時間の比較のため、さらに BUGS を用いてθ-uniform の事後分布から 302,424 セットのサンプリングを得た。計算時間は burn-in を含めると、2.4 倍になった。また BUGS で生成された値は自己相関していたが、rejection sampling で生成されたものはそれがなく独立であった（BUGS の計算では有効なサンプルサイズは小さくなっていた）。

Table 5.4 は生存率ϕ_iの平均と標準偏差である
rejection sampling のθ-uniform でも、BUGS で行うθ-uniform でも、結果はほぼ同じ。
この近似を使うと（つまり rejection sampling なら）6 倍計算を早くできる。

最後に Rejection sampling により、ψ-uniform の事後分布のサンプルを θ-uniform の事後分布のサンプルに変換した方法を示す。(5.18)と(5.17)より、

$$w(\theta) \propto \frac{\lambda_1}{\lambda_{t-1}\prod_{i=1}^{t-2}\phi_i} \qquad (5.19)$$

λで記述してあるが、$w(\theta)$は ϕとp の関数である($\theta = \{\phi, p\}$)。ψ-uniform な事後分布からの 2,001,676 セットの値で右辺を計算した。
結果の分布は偏っていて、99.9^{th} percentile の 4 倍の値が最大値であった。
(5.19)の右辺をスケーリングした方がよさそうに見えるが、あまり結果には影響しない。

5.5.3 CJS Model: Summary

ここでは、ベイズ推定の優れた点を例示するために CJS モデルの異なるパラメータ化を見てきた。
ベイズではパラメータの関数に関する推定が行いやすい。
パラメータ α の 事後分布を与えられれば、$\beta = g(\alpha)$ の事後分布は簡単に得られる。

α の 事後分布の数学的形が分かっていれば（$f_\alpha(\alpha)$）、βの事後分布は解析的に得られる（5.2.2 参照）。そうでなければ、α の 事後分布をシミュレーションで調べる。

我々の示した例では、derived parameter の分かりやすい使用により、計算の効率が大きく上がる。（期待される事後分布を正確に得られるように）Rejection sampling とともに用いると、59%計算時間が節約される。少しだけ異なる事前分布を使った事後分布を用いると 94%計算時間が節約できる。

5.6 POSTERIOR PREDICTIVE MODEL CHECKING

“all models are wrong, but some are useful” G.E.P. Box
自分のモデルがデータと一貫性を保っているか確かめる必要がある。そうでなければ、モデルとデータの不一致（inconsistency）が推論を偏らせていないかチェックするべきである。

事後予測分布はモデルとデータの一貫性を調べるのに便利な道具である。
データ $\boldsymbol{X}$：未知のパラメータθ以外、データ $\boldsymbol{X}$ の分布は特定されている。
モデル M、
仮想的な反復データセット $\boldsymbol{X}^{\mathrm{New}}$
事後予測分布　$[\boldsymbol{X}^{\mathrm{New}}|\boldsymbol{X}]$

$f(\boldsymbol{X})$がどの程度仮想的な反復の値 $f(\boldsymbol{X}^{\mathrm{New}})$と一致するか？

例　Fisher’s tick data　(4.1.2 で既出)
60 頭の羊についているノミのカウントデータ　→　ポアソン分布でよさそう。
（データ及びモデルの BUGS コードは Panel 5.5）
λ の事前分布にガンマ分布 Ga(0.001,0.001)　(近似的には Jeffreys prior、6.2.3 参照)

ポアソン分布が適切かどうか調べる。
`stat.x`（分散と平均の比）を計算。
ノミ数の標本分散は標本平均の 1.89 倍大きい。→　ポアソンではだめらしい。
しかし、分散も平均もランダムな変数なので 1.89 が 1 とどれくらい違うのか調べなければならない。
$f(\boldsymbol{X})$の分布は未知のパラメータλに依存するので、これを調べるのは簡単ではない。
→λ を固定してシミュレーション parametric bootstrap が可能。
→　しかし、λ の不確実性がうまく表現できない。

ベイジアンなら、 λ がばらついていても扱える。
事後分布$[\lambda|\boldsymbol{X}]$から未知のパラメータλ をサンプリングする。
分布$[\boldsymbol{X}^{\mathrm{New}}|\lambda]$の平均の値を事後分布にかけたもの（事後予測分布）からサンプリングする。
事後予測分布は

$$[\boldsymbol{X}^{\mathrm{New}}|\boldsymbol{X}] = \int [\boldsymbol{X}^{\mathrm{New}}|\lambda][\lambda|\boldsymbol{X}]d\lambda$$

$[\boldsymbol{X}^{\mathrm{New}}|\boldsymbol{X}]$からサンプリングされたそれぞれの$\boldsymbol{X}^{\mathrm{New}}$から（分散/平均）を計算する。
これがオリジナルデータと比べて大きくなっているかどうか調べる。

Bayesian p-value：仮想データの（分散/平均）値がデータの標本値と同じかそれ以上になっている確率

$$\Pr\{f(\boldsymbol{X}^{\mathrm{New}}) \geq f(\boldsymbol{X})|\boldsymbol{X}\}$$

この例では 1/4000 (100 000 回中 215 回は観測値の 1.89 より小さい)
→　ポアソンモデルは適切ではないと判断する強い証拠。

ノミ数の比較
ポアソンモデルでは 4.0%が 0 であり、1.9%が 7 より大きい。
Fisher’s tick data では、11.7%が 0、5.0%が 7 より大きい。
→　予測の上でもあまりうまくいかない。

5.6.1 Bayesian and Frequentist *p*-Value

Bayesian *p*-value は頻度主義の *p* 値と同じく、モデルがデータ X を記述するのに使われる。
注目しているデータの統計量 $f(X)$をモデルから計算された値と比べる。
モデルを観測データと比べるときには $f(X)$を注意深く選ぶ必要がある。
モデル M の特異的な（他のモデルにない）特徴を反映した統計量を選ぶ。

ポアソン分布の場合には、（分散/平均）を用いた。
しかし他の分布でも分散＝平均となる場合がある。e.g. discrete uniform distribution（T=4 のとき。テキスト参照）
→選んだ $f(X)$に不一致が見られなかったとしても、モデルの正当性が約束されるわけではない。

確率分布が未知のパラメータに依存しているような場合には、モデルから計算される仮想的反復の値をどのように評価すべきかが難しい。頻度主義でこれを解決するには、未知のパラメータ全てについて仮想的な反復の確率論的評価が正しい、ということを示す必要がある。もしすべての$\boldsymbol{\theta}$（未知のパラメータ）について、以下のように言った時には、

$$\Pr\{f(\boldsymbol{X}) > f(\boldsymbol{X}^{\mathrm{New}})|\boldsymbol{\theta}\} \leq 0.023$$

0.0023 は頻度主義の *p*-value である。

一方でベイジアンはパラメータの不確実性を、平均化することによって解決する。
理想的にはモデルの批判・評価に周辺分布（marginal distribution）を用いる。

$$[\boldsymbol{X}^{\mathrm{New}}] = \int [\boldsymbol{X}^{\mathrm{New}}|\boldsymbol{\theta}][\boldsymbol{\theta}]d\boldsymbol{\theta} \qquad (5.20)$$

これには informative な事前分布を使うのがよい。しかしあいまいな（vague）事前分布を objective analysis に用いると、周辺分布がとても平らに（diffuse）になり、判別力がなくなる可能性がある。

主観的プロセスの問題を（使用者が）認識していれば、予測分布は不適切なモデルを排除するのに十分有効な道具である。Bayesian *p*-value に基づいてモデルが有効でないと判断されれば、それは信頼できる判断である。8.5.3 と 9.3.2 に Bayesian *p*-value の使用例がある。