

# MCMCプロシジャを用いたNormalized Power Priorの実用的な実装

武田 純

アステラス製薬株式会社 開発本部 データサイエンス部

## Practical Implementation of Normalized Power Prior with MCMC Procedure

Jun Takeda

Data Science, Global Development, Astellas Pharma Inc.

## 要旨

MCMCプロシジャ, FCMPプロシジャ, 及びモンテカルロシミュレーションを組み合わせることにより, Normalized Power Priorを実現する実用的な方法を提示する。

キーワード: Bayesian modeling, historical data, Normalized Power Prior, Modified Power Prior, MCMCプロシジャ, FCMPプロシジャ

## 発表内容

1. Bayes流モデリング
2. Power Priorの概要
3. MCMCプロシジャによるPower Priorの実装
4. 数値例
5. おわりに

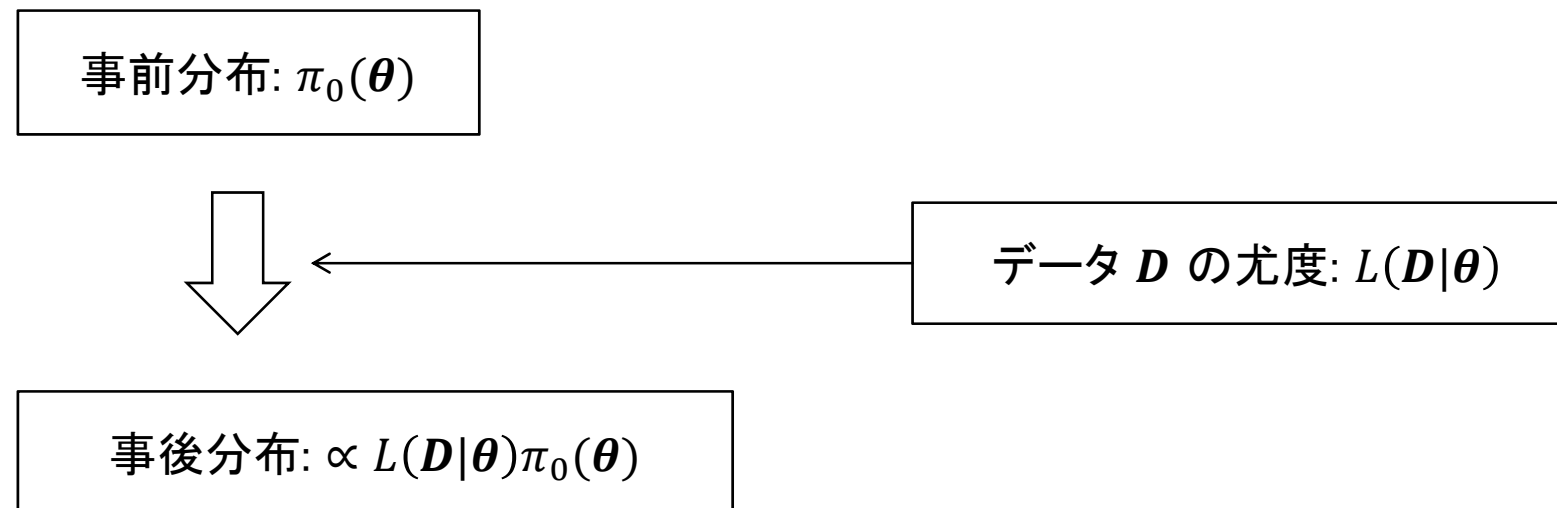
# 1. Bayes流モデリング

## Bayesの定理

$\theta$ : 推論の対象となるパラメータから成るベクトル

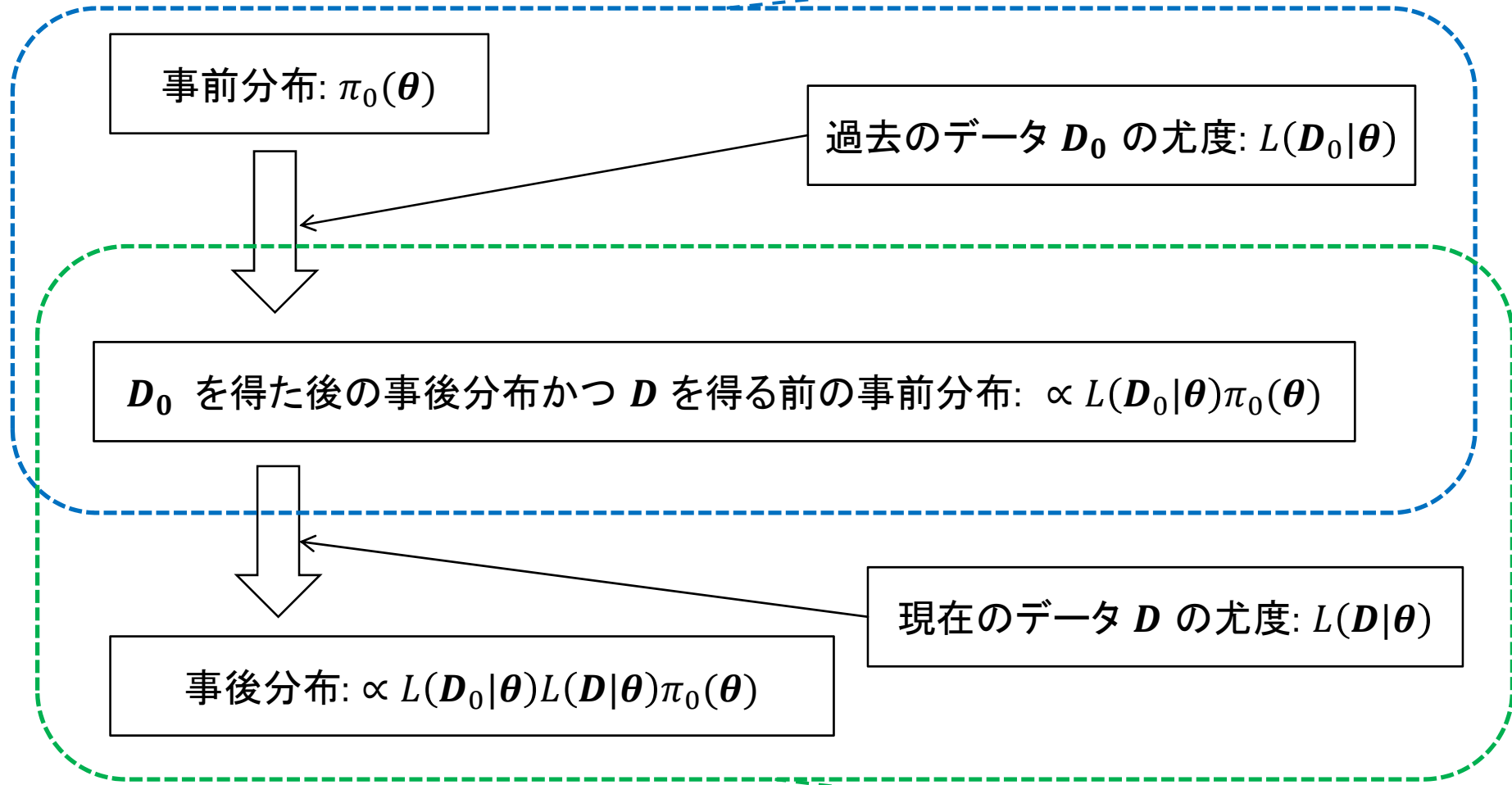
$D$ : データ (共変量も含む)

$L$ : データ生成に関する尤度 (確率関数/確率密度関数)



# Bayes更新

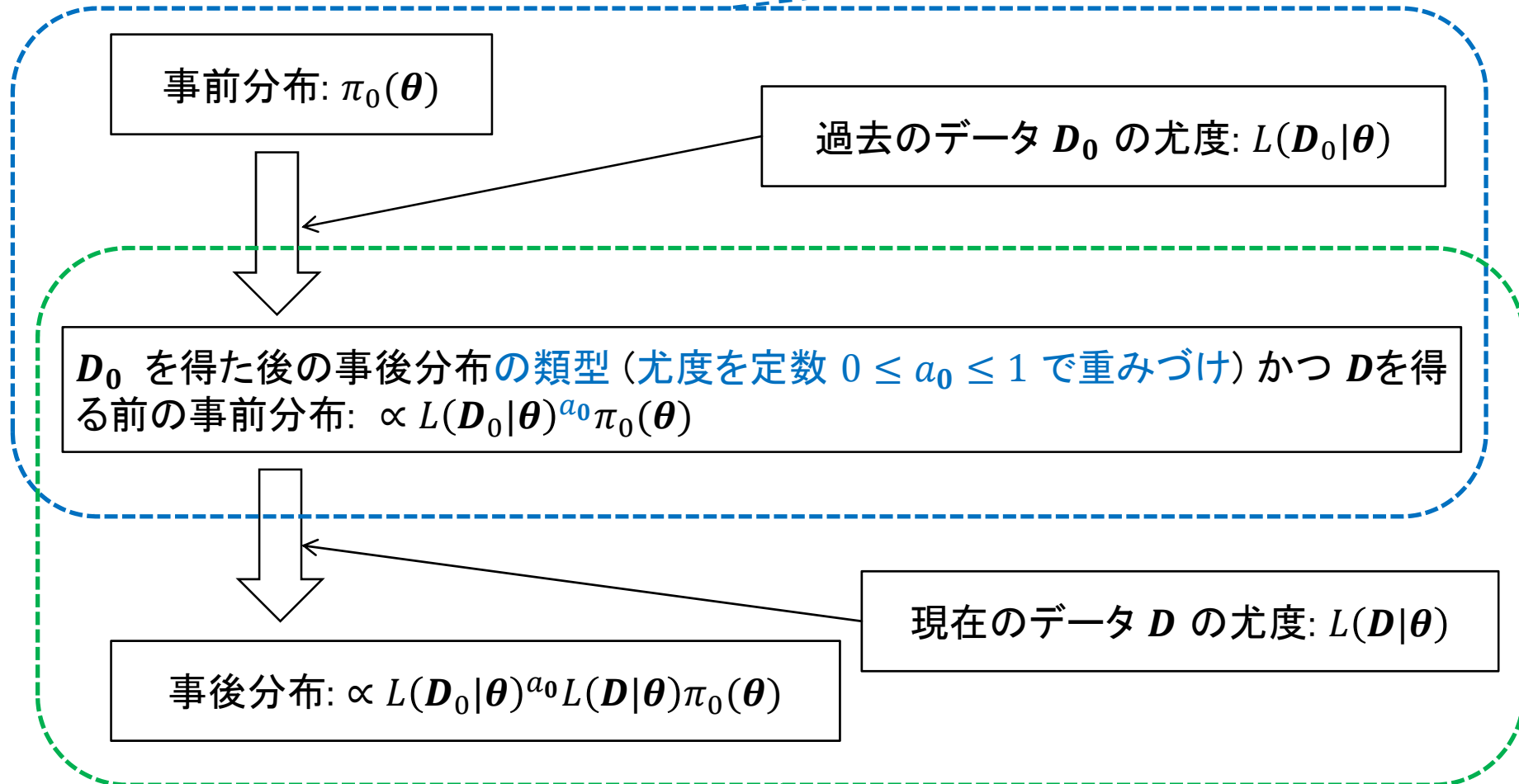
Bayesの定理1回目



Bayesの定理2回目

情報を割り引いて過去のデータを利用

Bayesの定理の類型



Bayesの定理

$L(D_0|\theta)^{a_0}\pi_0(\theta)$  は特に条件付きPower Priorと呼ばれる ([5]など)

## 2. Power Priorの概要



## 条件付きPower Prior ([5]など)

$a_0$  ( $0 \leq a_0 \leq 1$ ) を定数としたもとの、条件付きPower Priorは以下に定義される:

$$\pi_C(\boldsymbol{\theta} | \mathbf{D}_0, a_0) \propto L(\mathbf{D}_0 | \boldsymbol{\theta})^{a_0} \pi_0(\boldsymbol{\theta})$$

基準化定数を含めた式は以下となる:

$$\pi_C(\boldsymbol{\theta} | \mathbf{D}_0, a_0) = \frac{L(\mathbf{D}_0 | \boldsymbol{\theta})^{a_0} \pi_0(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} L(\mathbf{D}_0 | \boldsymbol{\theta})^{a_0} \pi_0(\boldsymbol{\theta}) d\boldsymbol{\theta}} = \frac{L(\mathbf{D}_0 | \boldsymbol{\theta})^{a_0} \pi_0(\boldsymbol{\theta})}{g(a_0 | \mathbf{D}_0)}$$

ここに  $g(a_0 | \mathbf{D}_0) = \int_{\boldsymbol{\theta}} L(\mathbf{D}_0 | \boldsymbol{\theta})^{a_0} \pi_0(\boldsymbol{\theta}) d\boldsymbol{\theta}$ 。

- $a_0 = 1$  とした場合:  $\mathbf{D}_0$  の情報を最大限に利用。
- $a_0 = 0$  とした場合:  $\mathbf{D}_0$  の情報は一切用いられない。

## Ibrahim-Chen型Power Prior ([5]など)

$a_0$  ( $0 \leq a_0 \leq 1$ ) を確率変数としたもとで, Ibrahim-Chen型Power Priorは以下に定義される:

$$\pi_{IC}(\boldsymbol{\theta}, a_0 | \mathbf{D}_0) \propto L(\mathbf{D}_0 | \boldsymbol{\theta})^{a_0} \pi_0(\boldsymbol{\theta}) \pi(a_0)$$

ここに  $\pi(a_0)$  は  $a_0$  に関する事前分布。

基準化定数を含めた式は以下となる:

$$\pi_{IC}(\boldsymbol{\theta}, a_0 | \mathbf{D}_0) = \frac{L(\mathbf{D}_0 | \boldsymbol{\theta})^{a_0} \pi_0(\boldsymbol{\theta}) \pi(a_0)}{\int_0^1 \int_{\boldsymbol{\theta}} L(\mathbf{D}_0 | \boldsymbol{\theta})^{a_0} \pi_0(\boldsymbol{\theta}) \pi(a_0) d\boldsymbol{\theta} da_0}$$

- $a_0$  の事後分布が小さな値を取りやすい → 過去のデータを推論に活用する枠組みであるにもかかわらず, 実際には過去のデータが  $\boldsymbol{\theta}$  の事後分布にほとんど影響を与えていない ([2][3][10])。
- $L(\mathbf{D}_0 | \boldsymbol{\theta})$  を定数倍すると異なる分布を導く (尤度原理を満たさない, [2])。

## Normalized Power Prior (別名 Modified Power Prior, [2][3][10])

$a_0$  ( $0 \leq a_0 \leq 1$ ) を確率変数としたもとで, Normalized Power Priorは以下に定義される:

$$\pi_{\text{NPP}}(\boldsymbol{\theta}, a_0 | \mathbf{D}_0) = \frac{L(\mathbf{D}_0 | \boldsymbol{\theta})^{a_0} \pi_0(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} L(\mathbf{D}_0 | \boldsymbol{\theta})^{a_0} \pi_0(\boldsymbol{\theta}) d\boldsymbol{\theta}} \pi(a_0) = \frac{L(\mathbf{D}_0 | \boldsymbol{\theta})^{a_0} \pi_0(\boldsymbol{\theta})}{g(a_0 | \mathbf{D}_0)} \pi(a_0)$$

- 基準化された条件付きPower Priorに  $\pi(a_0)$  を乗じた式。
- Ibrahim-Chen型ほど  $a_0$  の事後分布が小さい値を取る傾向にない。
- $L(\mathbf{D}_0 | \boldsymbol{\theta})$  を定数倍しても分布は変わらない。
- MCMCサンプラー内で  $a_0$  が更新される度に  $g(a_0 | \mathbf{D}_0)$  を求める必要がある。その度に 数値積分を行うのでは計算が高負荷になる。
- $\boldsymbol{\theta}$  が多次元の場合, 数値積分は1次元の場合に比べて格段に難しくなる。

次節で上の2つの問題を回避する方法論を提示する。

### 3. MCMCプロシジャによるPower Priorの実装

## MCMCプロシジャによるPower Priorの実装

以降、一般論と2項分布モデルのもとでの具体的な例について述べる。

2項分布モデル

現在のデータ:  $D = \{y_1, \dots, y_N, n_1, \dots, n_N\}$ ,  $n_i$  は  $i$  番目のデータの組における試行の数,  $y_i$  はそのうち成功した数,  $N$  はそのようなデータの組の数。

過去のデータ:  $D_0 = \{y_{01}, \dots, y_{0N_0}, n_{01}, \dots, n_{0N_0}\}$ , 現在のデータと同様の定義。

確率分布は以下に従うものとする:

$p \sim B(1, 1) [\pi_0(p)]$	試行の成功確率 $p$ についての事前分布。 パラメータ (1, 1) の $\beta$ 分布。
$y_{0i} \sim \text{Bin}(n_{0i}, p)$	成功確率 $p$ , 試行数 $n_{0i}$ の2項分布。
$y_i \sim \text{Bin}(n_i, p)$	成功確率 $p$ , 試行数 $n_i$ の2項分布。
$a_0 \sim B(1, 1) [\pi(a_0)]$	べき数に関する事前分布。 パラメータ (1, 1) の $\beta$ 分布。

## データ構造

### 一般論

- 過去のデータと現在のデータを合わせて1つのデータセットとする。
- オブザベーションが過去、現在のどちらに属するのかを特定する変数を追加する。
- オブザベーションのキー項目を追加する。

### 2項分布モデル

以下にデータ例:

$$N_0 = 3, \mathbf{D}_0 = \{y_{01} = 10, y_{02} = 5, y_{03} = 5, n_{01} = 25, n_{02} = 50, n_{03} = 25\},$$

$$N = 2, \mathbf{D} = \{y_1 = 5, y_2 = 15, n_{01} = 50, n_{02} = 50\}$$

データセット ALLDATA

OBS	HIST_CUR	REC_ID	Y	N
1	0	1	10	25
2	0	2	5	50
3	0	3	5	25
4	1	1	5	50
5	1	2	15	50

HIST\_CUR: 過去／現在の特定  
REC\_ID: キー項目

## 条件付きPower Priorの実装

### 一般論

- modelステートメントはプロシジャ内で指定された対数尤度をgeneral関数により呼び出す。
- データセットのオブザベーションごとの対数尤度はlogpdf関数により指定する。
- 過去のデータに対応するオブザベーションについては、対数尤度を  $a_0$  倍する。

### 2項分布モデル

$a_0 = 0.3$  のもとでのMCMCプロシジャのサンプルコード

```
proc mcmc data=ALLDATA;
  parms P 0.5;
  begincnst; A0=0.3; endcnst;----- 定数として  $a_0 = 0.3$  を指定
  prior P ~ beta(1,1);
  if (HIST_CUR=0) then do; LLIKE=A0*logpdf('binomial', Y, P, N); end;
  if (HIST_CUR=1) then do; LLIKE=logpdf('binomial', Y, P, N); end;
  model general(LLIKE);
run;
```

以上の方法論は SAS Institute Inc. (2012). Bayesian binomial model with power prior using the MCMC procedure. (SAS Institute Inc. のウェブサイト) に記載 ([11])。

## Ibrahim-Chen型Power Priorの実装

### 一般論

- parmsステートメントで  $a_0$  が確率変数であることを明示する。
- priorステートメントにより  $a_0$  の事前分布を指定する。

### 2項分布モデル

#### MCMCプロシジャのサンプルコード

```
proc mcmc data=ALLDATA;  
  parms P 0.5;  
  parms A0 0.5;  
  prior P ~ beta(1,1);  
  prior A0 ~ beta(1,1);  
  if (HIST_CUR=0) then do; LLIKE=A0*logpdf('binomial', Y, P, N); end;  
  if (HIST_CUR=1) then do; LLIKE=logpdf('binomial', Y, P, N); end;  
  model general(LLIKE);  
run;
```

べき乗に2項係数  $n!/(x!(n-x)!)$  を含めない場合

```
if (HIST_CUR=0) then do; A0*(logpdf('binomial', X, P, N)-lcomb(N, X)) end;
```



## Normalized Power Priorの実装: (a) 積分近似 [一般論]

1. 十分大きな  $M_\theta$  に対して,  $\pi_0(\boldsymbol{\theta})$  に従う独立な確率変数ベクトルの列  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{M_\theta}$  をモンテカルロシミュレーションにより生成する。
2. 十分大きな  $M_a$  に対して,  $a_{0,i} = i/M_a$  ( $i = 0, \dots, M_a$ ) を定義し, 近似により以下を得る:

$$g(a_{0,i}|\mathbf{D}_0) = \int_{\boldsymbol{\theta}} L(\mathbf{D}_0|\boldsymbol{\theta})^{a_{0,i}} \pi_0(\boldsymbol{\theta}) d\boldsymbol{\theta} \approx \frac{1}{M_\theta} \sum_{j=1}^{M_\theta} L(\mathbf{D}_0|\boldsymbol{\theta}_j)^{a_{0,i}} \stackrel{\text{def}}{=} g_{0,i}$$

3. 第  $i$  オブザベーションに  $g_{0,i-1}$  が格納されるような  $(M_a + 1)$  レコードを持つデータセットを作成する。

## Normalized Power Priorの実装: (a) 積分近似 [2項分布モデル]

$a_{0,i}=0.01$ のもとでの1つの  $p (= \theta)$  の実現値に対する  $L(\mathbf{D}_0|p)^{a_{0,i}}$  を求めるサンプルコード

```
proc sort data= ALLDATA out=HISTDAT; by HIST_CUR REC_ID; where HIST_CUR=0; run;
data LIKE_A0;
  set HISTDAT; by HIST_CUR;
  retain LLIKE;
  A0=0.01; THETA_P=rand("beta", 1, 1);
  if first.HIST_CUR then do; LLIKE=0; end;
  LLIKE0=logpdf('binomial', X, THETA_P, N);
  LLIKE=LLIKE+LLIKE0;
  if last.HIST_CUR then do;
    LIKE_A0=exp(A0*LLIKE);
  output;
end;
run;
```

$p$  の実現値  
 過去のデータの抽出  
 対数尤度の計算  
 対数尤度  $\rightarrow$  尤度の  $a_{0,i}$  乗

1. 以上の処理を  $M_a$  ( $a_{0,i}$  の刻みの細かさ)  $\times$   $M_\theta$  (シミュレーションの回数) 回反復する。
2.  $a_{0,i}$  ごとに平均を求めて  $g_{0,i}$  に格納。
3. 更に  $g_{0,0}, g_{0,1}, \dots, g_{0,M_a}$  を縦積みにしたデータセットMAT1を作成。

Normalized Power Priorの実装:

(b) FCMPプロシジャによる関数の定義 [一般論]

- 関数の引数は  $a_0$  ( $0 \leq a_0 \leq 1$ ) とする。
- read\_array 関数を用いて、データセットに格納された  $g_{0,0}, \dots, g_{0,M_a}$  の値をFCMPプロシジャに渡す。
- 関数の返り値を線形補完により求める。
  1.  $a_{0,i'} < a_0 < a_{0,i'+1}$  となる  $i'$  を特定する。
  2.  $(a_0 - a_{0,i'}) \times (g_{0,i'+1} - g_{0,i'}) / (a_{0,i'+1} - a_{0,i'}) + g_{0,i'}$  を返り値とする  
( $a_0 = a_{0,i'}$  の場合は  $g_{0,i'}$  を返り値とする)。

Normalized Power Priorの実装:

(b) FCMPプロシジャによる関数の定義 [2項分布モデル]

以下FCMPプロシジャにより,  $g(a_0|\mathbf{D}_0)$  を近似する関数  $\hat{g}(a_0|\mathbf{D}_0)$  が定義される。

```
proc fcmp outlib=WORK.TEMP.G;
function G(A0);
  array MAT2[%eval(&M_A.+1), 1] / nosymbols;
  RC=read_array("MAT1", MAT2);
  I_A0=A0*(&M_A.)+1;
  I_A0_R=ceil(A0*(&M_A.) +1);
  I_A0_L=floor(A0*(&M_A.)+1);
  Y_L=MAT2[I_A0_L, 1];
  Y_R=MAT2[I_A0_R, 1];
  if I_A0_L=I_A0_R then do; Y=Y_L; end;
  else do; Y=(I_A0-I_A0_L)*(Y_R-Y_L)/(I_A0_R-I_A0_L)+Y_L; end;
  return(Y);
endsub;
run;
options cmplib = WORK.TEMP;
```

MAT2は  $(M_a + 1) \times 1$  行列  
 $g_{0,0}, g_{0,1}, \dots, g_{0,M_a}$  の値を行列  
MAT2として読み込む。

線形補完

## Normalized Power Priorの実装:

### (c) MCMCプロシジャにおけるの実装 [一般論]

- 対数尤度において  $\log [\hat{g}(a_0 | \mathbf{D}_0)]$  の引き算が行われるようにする。
- $\hat{g}(a_0 | \mathbf{D}_0)$  は過去のデータ全体に対する尤度の基準化であることから、MCMCプロシジャに渡されるデータセットのレコードが複数であってもこの引き算は1回のみ行われるように指定する。

## Normalized Power Priorの実装:

## (c) MCMCプロシジャにおけるの実装 [2項分布モデル]

## MCMCプロシジャのサンプル

```
proc mcmc data=ALLDATA;
  parms P 0.5;
  parms A0 0.5;
  prior P ~ beta(1,1);
  prior A0 ~ beta(1,1);
  if (HIST_CUR=0) then do; LLIKE=A0*logpdf('binomial', Y, P, N); end;
  if (HIST_CUR=0 and REC_ID=1) then do; LLIKE=LLIKE-log(G(A0)); end;
  if (HIST_CUR=1) then do; LLIKE=logpdf('binomial', Y, P, N); end;
  model general(LLIKE);
run;
```

この行のみ追加

過去のデータの第1レコードのみにおいて  $\hat{g}(a_0 | \mathbf{D}_0)$  による割り算 (対数尤度上では引き算) が行われる。

## 4. 数値例

## シナリオ（[10]のシナリオに追加）

過去、現在それぞれ2項分布モデルのもと1レコードで表現されたデータ

シナリオ	過去のデータ ( $x_0/n_0$ )	現在のデータ ( $x/n$ )	解釈
1-1	6/30	6/30	過去と現在のデータが 同じ性質を持つもとで データ量が異なる。
1-2	20/100	20/100	
1-3	60/300	60/300	
1-4	200/1000	200/1000	
2-1	3/30	6/30	過去と現在のデータが 異なる性質を持つもとで データ量が異なる。
2-2	10/100	20/100	
2-3	30/300	60/300	
2-4	100/1000	200/1000	



## 比較するPower Prior

1. 2項係数を含まないIbrahim-Chen型Power Prior。以下の尤度を用いる ([10]で使用):

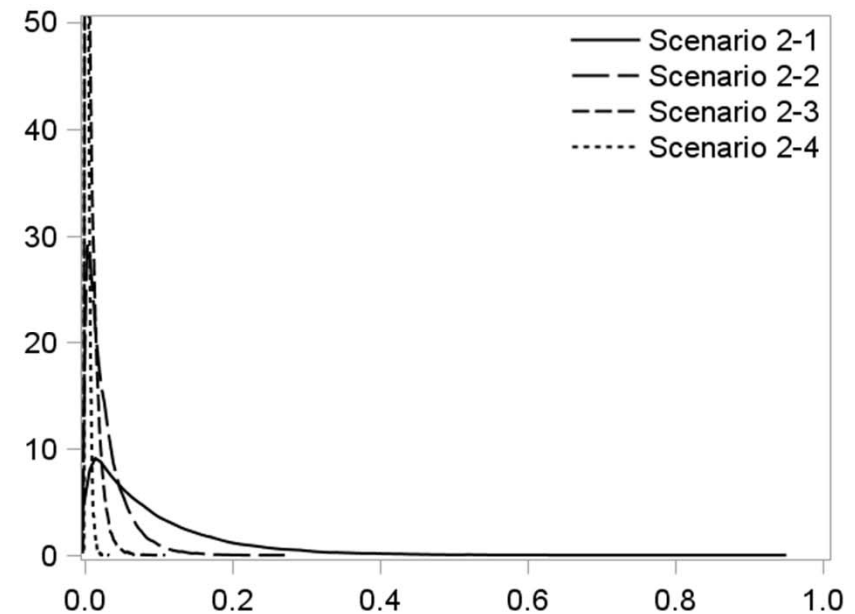
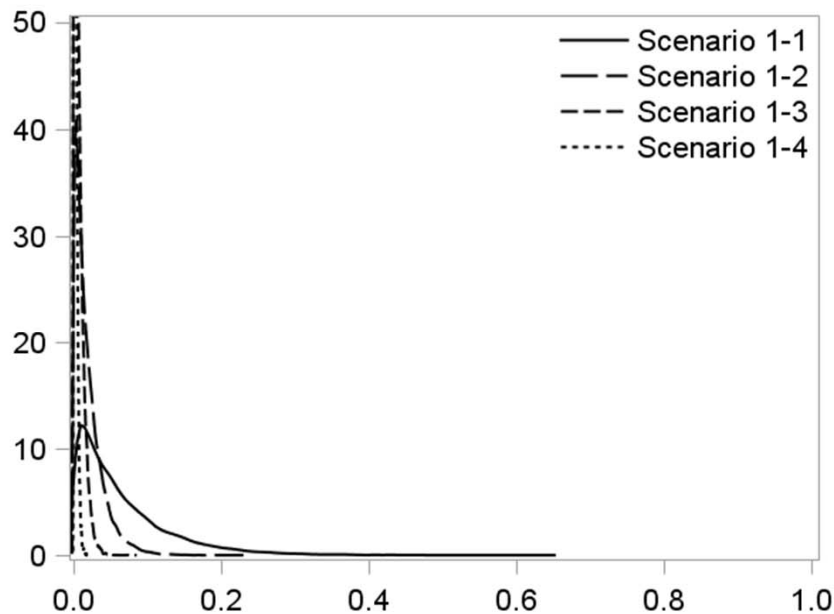
$$L_1(\mathbf{D}_0|p) = p^{y_0}(1-p)^{n_0-y_0}$$

2. 2項係数を含むIbrahim-Chen型Power Prior。以下の尤度を用いる:

$$L_2(\mathbf{D}_0|p) = \frac{n_0!}{y_0!(n_0-y_0)!} p^{y_0}(1-p)^{n_0-y_0}$$

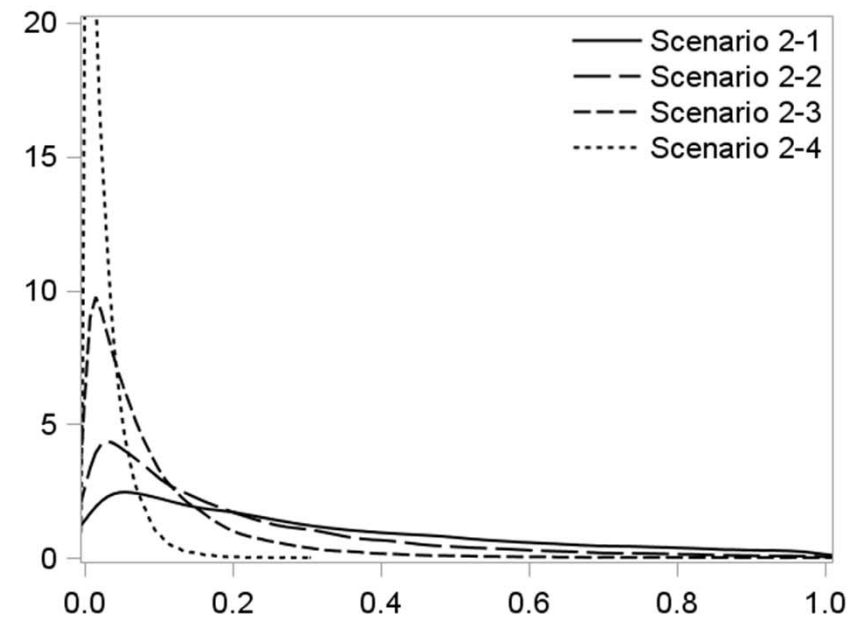
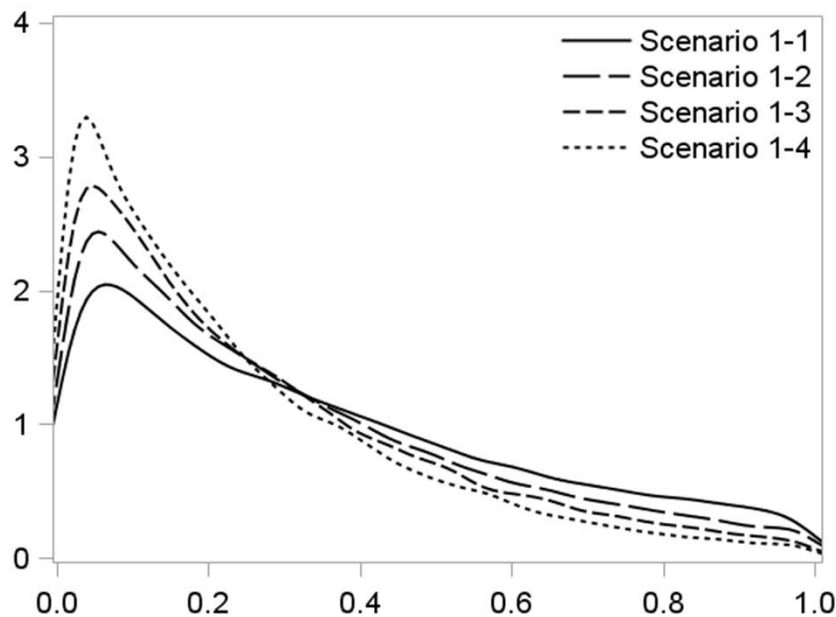
3. Normalized Power Prior ([10]で使用)。理論上  $L_1(\mathbf{D}_0|p)$ ,  $L_2(\mathbf{D}_0|p)$  のいずれを用いても同じ分布。数値計算においては  $L_2(\mathbf{D}_0|p)$  を使用。

べき数  $a_0$  の事後分布のカーネル密度推定:  
2項係数を含まないIbrahim-Chen型Power Prior



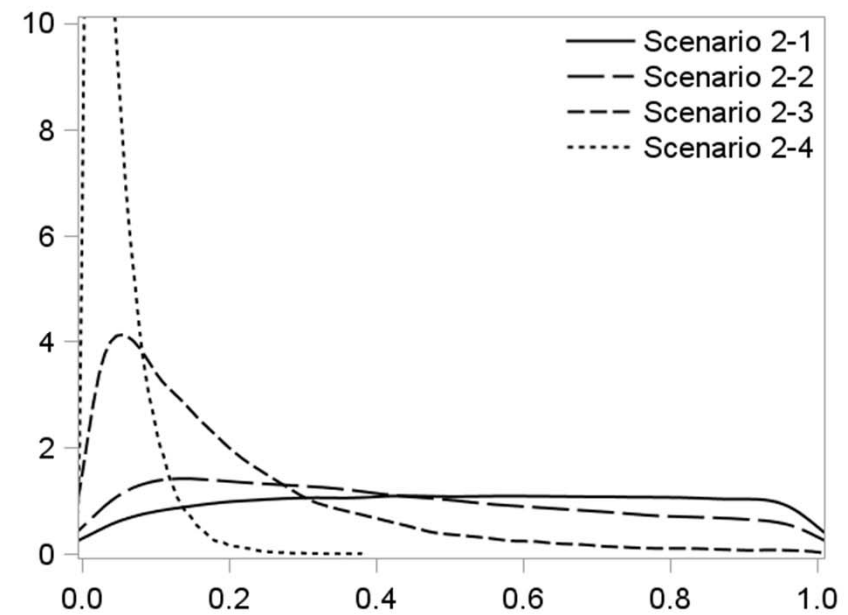
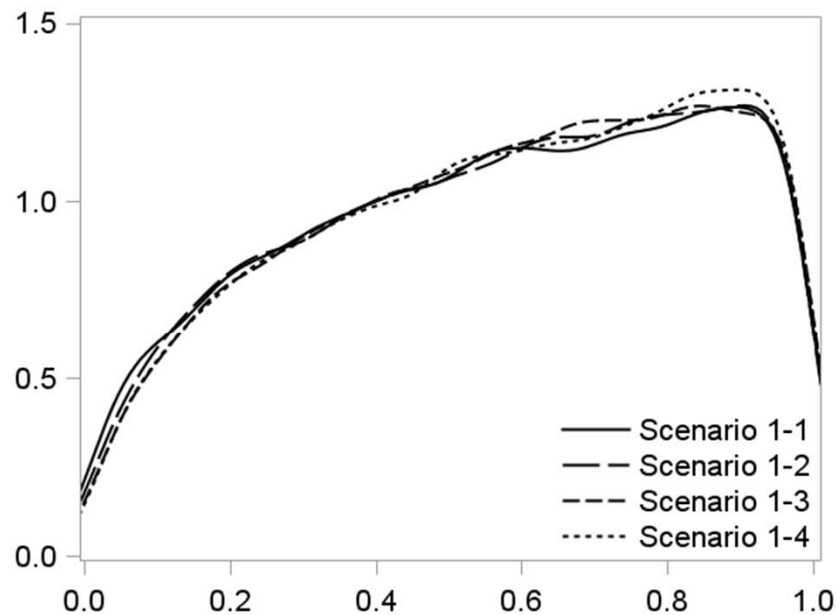
0付近に極端に集中している。過去のデータの情報はほとんど用いられていない。  
この傾向はデータ量が多いほど顕著。

べき数  $a_0$  の事後分布のカーネル密度推定:  
2項係数を含むIbrahim-Chen型Power Prior



2項係数を含まない場合ほどではないが、0付近への集中がみられる。

べき数  $a_0$  の事後分布のカーネル密度推定:  
Normalized Power Prior



シナリオ1-1～1-4ではほぼ変わらない分布。0への集中は起きていない。  
シナリオ2-1～2-4で、情報量の増大に対して急速に0へ収束。

2項比率  $p$  の事後分布の平均と95%信用区間

シナリオ	2項係数を含まない Ibrahim-Chen 型 Power Prior	2項係数を含む Ibrahim-Chen型Power Prior	Normalized Power Prior
1-1	0.217 (0.098, 0.367)	0.215 (0.103, 0.353)	0.213 (0.110, 0.339)
1-2	0.206 (0.134, 0.289)	0.205 (0.139, 0.280)	0.204 (0.144, 0.273)
1-3	0.202 (0.159, 0.249)	0.202 (0.162, 0.244)	0.201 (0.166, 0.240)
1-4	0.201 (0.176, 0.226)	0.200 (0.178, 0.224)	0.201 (0.181, 0.221)
2-1	0.210 (0.093, 0.361)	0.195 (0.088, 0.338)	0.182 (0.086, 0.312)
2-2	0.203 (0.131, 0.285)	0.191 (0.124, 0.271)	0.177 (0.116, 0.256)
2-3	0.201 (0.157, 0.248)	0.194 (0.151, 0.241)	0.187 (0.143, 0.234)
2-4	0.200 (0.176, 0.226)	0.198 (0.174, 0.224)	0.196 (0.172, 0.222)

Normalized Power Priorの特徴:

- シナリオ1-1～1-4では過去のデータの情報を最も活用しているため信用区間幅が狭い。
- シナリオ2-1～2-4とデータ量が増えるにつれ、過去のデータの活用を避ける。

## 5. おわりに

## Normalized Power Priorの実用的な実装

- FCMPプロシジャで定義された関数をMCMCプロシジャから呼び出すことが可能であることを活用した。
- 以下により関数呼び出しの度に数値計算を行うことを回避した:
  - 指定された値に対応する積分値を事前に計算しておく。
  - 関数呼び出しの際に求めてある積分値を線形補完して返り値を求める。
- 事前に求めておく積分値は、被積分関数の性質を利用しモンテカルロシミュレーションにより求めた。

## Normalized Power Priorの性質（2項分布モデルの数値例より）

- 過去と現在のデータの性質が同じ場合は、Normalized Power PriorがIbrahim-Chen型Power Priorと比べて、過去のデータの情報をより活用する。
- 過去と現在のデータの性質が異なる場合、それぞれの情報量が増えるにつれ、Normalized Power Priorも過去のデータの情報を活用しなくなる。
- 以上をまとめると、Normalized Power Priorはデータ適応的な重みづけを行う性質を持つ。



## 参考文献 (1)

- [1] Hobbs BP, Carlin BP, Mandrekar SJ, and Sargent DJ. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, **67**, 1047-1056.
- [2] Duan Y, Ye K and Smith EP. (2006). Evaluating water quality using power priors to incorporate historical information. *Environmetrics*, **17**, 95–106.
- [3] Duan Y, Smith EP, and Ye K. (2006). Using power priors to improve the binomial test of water quality. *Journal of Agricultural, Biological and Environmental Statistics*, **11**, 151–168.
- [4] Gajewski BJ. (2010). Comments on ‘A note on the power prior’ by Neuenschwander B, Branson M and Spiegelhalter DJ. *Statistics in Medicine*, **29**, 708-710.
- [5] Ibrahim JG and Chen MH. (2000). Power prior distributions for regression models. *Statistical Science*, **15**, 46–60.
- [6] Ibrahim JG, Chen MH, and Chu H. (2012). Bayesian methods in clinical trials: a Bayesian analysis of ECOG trials E1684 and E1690. *BMC Medical Research Methodology*, **12**, 183.
- [7] 小西貞則, 越智義道, 大森裕浩 (2008). 計算機統計学の方法 –ブートストラップ・EMアルゴリズム・MCMC–. 朝倉書店.

## 参考文献 (2)

- [8] Lunn D, Jackson C, Best N, Thomas A, and Spiegelhalter D (2012). *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Chapman and Hall.
- [9] Neelon B and O'Malley AJ. (2010). Bayesian analysis using power priors with application to pediatric quality of care. *Journal of Biometrics & Biostatistics*, **1**, 103.
- [10] Neuenschwander B, Branson M, and Spiegelhalter DJ. (2009). A note on the power prior. *Statistics in Medicine*, **28**, 3562-3566.
- [11] SAS Institute Inc. (2012). Bayesian binomial model with power prior using the MCMC procedure. (2014/06/27に SAS Institute Inc. のウェブサイトにて存在を確認).