

ベイズ統計における有効サンプルサイズ(ESS)のSASによる実装

○玉乃井峻太¹

(シミック (株) データサイエンス事業本部)

Implementation of Prior Effective Sample Size (ESS) in Bayesian Statistics Using SAS

Ryota Tamanoi

Data Science Div., CMIC Co.,Ltd.

要旨

近年、医療統計分野においては、ベイズ統計の応用が活発に進展している。ベイズ統計では観察データに加え外部情報を「事前分布」として組み込むことができる他、複雑な試験デザインにも柔軟に対応できるなど、多くの利点がある。一方、事前分布の情報量が過度に大きい場合、解析結果が事前分布に大きく依存し、観察データの影響が十分に反映されないという課題も指摘されている。したがって適切な情報量の事前分布を選択することが重要である。このとき有効サンプルサイズ (Effective Sample Size: ESS) を事前分布の情報量の指標として用いることができる。ESS は事前分布の情報量を観察データの人数に換算して定量化するものであり、2010年の医療機器におけるベイズ統計のFDAガイダンスでは「プロトコルに含めることを推奨する情報」の一つとして挙げられている。近年複数のESSが提案されており、それぞれ定義や適用範囲など特徴が異なる。本稿ではまずベイズ統計の基礎の説明を行い、その後3つのESSの解説、SASによる実装を行う。

キーワード: ベイズ統計, 事前分布, 有効サンプルサイズ, ESS, 無情報事前分布

1. ベイズ統計とESSの導入

本節ではベイズ統計になじみのない読者を対象に基礎的な内容から紹介し、事前分布のESSを導入する。

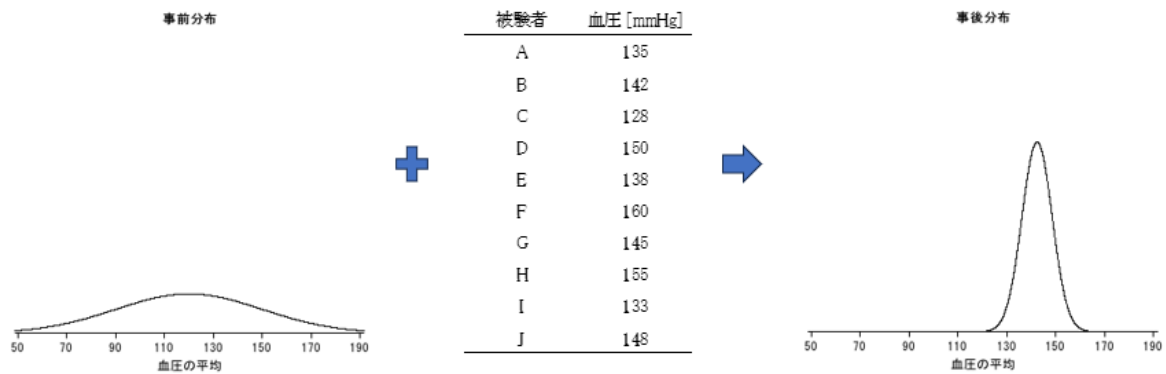
1.1. ベイズ統計と事前分布

ベイズ統計では一般的な統計と異なり、事前分布とよばれる確率分布を観察データとは別に用意する必要がある。事前分布と観察データの結果から事後分布を導出し、この事後分布から信頼区間やp値に対応する値を算出することができる。

図1は事前分布と観察データから事後分布を求める流れを示している。この例では観察データを得る以前には、血圧の真の平均は120前後、80から160程度であろうと考えていることがわかる。ここに観察データを加えた結果、真の平均は140前後、130から150程度の範囲にあると予測できる。

事前情報がない場合には無情報事前分布 (non-informative prior. 情報の少ない事前分布; uninformative prior, 曖昧な事前分布; vague prior とよぶこともある) を用いることで事前分布の影響を制限できる。どのような事

図 1 事前分布 (左図) と観察データ (中央表) から事後分布 (右図) が算出される。事前分布は平均 120, 標準偏差 30 であり, 事後分布は平均 142.4, 標準偏差 6.2 である。



前分布が無情報といえるかはモデルによって異なり課題も多い [1]が一般的には Jefferys prior [2]を含む reference prior [3]がよく用いられる。

ベイズ統計のメリットとして, 中間解析の結果やリアルワールドデータなどの事前情報がある場合, これを事前分布にすることで意思決定に用いるデータを増やすことができる他, 少ない症例数で検出力を確保できる場合がある [4]. また事前情報がない場合にも, 事後分布を用いた治療効果の定量的評価や中間解析による柔軟な意思決定が可能になる点で有用である [5]. 一方ベイズ統計のデメリットとして, 理論的な専門性や難しさの他に, 事前分布の設定の問題が指摘されている [4, 5, 6]. 情報のある事前分布を用いる場合, その選択によって解析結果が大きく変わりうる. 例えば事前分布の情報量が観察データに比べて過剰に大きい場合, 解析結果が事前分布に大きく依存し, 観察データの影響が十分に反映されない. そのため適切な情報量を持つ事前分布を選択することが重要となる. 平川ら (2021) [5]では情報のある事前分布を用いる場合, 感度分析として無情報事前分布を用いた結果との比較が提案されている。

1.2. ESS について

事前分布の情報量を「観察データ何人分であるか」で表した指標が事前分布の有効サンプルサイズ (prior effective sample size; ESS) である. すなわち ESS=1 の事前分布は一人分の情報を持ち, ESS=100 の事前分布は 100 人分の情報を持っていることを表す. FDA2010 [4]では Suggested Information to Include in Your Protocol の一つとして Effective Sample Size が挙げられている。

どの程度の ESS が適切であるかは試験によって異なる. 事前情報と試験の類似性が高い場合は ESS の大きい事前分布によって試験効率を向上することができるが, そうでない場合は ESS の小さい事前分布を用いることで事前分布と観察データの衝突 (prior-data-conflict) を防ぐことができる. FDA (2010) [4]では事前分布の ESS が試験の症例数を上回る場合事前分布は情報過多である可能性があり, 修正する必要があると述べられている。

注意として ESS という単語の意味はいくつかの文脈で用いられる. 本稿では事前分布の有効サンプルサイズを指しているが, FDA (2010) [4]では観察データを加えた後の事後分布の有効サンプルサイズを ESS とよんでいる. また PROC MCMC でも ESS が output に表示されるがこれは事前分布の ESS とは関係がない. 混乱が生じる場合は prior ESS などと示すべきである. 本稿では特に指定がない場合 ESS は事前分布の ESS を指す。

2. 様々な ESS の定義と性質

ESS はベイズ統計で古くから用いられており, Novick and Hall (1965) [7]では二項モデルに対するベータ分布の ESS (論文内では **prior information** とよばれている) について言及されている. 一方で一般の事前分布について ESS を求めることは容易ではなく, 近年も複数の ESS が提案されている. 本稿では提案されている ESS の定義をいくつか挙げ, 特に ESS_{MTM} , ESS_{ELIR} , ESS_{E-F} の説明と SAS での実装を行う.

2.1. 共役事前分布における ESS

ベイズ統計において古典的に用いられている事前分布に共役事前分布とよばれる事前分布がある. これはモデルに指数型分布族を指定している場合に与えられる事前分布(族)であり, 事後分布が容易に計算できる特徴がある. 表 1 は代表的な共役事前分布の一覧である.

表 1: 代表的な共役事前分布 代表的なモデルと共役事前分布の組とその ESS の一覧 ここで **N, invGa, Ber, Beta, Exp, Ga, Pois** はそれぞれ正規分布, 逆ガンマ分布, ベルヌーイ分布, ベータ分布, ガンマ分布, ポアソン分布を指す.

| 尤度分布 | パラメータ | 共役事前分布とハイパーパラメータ | 古典的に与えられている ESS の例 |
|----------------|---|---|---|
| 正規分布 (分散既知) | $N(Y \mu, \sigma^2)$ (σ^2 既知) | $N(\mu m_0, s_0^2)$ | σ^2/s_0^2 |
| 正規分布 (平均既知) | $N(Y \mu, \sigma^2)$ (μ 既知) | $invGa(\sigma^2 \alpha, \beta)$ | 2α |
| 正規分布 | $N(Y \mu, \sigma^2)$ | $N(\mu m_0, \sigma^2/\kappa)$ $\times invGa(\sigma^2 \alpha, \beta)$ | $\kappa = 2\alpha$ のとき κ . そうでないとき ESS は古典的に与えられていない |
| ベルヌーイ分布 (二項分布) | $Ber(Y p)$ | $Beta(p \alpha, \beta)$ | $\alpha + \beta$ |
| 指数分布 | $Exp(Y \lambda)$ | $Ga(\lambda \alpha, \beta)$ | α |
| ポアソン分布 | $Pois(Y \lambda)$ | $Ga(\lambda \alpha, \beta)$ | β |
| 線形回帰 | $N(Y \sum_{k=1}^p \beta_k X_k, \sigma^2)$ | g prior | n/g ただし n は症例数 |

多くの共役事前分布に対しては ESS が一般的に用いられており, その値が表 1 の右側である. 例えば尤度分布がベルヌーイ分布, 事前分布にベータ分布 $Beta(\alpha, \beta)$ を用いるとき, 観察データとして $Y = 1$ を n_1 例, $Y = 0$ を n_0 例得たとすると, 事後分布は再びベータ分布 $Beta(\alpha + n_1, \beta + n_0)$ となる. (事後分布が, 事前分布のハイパーパラメータを変えるだけで得られることが共役事前分布の最大の特徴である.) ここから α, β はそれぞれ $Y = 1, 0$ の事前情報といえる. したがって ESS を合計 $\alpha + \beta$ と考えることができる.

注意として, ESS は文献によって 1, 2 程度異なる場合がある. 例えばベータ分布の ESS に関して無情報事前分布, すなわち $ESS = 0$ の基準として Haldane prior ($Beta(0, 0) \propto p^{-1}(1-p)^{-1}$) を用いれば $ESS = \alpha + \beta$ となるが, Jefferys' prior $Beta(0.5, 0.5)$ や一様分布 $Beta(1, 1)$ を用いれば ESS はそれぞれ $\alpha + \beta - 1, \alpha + \beta - 2$ となる.

2.2. 分散比を用いた ESS

前節の ESS は共役事前分布に対してしか与えられていない。一般の事前分布に対する ESS の定義は明らかではなく、状況に応じて選択することが必要になる。最初に, Malec (2001) [8]や FDA (2010) [4]で与えられている, 事後分布と「無情報事後分布」の分散の比を利用して ESS を算出する方法を紹介する。

ESS を求めたい事前分布を $p(\theta)$, 無情報事前分布を $\pi_0(\theta)$, m 例の観察データを Y_1, \dots, Y_m とおく。このとき, 無情報事後分布 $\pi_m(\theta) = \pi_0(\theta | Y_1, \dots, Y_m)$ と $p(\theta)$ の事後分布 $p(\theta | Y_1, \dots, Y_m)$ を比較する。 $\pi_m(\theta)$ は無情報事前分布の事後分布であるから ESS はちょうど m であると考えられる。 $p(\theta | Y_1, \dots, Y_m)$ の情報量がこの何倍であるかによって事後分布の ESS を算出し, 事前分布の ESS を求める。

$\pi_m(\theta)$ の分散を V_1 , 事後分布 $p(\theta | Y_1, \dots, Y_m)$ の分散を V_2 とおき, ESS を次で与える:

$$ESS = m \cdot (V_1/V_2) - m$$

正規モデルにおいて, この定義はうまく機能する。分散の比が無情報事後分布と事後分布の情報量の比になると考えれば, $\pi_m(\theta)$ の ESS は m であるから, $m \cdot (V_1/V_2)$ が事後分布 $p(\theta | Y_1, \dots, Y_m)$ の ESS である。したがって事前分布の ESS は $m \cdot (V_1/V_2) - m$ と考えることができる。一方でこの定義はいくつかの問題がある。

- 観察データ Y_1, \dots, Y_m によって ESS が変わるため計画段階で ESS を算出することが難しい。
- 分散の比は変数変換によって変わるため, 本質的に同じ分布であっても異なる ESS が与えられる。例えば対数変換によって ESS が変わる。
- モデルによっては分散を情報量の指標として用いることはできない。モデルごとに適切な指標を考える必要がある。特に多変数パラメータの場合これは簡単でない。

2.3. モデルベースの ESS

この節では観察データによらず, モデルによって与えられる ESS を考える。

2.3.1. ESS_{MTM} の定義

ESS_{MTM} は Morita et al. (2008) [9]で与えられた ESS である。この ESS ではまず, 事前分布 $p(\theta)$ の情報量を次の式で定義する。パラメータ θ は多変数でもよい:

$$i(p(\bar{\theta})) = \text{tr} \left\{ -\frac{\partial^2}{\partial \theta^2} \log p(\theta) \right\}_{\theta=\bar{\theta}}$$

ここで, 記法は以下のとおり。

- tr は行列の対角成分の和を意味する。
- $\bar{\theta}$ は事前分布の期待値であり, $\theta = \bar{\theta}$ は偏導関数 $\partial \theta^2 \log p(\theta)$ の $\theta = \bar{\theta}$ における値であることを意味する。

この情報量において, m 例のデータを与えた無情報事後分布 $\pi_m(\theta)$ と $p(\theta)$ の情報量の差が最も小さくなる m を ESS_{MTM} とおく:

$$ESS_{MTM} = \underset{m}{\text{argmin}} \left| i(p(\bar{\theta})) - E_{Y^m} [i(\pi_0(\bar{\theta} | Y_1, \dots, Y_m))] \right|$$

ここで E_{Y^m} は事前分布の予測分布 $\int f(Y | \theta) p(\theta) d\theta$ による期待値。これを式変形すると次を得る [10].

$$ESS_{MTM} = \frac{i(p(\bar{\theta})) - i(\pi_0(\bar{\theta}))}{i_{MODEL}(\bar{\theta})}$$

ここで $i_{MODEL}(\bar{\theta})$ はモデル $f(Y | \theta)$ の $\bar{\theta}$ における情報量である:

$$i_{MODEL}(\bar{\theta}) = -E_{Y^1} \left[tr \left\{ \frac{\partial^2}{\partial \theta^2} \log f(Y | \theta) \right\}_{\theta=\bar{\theta}} \right].$$

すなわち ESS_{MTM} は「事前分布の情報量」を「モデルの一人分の情報量」で割った値であるといえる。この定義は多くのモデルに当てはめることが可能である。一方、問題点として $\theta = \bar{\theta}$ の周りの局所的な値のみによって ESS を定義していることが挙げられる。事前分布が正規分布であれば $i(p(\theta))$ は任意の θ で一定であり問題ないが、例えば t 分布のような裾の広い分布では $i(p(\bar{\theta}))$ は周辺の $i(p(\theta))$ より大きい値であり、 ESS を過大評価することになる。

2.3.2. ESS_{ELIR} の定義

Neuenschwander et al. (2020) [10] は θ が一変数の場合に ESS_{MTM} を改善し、*expected local-information-ratio ESS* (以下 ESS_{ELIR}) を与えた。これは次で定義される:

$$ESS_{ELIR} = E_{\theta} \left[\frac{i(p(\theta))}{i_F(\theta)} \right]$$

ここで E_{θ} は事前分布 $p(\theta)$ による期待値であり、 $i_F(\theta)$ はフィッシャー情報量:

$$i_F(\theta) = -E_{Y|\theta} \left[\frac{d^2}{d\theta^2} \log f(Y | \theta) \right]$$

である。 $i_{MODEL}(\theta)$ との違いは予測分布 $\int f(Y | \theta) p(\theta) d\theta$ による期待値を用いるか、モデル $f(Y | \theta)$ による期待値を用いるかの差である。 ESS_{ELIR} は各 θ での「事前分布の情報量と一人分の情報量の比」(local-information ratio) の期待値であり、大域的な値になっている。また次の予測的無矛盾性 (predictive consistency) を持つことが証明できる [10].

予測的無矛盾性 (predictive consistency)

予測分布に従う観察データ $\mathbf{Y}^m = Y_1, Y_2, \dots, Y_m$ を与えるとき、事後分布の ESS_{ELIR} の期待値は事前分布の $ESS_{ELIR} + m$ となる。

$$E_{\mathbf{Y}^m} [p(\theta | \mathbf{Y}^m) \text{ の } ESS_{ELIR}] = (p(\theta) \text{ の } ESS_{ELIR}) + m$$

すなわち「観察データを m 例与えると ESS が m 大きくなる」という性質であり、自然である。このように θ が 1 次元の場合 ESS_{ELIR} は自然な値といえるが、一方で回帰分析のような多次元の場合に定義されていない。

2.3.3. Clarke (1996) における ESS の定義

予測的無矛盾性を持ち一部の多変数パラメータモデルで定義できる ESS として ESS_{E-F} がある。この定義の前に Clarke (1996) [11] で定義された ESS について述べる。簡単にいえばこれは「無情報事後分布 $\pi_m(\theta | \mathbf{Y}^m)$ のうち最も事前分布 $p(\theta)$ に似ているもの」を探索し、その m を ESS として定義する。ここで似ているとは Kullback-Leibler divergence [12] (以下 KLd) を最小にする \mathbf{Y}^m を選択することを指す:

$$KL(p(\theta) || \pi_m(\theta | \mathbf{Y}^m)) = \int p(\theta) \log \frac{p(\theta)}{\pi_m(\theta | \mathbf{Y}^m)} d\theta$$

KLd は 0 以上の実数であり 0 に近いほど $p(\theta)$ と $\pi_m(\theta | \mathbf{Y}^m)$ が近いことを示す指標として統計学でよく用いられる。すなわち Clarke (1996) は effective sample を

$$\text{effective sample} = \underset{\mathbf{Y}^m}{\operatorname{argmin}} KL(p(\theta) || \pi_m(\theta | \mathbf{Y}^m))$$

で与え、 ESS を effective sample の例数 m として与える。

Clarke (1996) の ESS は簡明であり任意のモデルに対して用いることができる。一方で ESS が大きいとき計算

量が非常に大きくなる. 実際 m が大きくなると \mathbf{Y}^m の取り方も多くなり, KLd についての最適化問題は莫大な量の計算を必要とする. また $KL\langle p(\boldsymbol{\theta}) || \pi_m(\boldsymbol{\theta} | \mathbf{Y}^m) \rangle$ は「 $\pi_m(\boldsymbol{\theta} | \mathbf{Y}^m)$ を基準とする $p(\boldsymbol{\theta})$ のエントロピー」を意味しており, \mathbf{Y}^m によって基準が変わる. これが一因しこの ESS は予測的無矛盾性を持たない.

2.3.4. ESS_{E-F} の定義

ESS for exponential family (以下 ESS_{E-F}) は Tamanoi (2024) [13]で定義された ESS であり, モデルが指数型分布族の場合に Clarke (1996) の問題点を解決したものである. 指数型分布族には観察データのモデルが正規分布や二項分布, 線形回帰, ロジスティック回帰の場合が含まれ, 逆にワイブル分布, プロビット回帰, コックス回帰, 混合モデルなどが含まれない. まず, 次の一人分の尤度を考える:

$$L(\boldsymbol{\theta} | \mathbf{Y}^m) = \left(\prod_{i=1}^m f(Y_i | \boldsymbol{\theta}) \right)^{1/m}$$

これは Y_1, \dots, Y_m の尤度を平均したものであり, 例えば 100 例のデータから 80 例分の情報を得るときなど (power prior) によく用いられる. $f(Y | \boldsymbol{\theta})$ が指数分布族の場合, この一人分の尤度は Y_1, Y_2, \dots, Y_m の十分統計量のみで与えられる [13]. よって十分統計量 t_1, t_2, \dots, t_K を用いて一人分の尤度を $L(\boldsymbol{\theta} | t_1, t_2, \dots, t_K)$ と表すことができる. 指数分布族の十分統計量の次元 K は高々パラメータの個数であるから, ESS が大きい場合にも計算量が大きくなることはない.

次に一人分の尤度を用いて次の事前分布 $p(\boldsymbol{\theta})$ から α 例分の情報を引いた分布を与える:

$$p_{-\alpha}(\boldsymbol{\theta} | t_1, t_2, \dots, t_K) = \frac{1}{Z_\alpha} p(\boldsymbol{\theta}) L(\boldsymbol{\theta} | t_1, t_2, \dots, t_K)^{-\alpha}$$

ここで Z_α は正規化定数である. 通常の事後分布と異なり $-\alpha$ 乗をとることで情報を引くことを実現している. しかし $Z_\alpha = \int p(\boldsymbol{\theta}) L(\boldsymbol{\theta} | t_1, t_2, \dots, t_K)^{-\alpha} d\boldsymbol{\theta}$ は一般に収束しない. そこで $p(\boldsymbol{\theta})$ の 95%の区間 (Bayesian evidence interval [14]) 上で $p_{-\alpha}(\boldsymbol{\theta} | t_1, t_2, \dots, t_K)$ を考える.

最後に無情報事前分布 $\pi_0(\boldsymbol{\theta})$ と $p_{-\alpha}(\boldsymbol{\theta} | t_1, t_2, \dots, t_K)$ の間の KLd を考え, これを最小にする α を ESS_{E-F} とおく:

$$ESS_{E-F} = \operatorname{argmin}_{\alpha} \left(\min_{t_1, \dots, t_K} KL\langle p_{-\alpha}(\boldsymbol{\theta} | t_1, \dots, t_K) || \pi_0(\boldsymbol{\theta}) \rangle \right)$$

この定義では固定された $\pi_0(\boldsymbol{\theta})$ に対する $p_{-\alpha}(\boldsymbol{\theta} | t_1, \dots, t_K)$ のエントロピーを考えており, ESS_{E-F} は予測的無矛盾性を持つ [13]. また, 定義の中で微分を用いていないためモンテカルロ積分による近似計算が可能であり, 複雑な事前分布に対しても尤度と事前分布に従うシミュレーションサンプル $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N$ ($N = 10000$ など) があれば近似可能である. 一方で以下の問題点が挙げられる.

- モデルが指数型分布族に含まれなければ定義できない.
- Clarke (1996) に比べて改善しているものの, 回帰分析などで K が大きいとき依然計算量が大きい.
- KLd の最小値は必ずしも 0 にならない. この値が十分小さくない場合 ESS_{E-F} は正当化できない.

KLd の最小値に関して, Tamanoi (2025) で RKL という指標が導入されている. これは次で定義される.

$$RKL = \frac{KL_{min}}{KL\langle p(\boldsymbol{\theta}) || \pi_0(\boldsymbol{\theta}) \rangle}$$

ここで KL_{min} はKLdの最小値である. すなわち RKL は KL の最小値と KL の初期値の比である. $0 \leq RKL \leq 1$ であり, $RKL = 0$ のとき理想的である. RKL の大小によって ESS_{E-F} の正当性を評価できるが, 許容される RKL の上限については議論されていない.

3. 実装例

この節では ESS_{MTM} , ESS_{ELIR} , ESS_{E-F} の実装例を紹介する. 添付の macro ファイルと sas ファイルによって再現が可能である. SAS version 9.4 を用いている.

3.1. 一変数の場合 (二項モデルにおけるベータ分布の ESS)

最も単純な例として, $Y = 0$ or 1 の二項モデルにおけるベータ分布 $Beta(a, b)$ の ESS を sas を用いて計算する. 前述の通りこのベータ分布の ESS は古典的に $a + b$ であることが知られている.

3.1.1. ESS_{MTM}

添付の ESS_MTM.mac に SAS マクロ関数「macro_ESS_MTM」が含まれており, これを用いて ESS_{MTM} を計算することが可能である. 計算するために以下の値が必要である:

- 事前分布の期待値 $\bar{\theta}$
- 事前分布の対数二階微分. 事前分布の情報量 $i(p(\bar{\theta}))$ の算出に用いる.
- モデルの対数二階微分. モデルの情報量 $i_{MODEL}(\theta)$ の算出に用いる.
- 無情報事前分布の対数二階微分. 無情報事前分布の情報量 $i(\pi_0(\theta))$ の算出に用いる.
- 予測分布に従うサンプルが 10000 例程度含まれるデータセット

事前分布の期待値, 対数二階微分については直接計算する必要がある. 対数二階部分は一様に簡単ではないが, Wolfram Alpha のようなオープンシステムによっても計算可能である. ベータ分布 $Beta(p | a, b) \propto p^{a-1}(1-p)^{b-1}$ について, 期待値は $a/(a+b)$, 対数二階微分は

$$\frac{d^2}{dp^2} \log Beta(p | a, b) = -\frac{a-1}{p^2} - \frac{b-1}{(p-1)^2}$$

である.

モデルの情報量について, モデル $f(Y | \theta)$ の対数二階微分の期待値をとる必要がある. 主要なモデルについては対数二階偏微分の一覧が ESS_MTM.sas に含まれている. 二項モデルの場合 (すなわちベルヌーイ分布 $Ber(Y | p) = p^Y \cdot (1-p)^{1-Y}$) では

$$\frac{\partial^2}{\partial p^2} \log Ber(Y | p) = -\frac{Y}{p^2} - \frac{1-Y}{(1-p)^2}$$

である. この対数二階微分の期待値をとる必要があるが, このためにモンテカルロ積分による近似を用いる. すなわち予測分布 $\int Ber(Y | p) Beta(p | a, b) dp$ に従うシミュレーションサンプル Y_1, Y_2, \dots, Y_n を 10000 例程度発生させることで次の近似が可能になる.

$$i_{MODEL}(p) \simeq \frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial p^2} \log f(Y_i | p)$$

このため macro_ESS_MTM では予測分布に従うサンプルが必要になる. 以下のコードによって計算可能である. 引数はそれぞれ次を意味する:

- DATA: 予測分布に従うサンプルが含まれるデータセット. 変数 Y にサンプルを与える.
- numParameter: パラメータの個数. 今回は p のみであるから 1.
- parameter_1: パラメータを表す記号. 以降の引数で用いる記号.
- priorExpectedValue_1: 事前分布の期待値. ここでは $a/(a+b)$.

- logModel_second_deriv_1: モデルの対数二階微分. ここでは $\frac{\partial^2}{\partial p^2} \log \text{Ber}(Y | p)$.
- logPrior_second_deriv_1: 事前分布の対数二階微分. ここでは $\frac{d^2}{dp^2} \log \text{Beta}(p | a, b)$.
- logNoninfo_second_deriv_1: 無情報事前分布の対数二階微分. ここでは $\frac{\partial^2}{\partial p^2} \log(p^{-1} \cdot (1-p)^{-1})$.

```

%inc 'ESS_MTM.mac' ; /* マクロの読み込み */
/* ハイパーパラメータを指定する. */
%let alpha = 3;
%let beta = 2; /* ESSはalpha + betaになることが知られている */
/* 予測分布に従うサンプルを10000例発生させる */
DATA priorPredictiveSamples;
DO n = 1 TO 10000;
    p = RAND("BETA", &alpha., &beta.);
    Y = RAND("BERNOULLI", p);
    OUTPUT;
END;
DROP n;
RUN;
/* マクロの実行 */
%macro ESS_MTM(
DATA = priorPredictiveSamples, /* 予測分布のサンプル */
numParameter = 1, /* パラメータの個数 */
parameter_1 = p, /* パラメータの記号.多変量パラメータではその分与える */
priorExpectedValue_1 = &alpha / (&alpha + &beta), /* 事前分布の期待値 */
logModel_second_deriv_1 = -Y/p**2 - (1-Y)/(1-p)**2, /* モデルの対数二階微分 */
logPrior_second_deriv_1 = -(&alpha.-1)/p**2 - (&beta.-1)/(1-p)**2, /* 事前分布の対数二階微分 */
logNoninfo_second_deriv_1 = 1/p**2 + 1/(1-p)**2 /* 無情報事前分布の対数二階微分 */
);

```

計算結果はデータセット _ESSMTM_ で与えられる.

| | ESS_MTM | Model Information | Prior Information | prior expected value of parameter_1 |
|---|--------------|-------------------|-------------------|-------------------------------------|
| 1 | 5.0205001077 | 4.149853 | 20.83333333 | 0.6 |

変数 ESS_MTM は ESS_{MTM} の計算結果であり, Model Information, Prior Information はそれぞれ $i_{MODEL}(\hat{\theta})$,

$i(p(\bar{\theta}))$ である。DATA のサンプル数を増やすことで精度を上げることができる。

logModel_second_deriv_1 や logNoninfo_second_deriv_1 について、主要なモデルでは一覧が与えられている。

3.1.2. ESS_{ELIR}

添付の ESS_ELIR.mac に SAS マクロ関数「macro_ESS_ELIR」が含まれており、これを用いて ESS_{ELIR} を計算できる。計算するために以下の値が必要である：

- 事前分布の情報量，すなわち事前分布の対数二階微分 $i(p(\theta))$
- モデルのフィッシャー情報量 $i_F(\theta)$
- 情報量の比の期待値を計算するために以下が必要
 - 事前分布の確率密度関数
 - パラメーターの積分範囲

$i(p(\theta))$ については ESS_{MTM} と同様である。主要なモデルのフィッシャー情報量については ESS_ELIR.sas に含まれている。与えられた情報量の比に対して期待値を計算する。 ESS_{ELIR} は一変数パラメータしか扱わないためこれは一次元の積分でありモンテカルロ積分による近似は必要ない。代わりに積分範囲と事前分布の確率密度関数が必要である。以下のコードによって計算可能である。引数はそれぞれ次を意味する：

- parameter: パラメータを表す記号。以降の引数で用いる記号。
- FisherInformation: モデルのフィッシャー情報量。
- logPrior_second_deriv: 事前分布の対数二階微分。
- from: 積分範囲の下限。
- to: 積分範囲の上限。
- breaks: 積分の近似に用いる分割の個数。
- priorDensity: 事前分布の確率密度関数。

```
/* マクロ読み込み*/
%inc 'ESS_ELIR.mac' ;
/* ハイパーパラメータの指定 */
%let alpha = 3;
%let beta = 2;
/* マクロを実行 */
%macro_ESS_ELIR(
  parameter = p, /* パラメータの記号 */
  FisherInformation = 1/p/(1-p), /* フィッシャー情報量 */
  logPrior_second_deriv = -(&alpha.-1)/p**2 - (&beta.-1)/(1-p)**2, /* 事前分布の対数二階微分 */
  from = 0, /* 積分範囲の下限 */
  to = 1, /* 積分範囲の上限 */
  breaks= 100, /* 積分の近似に用いる分割の個数 */
  priorDensity= pdf("BETA", p, &alpha, &beta) /* 事前分布の確率密度関数 */
);
```

データセット _ESSELIR_ に ESS_{ELIR} の計算結果が表示される。

| | ESS_ELIR | Length of interval used to calculate ESS_ELIR | Lower bound of integration | Upper bound of integration |
|---|----------|---|----------------------------|----------------------------|
| 1 | 4.995 | 1 | 0 | 1 |

ESS_ELIR は ESS_{ELIR} の計算結果であり、他は期待値の計算に用いた積分範囲を示している。代表的なモデルについてはフィッシャー情報量の一覧が ESS_ELIR.sas にあたえられている。

3.1.3. ESS_{E-F}

ESS_{E-F} では偏微分は用いない。一方で $KL(p_{-\alpha}(\boldsymbol{\theta} | t_1, \dots, t_K) || \pi_0(\boldsymbol{\theta}))$ を最小化する α, t_1, \dots, t_K を探索する必要があるためマクロの中身は非常に複雑である。主要なモデルについては簡易に計算可能なマクロをそれぞれ用意しており、二項モデルでは「macro_ESS_EF_Bernoulli」を用いて ESS_{E-F} を計算することが可能である。計算するため、以下の値が必要である。

- 事前分布に従うサンプル 10000 例程度
- Bayesian evidence interval の信頼水準. 95%が標準.
- 事前分布の対数確率密度関数
- t_1, t_2, \dots, t_k が与えられたとき $KL(p_{-\alpha}(\boldsymbol{\theta} | t_1, \dots, t_K) || \pi_0(\boldsymbol{\theta}))$ を最小にする α を求めるために以下が必要 (二分探索法を用いる)
 - α の初期値
 - 計算回数. すなわち二分割する回数
- $KL(p_{-\alpha}(\boldsymbol{\theta} | t_1, \dots, t_K) || \pi_0(\boldsymbol{\theta}))$ を最小にする t_1, t_2, \dots, t_k を Nelder Mead 法 [15] によって求める. そのために以下が必要:
 - simplex の初期値の設定方法. 自動で作成するか, 自分で作成するか選択が可能.
 - Nelder Mead 法のループ回数

注意として、計算過程で作業用のデータセットを複数用いる。同じ名前のデータセットがある場合削除される。以下のコードによって計算可能である。引数はそれぞれ次を意味する:

- data_priorSamples: 事前分布のサンプルが含まれるデータセット. parameter で指定する変数名を用いる.
- logPriorDensity: 事前分布の確率密度関数を対数変換したもの.
- initial_alpha: KLd を最小にする α を求めるための初期値.
- loop_depth: 二分探索法の分割回数.
- simplex_init_method: Nelder Mead 法で用いる simplex の初期値を指定する方法. RANDOM の場合マクロ内で作成する. そうでない場合はデータセット_current_simplex_を用意する必要がある.
- simplex_seed: RANDOM の場合シード値を指定可能. 指定しないことも可能.
- LoopTimes_NM: Nelder Mead 法のループ回数. パラメータが多いほどこの回数を増やす必要がある.
- sampleLimitingLevel: Bayesian evidence interval の信頼水準. デフォルトは 0.95.

```
/* マクロを読み込む */
```

```

%inc 'ESS_EF_given_t.mac';
%inc 'ESS_EF_NelderMead.mac';
%inc 'ESS_EF.mac';
/* ハイパーパラメータの設定 */
%let alpha = 3;
%let beta = 2;
/* 事前分布のサンプルを10000例発生させる */
DATA priorSamples;
  DO n = 1 TO 10000;
    p = RAND("BETA", &alpha, &beta);
    OUTPUT;
  END;
RUN;
/* マクロを実行 */
%macro ESS_EF_Bernoulli(
  data_priorSamples = priorSamples, /* 事前分布のサンプル */
  logPriorDensity = %str(logpdf("BETA", p, &alpha, &beta)), /* 事前分布の対数確率密度関数 */
  initial_alpha = 1, /* alphaの初期値 */
  loop_depth = 6, /* 各tにおける二分探索法のループ回数 */
  simplex_init_method = RANDOM, /* RANDOM or MANUAL */
  simplex_seed = , /* RANDOMの場合シードを指定できる */
  LoopTimes_NM = 10, /* Nelder Mead 法のループ回数 */
  sampleLimitingLevel = 0.95 /* Bayesian evidence intervalの水準 */
);

```

データセット _ESSEF_ に $ESS_{E,F}$ の計算結果が表示される。

| | Sufficient Statistic 1 | ESS_EF | _KL_ | _KL0_ | _RKL_ |
|---|------------------------|--------|--------------|--------------|--------------|
| 1 | 0.6 | 5 | 3.581979E-11 | 0.1462711301 | 2.448863E-10 |

Sufficient Statistics 1 は $KL(p_{-\alpha}(\boldsymbol{\theta} | t_1, \dots, t_K) || \pi_0(\boldsymbol{\theta}))$ を最小にする t_1, \dots, t_K の値であり, ESS_EF は $ESS_{E,F}$, $_KL_$ は KLd の最小値, $_KL0_$ は事前分布と無情報事前分布の間の KLd , $_RKL_$ は $RKL = KL_{min}/KL_0$ である. この例は $RKL=0$ であり, 理想的な例になっている.

3.2. 回帰モデルの場合 (線形回帰における多変量正規分布の ESS)

回帰モデルでは説明変数の分布が ESS に影響を与える. 例えば説明変数として年齢と投与群 (0 or 1) を用いる場合, 年齢の標準偏差は投与群の標準偏差より大きい. よって年齢の回帰係数 = 1 と投与群の回帰係数 =

1 では意味が全く異なる。したがって回帰係数の事前分布が同じであっても ESS は同じにはならない。説明変数の分布は試験前に予測することは一般に難しい。そのため説明変数に合わせて事前分布を調整するか、各説明変数を正規化、すなわち平均 0, 分散 1 になるようにスケールを直してから計算を行う必要がある。

本節では線形回帰の事前分布の ESS を計算するマクロを紹介する。上述のように説明変数の分布が影響することに注意が必要である。説明変数 x_1, x_2, x_3 に対して応答変数 Y が

$$f(Y | \beta_0, \beta_1, \beta_2, \beta_3) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_p x_p + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$

に従うとする。説明変数は互いに独立かつ標準正規分布に従うとし、事前分布は

$$\beta_i \sim N(0, 1^2)$$

とする。また簡単のためここでは σ^2 を既知とするが、未知とした場合も計算可能である。

3.2.1 ESS_{MTM}

基本的に 4.1.1 節と大きく変わらないが次の 2 点に注意が必要である。

- 予測分布のサンプルを作成する際、説明変数のサンプルを同時に作成する必要がある。
- パラメータが 4 変数であるから偏微分や期待値もその分計算する必要がある。引数 `logModel_second_deriv_1, logModel_second_deriv_2, logModel_second_deriv_3, …` のように増やせばよい。

```
%let model_var = 1;
DATA priorPredictiveSamples;
DO n = 1 TO 10000;
  x_1 = RAND("NORMAL", 0, 1); x_2 = RAND("NORMAL", 0, 1); /* 説明変数のサンプル */
  beta_0 = RAND("NORMAL", 0, 1);
  beta_1 = RAND("NORMAL", 0, 1);
  beta_2 = RAND("NORMAL", 0, 1);
  Y = beta_0 + x_1 * beta_1 + x_2 * beta_2 + RAND("NORMAL", 0, sqrt(&model_var.));
OUTPUT;
END;
DROP n beta_0 beta_1 beta_2;
RUN;
%macro ESS_MTM(
  DATA                = priorPredictiveSamples, /* 予測分布のサンプルYと説明変数のサンプルx_nを含む */
  numParameter         = 3, /* 切片を含む */
  parameter_1          = beta_0,
  parameter_2          = beta_1,
  parameter_3          = beta_2,
  priorExpectedValue_1 = 0,
  priorExpectedValue_2 = 0,
  priorExpectedValue_3 = 0,
  logModel_second_deriv_1 = - 1 / &model_var.,
  logModel_second_deriv_2 = - x_1**2 / &model_var.,
```

```

logModel_second_deriv_3 = - x_2**2 / &model_var.,
logPrior_second_deriv_2 = - 1 / 1,
logPrior_second_deriv_1 = - 1 / 1,
logPrior_second_deriv_3 = - 1 / 1,
logNoninfo_second_deriv_1 = 0, /* 無除法事前分布は一様分布 */
logNoninfo_second_deriv_2 = 0, /* 無除法事前分布は一様分布 */
logNoninfo_second_deriv_3 = 0 /* 無除法事前分布は一様分布 */
);

```

3.2.2 ESS_{E-F}

前述の通り ESS_{E-F} はモデルが指数分布族であるときにしか定義されないが、回帰分析の場合これは正準リンク (canonical link) を持つ一般化線形モデルに対応する。すなわち

$$f(Y | \boldsymbol{\beta}, \mathbf{X}) = \exp \left\{ \frac{Y \cdot \mathbf{X}^T \boldsymbol{\beta} - a(\mathbf{X}^T \boldsymbol{\beta})}{\phi} + c(Y, \phi) \right\}$$

で表されるモデルである。例えば線形回帰、ロジスティック回帰、ポアソン回帰は含まれるが、Cox 回帰は一般化線形モデルではないため含まれず、プロビット回帰は正準リンクではないため含まれない。

回帰分析では説明変数のサンプルが含まれるデータセットを用意する必要がある。

```

/* 説明変数のサンプル */
DATA explanatory_vars;
DO n = 1 TO 1000;
  x0 = 1; x1 = RAND("NORMAL", 0, 1); x2 = RAND("NORMAL", 0, 1); /* x0は切片 */
OUTPUT;
END;
RUN;

/* 事前分布のサンプル */
DATA PriorSamples;
DO n = 1 TO 10000;
  beta0 = RAND("NORMAL", 0, 1); beta1 = RAND("NORMAL", 0, 1); beta2 = RAND("NORMAL", 0, 1);
OUTPUT;
END;
RUN;

/* マクロを実行 */
%macro ESS_EF_linearRegression(
  variance_error = 1, /* 既知の分散の値 */
  data_priorSamples = PriorSamples, /* 事前分布のサンプル */
  data_explanatory_vars = explanatory_vars, /* 説明分布のサンプル */
  logPriorDensity = logpdf("NORMAL", beta0, 0, 1) + logpdf("NORMAL", beta1, 0, 1) + logpdf("NORMAL",
beta2, 0, 1),
  initial_alpha = 0.1, /* alphaの初期値 */

```

```

loop_depth = 6, /* 各tにおける二分探索法のループ回数 */
simplex_init_method = RANDOM, /* RANDOM or MANUAL */
simplex_seed = , /* RANDOMの場合シードを指定できる */
NumParameter = 3, /* パラメータの個数 */
parameter_1 = beta0, /* パラメーターの記号. 例えば切片 */
parameter_2 = beta1, /* パラメーターの記号. 例えば説明変数1の傾き */
parameter_3 = beta2, /* パラメーターの記号. 例えば説明変数2の傾き */
explanatory_var_1 = x0, /* 説明変数の記号. 切片 x0 = 1 */
explanatory_var_2 = x1, /* 説明変数の記号. */
explanatory_var_3 = x2, /* 説明変数の記号. */
LoopTimes_NM = 50,
sampleLimitingLevel = 0.95 /* Bayesian evidence intervalの水準 */
);

```

4. まとめ

本稿でベイズ統計の簡単な説明と ESS の紹介, いくつかの ESS の定義, SAS による実装を行った. ESS は事前分布の解析結果への影響の大きさを評価するうえで重要な指標であるが, その定義は簡単ではなく, 状況に合わせて適切な ESS を選択する必要がある. 以下の表は本稿で紹介した ESS を簡単に比較したものである.

| | メリット | デメリット |
|---------------------|---|---|
| 分散比による ESS | <ul style="list-style-type: none"> 定義が簡明 観察データによって値を変えることができる | <ul style="list-style-type: none"> 観察データによって値が変わらない 分散が指標として適しているとは限らない 多変数パラメータに対して定義されていない |
| ESS_{MTM} | <ul style="list-style-type: none"> 観察データによって値が変わらない 多変数パラメータにも定義できる | <ul style="list-style-type: none"> 事前分布の期待値周辺のみによって値が決まり, 事前分布が正規分布でないとき, ESS を過大評価, 過小評価する場合がある |
| ESS_{ELIR} | <ul style="list-style-type: none"> 観察データによって値が変わらない 任意の事前分布で評価ができる 予測的無矛盾性を持つ | <ul style="list-style-type: none"> 多変数パラメータでは定義されない. |
| ESS in Clarke(1996) | <ul style="list-style-type: none"> 観察データによって値が変わらない 定義が簡明 多変数パラメータにも定義できる | <ul style="list-style-type: none"> ESS が大きいとき計算量が非常に大きい KLd が小さくならないとき, ESS が正しい値である保証がない. |
| ESS_{E-F} | <ul style="list-style-type: none"> 観察データによって値が変わらない 多変数パラメータにも定義できる 予測的無矛盾性を持つ | <ul style="list-style-type: none"> モデルが指数型分布族でないとき定義されない KLd が小さくならないとき ESS が正しい値である保証がない. |

ESS はこれらがすべてではない. Morita et al. (2012) [16]は conditionally independent hierarchical models (CIHMs) とよばれる階層ベイズにおける ESS を定義している. また, Wiesenfarth and Calderazzo (2020) [17]は事前分布

と観察データの傾向が大きく異なる場合 (prior-data-conflict) に同じ ESS の事前分布でも解析結果への影響が異なることに注目し, effective current sample size (ECSS) を定義した.

SAS マクロについて, ESS_{MTM} や ESS_{ELIR} ではモデルや事前分布の二階偏微分が必要になる. 二階微分の近似は簡単ではなく, 本マクロには含まれていないためユーザーが計算する必要がある. ESS_{E-F} では計算に積分しか用いないため近似計算が可能である. 一方で KLd に関する最適化問題を解く必要があるためコードは非常に複雑である.

臨床開発が複雑化する今日, ベイズ統計の果たす役割は大きくなることが予想されている. 本稿がベイズ統計や ESS について理解する一助になれば幸いである.

参考文献

- [1] 柳本武美, “望ましい無情報事前分布の基準,” 数理解析研究所講究録, 京都大学, 2022.
- [2] H. Jeffreys, "An Invariant Form for the Prior Probability in Estimation Problems," *Proceedings of the Royal Society A*, vol. 186, no. 1007, p. 453–461, 1946.
- [3] J. M. Bernardo, "Reference Posterior Distributions for Bayesian Inference," *Journal of the Royal Statistical Society. Series B*, vol. 41, no. 2, pp. 113-147, 1979.
- [4] FDA, "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials," The United States Food and Drug Administration, Rockville, Maryland, 2010.
- [5] 平川晃弘, 佐藤宏征, 井桁正堯, 藤川桂, 堀口剛, 大門貴志, 手良向聡, “希少疾患領域の治験におけるベイズ流アプローチの利用可能性と留意事項,” *薬理と治療*, 第 49 巻, 第 1, pp. 72-79, 2021.
- [6] FDA, "Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products," The United States Food and Drug, Rockville, Maryland., 2020.
- [7] M. R. Novick and W. J. Hall, "A Bayesian Indifference Procedure," *Journal of the American Statistical Association*, vol. 60, no. 312, pp. 1104-1117, 1965.
- [8] D. Malec, "A closer look at combining data among a small number of binomial experiments," *Statistics in Medicine*, vol. 20, no. 12, p. 1811–1824, 2001.
- [9] S. Morita, P. F. Thall and P. Müller, "Determining the effective sample size of a parametric prior," *Biometrics*, vol. 64, no. 2, p. 595–602, 2008.
- [10] B. Neuenschwander, S. Weber, H. Schmidli and A. O'Hagan, "Predictively consistent prior effective sample sizes," *Biometrics*, vol. 76, no. 2, p. 578–587, 2020.
- [11] B. Clarke, "Implications of reference priors for prior information and for sample size," *Journal of the American Statistical Association*, vol. 91, no. 433, p. 173–184, 1996.
- [12] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, p. 79–86, 1951.
- [13] R. Tamanoi, "Prior effective sample size for exponential family distributions with multiple parameters," *Statistical*

Analysis and Data Mining, vol. e11685, 2024.

- [14] R. Kelter, "The evidence interval and the Bayesian evidence value: On a unified theory for Bayesian hypothesis testing and interval estimation," *British Journal of Mathematical and Statistical Psychology*, vol. 75, no. 3, p. 550–592, 2022.
- [15] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, p. 308–313, 1965.
- [16] S. Morita, P. F. Thall and P. Müller, "Prior effective sample size in conditionally independent hierarchical models," *Bayesian Analysis*, vol. 7, no. 3, p. 591–614, 2012.
- [17] M. Wiesenfarth and S. Calderazzo, "Quantification of prior impact in terms of effective current sample size," *Biometrics*, vol. 76, no. 1, p. 326–336, 2020.