2017年12月10日 日本テスト学会賞記念講演
ワークショップ 明治学院大学

ベイズ統計的アプローチによる項目反応モデルの拡張

専修大学 人間科学部
岡田 謙介
(ken@psy.senshu-u.ac.jp)

全体の流れ

1. ベイズ統計学の考え方
   - 未知の・知りたいことを確率によって表す
   - データを使って確率を更新する
   - JAGSやStanによって実装・推定する
   - この枠組みで，適用場面や研究関心，利用可能な情報によってIRTモデルを柔軟に拡張し，推定・解釈することが可能になる
2. 研究例A：反応スタイルのモデリング
   - 係留ビネット法と２次元IRTモデルの利用
3. 研究例B：反応時間情報の取り入れ
   - 数理心理学で発展したLBAモデルとIRTの統合
4. まとめと展望

1. ベイズ統計学の考え方

ベイズ統計学の考え方

- 知りたいこと・不確実なことを**確率**で表現する
- **データ**の情報から，その確率を**更新**する

■ **推定**

手に入ったデータy

知りたいこと
パラメータθ

尤度　事前確率(分布)

$$p(\theta|y) = \frac{\boxed{p(y|\theta)\,p(\theta)}}{p(y)}$$

統計モデル

事後確率(分布)

▶ ベイズの定理

■ **予測**

事後予測確率(分布)

$$p(\tilde{y}|y) = \int_{\Theta} p(\tilde{y}|\theta)p(\theta|y)d\theta$$

尤度　事後確率(分布)

将来のデータ$\tilde{y}$

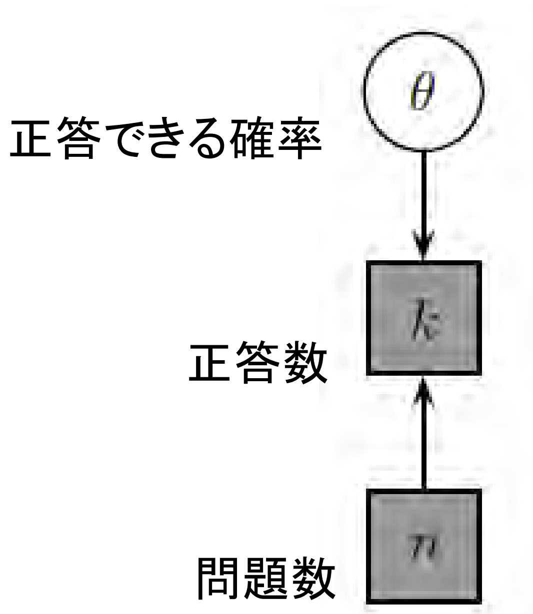
■ **モデル比較** → ベイズファクター（後ほど）

例：比率（正答確率）の推定

- 難易度が同じ課題が独立に20問提示される
- 結果は正答か誤答かのどちらか 全20問
- 個体の正答確率θを推定したい

- モデル：

θ ~ Beta(1, 1)
とも書ける



θ ~ Uniform(0, 1)
（一様分布） 事前分布

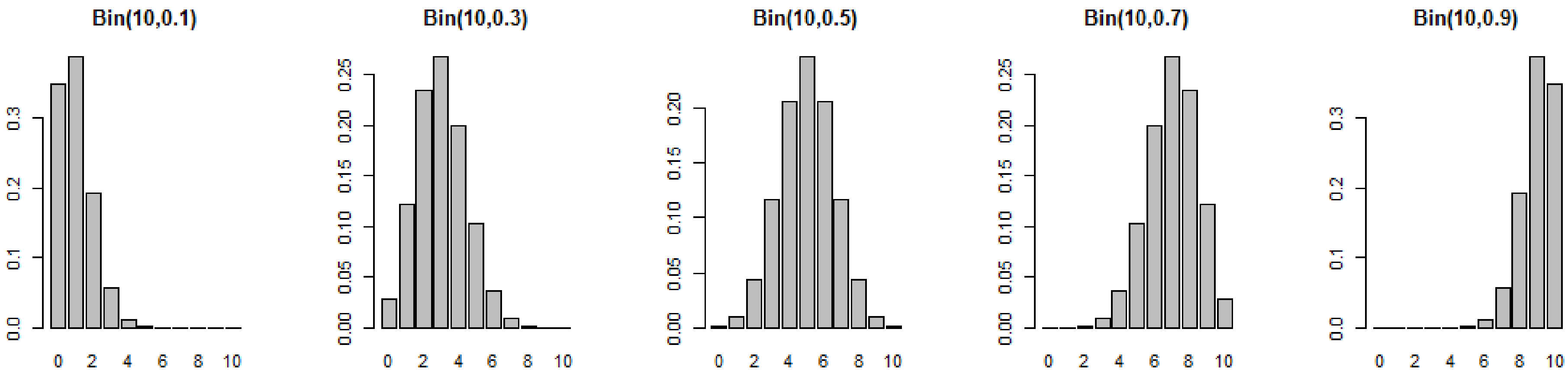
k ~ Binominal(n, θ)
（二項分布） 尤度

□：離散変数
○：連続変数
色つき：観測
色なし：非観測

二項分布

- 成功確率θの試行を独立にn回繰り返したときの成功回数kの確率分布

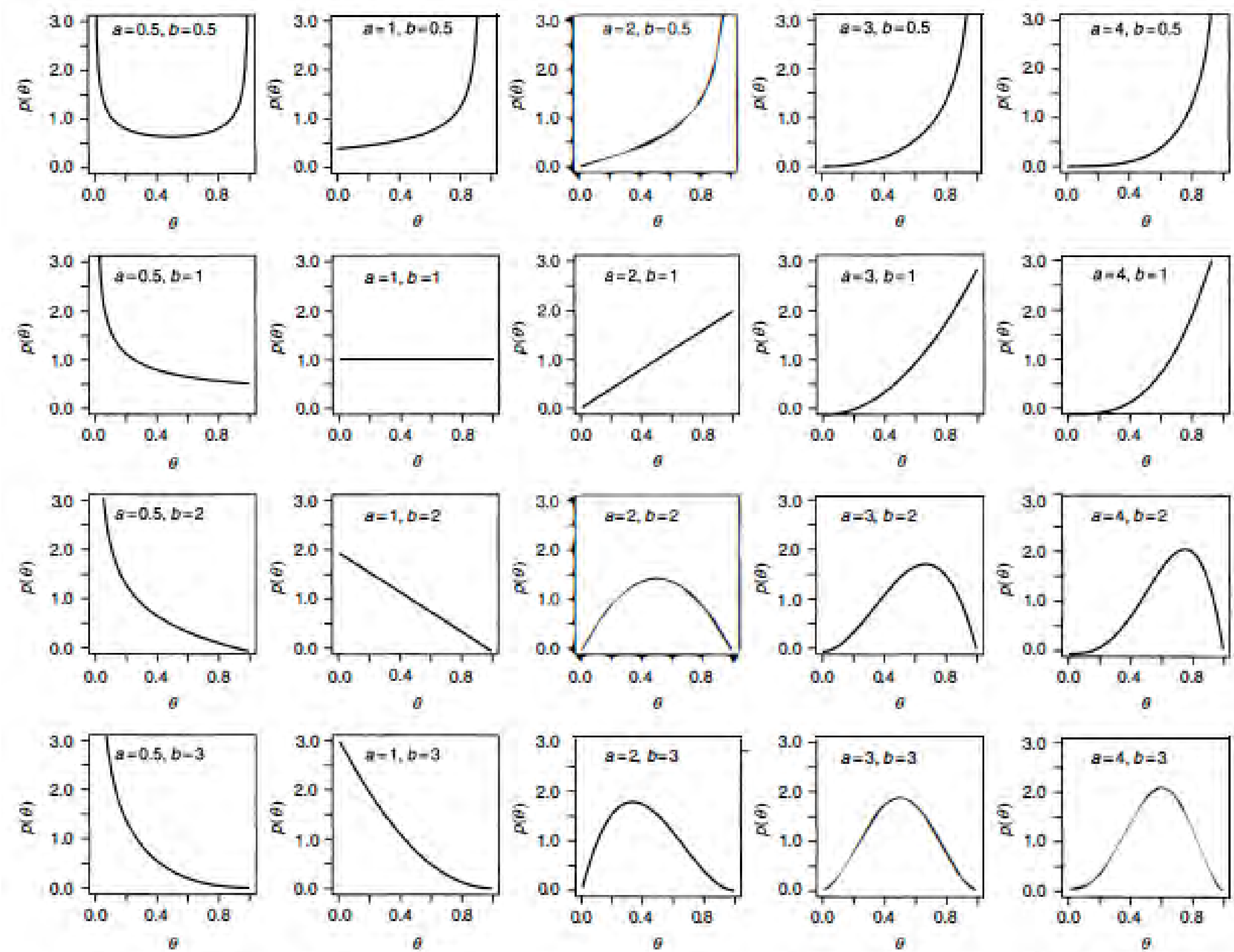
$$k \sim Binomial(n, \theta) \quad \Leftrightarrow \quad P(k|n, \theta) = {}_nC_k\, \theta^k\, (1-\theta)^{(n-k)}$$



ベータ分布

$$\theta \sim Beta(\alpha, \beta) \quad \Leftrightarrow \quad p(\theta|\alpha,\beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$$

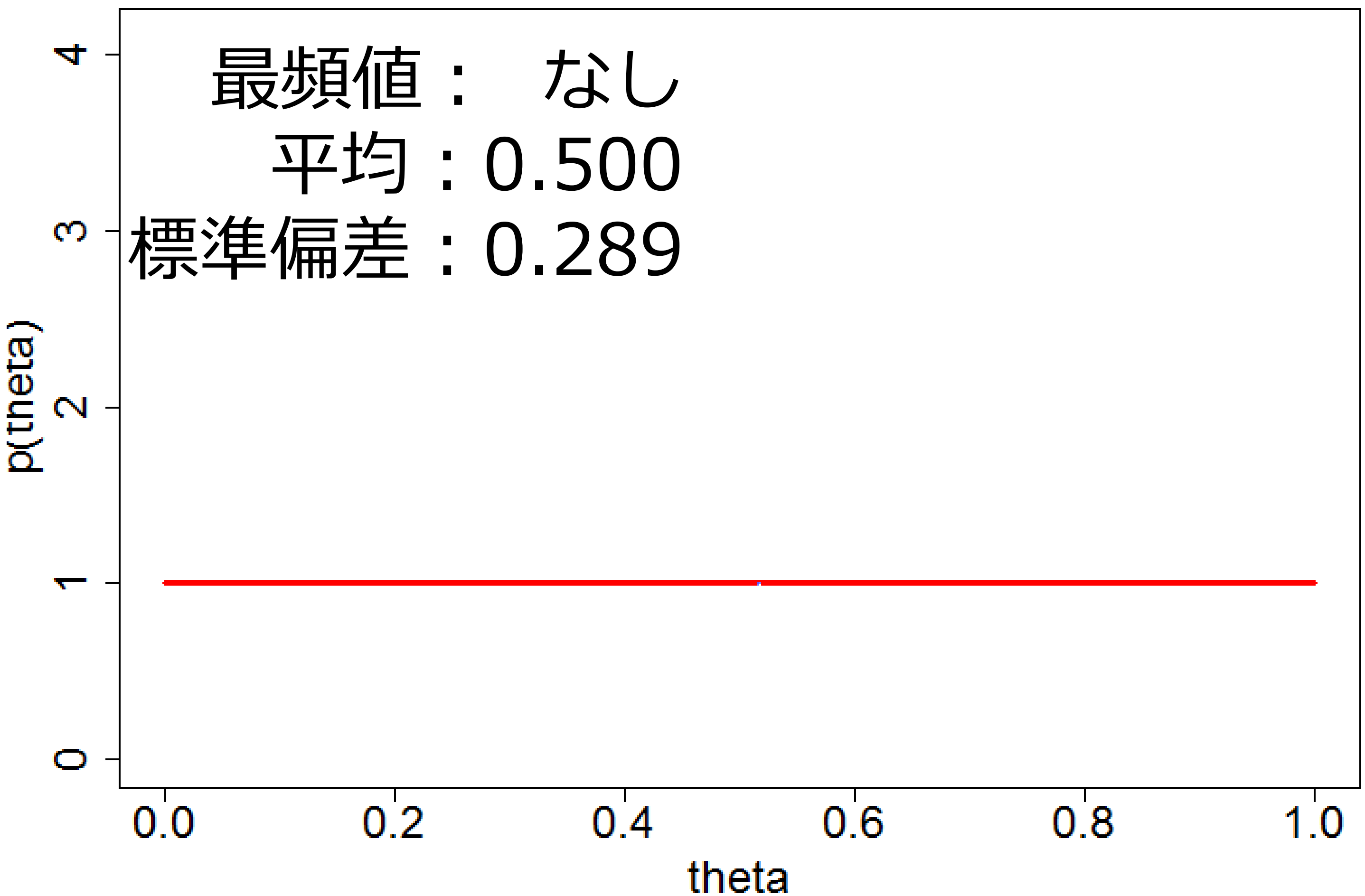
ベータ関数



ベイズ更新の例

- データが得られる前の状態は，能力θに対する一様事前分布で表現できるとする

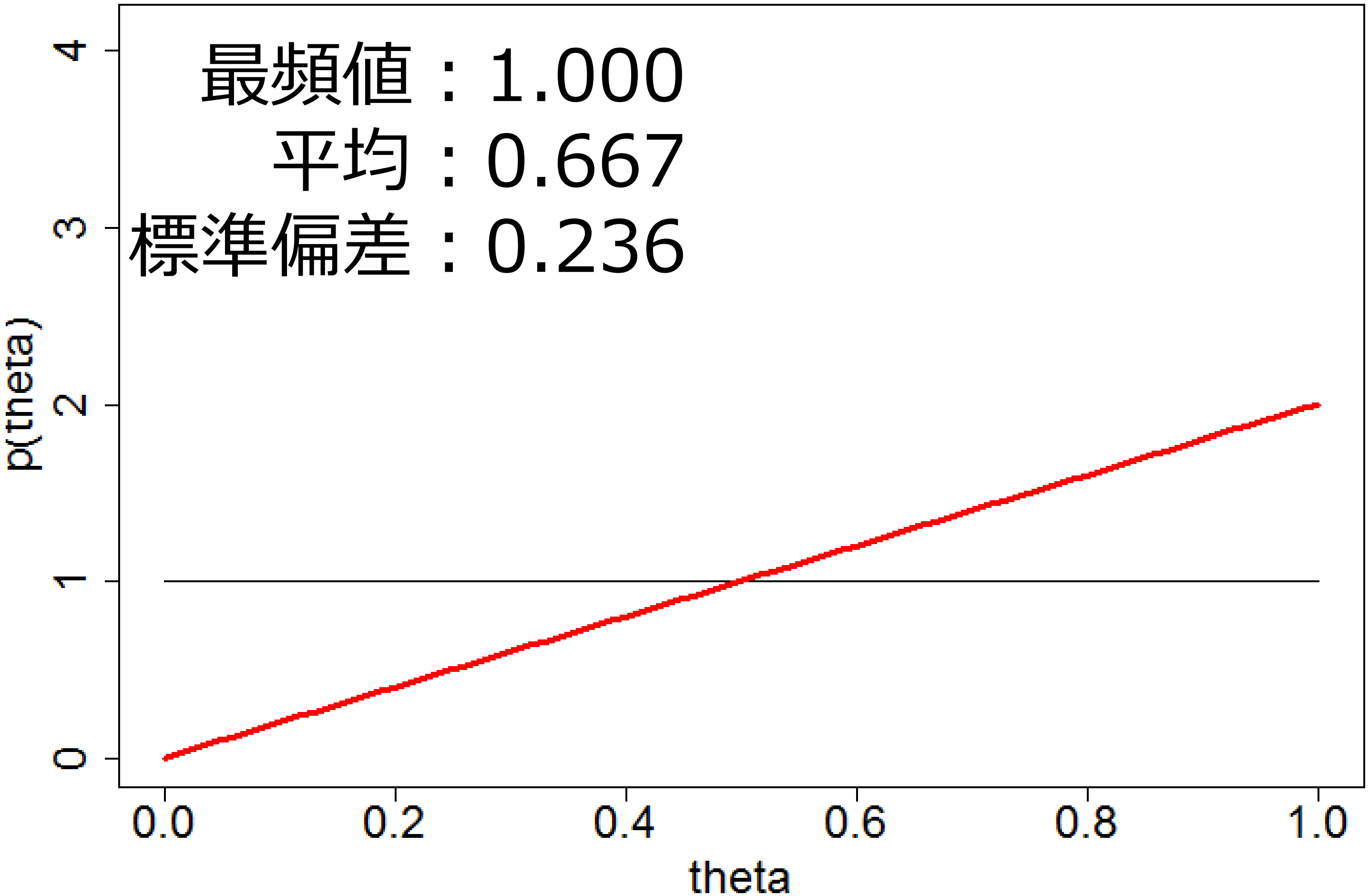
$$\theta \sim Beta(1,1)$$



ベイズ更新の例

- 1問め:**正答**

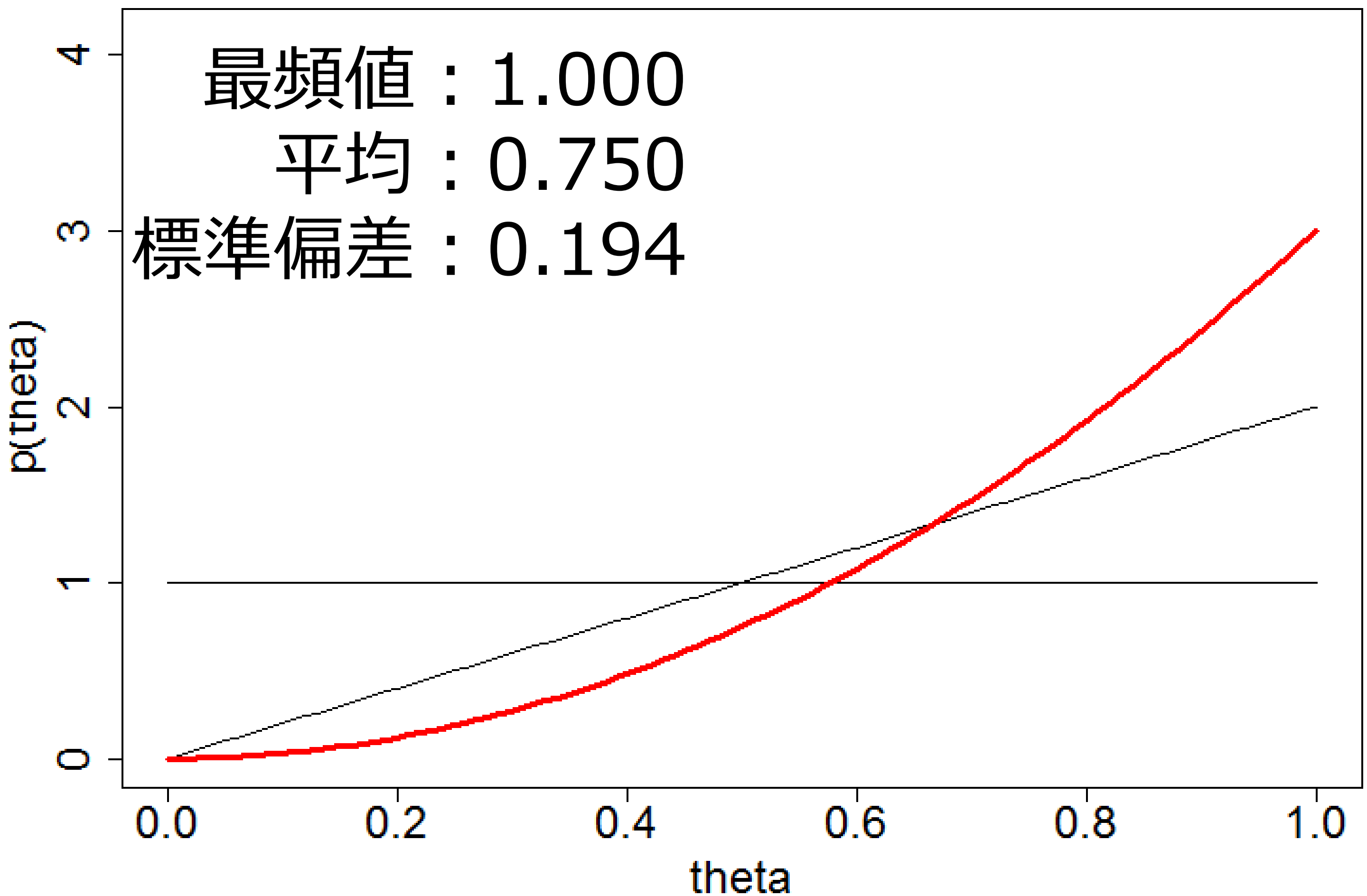
$\theta \sim Beta(2,1)$



ベイズ更新の例

- 1問め:**正答**→2問め:**正答**

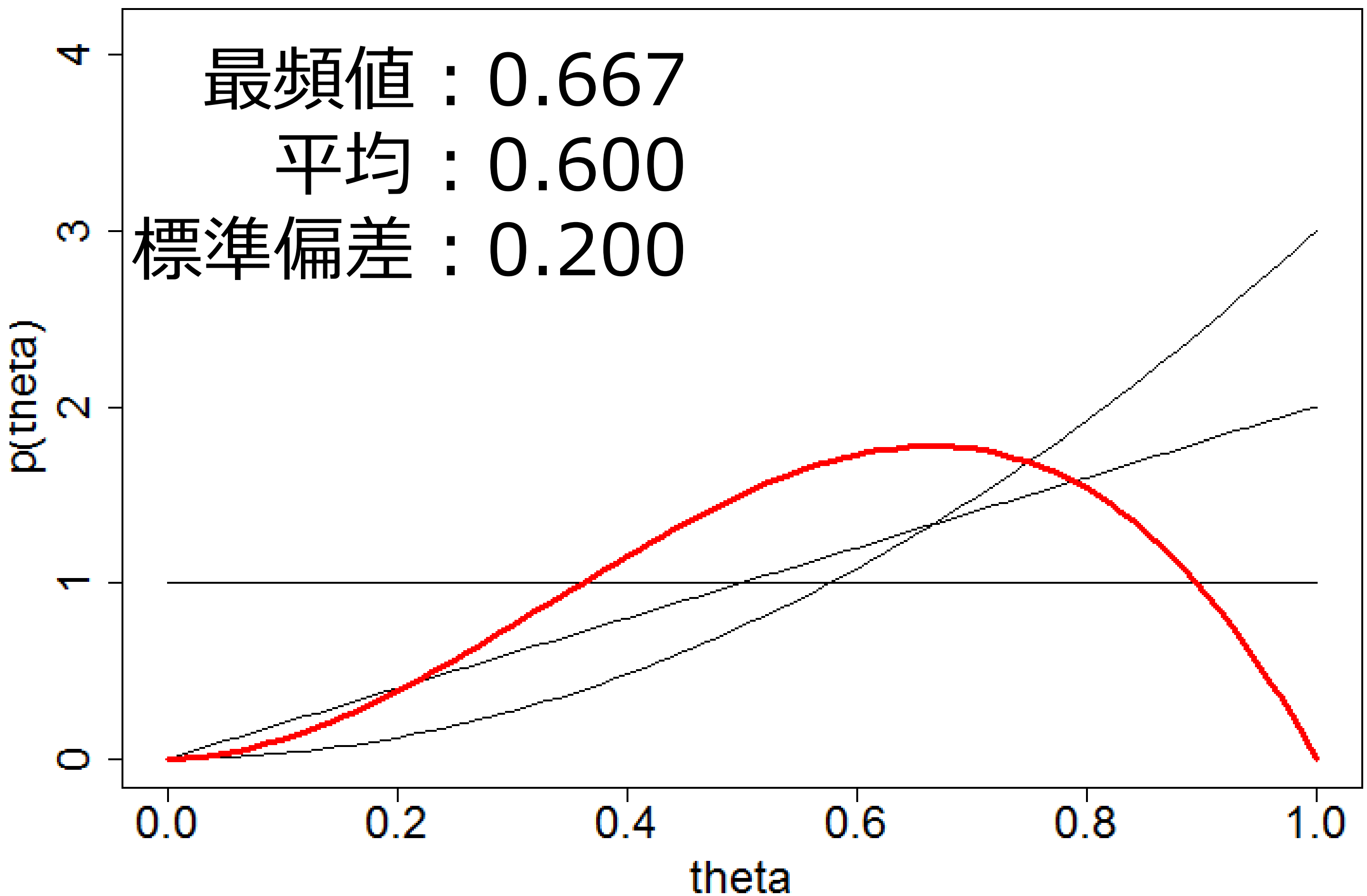
$\theta \sim Beta(3,1)$



ベイズ更新の例

- 1問め:**正答**→2問め:**正答**→3問め:**誤答**

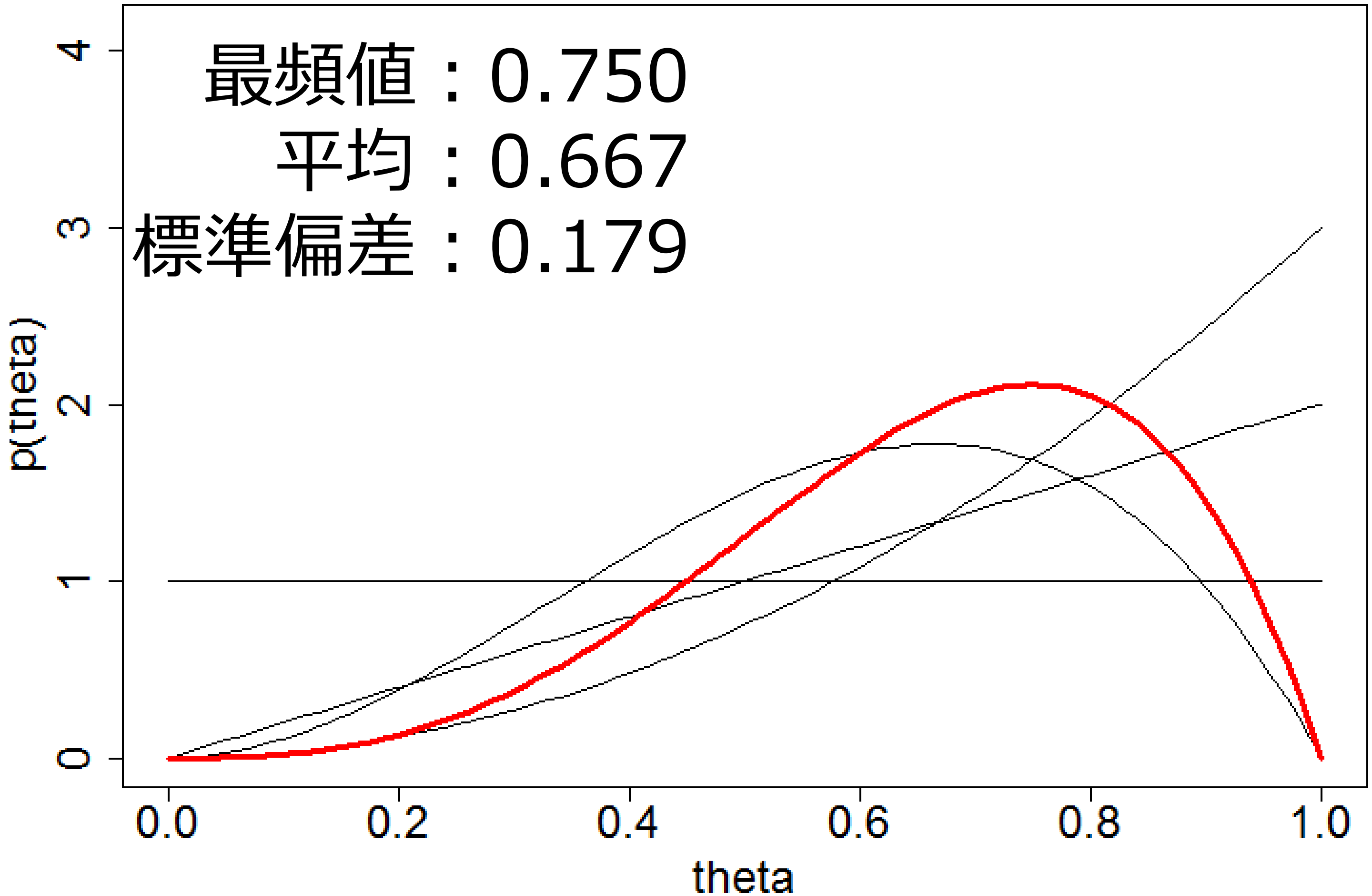
$$\theta \sim Beta(3,2)$$



ベイズ更新の例

- 1問め:**正答**→2問め:**正答**→3問め:**誤答**→4問め:**正答**

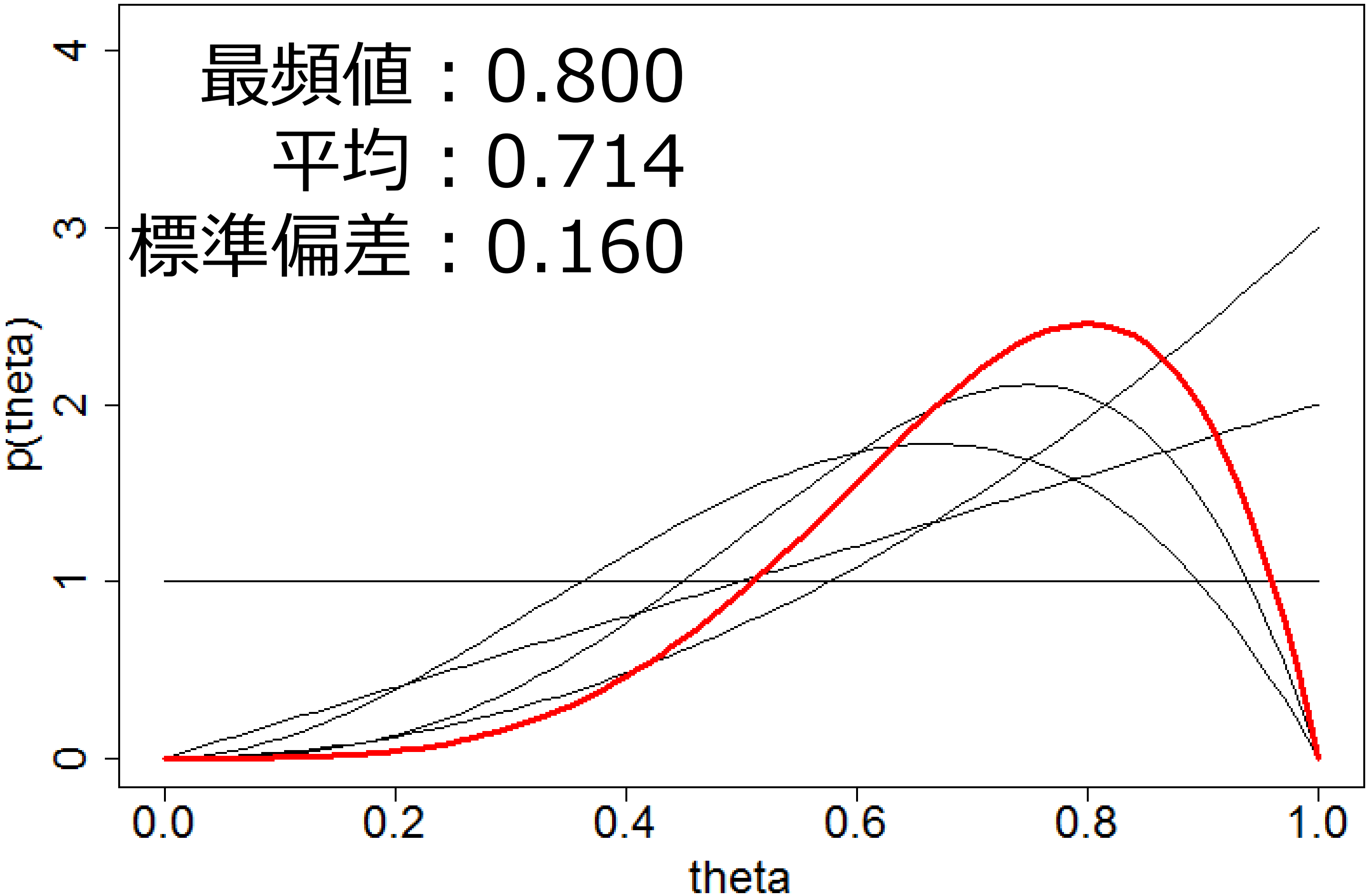
$\theta \sim Beta(4,2)$



ベイズ更新の例

- 1問め:**正答**→2問め:**正答**→3問め:**誤答**→4問め:**正答**→5問め:**正答**

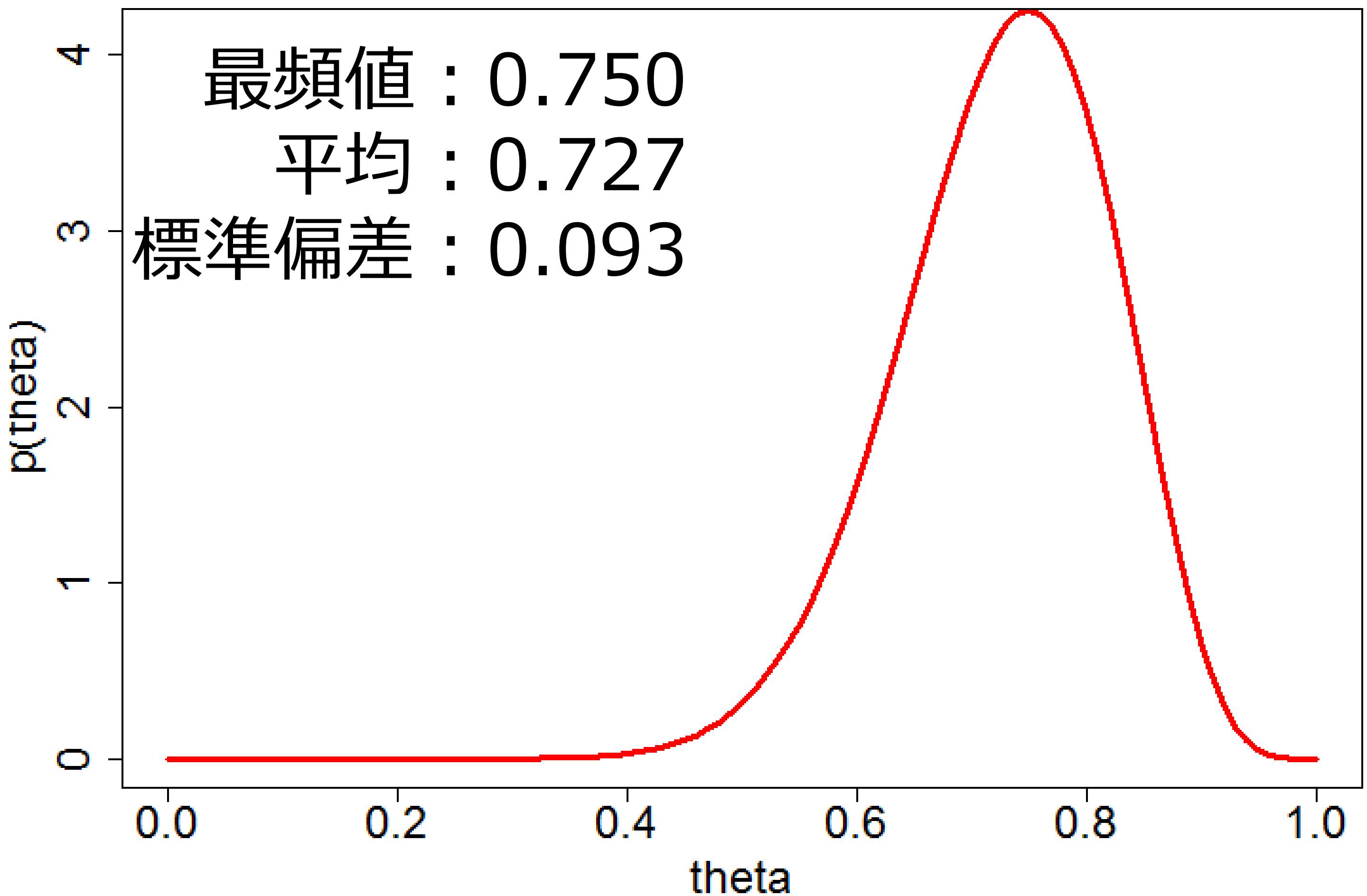
$\color{red}{\theta} \sim Beta(5,2)$



ベイズ更新の例

- 1問め:**正答**→2問め:**正答**→3問め:**誤答**→4問め:**正答**→5問め:**正答** → … →20問回答後（**正答**15回，**誤答**5回）

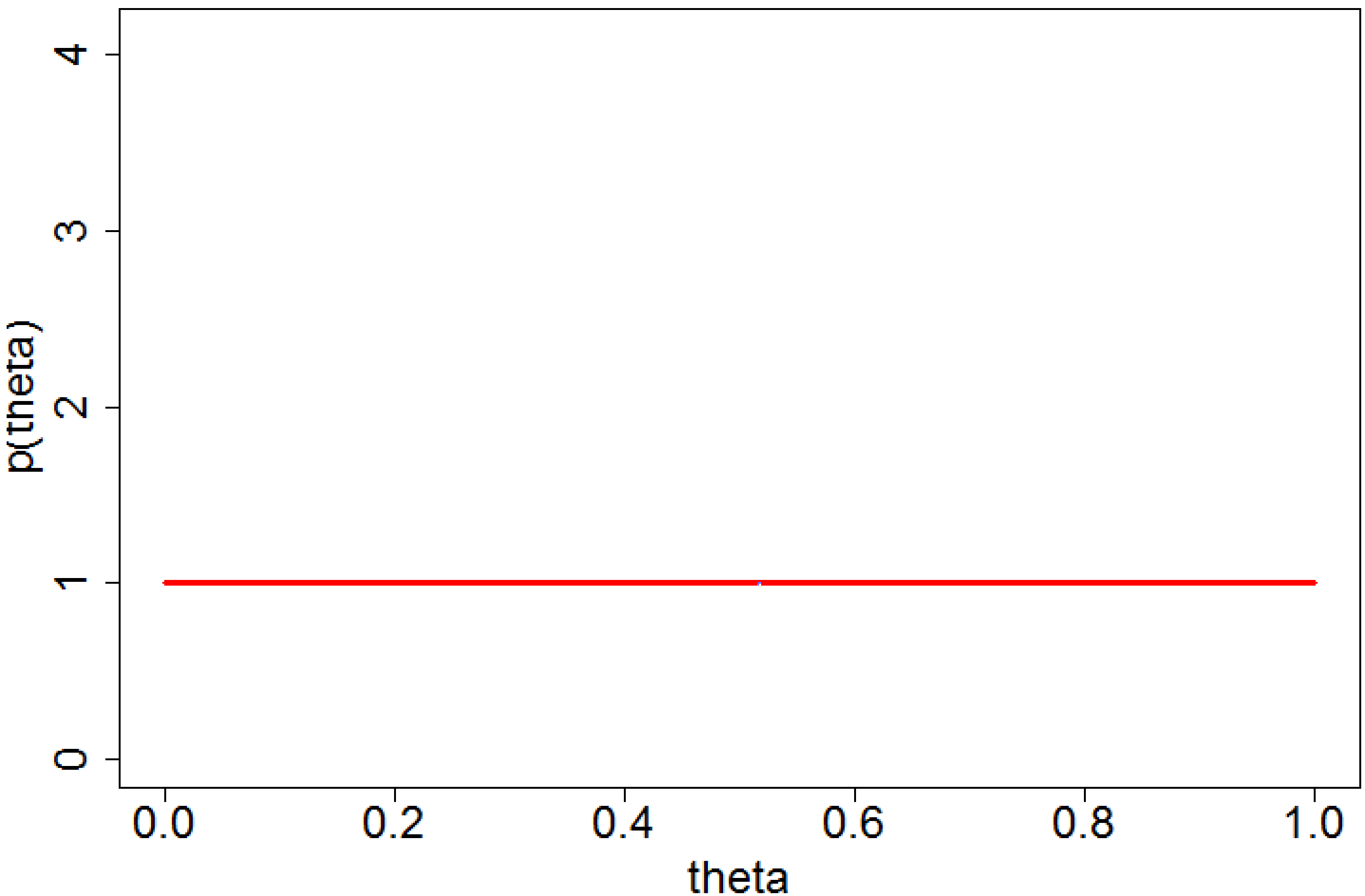
$\theta \sim Beta(16,6)$



ということは（少し余談）

- データが得られる前の状態であるθについての一様事前分布は，事前に「成功1回，失敗1回」ぶんの情報があったと解釈できる

$$\theta \sim Beta(1,1)$$



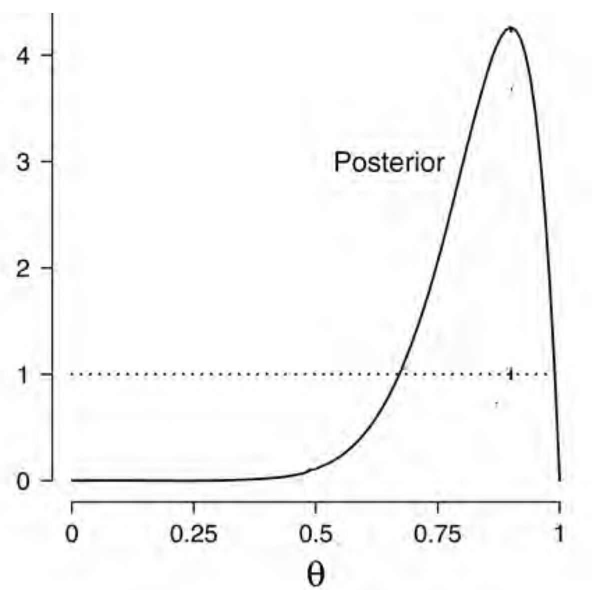
ベイズ統計学の考え方

- 情報（データy）が得られたら，ベイズの定理によって関心のある量（パラメータθ）の確率を更新する
- 結果は θ についての事後確率（分布）として得られる。事後確率分布がすべての情報を持つ
  - これを必要に応じて要約して使う

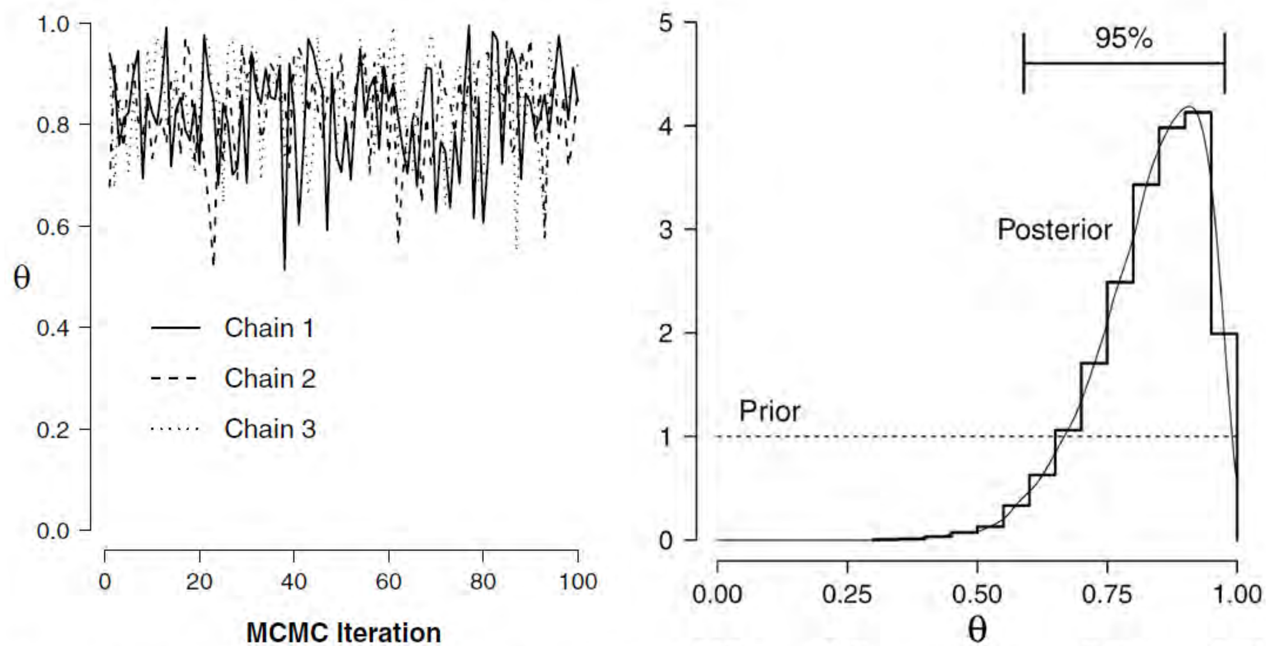
- しかし，モデルが複雑になると事後分布$p(\theta|y)$が解析的に求められなくなってしまい，事実上計算ができなくなってしまっていた（80年代までのベイズ統計学）
- マルコフ連鎖モンテカルロ（Markov chain Monte Carlo, MCMC）法の登場によりこの状況が一変した

マルコフ連鎖モンテカルロ(MCMC)法

- **事後分布**からの**乱数を発生**させる汎用性の高い方法
- MCMC法により事後分布から得られた乱数を集めることで，**事後分布**やその**要約統計量（平均・分散など）**を実用上十分な精度で得ることができる



事後分布



MCMC法で得られた乱数列と, そのヒストグラム

(Lee & Wagenmakers, 2013 井関訳, 2017 図1.3, 1.4)

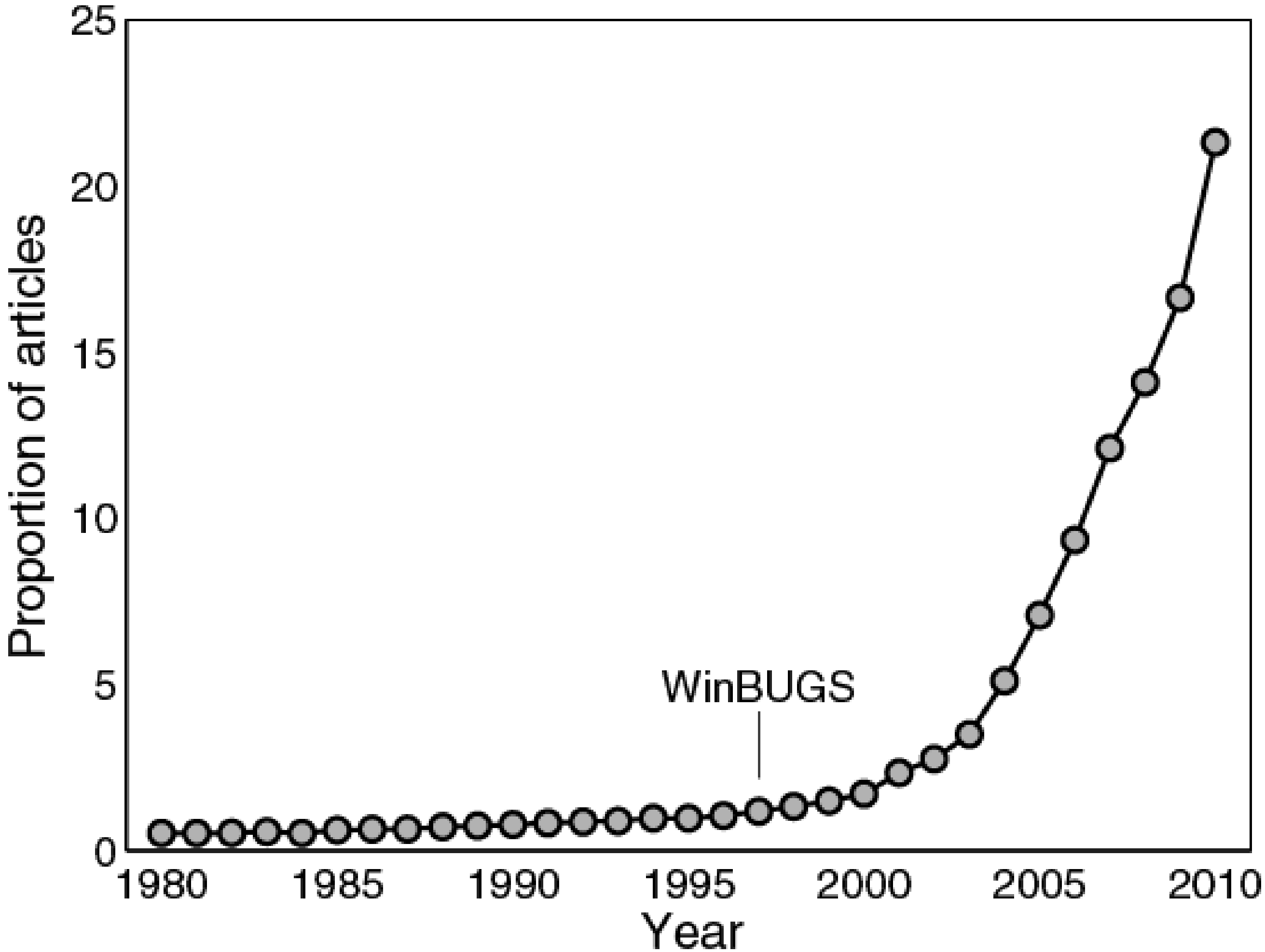


Fig. 1.2 A Google Scholar perspective on the increasing popularity of Bayesian inference, showing the proportion of articles matching the search “bayes OR bayesian - author: bayes” for the years 1980 to 2010.

(Lee & Wagenmakers, 2013 井関訳, 2017 図1.2)

現代のベイズ推論のための道具

- 統計環境

https://cran.r-project.org/

＋

- MCMCエンジン

JAGS

https://sourceforge.net/projects/mcmc-jags/

Stan

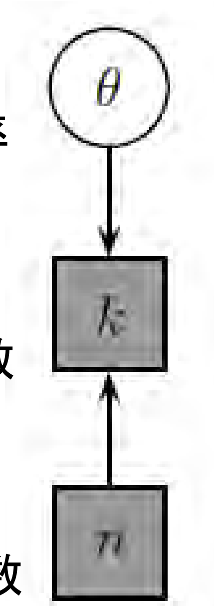
http://mc-stan.org/

例：比率の推定（再掲）

- 難易度が同じ課題が独立に20問提示される
- 結果は正答か誤答かのどちらか 全20問
- 個体の正答確率(能力)θを推定したい

- モデル：

正答できる確率



正答数

問題数

□ : 離散変数
○ : 連続変数
色つき : 観測
色なし : 非観測

θ ~ Beta(1, 1)
とも書ける

θ ~ Uniform(0, 1)
（一様分布）
事前分布

k ~ Binominal(n, θ)
（二項分布）
尤度

例１: JAGSでの記述

```
model{
        k ~ dbin(theta,n)
        theta ~ dbeta(1,1)
}
```

例 1 : Stanでの記述

```
data {
        int<lower=0> n;
        int<lower=0> k;
}
parameaters {
        real<lower=0,upper=1> theta;
}
model {
        k ~ binomial(n,theta);
        theta ~ beta(1,1);
}
```

JAGSとStan

- **JAGS**
  - 標準的な複数のMCMCアルゴリズムを実装
  - WinBUGS, OpenBUGSとほぼ同様の記法を採用しており，安定した推定ができることが多い
  - 工夫によって（例：パラメータ拡大法）単純にはうまく推定できない場合にも対処可能
- **Stan**
  - 新しいMCMCアルゴリズムを実装し，潜在変数や階層構造を含むときでも効率のよい推定が可能
  - 記法はJAGSよりは複雑だが，さほど難しくはない
  - 充実したマニュアル，サポート体制，コミュニティ
  - カテゴリカルな潜在変数は直接は扱えない

項目反応理論へ

- いま扱ったのは一人の受験者のθだった
- 応用上，$i = 1, \ldots, N$人の受験者の個人差$\{\theta_i\}$を知りたい
  → θの事前分布を標準正規分布にする
- 項目jも識別力a_j，困難度b_jという2つのパラメータを持つとし，ロジスティックリンク関数を利用する
  →項目反応理論モデルの尤度が得られる

$$P(y_{ij} = 1) = \frac{\exp\left(a_j(\theta_i - b_j)\right)}{1 + \exp\left(a_j(\theta_i - b_j)\right)}$$

- この2PLをはじめ，各種項目反応理論モデルのJAGSやStanでの実装は，たとえばCurtis (2010, *J Stat Soft*)やLuo & Jiao (2017, *Educ Psychol Meas*)を参照できる

Journal of Statistical Software

August 2010, Volume 36, Code Snippet 1. *http://www.jstatsoft.org/*

BUGS Code for Item Response Theory

S. McKay Curtis
University of Washington

Abstract

I present **BUGS** code to fit common models from item response theory (IRT), such as the two parameter logistic model, three parameter logisitic model, graded response model, generalized partial credit model, testlet model, and generalized testlet models. I demonstrate how the code in this article can easily be extended to fit more complicated IRT models, when the data at hand require a more sophisticated approach. Specifically, I describe modifications to the **BUGS** code that accommodate longitudinal item response data.

Keywords: education, psychometrics, latent variable model, measurement model, Bayesian inference, Markov chain Monte Carlo, longitudinal data.

(Curtis, 2010, *J Stat Soft*)

JAGS(BUGS)による2PL IRTモデル

```
model{
  for (i in 1:n){
    for (j in 1:p){
      Y[i, j] ~ dbern(prob[i, j])
      logit(prob[i, j]) <- alpha[j] * (theta[i] - delta[j])
    }
    theta[i] ~ dnorm(0.0, 1.0)
  }

  for (j in 1:p){
    delta[j] ~ dnorm(m.delta, pr.delta)
    alpha[j] ~ dnorm(m.alpha, pr.alpha) I(0, )
  }
  pr.delta <- pow(s.delta, -2)
  pr.alpha <- pow(s.alpha, -2)
}
```

Table 1: Two parameter logistic IRT model.

(Curtis, 2010, *J Stat Soft*)

JAGS(BUGS)による より複雑なIRTモデル

```
model{
  for (t in 1:T){
    for (i in 1:n){
      for (j in 1:p){
        Y[i, j, t] ~ dbern(prob[i, j, t])
        logit(prob[i, j, t]) <- alpha[j] * (theta[i, t] -
           delta[j])
      }
    }
  }

  for (i in 1:n){
    theta[i, 1:T] ~ dmnorm(mu.theta[], Pr.theta[,])
  }

  mu.theta[1] <- 0.0
  for (t in 2:T){
    mu.theta[t] ~ dnorm(m.mu.theta, pr.mu.theta)
  }
  pr.mu.theta <- pow(s.mu.theta, -2)

  sigsq.theta[1] <- 1.0
  Sigma.theta[1, 1] <- 1.0
  for (i in 2:T){
    sigsq.theta[i] ~ dgamma(a.sigsq.theta, b.sigsq.theta)
    Sigma.theta[i, i] <- sigsq.theta[i]
    for (j in 1:(i-1)){
      Sigma.theta[i, j] <- sqrt(sigsq.theta[i]) * sqrt(sigsq.
         theta[j]) * pow(rho, i - j)
      Sigma.theta[j, i] <- Sigma.theta[i, j]
    }
  }
  Pr.theta[1:T, 1:T] <- inverse(Sigma.theta[,])
  rho ~ dunif(-1.0, 1.0)

  for (j in 1:p){
    alpha[j] ~ dnorm(m.alpha, pr.alpha) I(0, )
    delta[j] ~ dnorm(m.delta, pr.delta)
  }
  pr.alpha <- pow(s.alpha, -2)
  pr.delta <- pow(s.delta, -2)
}
```

Table 8: Longitudinal two parameter logistic regression model with heterogeneous AR(1) covariance structure.

(Curtis, 2010, *J Stat Soft*)

Stanによるラッシュ(1PL)モデルの実装

1PL (Rasch) Model

The 1PL item-response model, also known as the Rasch model, has one parameter (1P) for questions and uses the logistic link function (L).

The model parameters are declared as follows.

```
parameters {
  real delta;           // mean student ability
  real alpha[J];        // ability of student j - mean ability
  real beta[K];         // difficulty of question k
}
```

The parameter `alpha[j]` is the ability coefficient for student `j` and `beta[k]` is the difficulty coefficient for question `k`. The non-standard parameterization used here also includes an intercept term `delta`, which represents the average student's response to the average question.[5] The model itself is as follows.

```
model {
  alpha ~ normal(0, 1);          // informative true prior
  beta ~ normal(0, 1);           // informative true prior
  delta ~ normal(0.75, 1);       // informative true prior
  for (n in 1:N)
    y[n] ~ bernoulli_logit(alpha[jj[n]] - beta[kk[n]] + delta);
}
```


(Stan Development Team 2017, Stan User's Guide and Reference Manual)

Article

Educational and Psychological Measurement
1–25

DOI: 10.1177/0013164417693666
journals.sagepub.com/home/epm
SAGE

Using the Stan Program for Bayesian Item Response Theory

Yong Luo[1] and Hong Jiao[2]

Abstract

Stan is a new Bayesian statistical software program that implements the powerful and efficient Hamiltonian Monte Carlo (HMC) algorithm. To date there is not a source that systematically provides Stan code for various item response theory (IRT) models. This article provides Stan code for three representative IRT models, including the three-parameter logistic IRT model, the graded response model, and the nominal response model. We demonstrate how IRT model comparison can be conducted with Stan and how the provided Stan code for simple IRT models can be easily extended to their multidimensional and multilevel cases.

Keywords

item response theory (IRT), Markov chain Monte Carlo (MCMC), Bayesian

(Luo & Jiao, 2017, *Educ Psych Meas*)

Listing 1: Stan Code for the 3PL IRT Model

```
data {
int<lower=0> n_student;
int<lower=0> n_item;
int<lower=0,upper=1> Y[n_student,n_item];
}
parameters {
vector[n_student] theta;
vector <lower=0> [n_item] alpha;
vector[n_item] beta;
vector<lower=0,upper=1> [n_item] gamma;    //item pseudo-guessing
real mu_beta;
real<lower=0> sigma_alpha;
real<lower=0> sigma_beta;
}
model {
theta ~ normal(0,1);
beta ~ normal(mu_beta,sigma_beta);
mu_beta ~ normal(0,5);
sigma_beta ~ cauchy(0,5);
alpha ~ lognormal(0,sigma_alpha);
sigma_alpha ~ cauchy(0,5);
gamma ~ beta(5,23);
for(i in 1:n_student){
for (j in 1:n_item){
```

(continued)

(Luo & Jiao, 2017, *Educ Psych Meas*)

Listing 2: Stan Code for the Graded Response Model

```
data{
int<lower=2, upper=4 > K;    //number of categories
int <lower=0 > n_student;
int <lower=0 > n_item;
int<lower=1,upper=K > Y[n_student,n_item];
}
parameters {
vector[n_student] theta;
real<lower=0 > alpha [n_item];
ordered[K-1] kappa[n_item];    //category difficulty
real mu_kappa;    //mean of the prior distribution of category difficulty
real<lower=0 > sigma_kappa;    //sd of the prior distribution of category difficulty
}
model{
alpha ~ cauchy(0,5);
theta ~ normal(0,1);
for (i in 1: n_item){
for (k in 1:(K-1)){
kappa[i,k] ~ normal(mu_kappa,sigma_kappa);
}}
mu_kappa ~ normal(0,5);
sigma_kappa ~ cauchy(0,5);
for (i in 1:n_student){
for (j in 1:n_item){
Y[i,j] ~ ordered_logistic(theta[i]*alpha[j],kappa[j]);
}}
}
```

(Luo & Jiao, 2017, *Educ Psych Meas*)

ベイズモデルの限界は想像力だけ

- As the diversity of the recent applied Bayesian references attests, the MCMC approach is so generally applicable and easy to use that **the class of candidate models for a given data set now appears limited only by the user's imagination.** (Carlin & Chib, 1995, *JRSS-B,* p.473)
- 最近のさまざまな応用ベイズ統計学の文献が示すように、MCMCアプローチは高い汎用性を持ち、容易に利用できる。いまや**データに対してどんなモデルを利用するか、ということの限界を決めるのは、ユーザーの想像力だけ**のようだ

ベイズ統計学の実践のための書籍の一部

4. まとめ

全体の流れ（再掲）

1. ベイズ統計学の考え方
   - 未知の・知りたいことを確率によって表す
   - データを使って確率を更新する
   - JAGSやStanによって実装・推定する
   - この枠組みで，適用場面や研究関心，利用可能な情報によってIRTモデルを柔軟に拡張し，推定・解釈することが可能になる
2. 研究例A：反応スタイルのモデリング
   - 係留ビネット法と２次元IRTモデルの利用
3. 研究例B：反応時間情報の取り入れ
   - 数理心理学で発展したLBAモデルとIRTの統合
4. まとめと展望

ベイズ統計学の隆盛の理由

(Okubo & Okada, *in revision*)

- **理論的整合性**
  - 「関心のある量を確率によって表現する」「データに基づき，確率を更新する」という原則をあらゆる場面で適用する，汎用性の高い枠組み
  - 直感にあった解釈（ ←→ *p*値・信頼区間 ）
- **柔軟性**
  - 潜在変数モデリング・階層モデリングの考え方を用いて，個人差を考慮しデータ生成メカニズムを表現する柔軟なモデルを構成でき，予測に役立つ
- **実践性**
  - BUGS, JAGS, Stanといった無償で利用でき汎用性の高いマルコフ連鎖モンテカルロ(MCMC)法の推定ソフトウェアが利用できる

心理尺度にもっとベイズIRTは使えるはず

- 因子分析モデルはリッカート尺度データに対して生成モデルを与えないが，IRTモデルは生成・予測モデルとして利用できる
- さまざまな背景情報を共変量として取り込んだり，データの階層性や群構造をモデルに取り入れたりすることが容易にできる
- IRTは教育テストのためのモデルという印象があるかもしれないが，心理尺度のモデリング，さまざまな社会科学分野への利用可能性も大きいはず

心理統計学の基本的ロジック

- 「**こころ**」は直接見えないが，何らかの形で**行動に反映**されている
- 「行動」の結果は**データ**として（数値で）記録し，定量的に扱うことができる
- 「行動」データを**統計学的に扱う**こと，**モデリング**することで，こころのありようについて推論することができる

ベイズ統計モデリングは，「こころ」と観測「データ」をつなぐ過程についての統計モデルを柔軟かつ実践的に構築することができ，推定結果を解釈しやすい確率分布の形で得られる

(南風原, 2015)