日本教育心理学会第58回総会自主企画シンポジウム

ベイズ統計学と歩む，これからの心理学研究

ー研究仮説が正しい確率と階層モデリングを用いたアプローチー

高松サンポートホール　54会議場A
2016年10月10日(月)10:00－12:00
企画：長尾圭一郎（早稲田大学大学院）
司会：豊田秀樹（早稲田大学）
話題提供：久保沙織（早稲田大学）
　秋山隆（早稲田大学）
　長尾圭一郎（早稲田大学大学院）
　磯部友莉恵（早稲田大学大学院）
　吉上諒（早稲田大学大学院）
指定討論：岡田謙介（専修大学)

リンク関数・プレート表現・BART モデル

秋山隆 [1]

早稲田大学 [1]

2016 年 10 月 10 日

教育心理学会総会　自主シンポジウム

Contents

リンク関数のベイジアンモデリング

- 事前の知識をモデルに反映可能
- 確信区間を参照可能
- 自身の要求に合わせたモデルの拡張が容易

データの記述：分布モデル

単純な分布モデル：データに対する性質を考慮した理論分布による記述

例

ポアソン分布：単位時間における窓口への問い合わせ件数・ある場所での交通事故発生件数

ベルヌイ分布：コインを投げて表裏どちらが出るかを記録する試行結果

仮定した分布の母数を推定し，推論を行う。

例えばデータ y が 100 点満点のテスト得点であり，その平均 μ の周りで正規分布していると仮定する。(標準偏差 σ)

$$\text{モデル式は } y_i \sim \mathrm{Normal}(\mu, \sigma), \quad i \text{ は個人を表す} \tag{1}$$

平均値や標準偏差の推定結果から，集団の性質を推論する

構造の仮定

母数について，関連する変数による線形構造を仮定することで，データの更なる説明や予測を試みることも可能

例

データ y_i が 100 点満点のテスト得点であり，説明変数 x_i が勉強時間であるとする。このとき，期待値に対する線形構造は，

$$y_i \sim \mathrm{Normal}(\mu_i, \sigma), \quad \mu_i = b + ax_i \tag{2}$$

と表される。ここで b は切片，a は係数である。

上記モデルでは将来新たなデータを取得した場合，当該データが線形構造によって構成される期待値 (予測値) の周りで正規分布することが仮定される。

成り立たない場合には (2) 式による線形構造を仮定できない。

リンク関数を用いた線形構造の仮定

方法

正規分布以外のデータには，リンク関数 (link function) と呼ばれる関数を用いて変換した分布の期待値に対して線形構造を仮定する。

本発表ではデータが期待値で条件づけられた際に，正規分布以外に従っていると仮定される場合に，分布の期待値に対して線形構造を導入する方法を紹介する。

ベルヌイ分布の場合

確率が一定の下で，互いに独立な試行を行い，結果が 2 値で表されるとき，これをベルヌイ試行といい，その試行結果に関する確率変数が従う分布をベルヌイ分布と呼ぶ。
ベルヌイ分布の確率密度関数は以下

$$f(y|\theta) = \theta^y (1-\theta)^{1-y}, \quad y \in \{1, 0\} \tag{3}$$

表 1 は「女性は家庭の切り盛りに注力し，社会の運営は男性に任せるべきである」という質問に対する，2871 名 (男性 1305 名，女性 1566 名) の「賛成—反対」回答データ

表 1: 「性役割」データ

	1	2	3	4	5	6	7	8	···	1304	1305
被教育年数	0	0	0	0	1	1	2	2	···	20	20
性別	男	男	男	男	男	男	男	男	···	男	男
賛否	賛	賛	賛	賛	賛	賛	賛	賛	···	反	反
	1	2	3	4	5	6	7	8	···	1565	1566
被教育年数	0	0	0	0	1	3	3	3	···	20	20
性別	女	女	女	女	女	女	女	女	···	女	女
賛否	賛	賛	賛	賛	賛	賛	賛	賛	···	反	反

表 1 は回答者の属性ごとに並び替えたデータの一部を示している。賛成を 1，反対を 0 として扱うと，ベルヌイ試行と見なすことができる。

「性役割」(ベルヌイ) データの「賛否」に対し，各回答者の「被教育年数」による線形構造を仮定し，賛否に対する説明を試みる。(2) 式を仮定。

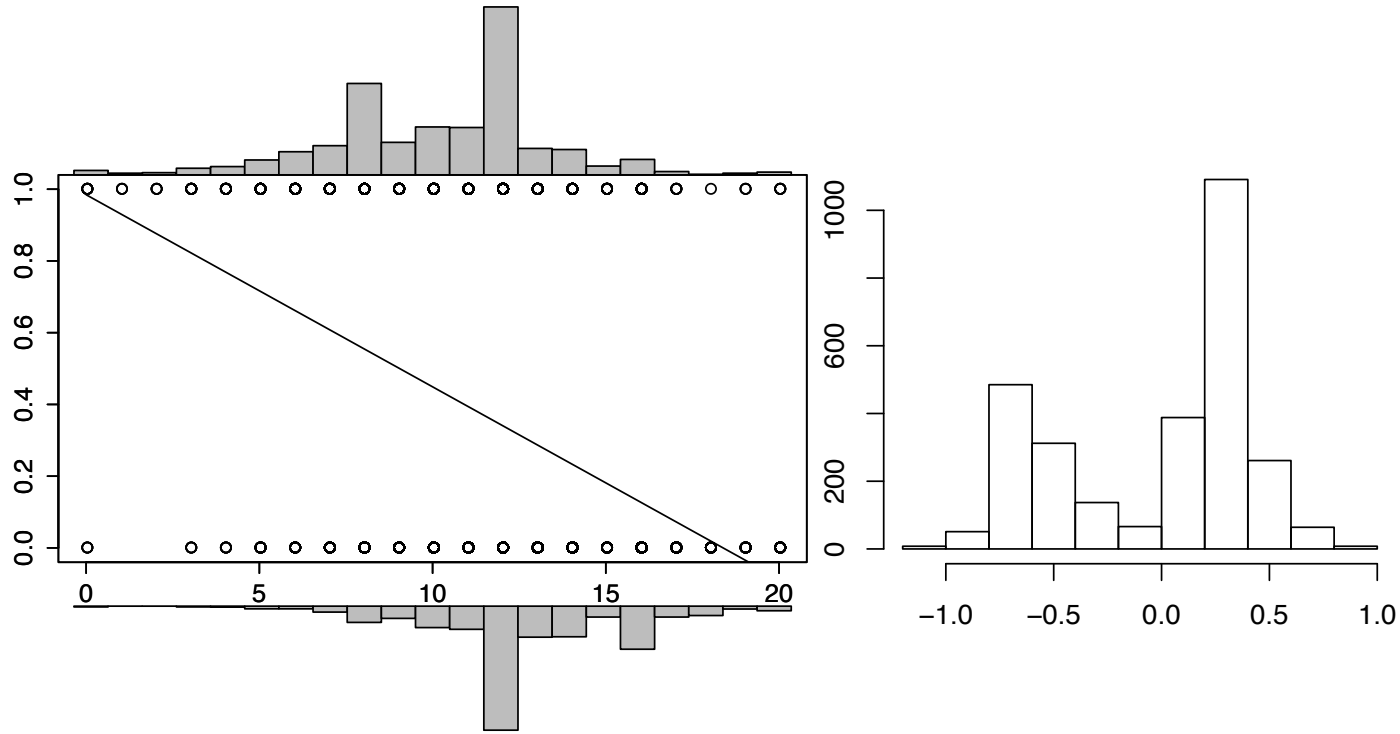


図. 1: 2 値データへの線形構造の当てはめ結果

図 2 左図は線形構造を当てはめた場合の推定結果 $\hat{y} = 0.984 - 0.054x_i$ を，また右図は残差 $y_i - \hat{y}_i$ のヒストグラム示している。

正規分布を仮定した場合の単回帰モデル

$$y_i \sim \mathrm{Normal}(\mu_i, \sigma), \quad \mu_i = b + a x_i \tag{4}$$

Stan コード

```
model{
  for(i in 1:N){
    y[i] ~ normal(b + a * edu[i], sigma);
  }
}
```

推定結果

```
        mean   post.sd   2.5%   97.5%
b      0.984     0.031  0.924   1.045
a     -0.054     0.003 -0.059   0.049
sigma  0.448     0.006  0.437   0.459
```

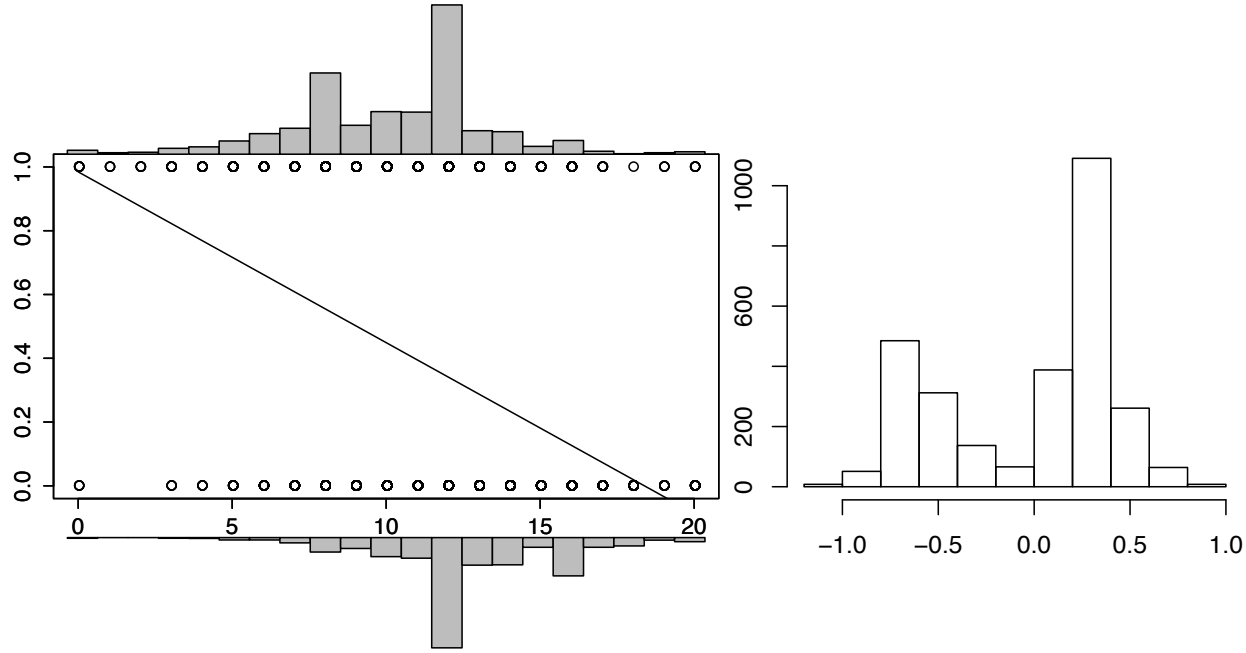


図. 2: 2 値データへの線形構造の当てはめ結果

図 2 左図上下の棒グラフは「被教育年数」ごとの人数を表している。推定結果から「被教育年数」が 19 年以上になると，予測値が 0(反対) 以下の値となることが分かるが，これは妥当な予測結果とはいえない。残差も，期待値の周りでデータが正規分布しているとは見なせない。

線形構造を用いて説明，予測を行うには，まずは従属変数となるデータの性質を考慮して，妥当なモデルを特定する必要がある。

本例の場合は期待値で条件づけた際にベルヌイ分布に従うことを仮定するため，モデルは以下となる。

$$y_i \sim \text{Bernoulli}(\theta_i), \quad \theta_i = b + a \times x_i, \quad 0 \leq \theta \leq 1 \tag{5}$$

2 値の回答結果は回答者 i ごとに母数 θ_i をもつベルヌイ分布に従っており，θ_i に対して線形構造を仮定し，「賛否」の確率に対する説明を試みるモデル。実際には，線形構造部分は $\theta_i =$ 切片 $b +$ 係数 $a \times$ 被教育年数$_i$ となる。

しかし・・・

θ_i は 0 から 1 の範囲でなければならないため，このままでは (5) 式も θ の構造として用いることはできない。

期待値と線形構造のリンク

y_i の期待値 $E[y_i] = \mu_{y_i}$ について，ある変換 $g(\cdot)$ を施した $g(\mu_{y_i})$ に線形構造

$$g(\mu_{y_i}) = b + ax_i \tag{6}$$

を仮定する [1]。変換のための関数 $g(\cdot)$ には逆関数

$$\mu_{y_i} = g^{-1}(b + ax_i) \tag{7}$$

が存在するものとする。このとき，変換 $g(\cdot)$ はリンク関数と呼ばれる。リンク関数にはいくつかの種類があり，データの性質に依って適宜選択する [2]。

---

[1] 内部に仮定する構造が線形なのであり，モデル全体に線形性を仮定するわけではない。

[2] リンク関数の観点からは，期待値で条件付けられた y が正規分布に従っている場合，μ_{y_i} に対する線形構造は，$g(\cdot)$ として恒等変換 $\mu_{y_i} = b + ax_i$ を選択しているものと見なせる。

期待値で条件付けられた y_i がベルヌイ分布に従う場合，$\mu_{y_i} = \theta_i$ のリンク関数としてロジット (logit) 変換を用いることが可能。
変換された θ_i と線形構造は

$$logit(\theta_i) = \log\left(\frac{\theta_i}{1-\theta_i}\right) = b + ax_i, \quad 0 \leq \theta \leq 1 \tag{8}$$

と連結される。ロジットについて逆関数をとると (ロジスティック (logistic) 変換),

$$\theta_i = \frac{1}{1+\exp(-(b+ax_i))} \tag{9}$$

となり，θ_i がロジスティック関数で表される。するとモデル式は以下に表される。

$$y_i \sim \text{Bernoulli}(\theta_i) \tag{10}$$

$$\theta_i = \frac{1}{1+\exp(-z_i)} = \frac{\exp(z_i)}{1+\exp(z_i)}, \quad z_i = b + a \times \text{被教育年数}_i \tag{11}$$

Stan ではベルヌイ分布とロジットリンク関数がセットになったサンプリング関数が用意されている。

(10) 式，(11) 式は Stan コードの `model` ブロック内で，以下とする。

```
for(i in 1:N){ //N は人数を表す
  y[i] ~ bernoulli_logit(b + a1 * edu[i]);
}
```

ハミルトニアン・モンテカルロ (HMC) 法により，サンプリング回数 11000 回，バーンイン期間を 1000 回，チェーン数を 5 とし，50000 個の事後標本を用いて推測を行った。

表 2: 推定結果

	EAP	post.sd	2.5%	97.5%
b	2.513	0.182	2.161	2.872
a_1「被教育年数」	-0.271	0.016	-0.302	-0.241

回帰係数の EAP 推定値から，回答者の教育年数が長くなるほど，性役割に関する当該質問に反対の立場をとる傾向が強まることが示唆された。

表 3: 推定結果

	EAP	post.sd	2.5%	97.5%
b	2.513	0.182	2.161	2.872
a_1「被教育年数」	-0.271	0.016	-0.302	-0.241

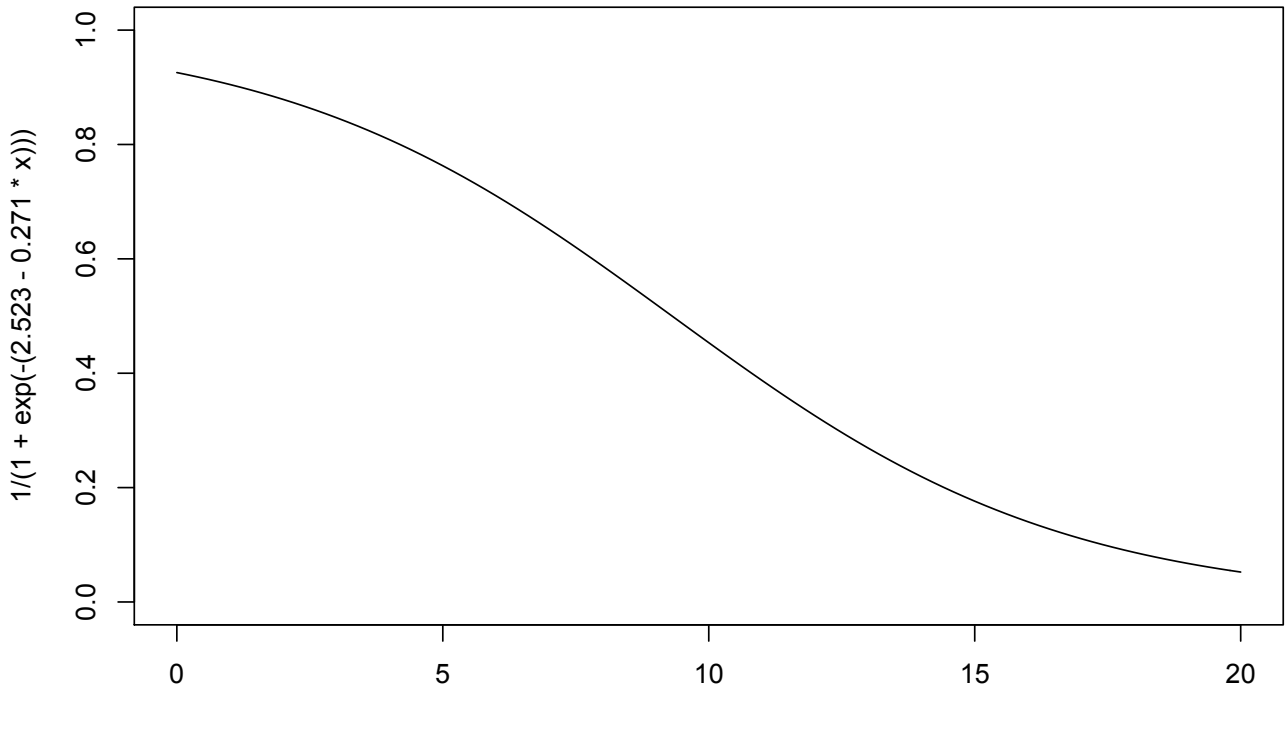


図. 3: 2 値データへのリンク関数を用いた線形構造の当てはめ結果

回帰係数の EAP 推定値から，回答者の教育年数が長くなるほど，性役割に関する当該質問に反対の立場をとる傾向が強まることが示唆された。

付録（リンク関数）：ポアソン分布

ある一定時間（単位時間）における事象の発生件数を記述するために広く利用されている分布

$$\text{ポアソン分布の確率密度関数：} \quad f(y|\lambda) = \frac{\exp(-\lambda)\lambda^y}{y!}, \quad \lambda \geq 0 \tag{12}$$

y は発生件数，λ は平均である。ポアソン分布はただ 1 つの母数 λ によって特徴づけられる。

表 4: 「大腸ポリープ」データ

個数 y	63	2	28	17	61	1	7	15	44	25
処遇 treat	0	1	0	1	0	1	0	0	0	1
年齢 age	20	16	18	22	13	23	34	50	19	17
個数 y	3	28	10	40	33	46	50	3	1	4
処遇 treat	1	0	0	0	1	0	0	1	1	1
年齢 age	23	22	30	27	23	22	34	23	22	42

処遇：薬 $= 1$，プラセボ $= 0$

表 4 は 20 名の実験参加者 [3] を薬効治療群 [4] とプラセボ群とに分け，12 カ月後の大腸ポリープの発生個数を計測したカウントデータを示している [5]。

[3] ポリープが多発する家族性大腸腺腫症 (familial andenomatous polyposis) 患者。
[4] 薬は非ステロイド性抗炎症薬 (non-steroidal anti-inflammatory drug)。
[5] R パッケージ `HSAUR` に含まれるデータセット `polyps` を元にしている

期待値で条件付けられたカウントデータ y_i の背後にポアソン分布を仮定し，期待値 λ_i と線形構造をリンクさせる。J 個の説明変数 x_j を用いて λ_i に線形構造 $\lambda_i = b + a_1 x_{i1} + \cdots + a_J x_{iJ}$ を導入。
ただし，ポアソン分布の定義より，$\lambda \geq 0$ であるため左辺の線形構造部分も非負の値をとる必要がある。ポアソン分布の場合，リンク関数として対数 $\log(\lambda)$ や平方根 $\sqrt{\lambda}$ をリンク関数として用いることができる。
ここでは「大腸ポリープ」データに対して対数リンク関数を用いてモデル化を行う。

$$y_i \sim \text{Poisson}(\lambda_i) \tag{13}$$

$$\log(\lambda_i) = b + a_1 \text{age}_i + a_2 \text{treat}_i, \lambda_i = \exp(b + a_1 \text{age}_i + a_2 \text{treat}_i) \tag{14}$$

(13) 式では大腸ポリープの発生数が平均 $\lambda_i (i = 1, \cdots, N)$ に従っており，さらに平均の対数に対して変数「年齢」と「処遇」による線形構造があることを仮定する。b, a_1, a_2 はそれぞれ切片項，「年齢」の係数，「処遇」の係数を表している。
「処遇」は，プラセボ群に所属する協力者の場合は 0 であるため，(14) 式の表現では，第 3 項 $a_2 \times \text{treat}_i$ が消え，切片と「年齢」に関する係数のみが残ることに注意。

(13) 式を Stan では model ブロックにおいて，以下のように記述。

```
for(i in 1:N){
  y[i] ~ poisson_log(b + a1 * age[i] + a2 * treat[i]);
}
```

表 5: 「大腸ポリープ」データ推定結果

	EAP	post.sd	95%下側	95%上側
b	4.532	0.147	4.245	4.820
a_1「年齢」	-0.039	0.006	-0.051	-0.027
a_2「処遇」	-1.366	0.118	-1.600	-1.140

a_1 は負の値 $-0.039[-0.051, -0.027]$ を示している。また，a_2 の EAP 推定値は $-1.366[-1.600, -1.140]$ となり，予測値に対して，薬効治療によって，プラセボ群よりもポリープ発生個数を抑えることが分かる。
家族性大腸腺腫症によって発生する大腸ポリープは放置すると加齢とともに大腸がん発生リスクが高まることが知られている。ポリープ発生を薬によって抑えることは治療の一環として有効である可能性。

付録（リンク関数）：負の 2 項分布

ある試行回数中に何回，試行が成功したかはしばしば 2 項分布で近似される。一方で，成功するまで試行を続けるものとして，成功までの失敗回数を記録する場合を考える。このとき，失敗回数は負の 2 項分布に従っているものと見なせる。負の 2 項分布にはいくつかの定式化があるが，ここでは μ と ϕ の 2 つの母数によって表現される確率密度関数を採用する。

$$f(y|\mu,\phi) = \left(\begin{array}{c} y+\phi-1 \\ y \end{array} \right) \left(\frac{\mu}{\mu+\phi} \right)^{y} \left(\frac{\phi}{\mu+\phi} \right)^{\phi}, \ \mu \geq 0, \phi \geq 0 \qquad (15)$$

y は失敗回数を表している。y の平均は μ，分散は $\mu + (\mu^2/\phi)$ によって表される。
表 6 に，1 度逮捕された人物 110 名について，釈放されてから再度逮捕されるまでの期間を週単位で記録したデータ (最大 50 週) を示した [6]。また，当該人物に関して，「財政援助の有無」，「配偶者の有無」および「被教育歴」も示している。「被教育歴」はアメリカにおける教育システムに基づいたカテゴリ分けがなされており，5 段階に分けられている。0：6 年生以下 (小学校相当)，1：7 年生から 9 年生 (中学校相当)，2：10 年生から 11 年生 (高校相当)，3：12 年生 (高校相当)，4：大学以上。

[6] R パッケージ `GlobalDeviance` に含まれる `Rossi` データを元にしている。元データは 432 名を 52 週に渡って追跡調査したデータである。本節では再犯をしなかった人物 (打ち切りデータ) に関しては，データから除いた。

表 6: 「再犯」データ

変数 \ 個人	1	2	3	4	5	6	···	110
再犯までの期間 (週)y	20	17	25	23	37	25	···	12
財政援助の有無 x_1	0	0	0	0	0	0	···	1
配偶者の有無 x_2	0	0	0	1	0	0	···	1
被教育歴 x_3	3	4	3	4	3	4	···	4

ここでは再度逮捕されるまでの週が期待値で条件付けられた際に負の 2 項分布に従っているものと仮定。また，リンク関数として対数関数を仮定し，「財政援助の有無」と「配偶者の有無」，「被教育歴」を用いて線形構造を

$$y_i \sim \text{NegativeBinomial}(\mu_i, \phi) \tag{16}$$

$$\log(\mu_i) = b + a_1 x_{1i} + a_2 x_{2i} + a_3 x_{3i} \tag{17}$$

のように仮定する。(16) 式から (17) 式を Stan では `model` ブロック内で以下のように表現[7]。

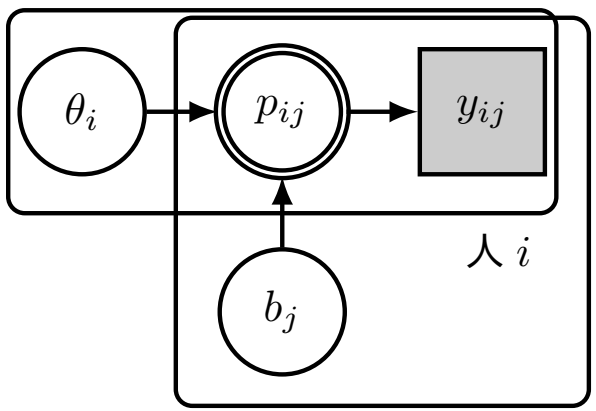
```
for(i in 1:N){
  y[i] ~ neg_binomial_2_log(b + a1*x1[i] + a2*x2[i] + a3*x3[i], phi);
}
```

[7](15) 式について，Stan では `neg_binomial_2` というサンプリング関数が用意されている。

プレート表現

- ベイジアンモデリングでは，モデルは数式を用いて表現される。
- モデルが複雑になると，母数や変数間の繋がりの把握が難しくなる。
- プレート表現 (plate notation) と呼ばれる方法によってモデルを図示することで，モデル内のデータ生成過程を視覚的に把握しやすくなる（グラフィカルモデル (graphical model)）とも呼ばれる)。

プレート表現では，ノード (node) と呼ばれる記号によって変数を表し，ノード間を矢印で繋ぐことで，データの生成過程を表現する。また，プレートと呼ばれる角が丸い長方形によってノードのまとまりを表す。



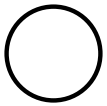
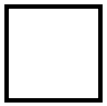
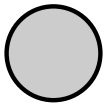
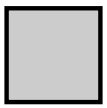
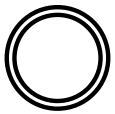
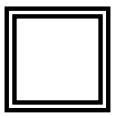
$$b_j \sim \text{Normal}(0, 2) \tag{18}$$

$$\theta_i \sim \text{Normal}(0, 1) \tag{19}$$

$$p_{ij} = \frac{1}{1 + \exp(\theta_i - b_j)} \tag{20}$$

$$y_{ij} \sim \text{Bernoulli}(p_{ij}) \tag{21}$$

本発表では変数の状態ごとにノードを次のスライドのように表現する。

ノード	変数の状態	使用例
	連続的かつ潜在的な変数	連続的な母数：正規分布の平均
	離散的かつ潜在的な変数	離散的な母数：トピックモデルにおける「トピック」
	連続的かつ顕在的な変数	連続的なデータ (時間や金額)
	離散的かつ顕在的な変数	離散的なデータ (状態や順序データ)
	連続的かつ生成量	モデル中で変数の状態に応じて生成される連続量：例えば確率。生成後，母数として用いられる場合もある。
	離散的かつ生成量	モデル中で変数の状態に応じて生成される指示変数の値。生成後，母数として用いられる場合もある。

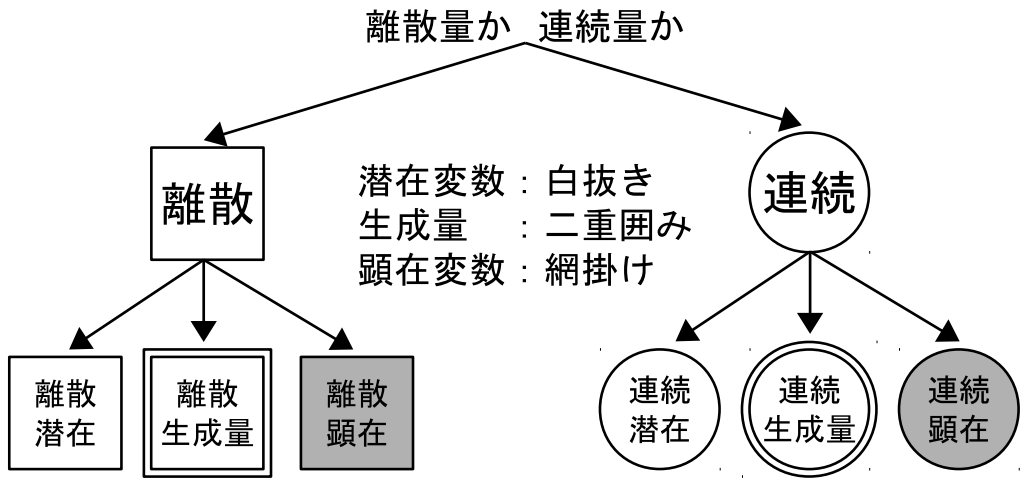
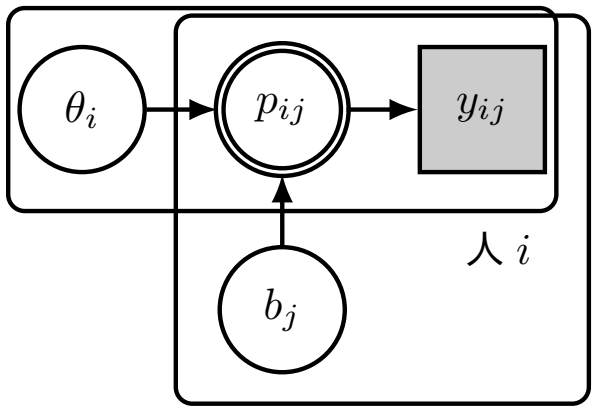


図. 4: プレート表現における変数の状態に応じたノード形状

項目反応理論における 1 母数ロジスティックモデルの表現



$$b_j \sim \text{Normal}(0, 2) \tag{22}$$

$$\theta_i \sim \text{Normal}(0, 1) \tag{23}$$

$$p_{ij} = \frac{1}{1 + \exp(\theta_i - b_j)} \tag{24}$$

$$y_{ij} \sim \text{Bernoulli}(p_{ij}) \tag{25}$$

項目反応理論における 1 母数ロジスティックモデルを考える。

項目に対する正誤 2 値データに対してベルヌイ分布を仮定し，更にベルヌイ分布の正答確率がロジスティック関数を通じて生成されるモデルを考える。モデルの数式による表現が (22) 式から (25) 式に当たる。ここでは (25) 式から数式を遡りつつ，プレート表現を確認していく。

まず，回答者 i の項目 j に対するデータ y_{ij} は離散的な 0–1 変数であるため，四角ノードかつ網掛けが施される。y_{ij} の背後には母数 p_{ij} をもつベルヌイ分布が仮定されている。

次に p_{ij} は項目 j に関する母数 b_j と回答者 i に関する母数 θ_i をもつロジスティック関数を通じて生成される。p_{ij} は連続的な生成量であるため二重丸ノードで表現される。生成量は数式上ではイコール $(=)$ で表されていることに注意 (文献によっては数式上で矢印を用いて，$p_{ij} \leftarrow \frac{1}{1+\exp(\theta_i - b_j)}$ と表す場合もある)。

最後に，(24) 式中の母数 b_j，θ_i は潜在的な連続変数であるため，白抜きの丸ノードで表される。ここで，事前分布として仮定されている正規分布の母数はプレート表現中に現れない点に注意。もしこれらの正規分布の平均や分散を未知母数として表す場合には，さらに対応するノードが追加される。

- 各ノード内には対応する変数名が記されている。
- ノード間はデータ生成過程における関係を考慮した矢印で繋がりが示される。
- また，変数間の添え字に応じて，角が丸い四角 (プレート) でノードが囲まれる。
- プレートは Stan コード中での `for` 文処理に対応している。

BART モデル

Balloon Analogue Risk Task（Lejuez, Read, Kahler, Richards, Ramsey, & Stuart, 2002）

リスクテイキング行動の個人差を，行動パフォーマンスの観点から検討するための実験・分析手法の一つ

リスクの存在：

交通事故や病気・自然災害

リスクを伴う行動：

信号無視・駆け込み乗車・パチンコ・競馬などの賭け事。バンジージャンプ・ロッククライミングなどの冒険的な行動。

リスクの認知の有無とは無関係にリスクを敢行する行動のことを，リスクテイキングという (森泉・臼井・中井, 2010)。リスクテイキング行動傾向を測定するために，これまで質問紙を利用した尺度作成が，寺崎・塩見・岸本・平岡 (1987) や小塩 (2001) や森泉ら (2010) によって行われてきた。

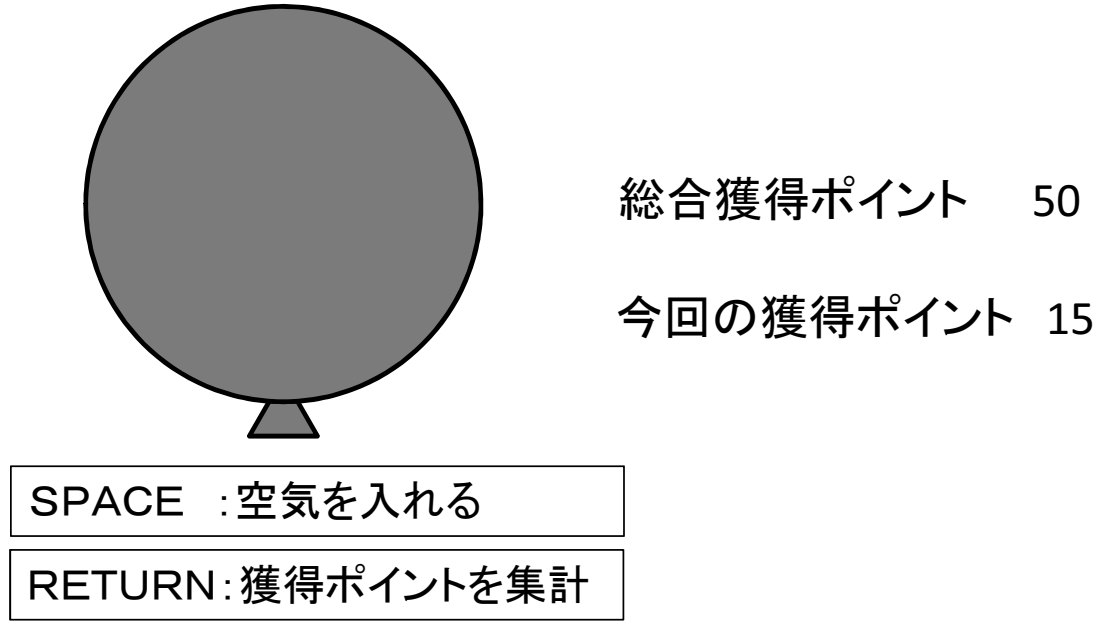


図. 5: Balloon Analogue Risk Task(BART)

BART では，PC の画面上で風船を膨らませるという課題を実験参加者に行ってもらう。

- 各試行において，参加者は風船を膨らませるか，風船の大きさに応じた金額やポイントを獲得するかのどちらかを選択する。
- 風船は，膨らませれば膨らませるほど得られる金額・ポイントは増えていくが，ある一定の大きさを超えると風船は破裂する。
- 風船が破裂した場合は，その試行で得られる金額やポイントはゼロになる。その点に気をつけつつ，なるべく多くの金額やポイントを獲得できるよう教示し，実験を行う。

表 7: BART の試行で得られるデータ例

		回数							
		1	2	3	4	5	6	···	16
	1	0	0	0	1	–	–	···	–
	2	0	0	0	0	0	–	···	–
試	3	0	0	0	0	0	0	···	–
行	4	0	0	0	0	0	1	···	–
	⋮				⋮				
	30	0	0	0	0	1	–	···	–

表 7 は，一人の参加者が 30 試行の実験に参加して得られたデータ例。

- 各行が各試行に対応し，0 が風船を膨らませたことを，1 が金額やポイントを獲得したことを表す。
  - ⇨データの 1 試行目 (1 行目) において，0 が 3 つ続いた後に 1 があるので，3 回膨らませた後に，金額・ポイントを獲得。
- 2 試行目 (2 行目) は，0 が 5 つ続いており 1 が出てきていない。
  - ⇨ 4 回膨らませた後に，5 回目も膨らませようとしたが，破裂したことを表す。

この際、獲得金額やポイントはゼロであることを意味する。

モデル化 1

BART モデルでは,

j 回目の試行における k 回目の判断 d_{jk}

$$d_{jk} = \begin{cases} 0 & \text{風船を膨らませる} \\ 1 & \text{ポイントを獲得する} \end{cases} \tag{26}$$

を，ベルヌイ分布を利用してモデル化する。

- Wallsten, Pleskac,& Lejuez(2005) は Lejuez et al.(2002) を拡張して，認知過程を考慮した BART モデルを提案
- Wallsten et al.(2005) では 10 個のモデルが紹介されている。
- ここでは，「リスクテイキング傾向」と「行動の一貫性」という 2 つの母数を取り入れた比較的単純なモデルを扱う。

2 つの母数

- それぞれリスクテイキング傾向を表す $\gamma(\gamma > 0)$ と，行動の一貫性を表す $\beta(\beta > 0)$ である。

風船の破裂確率 p

- 風船を膨らませていった際に破裂する確率を p とし，各試行で一定。
- 参加者には破裂する確率自体は伝えないが，一定であることは説明する。

パンプ回数の上限

- パンプ回数の上限は参加者には伝えない。

ここで，

参加者が最適だと思う膨らませる回数 (パンプ回数)ω は，破裂確率 p とリスクテイキングの傾向を表すパラメタ γ によって決定されると仮定し，

$$\omega = -\gamma / \log(1 - p) \tag{27}$$

と定式化する。

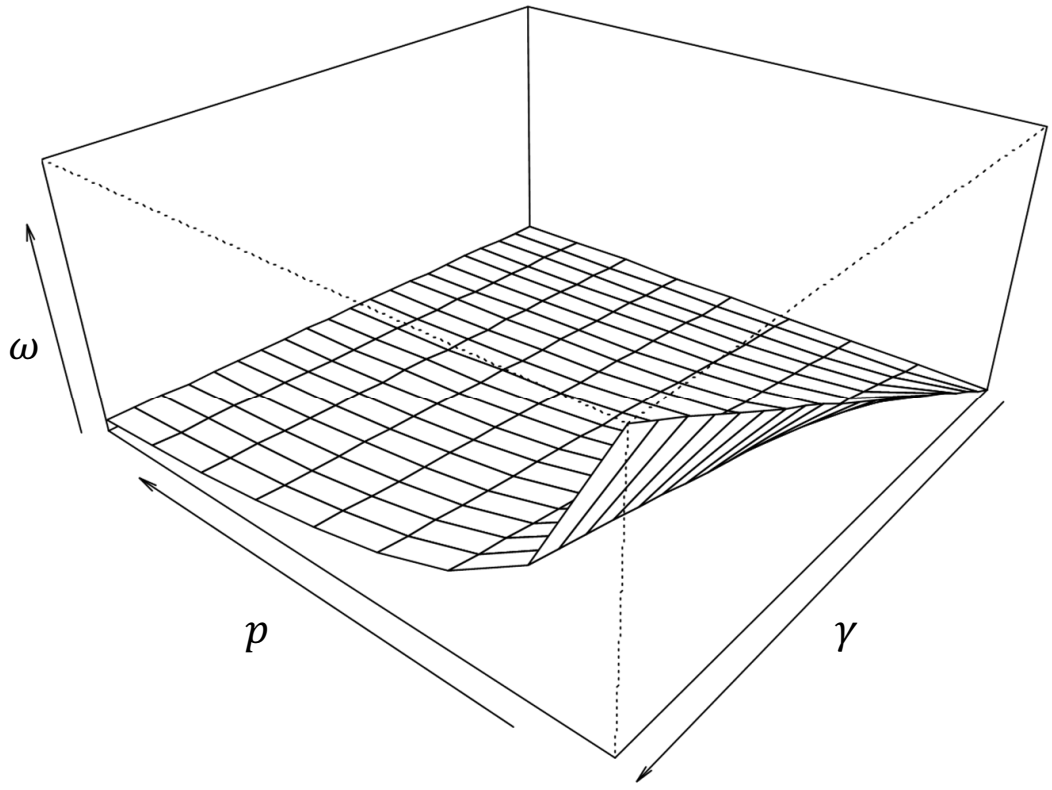


図. 6: γ と p の変化させたときの ω の値

γ の値が大きくなるにつれ，また，p の値が小さくなるにつれて，最適だと思うパンプ回数 ω は大きくなり，リスクをより追求することが分かる。

モデル化 2

j 回目の試行における k 回目の判断 (膨らませるか，ポイント獲得するか) は，参加者が風船を膨らませる確率 θ_{jk} によって決まると考える。

j 回目の試行における k 回目の判断の確率 θ_{jk} は，最適なパンプ回数 ω と参加者の行動の一貫性 β を用いてロジスティック関数により定式化を行う。

$$\theta_{jk} = \frac{1}{1 + \exp\{\beta(k - \omega)\}} \tag{28}$$

β は最適なパンプ回数 ω 付近における膨らませるかどうかの決定の一貫性を表す母数

$\beta = 0$ のときに，どの回においても $\theta_{jk} = 0.5$ となり，回を重ねても膨らませるかどうかは常に五分五分で判断することになる。β の値が大きくなると，最適なパンプ回数 ω 付近において，膨らませるかどうかの判断確率が急激に変化する。

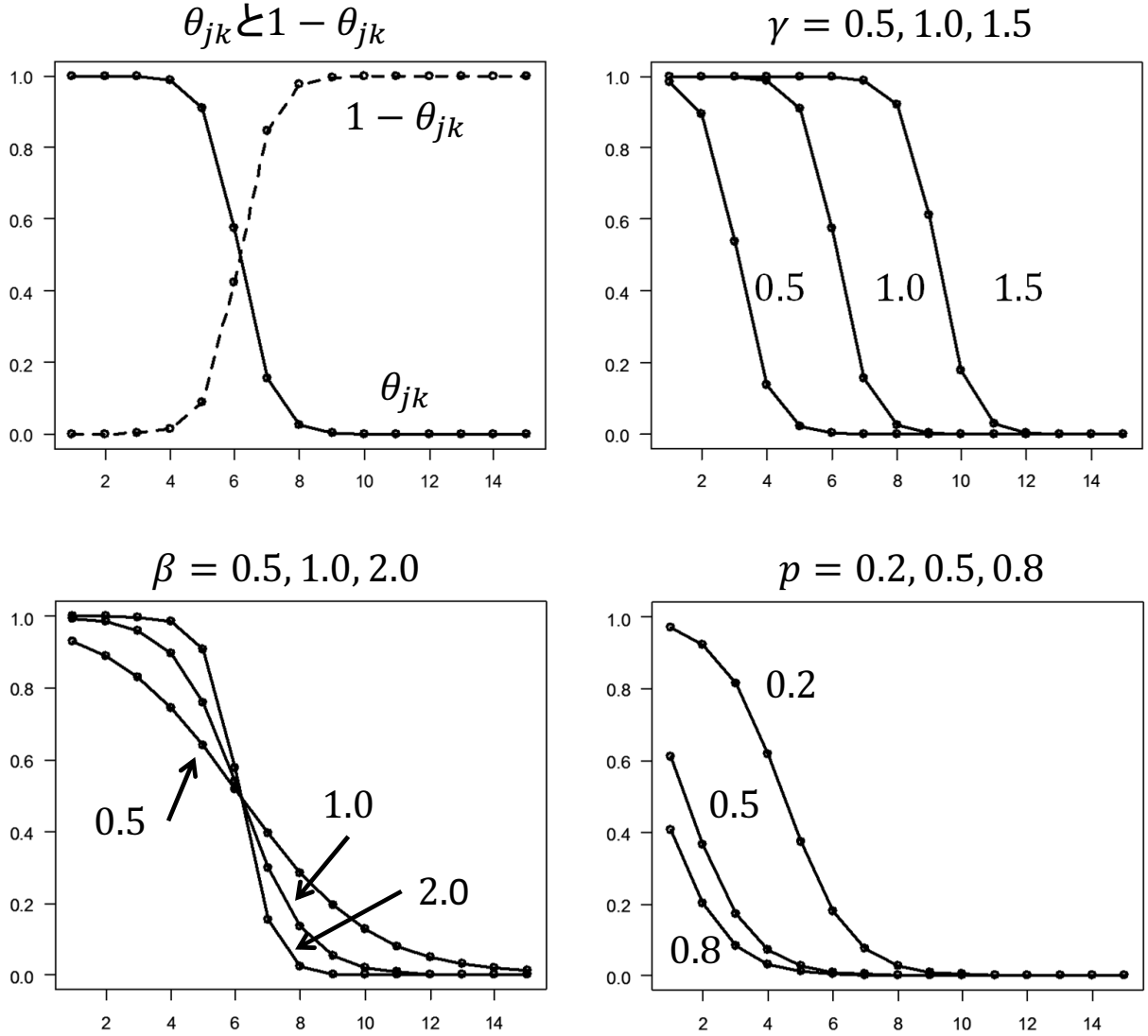


図. 7: 各母数による参加者が膨らませる確率 θ_{jk} の違い (縦軸：θ_{jk}，横軸：k)

図 7：リスクテイキング傾向を表す γ，行動の一貫性を表す β，風船が破裂する確率 p を変化させた際に，参加者が風船を膨らませる確率 θ_{jk} がどのように変化するかを示した図。

各図の横軸はパンプ回数，縦軸は確率 θ_{jk}

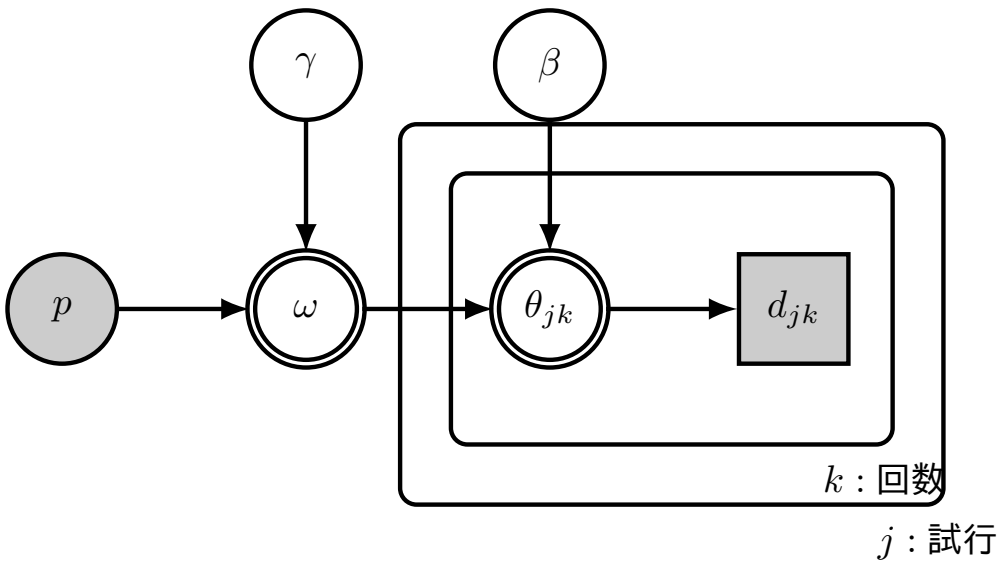
- 上段左は $\gamma = 1.0, \beta = 2.0, p = 0.15$ のときの θ_{jk} と $1 - \theta_{jk}$
- 上段右は $\beta = 2.0, p = 0.15$ と固定したときの $\gamma = (0.5, 1.0, 1.5)$ の違い
- 下段左は $\gamma = 1.0, p = 0.15$ と固定したときの $\beta = (0.5, 1.0, 2.0)$ の違い
- 下段右は $\gamma = 1.0, \beta = 1.0$ と固定したときの $p = (0.2, 0.5, 0.8)$ の違い

モデル化 3・BART モデルのプレート表現

実際に観察される j 回目の試行における k 回目の決定 d_{jk}($d_{jk}=0$：風船を膨らませる，$d_{jk}=1$:ポイントを獲得する) は，参加者が膨らませる確率を θ_{jk}，ポイントを獲得する確率を $\theta^*_{jk}=1-\theta_{jk}$ とすると，ベルヌイ分布を利用して以下のようにモデル化される。

$$d_{jk} \sim \text{Bernoulli}(\theta^*_{jk}) \tag{29}$$

2 つの母数の事前分布には 0 より大きな値を取る区間で一様分布を仮定。



$$\gamma \sim \text{Uniform}(0, u_\gamma)$$
$$\beta \sim \text{Uniform}(0, u_\beta)$$
$$\omega = -\gamma/\log(1-p)$$
$$\theta_{jk} = \frac{1}{1+\exp\{\beta(k-\omega)\}}$$
$$\theta^*_{jk} = 1-\theta_{jk}$$
$$d_{jk} \sim \text{Bernoulli}(\theta^*_{jk})$$

図. 8: BART モデルのプレート表現

BART モデル適用例

- 7 名からデータを収集し，破裂確率を $p = 0.15$ と設定して 30 試行の実験を行った。
- 最大のパンプ回数は 16 回とした。
- 分析では，4 つの連鎖を構成し，各連鎖で 5000 回サンプリングを行って，2500 個をバーンインとして破棄した。事後統計量は各連鎖のバーンイン期間後のサンプル 10000 個を利用した。

単一参加者の例として，ある 1 名のデータを分析する。

表 8: 母数の推定結果

	EAP	post.sd	2.5%	50%	97.5%
γ_1	2.553	0.076	2.429	2.544	2.728
β_1	1.231	0.256	0.778	1.215	1.776

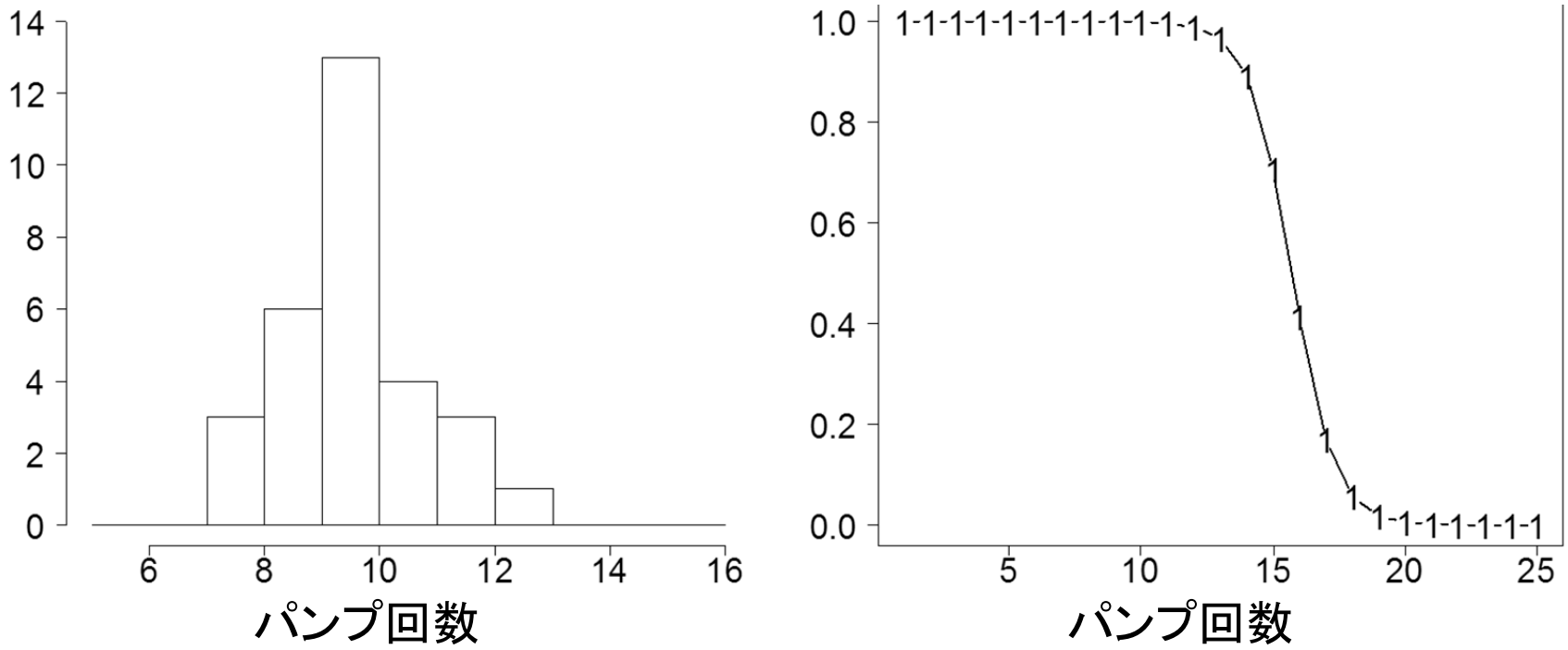


図. 9: ヒストグラム (上段左) と膨らませる確率 θ_{jk}(上段右)

単一参加者の例として，ある 1 名のデータを分析する。
図 9 左にはパンプ回数のヒストグラムを，図 9 右には，推定された母数の事後平均を利用して描いた風船を膨らませる確率 θ_{jk} を示した。図 9 右を見ると，12 回目まではほとんど確率は変わりませんが，それ以降減少していることが分かる。最大のパンプ回数が 16 回なので，比較的リスクテイキング傾向の高い参加者であると考えられる。

適用例 2：7 人のデータに対する BART モデルの適用

階層モデルを利用し，個人別のリスクテイキング傾向 γ_i と行動一貫性 β_i を推定。
γ_i と β_i の事前分布にはそれぞれ正規分布を仮定した。

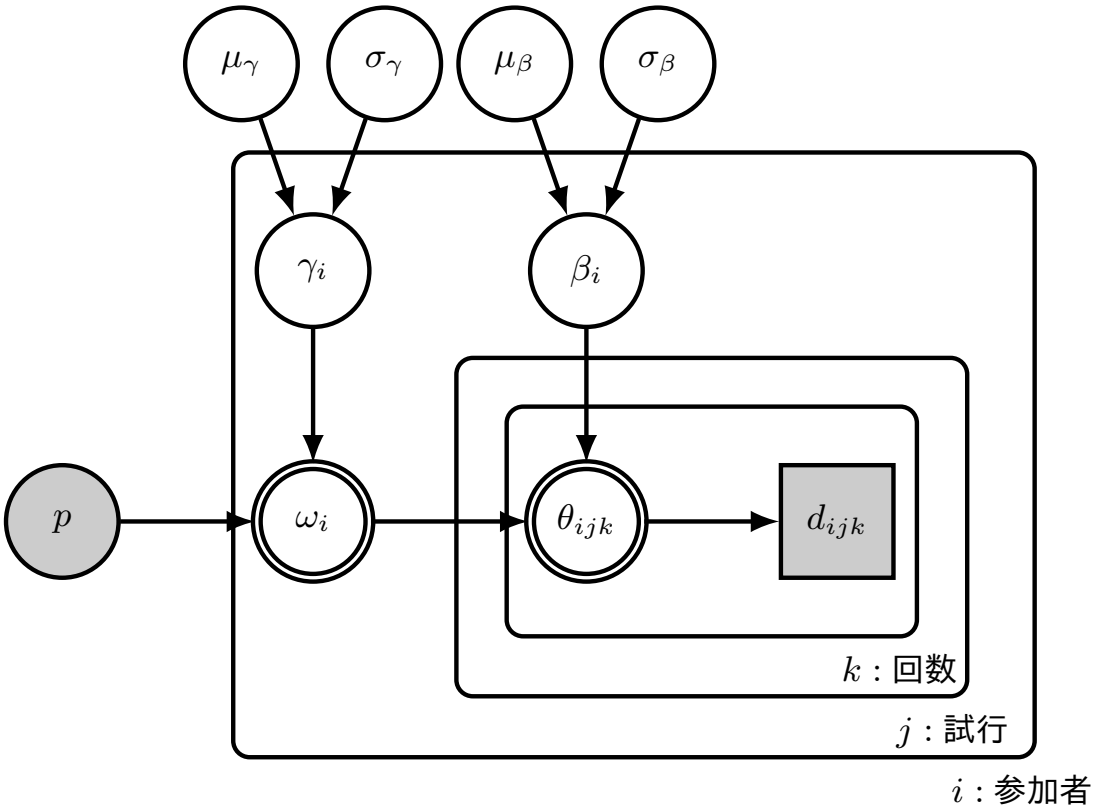


図. 10: 階層 BART モデルのプレート表現

$$
\begin{aligned}
\mu_\gamma &\sim \text{Uniform}(0, 10) \\
\sigma_\gamma &\sim \text{Uniform}(0, 10) \\
\mu_\beta &\sim \text{Uniform}(0, 10) \\
\sigma_\beta &\sim \text{Uniform}(0, 10) \\
\gamma_i &\sim N(\mu_\gamma, \sigma_\gamma^2) \\
\beta_i &\sim N(\mu_\beta, \sigma_\beta^2) \\
\omega_i &= -\gamma_i / \log(1 - p) \\
\theta_{ijk} &= \frac{1}{1 + \exp\{\beta_i (k - \omega_i)\}} \\
\theta^*_{ijk} &= 1 - \theta_{ijk} \\
d_{ijk} &\sim \text{Bernoulli}(\theta^*_{ijk})
\end{aligned}
$$

表 9: 母数の推定結果

	EAP	post.sd	2.5%	97.5%		EAP	post.sd	2.5%	97.5%
γ_1	2.572	0.071	2.436	2.717	β_1	1.086	0.151	0.845	1.439
γ_2	2.493	0.072	2.364	2.648	β_2	1.025	0.132	0.771	1.300
γ_3	2.582	0.073	2.451	2.738	β_3	0.998	0.126	0.745	1.260
γ_4	3.531	0.137	3.305	3.830	β_4	1.080	0.175	0.793	1.505
γ_5	2.312	0.063	2.200	2.455	β_5	1.092	0.149	0.855	1.442
γ_6	2.188	0.089	2.037	2.398	β_6	0.916	0.137	0.626	1.169
γ_7	2.126	0.061	2.012	2.254	β_7	1.080	0.140	0.844	1.395

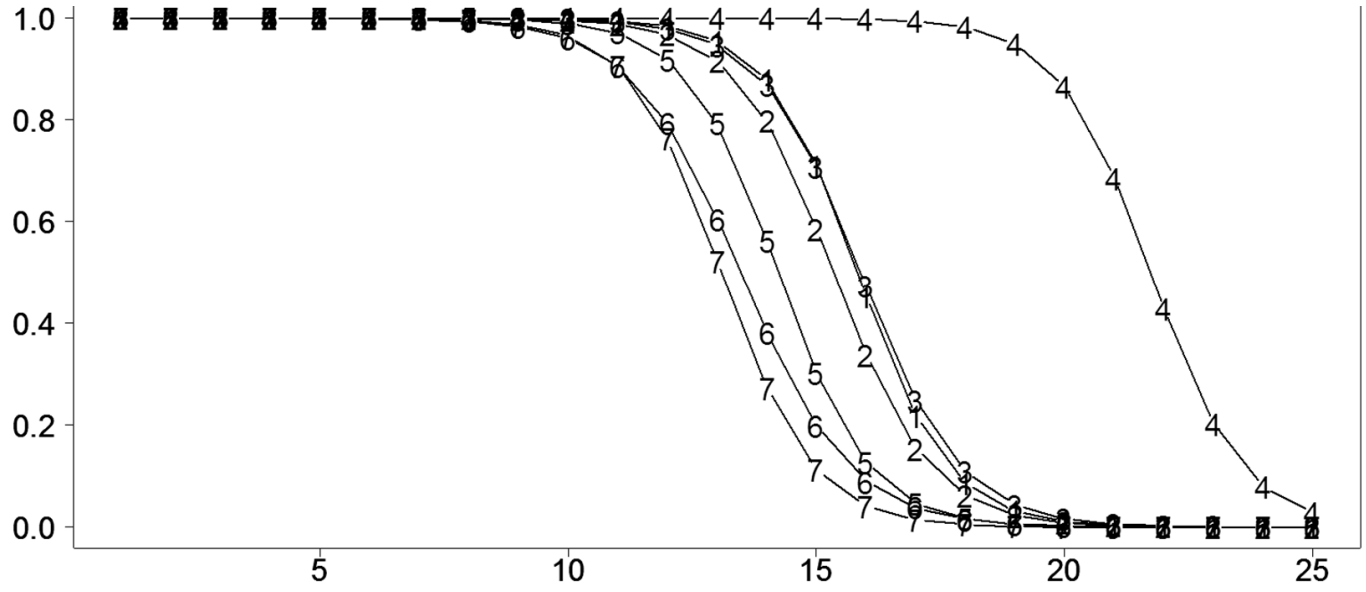


図. 11: 推定値から計算される確率

推定値より，最もリスクテイキング傾向の高い参加者は参加者 4 であり，もっとも低い参加者は参加者 7 であることが分かる。図 11 においても，参加者 7 の曲線 (7) が一番左にあり，10 回を超えた辺りから風船を膨らませる確率が減少していることが分かる。
一方，参加者 4 の曲線 (4) は一番右にあり，他の参加者と大きく離れている。最大回数の 16 回を超えても確率が減少していないことから，この実験においてはほとんどの試行において風船が破裂しており，リスクを冒してでもポイントを獲得しようとする傾向があると考えられる。

- リンク関数を利用することで，様々な分布モデルの母数に対して線形構造を仮定することができる。

- プレート表現を利用することで，ベイジアンモデルのデータ生成過程を視覚的に表現可能である。

- BART モデルを利用することで，参加者の行動の一貫性やリスクテイキング傾向を個人差を踏まえて推測することができる。

ベルヌイ・ロジットリンク

```
data{
  int<lower=0> N;
  real edu[N];
  int<lower=0> y[N];
}
parameters{
  real b;
  real a;
}
model{
  for(i in 1:N){ #下２行の表現は同等
    #y[i] ~ bernoulli(inv_logit(b + a * edu[i]));
    y[i] ~ bernoulli_logit(b + a * edu[i]);
  }
}
generated quantities{
  ## モデル比較を行う際に利用する変数
  real log_lik[N];
  for(i in 1:N)
    log_lik[i] = bernoulli_logit_lpmf(y[i]|b + a * edu[i]);
}
```

ポアソン・対数リンク

```
data{
  int<lower=0> N;
  int<lower=0> y[N];
  real age[N];
  real treat[N];
}
parameters{
  real b;
  real a1;
  real a2;
}
model{
  for(i in 1:N){
    y[i] ~ poisson_log(b + a1*age[i] + a2*treat[i]);
  }
}
```

負の 2 項・対数リンク

```
data {
  int<lower=0> N;
  int<lower=0> y[N];
  real<lower=0> x1[N];
  real<lower=0> x2[N];
  real<lower=0> x3[N];
}
parameters {
  real<lower=0> phi;
  real b;
  real a1;
  real a2;
  real a3;
}
model {
  for(i in 1:N){
    y[i] ~ neg_binomial_2_log(b+a1*x1[i]+a2*x2[i]+a3*x3[i], phi);
  }
}
```

BART モデル（適用例 1）

```
data {
  int<lower=1> ntrials;
  int<lower=1> maxpump;
  vector<lower=0,upper=1>[ntrials] p;
  int<lower=1> options[ntrials];
  int d[ntrials,maxpump];
}
parameters {
  real<lower=0,upper=10> gamma;
  real<lower=0,upper=10> beta;
}
transformed parameters {
  vector<lower=0>[ntrials] omega;
  for(j in 1:ntrials){ omega[j] = -gamma / log1m(p[j]); }
}
model {
  for (j in 1:ntrials) {
    for (k in 1:options[j]) {
      real theta;
      theta = 1 - inv_logit(-beta * (k - omega[j]));
      d[j,k] ~ bernoulli(theta);
    }
  }
}
```

BART モデル（適用例 2）

```
data {
  int<lower=1> N;
  int<lower=1> ntrials;
  int<lower=1> maxpump;
  vector<lower=0,upper=1>[ntrials] p;
  int<lower=1> options[N,ntrials];
  int d[ntrials,maxpump,N];
}
parameters {
  vector<lower=0,upper=10>[N] gamma;
  vector<lower=0,upper=10>[N] beta;
  real mu_g;
  real<lower=0> sigma_g;
  real mu_b;
  real<lower=0> sigma_b;
}
# 次のスライドに続く
```

```
transformed parameters {
  vector<lower=0>[ntrials] omega[N];
  for(i in 1:N){
    for(j in 1:ntrials){
      omega[i,j] = -gamma[i] / log1m(p[j]);
    }
  }
}
model {
  for(i in 1:N){
    for (j in 1:ntrials) {
      for (k in 1:options[i,j]) {
        real theta;
        theta = 1 - inv_logit(-beta[i] * (k - omega[i,j]));
        d[j][k,i] ~ bernoulli(theta);
      }
    }
  }
  gamma ~ normal(mu_g,sigma_g);
  beta ~ normal(mu_b,sigma_b);
}
#ここまで
```

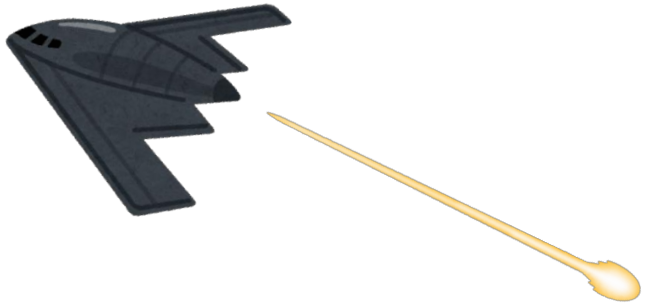
信号検出理論・トピックモデル

早稲田大学 グローバルエデュケーションセンター
久保 沙織

信号検出理論

信号検出理論

- 刺激の有無を判断する実験（Yes-No課題）において，刺激を正しく判別する検出力を測定するために用いられる



【適用場面】

✓単語・非単語識別課題

✓スクリーニング検査の精度

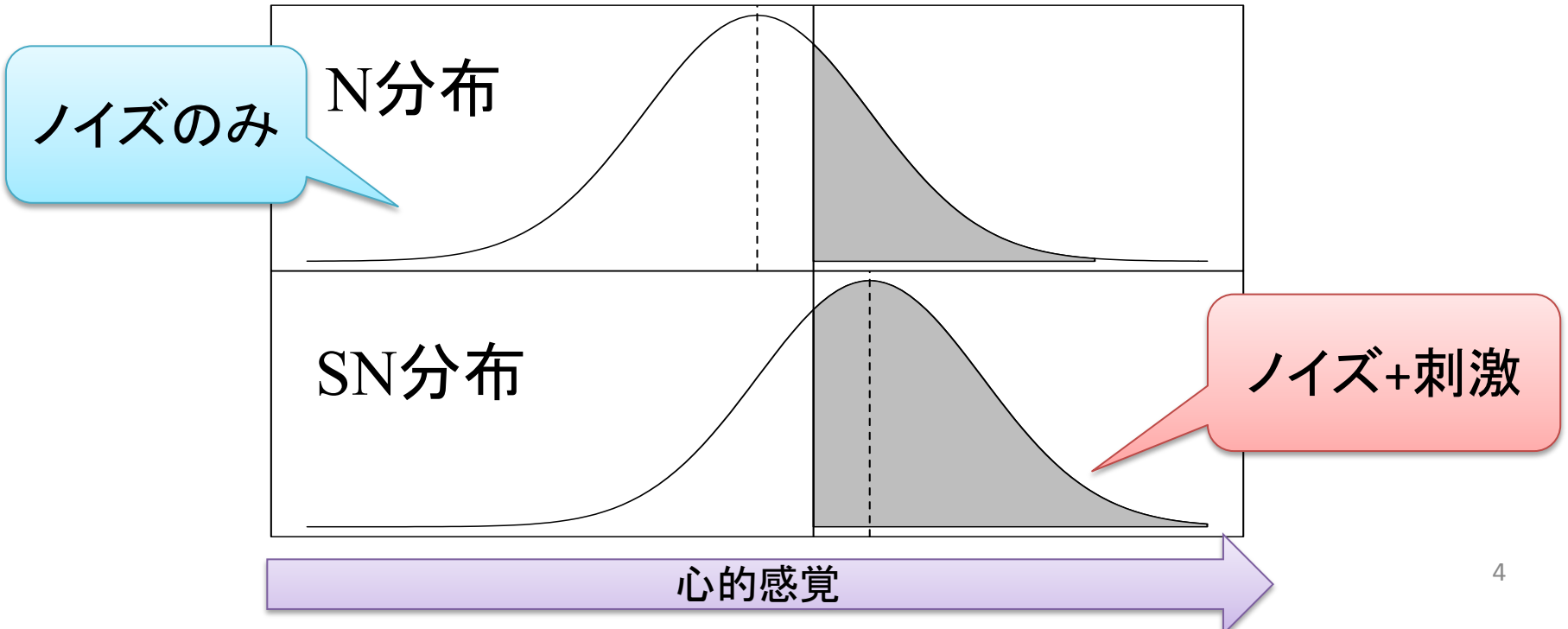
N分布とSN分布

反応 / 刺激	あり	なし
Yes	ヒット (hit)	誤警報 (false alarm)
No	ミス (miss)	正棄却 (correct rejection)

- それぞれの反応が得られる確率は？
  ⇒ ノイズ分布（noise distribution; N分布）と信号+ノイズ分布（signal-plus-noise distribution; SN分布）という2つの分布を仮定

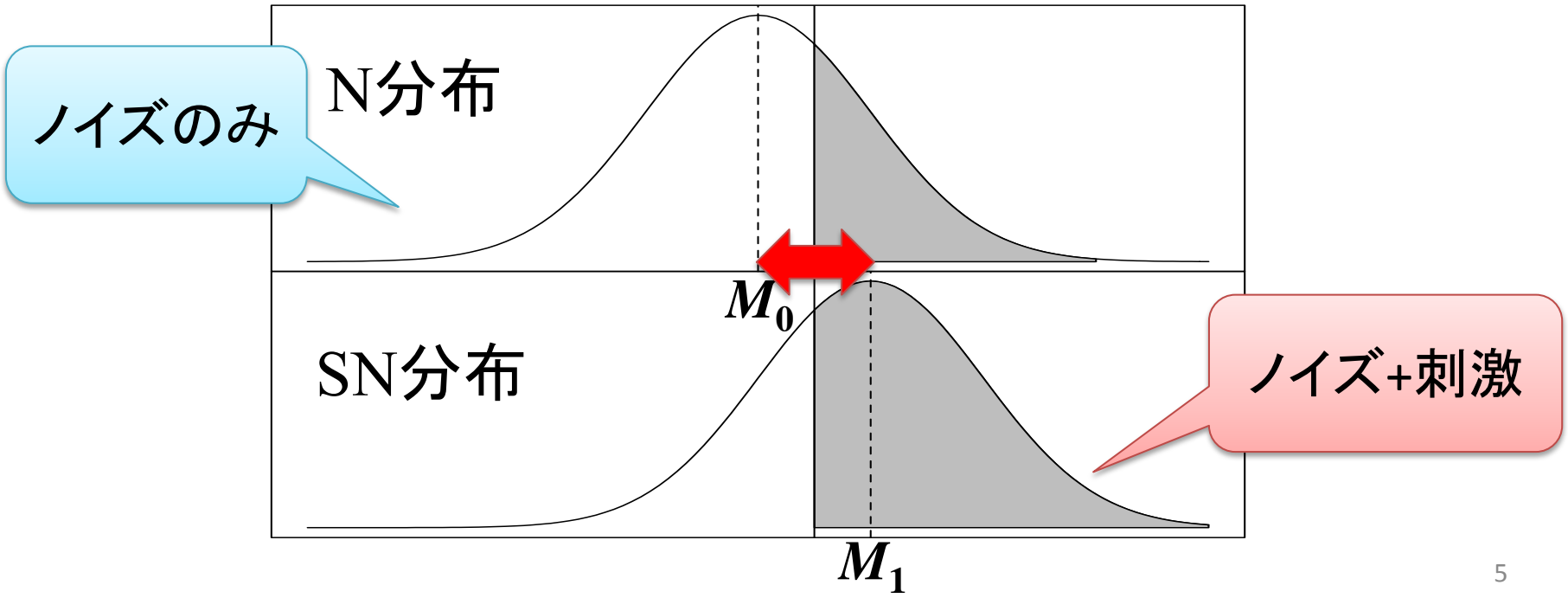
等分散を仮定した信号検出理論のモデル

- N分布とSN分布にはそれぞ正規分布を仮定し，それらの分散は等しいとする（equal-variance Gaussian SDT, Lee(2008)）



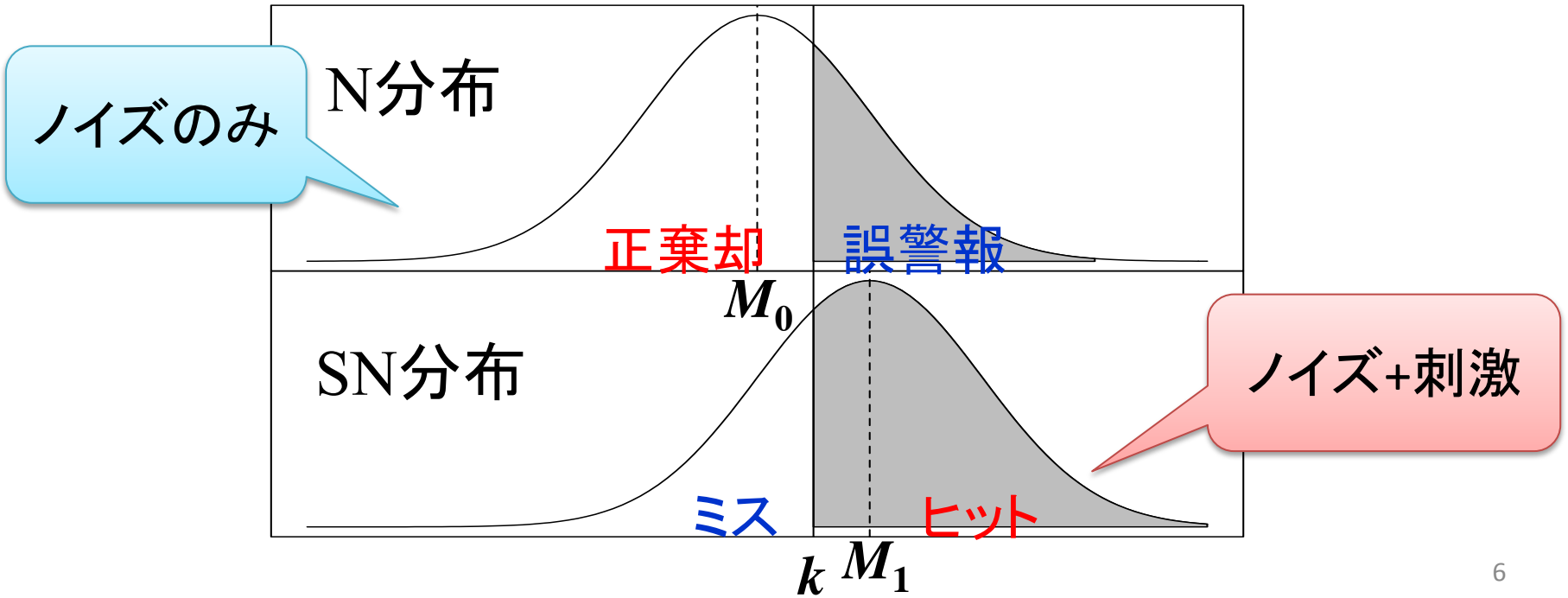
等分散を仮定した信号検出理論のモデル

- M_0：N分布における心的感覚の平均
- M_1：SN分布における心的感覚の平均
- M_0とM_1の差は刺激の強度に応じて変化



等分散を仮定した信号検出理論のモデル

- 実験参加者は心的感覚がある一定の値以上となったときに刺激がある(Yes)と判断する
- k: Yesと反応するか否かの判断基準の位置

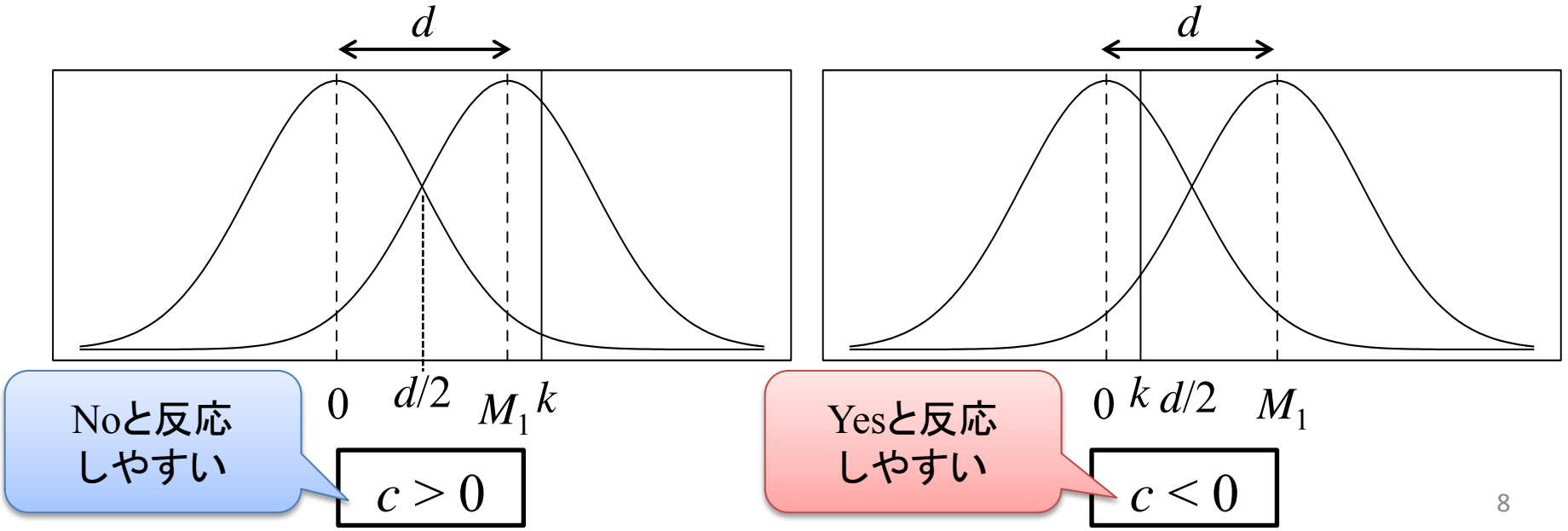


信号検出力

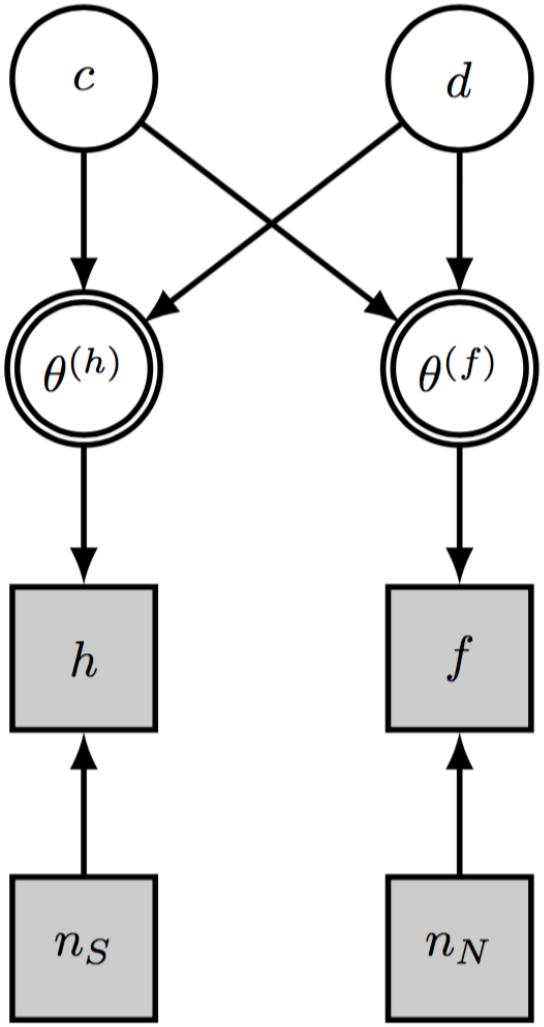
- 信号検出力d: $d = \dfrac{M_1 - M_0}{s}$
  （sは両分布に共通の標準偏差）
- N分布に標準正規分布$N(0,1)$を仮定すると，
  $M_0 = 1$，$s = 0$
  $\Rightarrow\ d = M_1$
- dは刺激の強度と感覚系の特性によって決まり，判断基準の位置kとは独立

反応バイアス

- 反応バイアスc: $c = k - \frac{d}{2}$
- $k = d/2$のときc=0
  ⇒ヒットの確率=正棄却の確率
  　誤警報の確率=ミスの確率



モデル1：非階層モデル



- n_S：刺激あり試行の数（ヒット+ミス）
- n_N：刺激なし試行の数（誤警報+正棄却）
- h：ヒットの度数
- f：誤警報の度数

$$h \sim \mathrm{Binomial}(\theta^{(h)}, n_S)$$

$$f \sim \mathrm{Binomial}(\theta^{(f)}, n_N)$$

$$\theta^{(h)} = \Phi\left(\frac{1}{2}d - c\right)$$

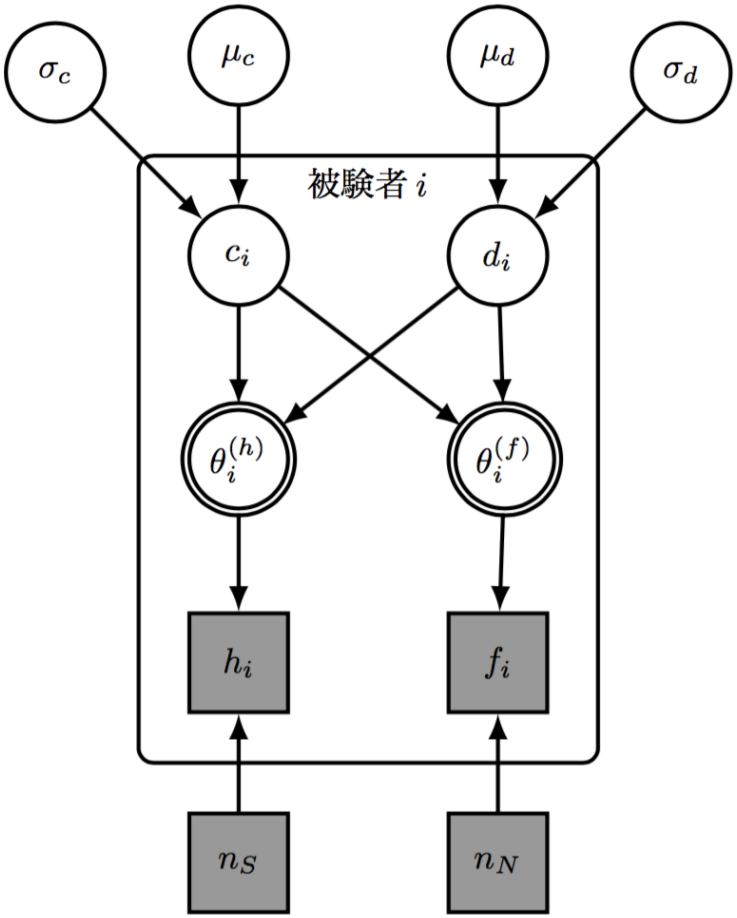
$$\theta^{(f)} = \Phi\left(-\frac{1}{2}d - c\right)$$

$$d \sim \mathrm{Normal}(0, \sqrt{2})$$

$$c \sim \mathrm{Normal}(0, 1/\sqrt{2})$$

モデル2：階層モデル

- 信号検出力dと反応バイアスcが個人ごとに異なることを表現



n_Sとn_Nはそれぞれすべての実験参加者を通して同じとする

$$h_i \sim \text{Binomial}(\theta_i^{(h)}, n_S)$$

$$f_i \sim \text{Binomial}(\theta_i^{(f)}, n_N)$$

$$\theta_i^{(h)} = \Phi\left(\frac{1}{2}d_i - c_i\right)$$

$$\theta_i^{(f)} = \Phi\left(-\frac{1}{2}d_i - c_i\right)$$

$$d_i \sim \text{Normal}(\mu_d, \sigma_d)$$

$$c_i \sim \text{Normal}(\mu_d, \sigma_d)$$

$$\mu_d, \mu_c \sim \text{Normal}(0, \sqrt{1000})$$

$$\sigma_d, \sigma_c \sim \text{Uniform}(0, \infty)$$

実験

1. 学習課題
   実験参加者に，合計60個の単語を呈示し，できるだけ多くの単語を記憶するよう教示。単語は，15単語から成る4つのリストに分けて順番に呈示し，各リストの学習時間は30秒とした。
   単語リストは，星野(2002)，堀田(2007)，宮地・山(2002)の実験刺激を参考にして作成し，リストの呈示順序を実験参加者ごとにランダムに変更した。
2. 再認課題
   学習課題の直後，実験参加者に48単語が12×4の配列で記載された用紙を渡し，学習段階で見たと思う単語すべてに丸を付けるよう教示。
   再認課題において呈示された単語の半分は学習段階で呈示した中に含まれる単語(Old)，残り半分は新規な単語(New)であったが，この比率は実験参加者には知らせていない。

実験結果（全体のクロス集計表）

実験参加者7名の反応の合計（n_S=n_N=168）

反応 / 刺激	Old	New
Yes	109 （ヒット）	15 （誤警報）
No	59 （ミス）	153 （正棄却）

このデータにモデル1を適用

分析結果1

	EAP	post.sd	95% 下側	95% 上側
$\theta^{(h)}$	0.647	0.036	0.575	0.717
$\theta^{(f)}$	0.094	0.022	0.055	0.142
d	1.709	0.164	1.390	2.042
c	0.475	0.083	0.314	0.639

実験結果（個人のクロス集計表）

実験参加者5の反応

反応 / 刺激	Old	New
Yes	10 （ヒット）	2 （誤警報）
No	14 （ミス）	22 （正棄却）

実験参加者6の反応

反応 / 刺激	Old	New
Yes	20 （ヒット）	4 （誤警報）
No	4 （ミス）	20 （正棄却）

これら個人のデータにモデル2を適用

分析結果2

実験参加者	$\theta^{(h)}$	$\theta^{(f)}$	d	c
1	0.645	0.100	1.695	0.469
2	0.558	0.064	1.734	0.717
3	0.647	0.074	1.882	0.555
4	0.722	0.099	1.931	0.362
5	0.523	0.074	1.569	0.726
6	0.756	0.136	1.846	0.209
7	0.679	0.092	1.845	0.448

トピックモデル

トピックモデル

- 文書データを定量的に分析し，そこに潜む意味をとらえることを目的とした分析手法
- トピックとは，文書の主題・テーマなどのこと
- 各文書が複数のトピックを持つと仮定

【適用場面】

✓アンケートの自由記述の分析

✓インタビューデータの分析

文書のデータ化

- 文書における単語の出現順序にはとらわれずに，文書ごとに，どのような単語が何回使われているかを整理する
- D: 全文書数
- N_d: d番目の文書に含まれる単語数
- N: 文書全体に含まれる単語の総数

$$N_1 + \cdots + N_d + \cdots + N_D$$

- V: 文書全体に含まれる単語の種類（Nから重複を除いた数）

文書のデータ化

- 各文書を構成する単語:

$$\boldsymbol{w}_d = (w_{d1}, \ldots, w_{dn}, \ldots, w_{dN_d})$$

w_{dn}: d(=1,…, D)番目の文書のn(=1, …, N_d)番目の単語

- D個の文書の集合:

$$\boldsymbol{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_d, \ldots, \boldsymbol{w}_D)$$

文書生成過程

①文書ごとに異なるトピック混合比率(トピック分布)に従って各単語の潜在的なトピックz_{dn}が選択される

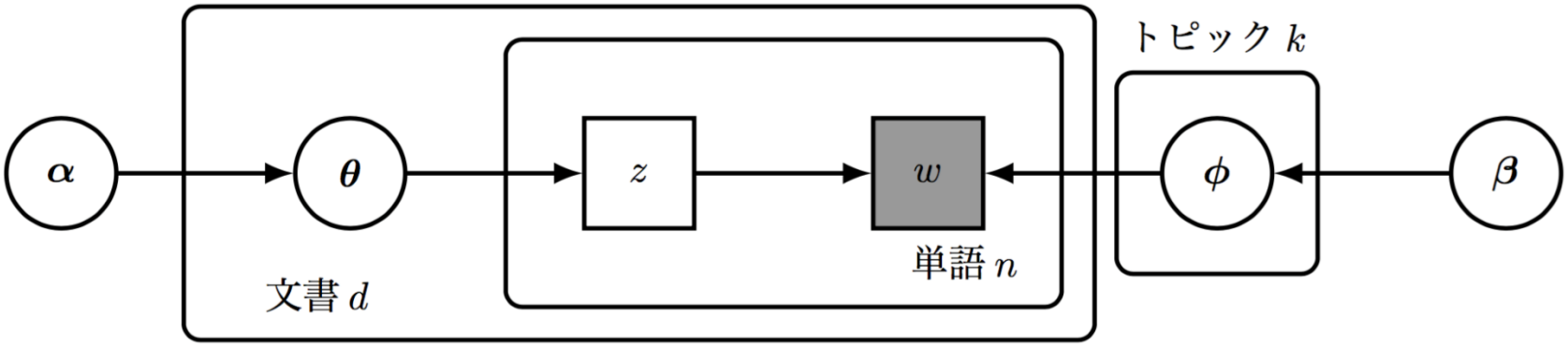
②トピックごとに異なる単語出現確率分布(単語分布)に従って単語w_{dn}が生成される
→文書全体が完成

- トピックを潜在変数として扱い，トピックの数Kは分析者が与える
- z_{dn}：単語w_{dn}に対応する潜在変数

潜在ディリクレ配分

- トピック分布の事前分布としてディリクレ分布(Dirichlet distribution) を仮定し，ベイズ推定する手法のことを潜在ディリクレ配分(Latent Dirichlet Allocation, LDA; Blei, Ng & Jordan (2003))モデルとよぶ

LDAによるトピックモデル



- $\boldsymbol{\theta}_d = (\theta_{d1}, \ldots, \theta_{dk}, \ldots, \theta_{dK})$: 文書dにおけるK個のトピックの混合比率
- $\boldsymbol{\phi}_k = (\phi_{k1}, \ldots, \phi_{kv}, \ldots, \phi_{kV})$: トピックkにおけるV種類の単語の出現確率

$$z_{dn} \sim \mathrm{Categorical}(\boldsymbol{\theta}_d),\ w_{dn} \sim \mathrm{Categorical}(\boldsymbol{\phi}_{z_{dn}})$$

$$\boldsymbol{\theta}_d \sim \mathrm{Dirichlet}(\boldsymbol{\alpha}) \quad (d = 1, \ldots, D)$$

$$\boldsymbol{\phi}_k \sim \mathrm{Dirichlet}(\boldsymbol{\beta}) \quad (k = 1, \ldots, K)$$

カテゴリカル分布(Murphy, 2012)

- 値がkとなる確率がθ_kであるK種類の離散値のうち1つの値が生じるような試行を1回行ったときの結果が従う確率分布

$$f(x = k|\boldsymbol{\theta}) = \mathrm{Categolical}(x|\boldsymbol{\theta}) = \theta_k$$
$$\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_k, \ldots, \theta_K\}'$$

ただし，$\theta_k \geq 0, \quad \sum_{k=1}^{K} \theta_k = 1$

- カテゴリカル分布の母数θの事前分布として一般的に用いられるのがディリクレ分布

ディリクレ分布

$$f(\boldsymbol{\theta}|\boldsymbol{\alpha}) = \mathrm{Dirichlet}(\boldsymbol{\theta}|\boldsymbol{\alpha}) = \frac{\Gamma\left(\sum_{k=1}^{K}\alpha_k\right)}{\prod_{k=1}^{K}\Gamma(\alpha_k)}\prod_{k=1}^{K}\theta_k^{\alpha_k-1}$$

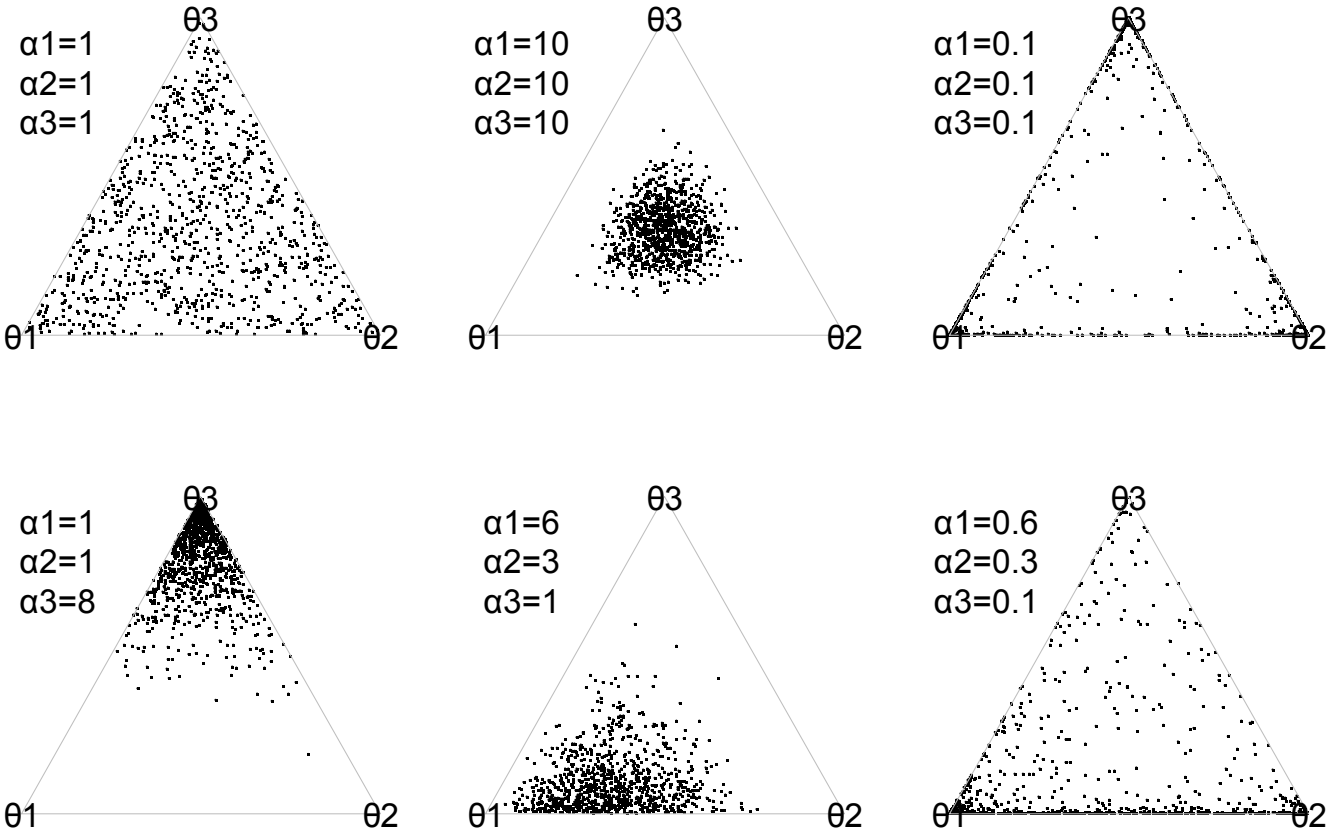
$$E[\boldsymbol{\theta}] = \left(\frac{\alpha_1}{\alpha},\cdots,\frac{\alpha_k}{\alpha},\ldots,\frac{\alpha_K}{\alpha}\right)$$

$$V[\boldsymbol{\theta}] = \left(\frac{\alpha_1(\alpha-\alpha_1)}{\alpha^2(\alpha+1)},\ldots,\frac{\alpha_k(\alpha-\alpha_k)}{\alpha^2(\alpha+1)},\ldots,\frac{\alpha_K(\alpha-\alpha_K)}{\alpha^2(\alpha+1)}\right)$$

ただし, $\alpha = \sum_{k=1}^{K}\alpha_k$

ディリクレ分布

- K=3のとき，ディリクレ分布に従う確率変数は，θ_1=(1,0,0)，θ_2=(0,1,0)，θ_3=(0,0,1)を頂点とする三角形内の点として表される



分析例

- 米国のAP通信(Associated Press)の記事データにLDAによるトピックモデルを適用
- Rのパッケージtopicmodelsに含まれるデータAssociatedPressより，100文書をランダム抽出して利用した
- 全文書数D=100，重複のない単語の種類V=10473
- トピック数はK=4として分析を実行

データファイル

```
K<-4          #トピック数
V<-10473      #単語の種類
D<-100        #文書数
N<-13131      #データサイズ
wordID<-
c(1,2,3,4,5,6,7,8,9,10,･･･,2750,742,4954,1795,213,
  3590,564,218,392,830)
freq<-c(1,2,2,1,1,1,1,1,1,2,･･･,2,1,1,1,1,1,1,1,1,2)
docID<-c(1,1,1,1,1,1,1,1,1,1,･･･,100,100,100,100,100,
         100,100,100,100,100)
alpha<-rep(1, 4)          #トピック分布の事前分布
beta<-rep(0.5, 10473)  #単語分布の事前分布
```

Stanコード

```
parameters {
  simplex[K] theta[D];
  simplex[V] phi[K];
}
model {
  for (d in 1:D)
    theta[d] ~ dirichlet(alpha);
  for (k in 1:K)
    phi[k] ~ dirichlet(beta);
  for (n in 1:N) {
    real gamma[K];
     for (k in 1:K)
       gamma[k] = log(theta[doc[n],k]) + log(phi[k,w[n]]);
    target += Freq[n]*log_sum_exp(gamma);
  }
}
```

simplex[K] theta[D]

によって，$\theta_k \geq 0, \quad \sum_{k=1}^{K} \theta_k = 1$

を満たす変数を定義

トピックは潜在変数であり，本来z_{dn}も母数としてサンプリングする必要があるが，StanのHMCでは離散変数のサンプリングが不可

⇒ target +=の右辺にモデルの対数尤度

対数尤度

- 事後分布

尤度

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}|\boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \propto p(\boldsymbol{w}|\boldsymbol{\theta}, \boldsymbol{\phi})p(\boldsymbol{\theta}|\boldsymbol{\alpha})p(\boldsymbol{\phi}|\boldsymbol{\beta})$$

$$p(\boldsymbol{w}|\boldsymbol{\theta}, \boldsymbol{\phi}) = \prod_{d=1}^{D}\prod_{n=1}^{N_d} p(w_{dn}|\boldsymbol{\theta}_d, \boldsymbol{\phi}) = \prod_{d=1}^{D}\prod_{n=1}^{N_d}\left\{\sum_{k=1}^{K} p(k, w_{dn}|\boldsymbol{\theta}_d, \boldsymbol{\phi})\right\}$$

$$= \prod_{d=1}^{D}\prod_{n=1}^{N_d}\left\{\sum_{k=1}^{K}(p(k|\boldsymbol{\theta}_d)p(w_{dn}|\boldsymbol{\phi}_k))\right\}$$

対数尤度

$$\log(p(\boldsymbol{w}|\boldsymbol{\theta}, \boldsymbol{\phi})) = \sum_{d=1}^{D}\sum_{n=1}^{N_d}\log\left\{\sum_{k=1}^{K}(\mathrm{Categorical}(k|\boldsymbol{\theta}_d) \times \mathrm{Categorical}(w_{dn}|\boldsymbol{\phi}_k))\right\}$$

分析結果(単語分布)

各トピックの出現確率上位5番目までの単語

	トピック 1	トピック 2	トピック 3	トピック 4
1	nbc(0.0031)	new(0.0047)	irs(0.0028)	people(0.0054)
2	available(0.0028)	million(0.0045)	returns(0.0020)	state(0.0042)
3	number(0.0020)	stock(0.0044)	program(0.0019)	south(0.0039)
4	new(0.0020)	company(0.0033)	income(0.0017)	i(0.0036)
5	ticketron(0.0020)	percent(0.0031)	corporations(0.0016)	states(0.0035)

米国のチケット電話
予約システム

企業による所得税
の確定申告

分析結果（トピック分布）

各文書のトピック分布

	文書 22	文書 23	文書 31	文書 57	文書 60
θ_1	0.420	0.010	0.933	0.008	0.006
θ_2	0.060	0.007	0.010	0.043	0.978
θ_3	0.066	0.009	0.007	0.801	0.006
θ_4	0.454	0.974	0.050	0.148	0.009
	トピック1 トピック4	トピック4	トピック1	トピック3	トピック2

参考文献(信号検出理論)

- Gescheider, G. A. (2002). 心理物理学—方法・理論・応用—上巻, 宮岡徹 監訳, 倉片憲治・金子利佳・芝崎朱美 訳　北大路書房.
- Gescheider, G. A. (2003). 心理物理学—方法・理論・応用—下巻, 宮岡徹 監訳, 倉片憲治・金子利佳・芝崎朱美 訳　北大路書房.
- 星野祐司 (2002). 関連語の学習による誤再生とリスト構成:ブロック呈示条件とランダム呈示条件の比較. 基礎心理学研究, **20**(2), 105-114.
- 堀田千絵 (2007). 虚再認における指示忘却の効果:活性化-モニタリング仮説の検討. 心理学研究, **78**(1), 57-62.
- Lee, M. D. (2008). BayesSDT: Software for Bayesian inference with signal detection theory. *Behavior Research Methods*, **40**(2), 450-456.
- Macmillan, N. A. & Creelman, C. D. (2005). *Detection Theory: A User's Guide*, *2nd ed*. Lawrence Erlbaum Associates.
- 宮地弥生・山祐嗣 (2002). 高い確率で虚記憶を生成するDRMパラダイムのための日本語リストの作成. 基礎心理学研究, **21**(1), 21-26.

参考文献（トピックモデル）

- Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet allocation, *Journal of Machine Larning*, **3**, 993-1022.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, **41** (6), 391-407.
- 石井健一郎・上田修功 (2014) 続・わかりやすいパターン認識―教師なし学習入門―, オーム社.
- 岩田具治 (2015). トピックモデル, 講談社.
- Murphy, K. P. (2012) Machine learning: a probabilistic perspective, p.35, The MIT Press.
- 奥村学 監修・佐藤一誠 著 (2015) トピックモデルによる統計的潜在意味解析, コロナ社.
- Stan Development Team. (2016). *Stan Modeling Language User's Guide and Reference Manual* Version 2.11.0

付録
（Stanコード）

◆ 信号検出理論モデル1（非階層モデル）

```
data {
  int<lower=0> h;
  int<lower=0> f;
  int<lower=0> MI;
  int<lower=0> CR;
}
transformed data{
  int<lower=0> s;
  int<lower=0> n;
  s = h + MI;
  n = f + CR;
}
```

```
parameters {
  real d;
  real c;
}
transformed parameters {
  real<lower=0,upper=1> thetah;
  real<lower=0,upper=1> thetaf;

    thetah = Phi(d / 2 - c);
    thetaf = Phi(-d / 2 - c);
}
model {
  d ~ normal(0, sqrt(2));
  c ~ normal(0, inv_sqrt(2));
  h ~ binomial(s, thetah);
  f ~ binomial(n, thetaf);
}
```

◆ 信号検出理論モデル2（階層モデル）

```
data {
  int<lower=1> k;
  int<lower=0> h[k];
  int<lower=0> f[k];
  int<lower=0> s;
  int<lower=0> n;
}
parameters {
  vector[k] d;
  vector[k] c;
  real mud;
  real muc;
  real<lower=0>
sigmad;
  real<lower=0>
sigmac;
}
```

```
transformed parameters {
  real<lower=0,upper=1> thetah[k];
  real<lower=0,upper=1> thetaf[k];

  for(i in 1:k) {
    thetah[i] = Phi(d[i] / 2 - c[i]);
    thetaf[i] = Phi(-d[i] / 2 - c[i]);
  }
}
model {
  mud ~ normal(0, sqrt(1000));
  muc ~ normal(0, sqrt(1000));

  d ~ normal(mud, sigmad);
  c ~ normal(muc, sigmac);
  h ~ binomial(s, thetah);
  f ~ binomial(n, thetaf);
}
```

◆トピックモデル

```
data {
  int<lower=2> K;
  int<lower=2> V;
  int<lower=1> D;
  int<lower=1> N;
  int<lower=1,upper=V> w[N];
  int<lower=1> Freq[N];
  int<lower=1,upper=D> doc[N];
  vector<lower=0>[K] alpha;
  vector<lower=0>[V] beta;
}
parameters {
  simplex[K] theta[D];
  simplex[V] phi[K];
}
```

```
model {
  for (d in 1:D)
    theta[d] ~ dirichlet(alpha);
  for (k in 1:K)
    phi[k] ~ dirichlet(beta);
  for (n in 1:N) {
    real gamma[K];
     for (k in 1:K)
       gamma[k] = log(theta[doc[n],k])
                  + log(phi[k,w[n]]);
    target += Freq[n]*log_sum_exp(gamma);
  }
}
```

隠れマルコフモデル
(汎用的解析ツール)
&
アイオワ・ギャンブリング課題
(心理モデル)

早稲田大学 大学院
長尾圭一郎

隠れマルコフモデル

1. 隠れマルコフモデル

- 「女心と秋の空」・・・
- 男性は女性が自分のことをどう思っているのかを女性の服装や仕草から推察するのでは
- 「ボディタッチ」 ⇒ 「自分に好意がある」
- 「髪型を気にしている」

  ⇒「自分のことが少し気になっている」
- 「スマートフォンを頻繁に使用している」

  ⇒「自分に興味がない」と解釈

- 男性の推察方法と似た統計手法が隠れマルコフモデル(hidden Markov model,HMM)
- 遷移する潜在変数の状態によって，観測される事象の確率分布が異なる確率モデルを表現
- 直接観測することができない潜在変数の状態を推測できる
- 心理学の分野でも応用できる(ex. 臨床系，マーケティング)

1.1 マルコフ連鎖

- 時間とともに変化する確率変数，あるいは現象のことを確率過程といい，時点を表す添え字$\tau(=1,\dots,T)$を用いて$Z^{(\tau)}$と表す
- $Z^{(\tau)}$の条件付き確率が直前の状態にのみ依存する確率過程を，特にマルコフ連鎖という
- マルコフ連鎖$Z^{(\tau)}=\{q_1,\dots,q_K\}$がとりうる状態$K$個のうち，状態$q_i$から状態$q_j$へ移る条件付確率を遷移確率といい，$a_{ij}$で表す

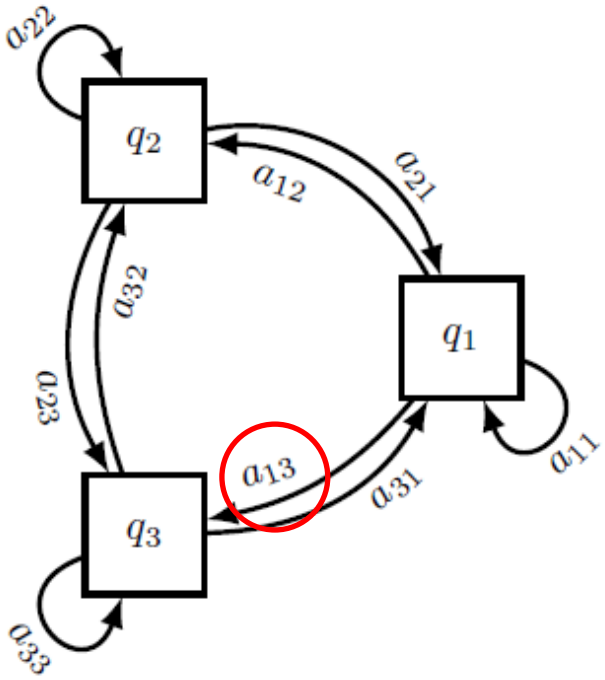


図1.1:マルコフ連鎖における
状態遷移図

列:状態q_j

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1K} \\ \vdots & \ddots & \vdots \\ a_{K1} & \cdots & a_{KK} \end{pmatrix}$$

行:状態q_i

- 冒頭の例でいえば，$K = 3$であり，「q_1:自分に好意がある」から「q_3:自分に興味がない」に遷移する確率はa_{13}と表される
- 遷移核:遷移確率を要素に持つ行列A

隠れマルコフモデルとは

- 潜在変数であるマルコフ連鎖の各状態において，確率変数である出力を備えたモデル
- 冒頭の例だと，「女性の心情」が潜在変数$Z^{(\tau)}$，「服装や仕草」が出力$X^{(\tau)}$

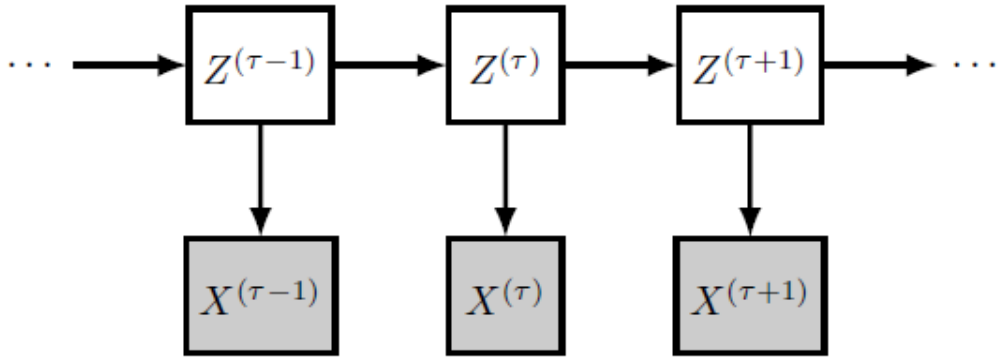


図1.2:隠れマルコフモデルの状態空間モデル

状態遷移系列：$Z^{(\tau)}$の系列
出力系列：$X^{(\tau)}$の系列

- $X^{(\tau)} = \{o_1, \ldots o_M\}$のように，出力の結果が$M$個に限られるとき，状態$q_j$から出力$o_l$が得られる確率を出力確率といい，$b_{jl}$で表す
- 出力確率行列: 出力確率を要素にもつ行列B

$$B = \begin{array}{c} \text{列}:\text{出力}o_l \\ \text{行}:\text{状態}q_j \end{array} \begin{pmatrix} b_{11} & \cdots & b_{1M} \\ \vdots & \ddots & \vdots \\ b_{K1} & \cdots & b_{KM} \end{pmatrix}$$

- モデルの同時確率は以下

$$\begin{aligned} & p\left(Z^{(1)}, \ldots, Z^{(T)}, X^{(1)}, \ldots, X^{(T)}\right) \\ & = p\left(Z^{(1)}\right) \left[\prod_{\tau=2}^{T} p\left(Z^{(\tau)} | Z^{(\tau-1)}\right)\right] \prod_{\tau=1}^{T} p\left(X^{(\tau)} | Z^{(\tau)}\right) \end{aligned} \qquad (1.1)$$

夕飯問題

- 結婚しても共働きのYさんは働いた後に夕飯の用意をせねばならず，大変!
- 夫のWさんは，Yさんのその日の疲労感によって，夕飯への手間の掛けようが違うと感じていた
- WさんはYさんの仕事後の疲労感を「q_1 :とても疲れている」，「q_2:やや疲れている」，「q_3:あまり疲れていない」で評価
- Yさんが用意する夕飯はおおまかに「o_1:出来合い物」，「o_2:鍋」，「o_3:炒め物」，「o_4:揚げ物」，「o_5:カレー・シチュー」「o_6:ハンバーグ」のどれか
- 30日間の夕食のデータから，以下の**RQ**を考える

表1.1:Yさんの仕事後の夕飯に関するデータ

日	1	2	3	4	5	6	7	8	9	10
疲労感	とても	やや	それほど	それほど	とても	とても	やや	それほど	それほど	とても
夕飯	出来合い	炒め	揚げ	ハンバーグ	鍋	出来合い	鍋	カレー	揚げ	出来合い
日	11	12	13	14	15	16	17	18	19	20
疲労感	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
夕飯	ハンバーグ	鍋	出来合い	出来合い	カレー	揚げ	炒め	鍋	ハンバーグ	鍋
日	21	22	23	24	25	26	27	28	29	30
疲労感	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
夕飯	出来合い	鍋	ハンバーグ	炒め	カレー	出来合い	炒め	揚げ	鍋	ハンバーグ

注:最初の10日間に関しては，Wさんが Yさんにその日の疲労感を何気なく聞いています。

- **RQ.1**　1ヶ月間のデータから推測されるYさんが「q_1:とても疲れている」ときに「o_1:出来合い物」を用意する確率はどれくらいでしょう?
- **RQ.2**　20日目はYさんが会社でプレゼンを行った日でした。この日，Yさんの疲労感はどの状態であったでしょうか?

1.2 教師あり学習モデル

- 全ての出力系列に関して，状態遷移系列が明らかである場合
- 単にマルコフモデルという
- 確率モデルはカテゴリカル分布を用いて記述される

$$
\begin{aligned}
z^{(\tau)} &\sim \mathrm{Categorical}(\boldsymbol{a}_i) && \tau = 2, \ldots, T \\
x^{(\tau)} &\sim \mathrm{Categorical}(\boldsymbol{b}_j) && \tau = 1, \ldots, T
\end{aligned}
\tag{1.2}
$$

- $\boldsymbol{a}_i$は遷移核Aにおいて$Z^{(\tau-1)} = q_i$であるときの遷移確率ベクトル，$\boldsymbol{b}_j$は出力確率行列Bにおいて，$Z^{(\tau)} = q_j$であるときの出力確率ベクトル
- 確率ベクトルの事前分布にはディリクレ分布を仮定

$$
\begin{aligned}
\boldsymbol{a}_i &\sim \mathrm{Dirichlet}(1, 1, 1) \\
\boldsymbol{b}_j &\sim \mathrm{Dirichlet}(0.1, 0.1, 0.1, 0.1, 0.1, 0.1)
\end{aligned}
\tag{1.3}
$$

1.3　前向きアルゴリズム

- 状態系列が未知である出力データから，遷移核と出力確率行列を推定する
- このために，(1.1)式を潜在変数$Z^{(\tau)}$に関して周辺化する方法として前向きアルゴリズム(forward algorighm)がある
- 出力系列$x^{(1)}$,…,$x^{(\tau)}$を出力して，時点τで状態q_jに到達する確率として

$$\alpha_j^{(\tau)} = p(x^{(1)}, \ldots, x^{(\tau)}, Z^{(\tau)} = q_j) \tag{1.4}$$

を定義し，これを前向き確率とよぶ

前向きアルゴリズム

1. $\alpha_i^{(1)}$については初期確率π_iを用いて，以下のように初期化する

$$\alpha_i^{(1)} = \pi_i b_{i,x^{(1)}} \tag{1.5}$$

2. 前向き確率$\alpha_j^{(\tau)}$は以下の式で求められる

$$\alpha_j^{(\tau)} = \left[\sum_{i=1}^{K} \alpha_i^{(\tau-1)} a_{ij}\right] b_{j,x^{(\tau)}} \tag{1.6}$$

3. 状態遷移系列が観測できない出力系列についての同時確率は以下のように求められる

$$p(x^{(1)}, \ldots, x^{(T)}) = \sum_{j=1}^{K} p(x^{(1)}, \ldots, x^{(T)}, Z^{(T)} = q_j) = \sum_{j=1}^{K} \alpha_j^{(T)} \tag{1.7}$$

1.4　半教師あり学習モデル

- 隠れマルコフモデルは完全な教師なしモデルとして母数の推定を行うことも可能
- 事後分布が多峰となることが多く，推定が不安定になる
- 状態遷移系列が明らかであるデータ(1日～10日)に関して(1.2)式を仮定し，残りのデータについて，前向きアルゴリズムを利用

RQ.1への回答

表1.2:出力確率のEAP

	出来合い	鍋	炒め	揚げ	カレー	ハンバーグ
とても	0.646	0.247	0.033	0.019	0.018	0.037
やや	0.029	0.454	0.415	0.016	0.052	0.034
あまり	0.021	0.028	0.018	0.335	0.221	0.377

- Yさんは「とても疲れている」ときに「出来合い物」を用意する傾向がある

表1.3:遷移確率のEAP

	とても	やや	あまり
とても	0.346	0.197	0.416
やや	0.481	0.152	0.292
あまり	0.173	0.651	0.292

- WさんはYさんが「やや疲れている」翌日は，代わりに夕飯の準備をすれば夫婦円満!

1.5 ビタビ・アルゴリズム

- 与えられた観測系列に対し，潜在変数の状態の最も確からしい系列を求める方法
- 出力系列$x^{(1)},\ldots,x^{(\tau)}$を生成して，時点τで状態$q_j$に到達する状態遷移系列は複数存在する
- このうち，最大の確率を与えるものだけを記憶すればよい
- 時点τで状態q_jに到達する状態系列に関して，最大の確率値を$\zeta_j^{(\tau)}$で表す

$$\zeta_j^{(\tau)} = \max_{Z^{(1)},\ldots,Z^{(\tau-1)}} p\left(Z^{(1)},\ldots,Z^{(\tau-1)},Z^{(\tau)}=q_j,x^{(1)},\ldots,x^{(\tau)}\right) \quad (1.8)$$

ビタビ・アルゴリズム

1. $\zeta_j^{(\tau)}$は以下のように計算する

$$\zeta_j^{(\tau)} = \max_i [\zeta_i^{(\tau-1)} a_{ij}] b_{j,x^{(\tau)}} \quad (1.9)$$

2. 状態系列を復元するために，最大の確率値$\zeta_j^{(\tau)}$を与える直前の状態q_iも同時に記憶

$$\psi_j^{(\tau)} = \operatorname*{argmax}_i [\zeta_i^{(\tau-1)} a_{ij}] \quad (1.10)$$

3. 各状態q_iについて，$\zeta_i^{(1)}$と$\psi_i^{(1)}$は以下のように初期化

$$\zeta_i^{(1)} = \pi_i b_{i,x^{(1)}}, \qquad \psi_i^{(1)} = 0 \quad (1.16)$$

4. 最後に，再帰計算を終了し，最適状態系列を復元する

$$\hat{p}(Z^{(1)}, \ldots, Z^{(T)}, x^{(1)}, \ldots, x^{(T)}) = \max_j \zeta_j^{(T)}$$

$$\hat{Z}^{(T)} = \operatorname*{argmax}_j \zeta_j^{(T)}, \qquad \hat{Z}^{(\tau)} = \psi_{\hat{Z}^{(\tau+1)}}^{(\tau+1)} \quad (1.11)$$

ベイズならでは

- 生成量を定義し，3つの研究仮説「20日目のYさんの疲労感は「q_1：とても疲れている」であった，「q_2：やや疲れている」であった，「q_3：あまり疲れていない」であった」が成り立つ確率を求める

$$u^{(t)}_{\hat{z}^{(20)}=q_1} = \begin{cases} 1 & \hat{z}^{(20),(t)} = q_1 \\ 0 & \text{それ以外の場合} \end{cases} \quad u^{(t)}_{\hat{z}^{(20)}=q_2} = \begin{cases} 1 & \hat{z}^{(20),(t)} = q_2 \\ 0 & \text{それ以外の場合} \end{cases}$$

$$u^{(t)}_{\hat{z}^{(20)}=q_3} = \begin{cases} 1 & \hat{z}^{(20),(t)} = q_3 \\ 0 & \text{それ以外の場合} \end{cases} \qquad (1.12)$$

RQ.2への回答

表1.3:研究仮説が正しい確率

	確率
$\hat{p}(z^{(20)} =$「とても疲れている」$)$	0.516
$\hat{p}(z^{(20)} =$「やや疲れている」$)$	0.437
$\hat{p}(z^{(20)} =$「あまり疲れてない」$)$	0.047

- 半教師あり学習モデルにビタビ・アルゴリズムを組み込んだモデルで推定
- プレゼン後のYさんは少なくとも疲労状態にあった

アイオワ・ギャンブリング課題

2. アイオワ・ギャンブリング課題(IGT)

- ギャンブルにおいて，人々は利益の最大化を目的とする
- ある程度経験を積むことで，利益の最大化に対する方略が決定し，探索行動は少なくなる
- このような意思決定のプロセスを評価する心理実験の一つがアイオワ・ギャンブリング課題(Iowa Gambling Tasks, IGT)

2.1 IGTとは

- 実験参加者は4つのカードのデッキ(A～D)から，カードを1枚引く試行を繰り返し行う
- カードの裏には報酬額と損失額が記載されており，それらの額と出現頻度は違う
- 参加者は純利益を最大化するよう教示される

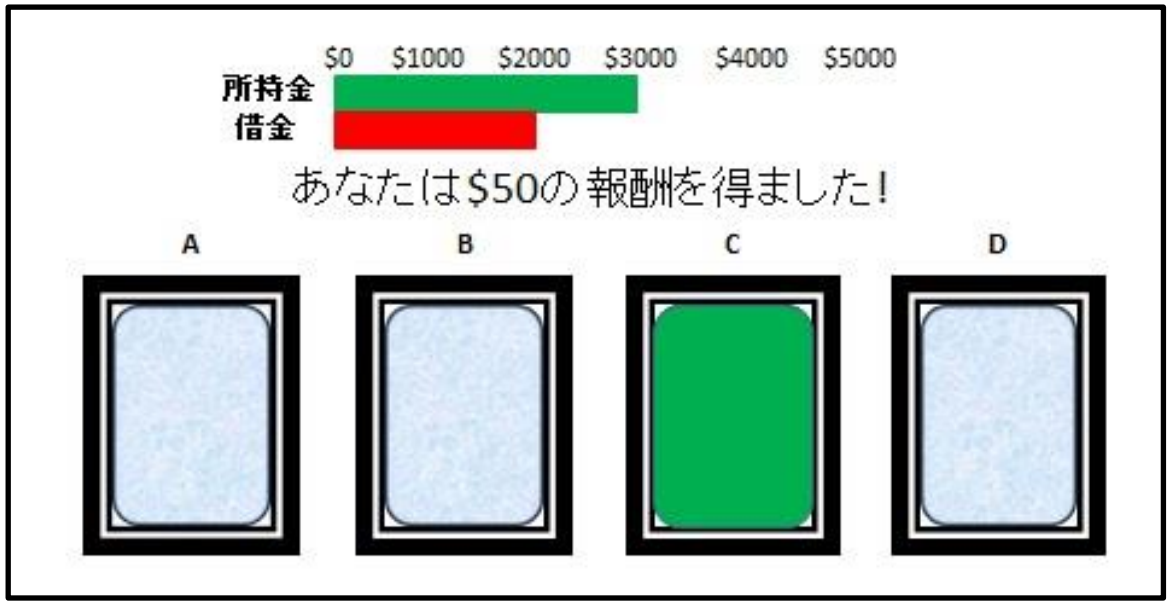


図2.1:実験イメージ

表2.1:デッキの性質

	A	B	C	D
試行 1 回あたりの報酬額 (単位：ドル)	100	100	50	50
試行 10 回あたりの損失回数	5	1	5	1
試行 10 回あたりの損失額 (単位：ドル)	1250	1250	250	250
試行 10 回あたりの純利益 (単位：ドル)	−250	−250	250	250

AとBのデッキを避け，CとDのデッキを選択することが望まれる！

- 健常群と比較して，アスペルガー症候群やコカイン中毒者といった臨床群ではこの方略を獲得する能力が低い(Wetzels et al. (2010))
- IGTは様々な臨床群における意思決定能力の欠如の度合いを評価するため用いられる

2.2 期待数価モデル

- IGTにおける意思決定を評価するモデル
- 過去の試行に対する強化学習のプロセスを表現
- Busemeyer & Stout(2002)を端緒に発展
- 3つの心理学的なステップを通してデッキの選択が行われていると仮定

期待数価モデルの他に,
効用関数を用いたモデルも
あります。

ステップ①

- 実験参加者は試行tで得られた得失の数価$v^{(t)}$を評価
- $v^{(t)}$は試行tにおいて経験した報酬$R^{(t)}$と損失$L^{(t)}$の重み付き関数として表現される

$$v^{(t)} = (1-w) \times R^{(t)} + w \times L^{(t)} \qquad (2.1)$$

- 母数$w(0,1)$：損失$L^{(t)}$に対する注意の度合い
- wの値が高い参加者は損失を軽視する傾向

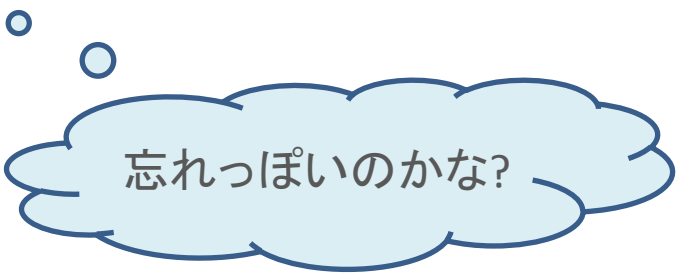
ギャンブラーはこれに該当！

ステップ②

- 得失の数価$v^{(t)}$から，次の試行$t+1$でデッキkから得られる期待数価$E_{v_k}^{(t+1)}$を評価

$$E_{v_k}^{(t+1)} = \begin{cases} (1-a) \times E_{v_k}^{(t)} + a \times v^{(t)} & \text{試行 } t \text{ でデッキ } k \text{ が選ばれた場合} \\ E_{v_k}^{(t)} & \text{それ以外の場合} \end{cases} \quad (2.2)$$

- 母数$a(0,1)$：1つ前の試行で得られた数価に対する注意の度合い(更新比率)
- aの値が高い参加者はそれまでに得られた経験の蓄積を軽視する傾向



ステップ③

- 試行tにおけるデッキkの選択を確率変数$Y^{(t)}$とおき，その選択確率をsoftmax関数で表現

$$p_k^{(t)} = p(Y^{(t)} = k) = \frac{\exp(\theta^{(t)} \times E_{v_k}^{(t)})}{\sum_{j=1}^{4} \exp(\theta^{(t)} \times E_{v_j}^{(t)})} \qquad (2.3)$$

- $\theta^{(t)}(0, \infty)$：期待数価を重視する度合い，0に近づくほどデッキの選択はランダムになる

ステップ③

- $\theta^{(t)}$には以下の変換を採用

$$\theta^{(t)} = (t/10)^c \tag{2.4}$$

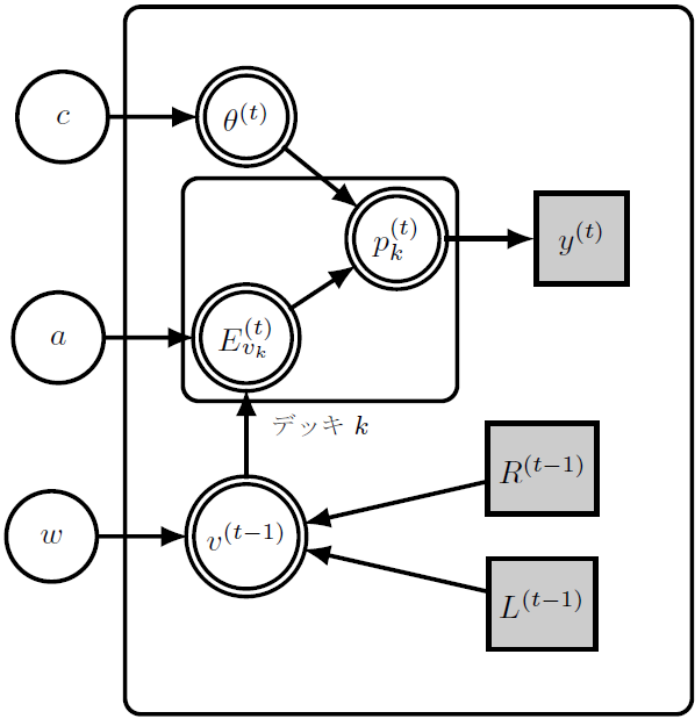
- 母数$c(-\infty, \infty)$：参加者の選択に対する一貫性
- $c > 0$のとき，初期の試行において探索行動が多く，試行を重ねるごとに期待数価を重視
- 実際に観察される試行tにおけるデッキの選択$y^{(t)}$はカテゴリカル分布を用いてモデル化

$$y^{(t)} \sim \mathrm{Categorical}\left(\boldsymbol{p}^{(t)}\right) \tag{2.5}$$

2.3 個人データの分析

表2.2：個人データ

試行	1	2	3	4	5	⋯	146	147	148	149	150
選択デッキ	A	B	C	D	A	⋯	D	D	D	D	D
報酬 (ドル)	100	100	50	50	100	⋯	50	50	50	50	50
損失 (ドル)	0	0	0	0	0	⋯	0	0	0	0	0



$$y^{(t)} \sim \text{Categorical}(\boldsymbol{p}^{(t)}) \quad p_k^{(t)} = \frac{\exp\left(\theta^{(t)} \times E_{v_k}^{(t)}\right)}{\sum_{j=1}^{4} \exp\left(\theta^{(t)} \times E_{v_j}^{(t)}\right)}$$

$$E_{v_k}^{(t)} = \begin{cases} \text{試行 } t-1 \text{ でデッキ } k \text{ が選ばれた場合} \\ (1-a) \times E_{v_k}^{(t-1)} + a \times v^{(t-1)} \\ \text{それ以外の場合} \\ E_{v_k}^{(t-1)} \end{cases}$$

$$v^{(t-1)} = (1-w) \times R^{(t-1)} + w \times L^{(t-1)}$$

$$\theta^{(t)} = (t/10)^c \qquad w \sim \text{Uniform}(0,1)$$

$$a \sim \text{Uniform}(0,1) \quad c \sim \text{Uniform}(-5,5)$$

図2.2：個人データのプレート表現

分析結果

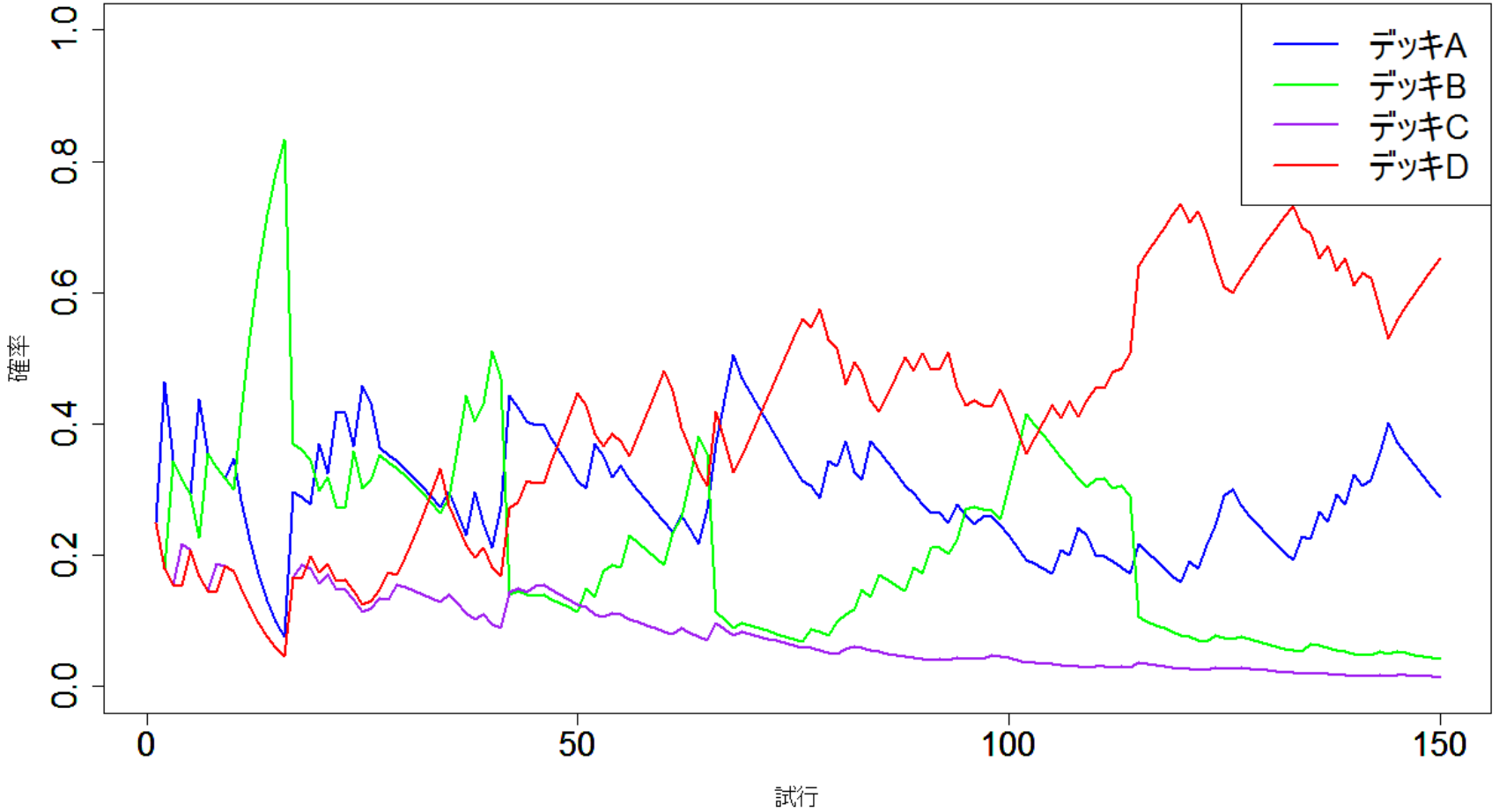
表2.3:母数のEAP

	EAP	post.sd	2.5 %	97.5 %
w：損失への注意	0.321	0.014	0.294	0.349
a：更新比率	0.008	0.003	0.003	0.014
c：一貫性	−0.291	0.152	−0.541	0.053

・損失に対して寛容

・試行を重ねるごとに期待数価を重視する傾向はない

図2.3：デッキの選択確率の推移



- 初期の試行において，デッキBを選択
- 徐々にデッキDに移行

2.4 階層モデルの適用

- 複数参加者の分析には階層ベイズモデルが適用される
- 階層モデルでは実験参加者を表す添え字iを用いて，w_i, a_i, c_iと表現し，各母数がさらにそれぞれ母数が未知である事前分布から発生するモデルを考える

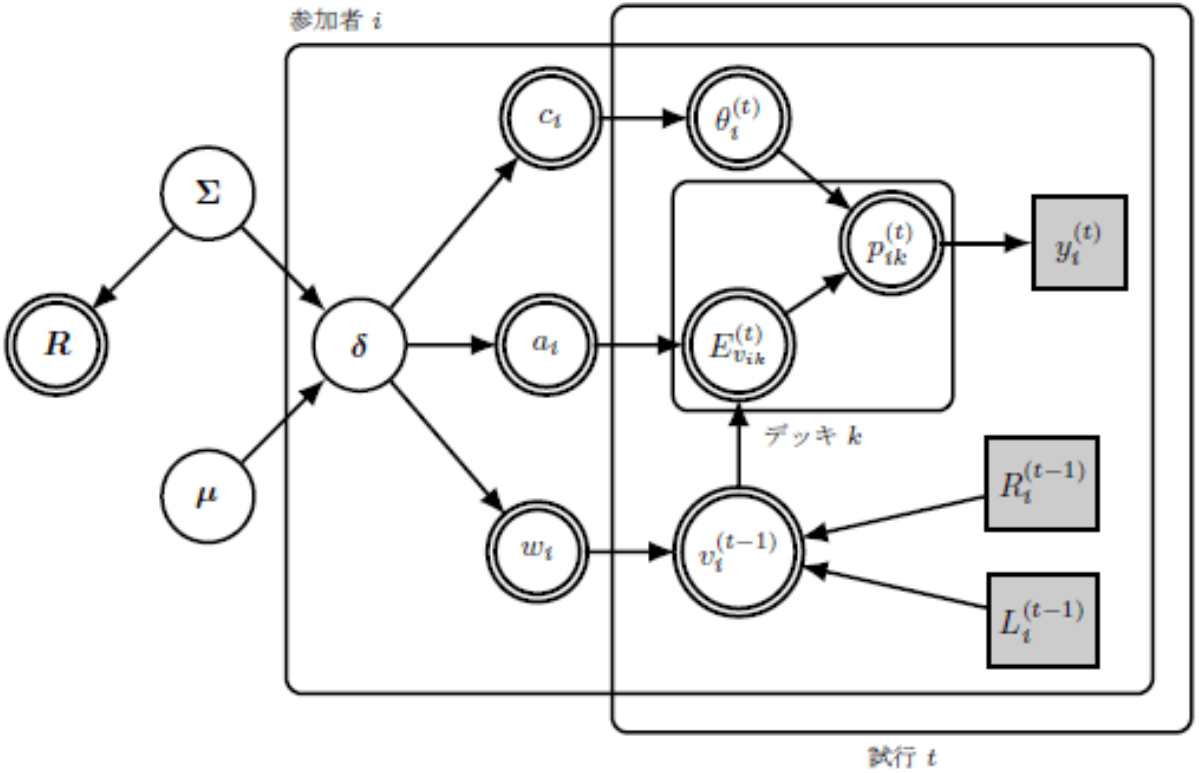


図2.4:階層モデルのプレート表現

- 心理学的な特性値w_i, a_i, c_iは参加者間で正規分布していると仮定
- これらの特性値の定義域は異なるため，まず，未知の平均ベクトルと共分散行列を持つ3変量正規分布から$\boldsymbol{\delta}$を発生させる

$$\boldsymbol{\delta} \sim \text{Multinormal}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \qquad (2.6)$$

- w_iとa_iに関しては$\boldsymbol{\delta}$の要素であるδ_1, δ_2を正規累積変換し，c_iはδ_3の値を用い，前述の定義域を踏襲

$$w_i = \Phi(\delta_1 | \mu_1, \sigma_1) \qquad a_i = \Phi(\delta_2 | \mu_2, \sigma_2)$$
$$c_i = \delta_3 \tag{2.7}$$

- 平均ベクトル$\boldsymbol{\mu}$の各要素と共分散行列$\boldsymbol{\Sigma}$にはそれぞれ正規分布と逆ウィシャート分布を事前分布として仮定

$$\mu_1, \mu_2, \mu_3 \sim \mathrm{Normal}(0, 1) \quad \boldsymbol{\Sigma} \sim \mathrm{Wishart}^{-1}(3, \boldsymbol{I}_3) \tag{2.8}$$

- 共分散行列$\boldsymbol{\Sigma}$からw_i, a_i, c_iに関する相関行列$\boldsymbol{R}$を生成し，母数の相関関係を検証

$$\boldsymbol{R} = \mathrm{diag}(\boldsymbol{\Sigma})^{-1/2} \times \boldsymbol{\Sigma} \times \mathrm{diag}(\boldsymbol{\Sigma})^{-1/2} \tag{2.9}$$

分析結果

表2.4:母数のEAP

実験参加者	w：損失への注意	a：更新比率	c：一貫性
1	0.554	0.009	0.532
2	0.324	0.006	−0.340
3	0.448	0.008	0.010
4	0.320	0.008	−0.281
5	0.394	0.002	−0.010
6	0.280	0.007	−0.300
7	0.144	0.001	−0.752

・「損失への注意」と「更新比率」に関して，1番と7番は対照的

・更新比率は皆優秀!

表2.5:相関行列$\boldsymbol{R}$のEAP

	w	a	c
w：損失への注意	1.000	0.186	0.488
a：更新比率	0.186	1.000	0.099
c 一貫性	0.488	0.099	1.000

中程度の正の相関

⇒前述の分析の結果を裏付ける

参考文献

- C.M.ビショップ(2012). パターン認識と機械学習 下，丸善出版株式会社.
- 石井健一郎・上田修功 (2014). 続・わかりやすいパターン認識ー教師なし学習入門ー， オーム社.
- 北 研二(1999). 言語と計算 4 確率的言語モデル，東京大学出版会.
- Bechara, A., Damasio, A. R., Damasio H., and Anderson, S (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, **50**, 7-15.
- Busemeyer, J. R., and Stout, J. C. (2002). A contribution of cognitive decision models to clinical
- assessment: Decomposing performance on the Bechara gambling task. *Psychological Assessment*, **14**, 253-262.
- Wetzels, R., Vandekerckhove, J., Tuerlinckx, F. and Wagenmakers, E. J. (2010) . Bayesian parameter estimation in the Expectancy Valance model of the Iowa gambling task. *Journal of Mathmatical Psychology*, **54**(1), 14-27.

付録
Stanコード

★隠れマルコフモデル
(半教師あり学習モデル+ビタビ・アルゴリズム)

```
data {
  int<lower=1> K;
  int<lower=1> V;
  int<lower=0> T;
  int<lower=1> T_unsup;
  int<lower=1,upper=V> x[T];
  int<lower=1,upper=K> z[T];
  int<lower=1,upper=V> x_u[T_unsup];
  vector<lower=0>[K] prior_a;
  vector<lower=0>[V] prior_b;
}
parameters {
  simplex[K] a[K];
  simplex[V] b[K];
}
model {
  for (i in 1:K)
    a[i] ~ dirichlet(prior_a);
  for (j in 1:K)
    b[j] ~ dirichlet(prior_b);
  for (t in 1:T)
    x[t] ~ categorical(b[z[t]]);
  for (t in 2:T)
    z[t] ~ categorical(a[z[t-1]]);
```

```
{
  real acc[K];
  real alpha[T_unsup,K];
  for (i in 1:K)
    alpha[1,i] = log(b[i,x_u[1]]);
  for (t in 2:T_unsup) {
    for (i in 1:K) {
          acc[i] = alpha[t-1,i] + log(a[i,j]) + log(b[j,x_u[t]]);
          alpha[t,j] = log_sum_exp(acc);
      }
    }
    target += log_sum_exp(alpha[T_unsup]);
  }
}

generated quantities {
  int<lower=1,upper=K> z_hat[T_unsup];
  real zeta;
  real best_logp[T_unsup,K];
  int  psi[T_unsup,K];
  real u1;
  real u2;
  real u3;
```

```
    for (i in 1:K)
      best_logp[1,K] = log(b[i,x_u[1]]);
    for (t in 2:T_unsup) {
      for (j in 1:K) {
        best_logp[t,j] = negative_infinity();
        for (i in 1:K) {
          real logp;
          logp = best_logp[t-1,i] + log(a[i,j]) + log(b[j,x_u[t]]);
          if (logp > best_logp[t,j]) {
            psi[t,j] = i;
            best_logp[t,j] = logp;
          }
        }
      }
    }
    zeta = max(best_logp[T_unsup]);
    for (j in 1:K){
      if (best_logp[T_unsup,j] == zeta)
        z_hat[T_unsup] = j;
    }
    for (t in 1:(T_unsup - 1)){
      z_hat[T_unsup - t] = psi[T_unsup - t + 1,
                               z_hat[T_unsup - t + 1]];
    }
    u1 = z_hat[10]==1 ? 1 : 0;
    u2 = z_hat[10]==2 ? 1 : 0;
    u3 = z_hat[10]==3 ? 1 : 0;

}
```

★IGT期待数価モデル
(階層モデル)

```
data{
    int<lower=1> T;
    int<lower=1> K;
    int<lower=1> N;
    matrix[3,3] R;
    int<lower=1,upper=K> y[T,N];
    int<lower=0,upper=100> W[T,N];
    int<lower=-1250,upper=0> L[T,N];
}
parameters{
    vector[3] Eta[N];
    vector[3] Mu;
    cov_matrix[3] Sigma;
}
transformed parameters{
    real<lower=0> theta[T,N];
    real v[T,N];
    vector[K] Ev[T,N];
    simplex[K] p[T,N];
    real<lower=0,upper=1> w[N];
    real<lower=0,upper=1> a[N];
    real c[N];
```

```
for(n in 1:N){
    w[n]=normal_cdf(Eta[n,1],Mu[1],sqrt(Sigma[1,1]));
    a[n]=normal_cdf(Eta[n,2],Mu[2],sqrt(Sigma[2,2]));
    c[n]=Eta[n,3];
}
for(t in 1:T){
   for(n in 1:N){
     theta[t,n]=pow((0.1*t),c[n]);
     v[t,n]=(1-w[n])*W[t,n]+w[n]*L[t,n];
   }
}
for(n in 1:N){
  for(k in 1:K){
    Ev[1,n,k]=0;
    p[1,n,k]=0.25;
  }
}
for(t in 2:T){
   for(n in 1:N){
     for(k in 1:K){
      if(y[t-1,n]==k)
        Ev[t,n,k]=Ev[t-1,n,k]+a[n]*( v[t-1,n]-Ev[t-1,n,k] );
```

```
          else
            Ev[t,n,k]=Ev[t-1,n,k];
          }
        }
    }
    for(t in 2:T){
       for(n in 1:N){
          p[t,n]=softmax(theta[t,n]*Ev[t,n]);
       }
    }
}
model{
    for(i in 1:3){
        Mu[i]~normal(0,1);
    }
    Sigma~inv_wishart(3,R);
    Eta~multi_normal(Mu,Sigma);

    for(n in 1:N){
       for(t in 1:T){
         y[t,n]~categorical(p[t,n]);

       }
    }
}
```

```
generated quantities{
    matrix[3,3] rho;

    for(i in 1:3){
      for(j in 1:3){
        rho[i,j]=Sigma[i,j]/sqrt(Sigma[i,i]*Sigma[j,j]);
      }
    }
}
```

階層ベイズ法による二項分布モデルとベータ二項分布モデル

早稲田大学大学院　文学研究科　吉上諒

発表の流れ

・心理学領域における有用性
・母比率の異質性について
・異質性を考慮した二項分布モデル
・待機児童データへの適応
・ベータ二項分布モデル
・ネズミの胎児死亡データへの適応

心理学領域における有用性

・観測対象全体に対する特定のカテゴリに属する観測対象の比率

（例）日本人全体に対するうつ病患者の比率

＋

・観測対象の属性

（例）日本人全体に対する職業別うつ病患者の比率

観測対象の属性を考慮したモデリングが可能に。

母比率の異質性

図１　フリースローのゴール数のデータ　※人工データ

	1	2	3	4	5	...	96	97	98	99	100
ゴール成功回数	1	0	2	9	9	...	5	8	3	0	5

人数：100名（i=100）　フリースロー回数：10回（n=10）

・二項分布を仮定した際の成功回数の期待値と分散

$$E\left[X\right] = n\theta, \quad V\left[X\right] = n\theta(1-\theta)$$

・データから計算される期待値と分散の予測値

$$\hat{\theta} = \frac{1}{1000}\sum_{i=1}^{N} x_i \qquad V\left[X\right] = 10\hat{\theta}(1-\hat{\theta})$$

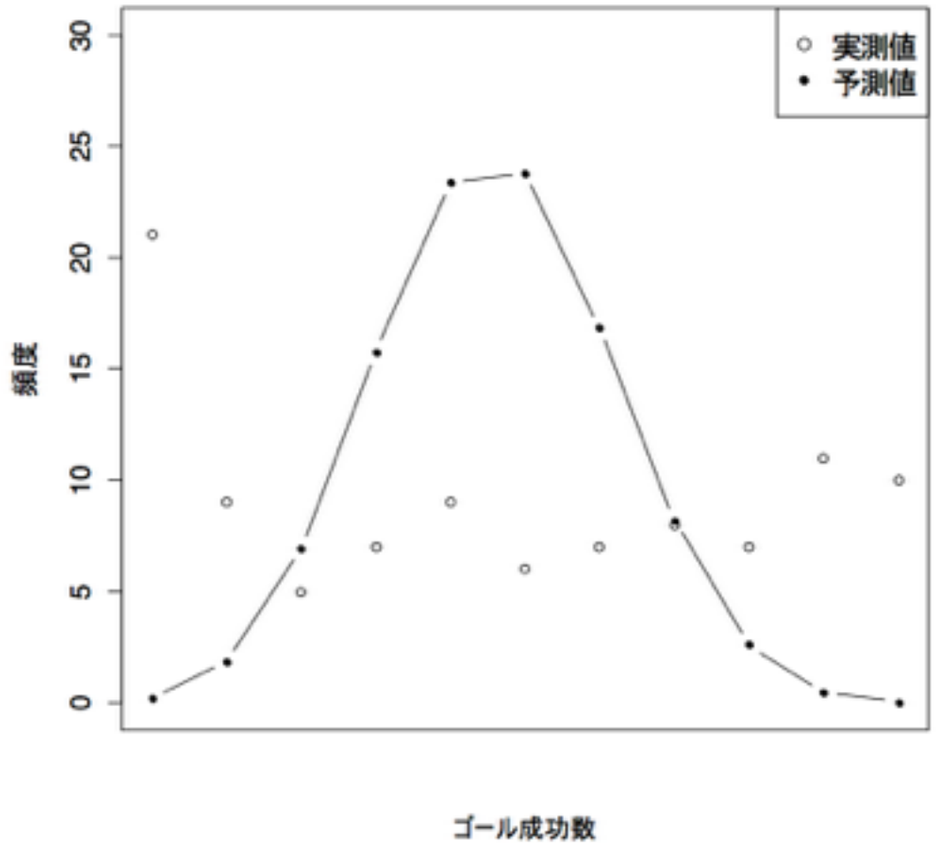
$$\hat{\theta} \simeq 0.46 \qquad = 2.48$$

母比率の異質性

分散の予測値　　約2.48
実際の標本分散　約12.75

・二項分布を仮定した場合の分散の予測値と比べて、実際のデータから求められる標本分散は約5.14倍も大きい。これを過分散という。

図２　予測値と実測値の比較



成功確率（母比率）が観測対象ごとに変わらないと仮定する二項分布モデルではデータを適切に表現できない（図２）

異質性を考慮した二項分布モデル

階層ベイズモデルを利用し、二項分布の母比率に
ベータ分布を仮定したモデル

ベータ分布の確率密度関数

$$f(x \mid \alpha_0, \beta_0) = B(\alpha_0, \beta_0)^{-1} x^{\alpha_0 - 1}(1-x)^{\beta_0 - 1}$$

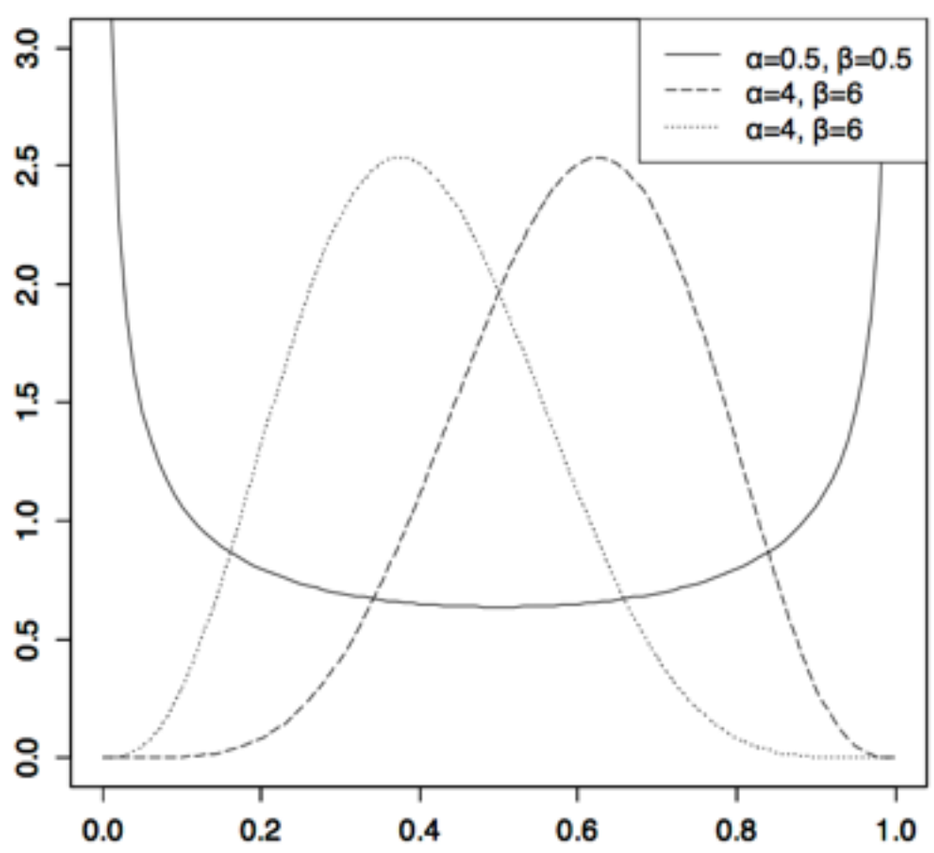
$$B(\alpha_0, \beta_0) = \int_0^1 x^{\alpha_0 - 1}(1-x)^{\beta_0 - 1} dx$$

ベータ分布に従う確率変数の期待値と分散

$$E[\Theta] = \frac{\alpha}{\alpha + \beta} = \mu \quad 0 \leq \mu \leq 1$$

$$V[\Theta] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

図３　母数の変化させた場合の
ベータ分布の密度関数



異質性を考慮した二項分布モデル

サンプリングの安定性を考慮し、Gelman, Carlin, Stern& Rubin(2015,110-111) ベータ分布の母数の再母数化を行い、κ に母数(0.1 1.5)のパレート分布を仮定する。

$$\alpha = \left(\frac{\alpha}{\alpha+\beta}\right)(\alpha+\beta) = \mu\kappa$$

$$\beta = (\alpha+\beta) - \left(\frac{\alpha}{\alpha+\beta}\right) = \kappa - \mu\kappa = \kappa(1-\mu)$$

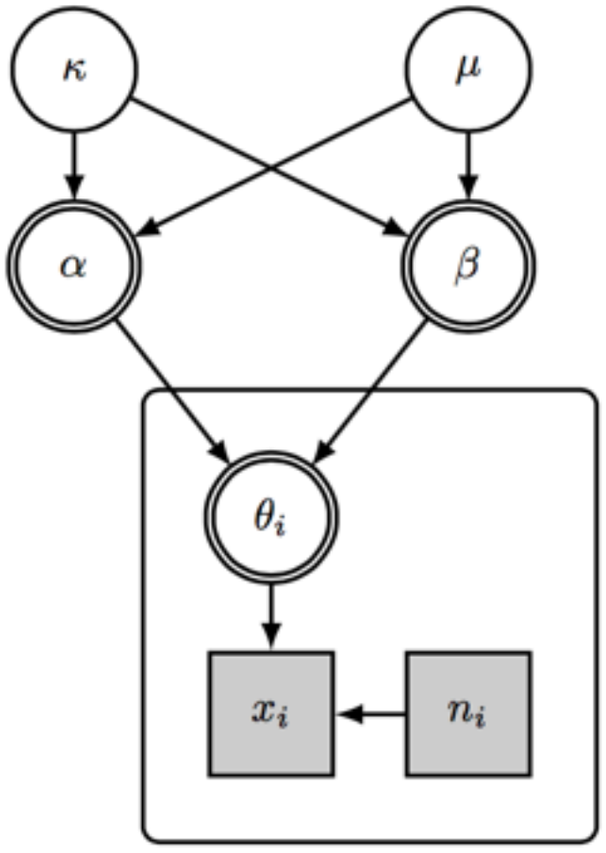
$$\kappa = \alpha + \beta$$

待機児童データへの適応

図４　各都道府県の待機児童データ

都道府県名	北海道	青森	岩手	...	宮崎	鹿児島	沖縄
全児童数	40383	26285	21664	...	18571	24963	30959
待機児童数	64	0	139	...	0	185	1721

図５　プレート図　　都道府県全体の平均待機児童率と各都道府県の待機児童率を求め、双方の大きさを比較する。



i 都道府県

$$\kappa \sim \text{Pareto}(0.1, 1.5)$$
$$\alpha = \mu\kappa, \quad \beta = \kappa(1-\mu)$$
$$\mu \sim \text{Uniform}(0, 1)$$
$$\theta_i \sim \text{Beta}(\alpha, \beta)$$
$$x_i \sim \text{Binomial}(\theta_i, n_i)$$

待機児童データへの適応

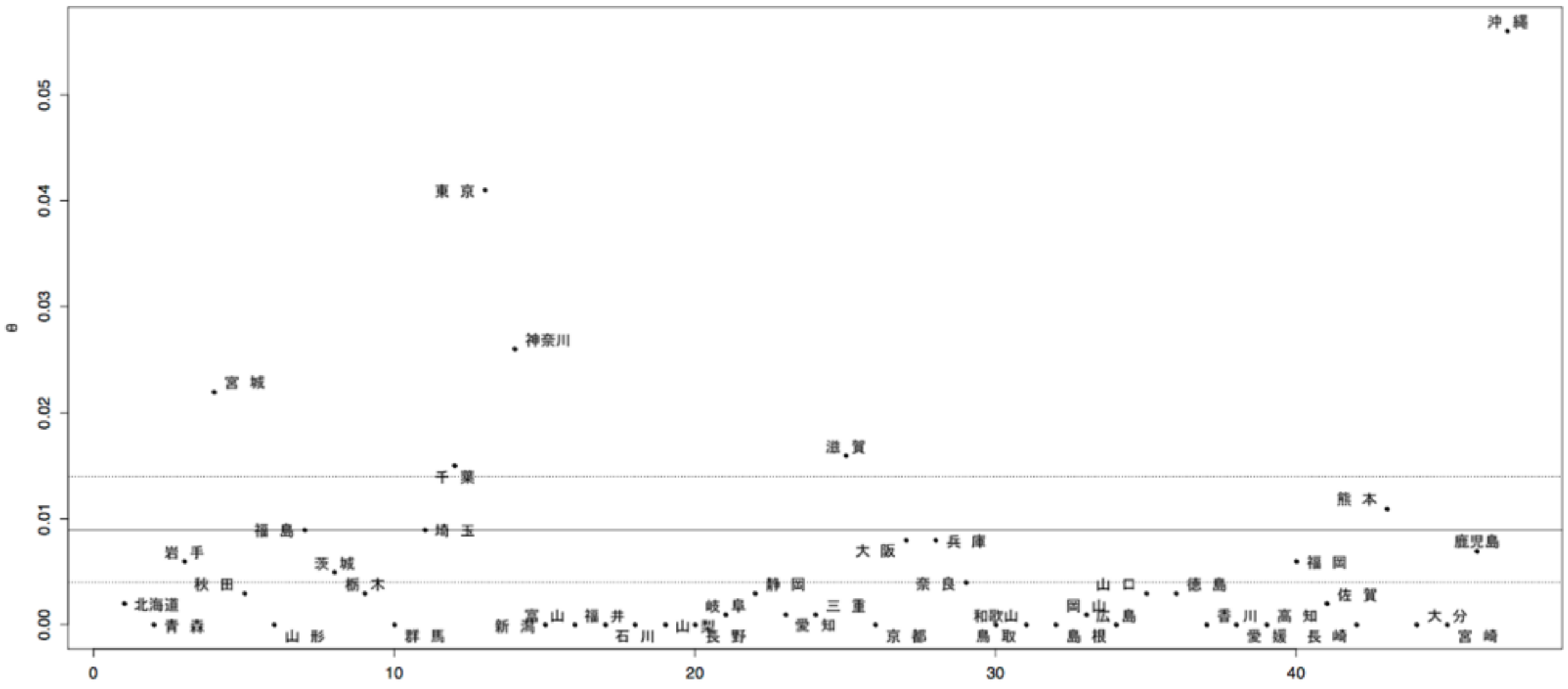
図６　分析結果

	EAP	post.sd	2.5%	50%	97.5%
μ	0.009	0.004	0.004	0.008	0.021
σ	0.001	0.001	0.000	0.000	0.002
α	0.197	0.039	0.129	0.194	0.282
β	25.135	11.506	7.536	23.563	51.859

北海道	0.002	東京	0.041	滋賀	0.016	香川	0.000
青森	0.000	神奈川	0.026	京都	0.000	愛媛	0.000
岩手	0.006	新潟	0.000	大阪	0.008	高知	0.000
宮城	0.022	富山	0.000	兵庫	0.008	福岡	0.006
秋田	0.003	石川	0.000	奈良	0.004	佐賀	0.002
山形	0.000	福井	0.000	和歌山	0.000	長崎	0.000
福島	0.009	山梨	0.000	鳥取	0.000	熊本	0.011
茨城	0.005	長野	0.000	島根	0.000	大分	0.000
栃木	0.003	岐阜	0.001	岡山	0.001	宮崎	0.000
群馬	0.000	静岡	0.003	広島	0.000	鹿児島	0.007
埼玉	0.009	愛知	0.001	山口	0.003	沖縄	0.056
千葉	0.015	三重	0.001	徳島	0.003		

待機児童データへの適応

図７　都道府県全体の平均と各都道府県の待機児童率の比較



都道府県ごとの待機児童率の異質性を考慮できている。

ベータ二項分布モデル

ベータ二項分布を利用し、二項分布の母比率の異質性を考慮したモデル

ベータ二項分布の確率密度関数

$$f(x_i \mid n, \boldsymbol{\theta}, \alpha_0, \beta_0) = f(x_i \mid n, \theta_i) \times p(\theta_i \mid \alpha_0, \beta_0)$$

$$f(x_i \mid n, \alpha_0, \beta_0) = \binom{n}{x_i} \frac{B(x_i + \alpha_0, n - x_i + \beta_0)}{B(\alpha_0, \beta_0)}$$

先のモデルと異なり、ベータ二項分布においては
二項分布の母比率を母数としない。

ネズミの胎児死亡データへの適応

図８　ネズミの胎児死亡データ

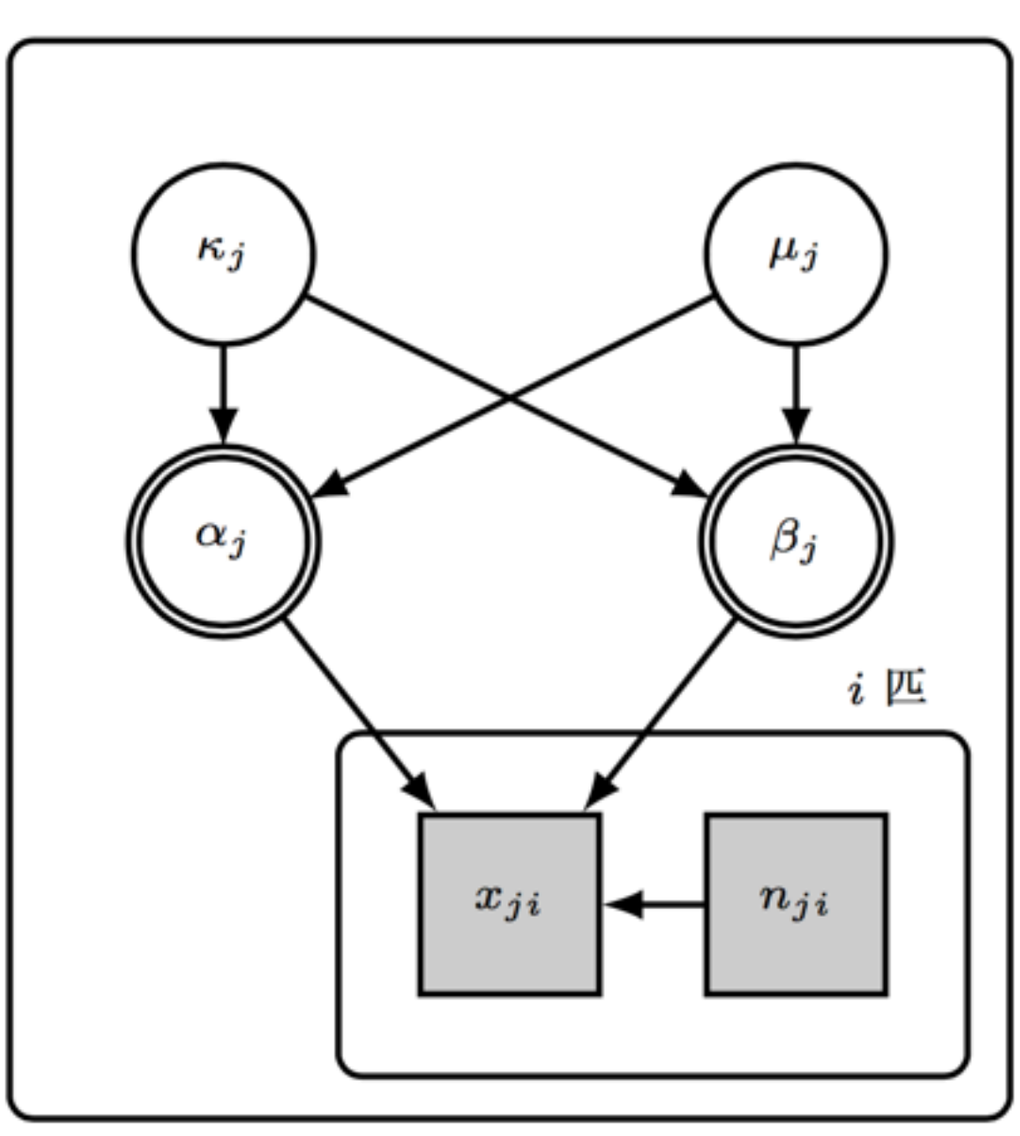
第 1 群	(4.1,10,1)	(3.2,11,4)	(4.7,12,9)	(3.5,4,4)	(3.2,10,10)
			⋮		
	(4.4,13,13)	(5.2,4,3)	(3.9,8,8)	(7.7,13,5)	(5.0,12,12)
第 2 群	(11.2,8,0)	(11.5,11,1)	(12.6,14,0)	(9.5,14,1)	(9.8,11,0)
第 3 群	(16.6,3,0)	(14.5,13,0)	(15.4,9,2)	(14.5,17,2)	(14.6,15,0)
	(16.5,2,0)	(14.8,14,1)	(13.6,8,0)	(14.5,6,0)	(12.4,17,0)

鉄欠乏のネズミに関して、（ヘモグロビンレベル、妊娠数、胎児死亡数）をまとめたデータ。ヘモグロビンレベルは個体によって異なる。
またネズミは以下の３群に分けられ、それぞれ３週間投薬を受けている。

・第１群　何も投薬をしない（31匹）
・第２群　初日と７日目に投薬を受ける（5匹）
・第３群　毎週投薬を受ける（10匹）

ネズミの胎児死亡データへの適応

図９　プレート図



$$\alpha_j = \mu_j \kappa_j$$
$$\beta_j = \kappa_j (1 - \mu_j)$$
$$\mu_j \sim \text{Uniform}(0, 1)$$
$$\kappa_j \sim \text{Pareto}(0.1, 1.5)$$
$$x_{ji} \sim \text{Beta\_Binomial}(\alpha_j, \beta_j, n_{ji})$$

ネズミの胎児死亡データへの適応

・各群の平均死亡率
・「第１群は第２群に比べて胎児死亡率が大きい」という仮説が正しい確率
・「第２群は第３群に比べて胎児死亡率が大きい」という仮説が正しい確率

なお仮説の正しい確率を算出するために、以下を定義する。

$$u_{\mu_1>\mu_2}^{(t)} = g(\mu_1^{(t)}, \mu_2^{(t)}) = \begin{cases} 1 & \mu_1^{(t)} - \mu_2^{(t)} > 0 \\ 0 & \text{それ以外の場合} \end{cases}$$

$$u_{\mu_2>\mu_3}^{(t)} = g(\mu_2^{(t)}, \mu_3^{(t)}) = \begin{cases} 1 & \mu_2^{(t)} - \mu_3^{(t)} > 0 \\ 0 & \text{それ以外の場合} \end{cases}$$

ネズミの胎児死亡データへの適応

図10　分析結果

	EAP	post.sd	2.5%	50%	97.5%
μ_1	0.763	0.049	0.658	0.765	0.849
μ_2	0.239	0.132	0.044	0.219	0.534
μ_3	0.165	0.090	0.038	0.150	0.378
$U_{\mu_1>\mu_2}$	1.000	0.020	1.000	1.000	1.000
$U_{\mu_2>\mu_3}$	0.671	0.470	0.000	1.000	1.000

- 各群の平均死亡率は第１群が0.763、第２群が0.239、第３群が0.165と推定された
- 「第１群は第２群に比べて胎児死亡率が大きい」という仮説は1.00の確率で正しいと言える。
- 「第２群は第３群に比べて胎児死亡率が大きい」という仮説は0.671の確率で正しいと言える。

参考文献

・Stan Development Team (2015).

Stan Modeling Language User's Guide and Reference Manual Stan Version 2.9.0.

・豊田秀樹 (2015)
基礎からのベイズ統計学-ハミルトニアンモンテカルロ法による
実践的入門-朝倉書店.

・厚生労働省 (2014)　保育所関連状況取りまとめ.

・李政元 (2012)　二項-ベータ階層ベイズモデルによる児童虐待相談対応率
の地域差に関する研究:都道府県政令指定都市別による多重比較
総合政策研究,41,29-36.

付録：stanコード①

```
data{
 int<lower=0> N;
 int<lower=0> S[N];
 int<lower=0> x[N];
}
parameters{
 real<lower=0,upper=1> mu;
 real<lower=0.01> kap;
 real<lower=0,upper=1-0.0001> theta[N];
}
transformed parameters{
 real<lower=0> a;
 real<lower=0> b;
 a = mu * kap;
 b = kap - a;
}
```

```
model{
 kap ~ pareto(0.1,1.5);
 for(i in 1:N){
   theta[i] ~ beta(a, b);
   x[i] ~ binomial(S[i],theta[i]);
 }
}
generated quantities{
 real sig;
 sig = a*b/((a + b)^2*(a + b +1));
}
```

付録：stanコード

```
data{
 int<lower=0> N1;
 int<lower=0> S1[N1];
 int<lower=0> x1[N1];
 int<lower=0> N2;
 int<lower=0> S2[N2];
 int<lower=0> x2[N2];
 int<lower=0> N3;
 int<lower=0> S3[N3];
 int<lower=0> x3[N3];
}
transformed data{
 real S1_m;
 real S2_m;
 real S3_m;
 S1_m = sum(S1)/N1;
 S2_m = sum(S2)/N2;
 S3_m = sum(S3)/N3;
}
```

```
parameters{
 real<lower=0,upper=1> mu1;
 real<lower=0.01> kap1;
 real<lower=0,upper=1> mu2;
 real<lower=0.01> kap2;
 real<lower=0,upper=1> mu3;
 real<lower=0.01> kap3;
}
transformed parameters{
 real<lower=0> a1;
 real<lower=0> b1;
 real<lower=0> a2;
 real<lower=0> b2;
 real<lower=0> a3;
 real<lower=0> b3;
 a1 = mu1 * kap1;
 b1 = kap1 - a1;
 a2 = mu2 * kap2;
 b2 = kap2 - a2;
 a3 = mu3 * kap3;
 b3 = kap3 - a3;
}
```

```
model{
 kap1 ~ pareto(0.1,1.5);
 kap2 ~ pareto(0.1,1.5);
 kap3 ~ pareto(0.1,1.5);
 for(i in 1:N1){
    x1[i] ~ beta_binomial(S1[i],a1,b1);
 }
 for(i in 1:N2){
    x2[i] ~ beta_binomial(S2[i],a2,b2);
 }
 for(i in 1:N3){
    x3[i] ~ beta_binomial(S3[i],a3,b3);
 }
}
generated quantities{
 real eta1;
 real eta2;
 real U1;
 real U2;
 real U3;
 real gamma;
 real omega1;
 real omega2;
 eta1 = mu1-mu2;
 eta2 = mu2-mu3;
 U1= int_step(eta1);
 U2= int_step(eta2);
 U3= U1*U2;
 gamma=(mu1*(1-mu3))/(mu3*(1-mu1));
 omega1 = S1_m*(a1*b1*(a1+b1+S1_m))/
                  ((a1+b1)^2*(a1+b1+1));
 omega2 = S1_m*(a1)/(a1+b1);
}
```

ご静聴ありがとうございました。

階層ベイズ法による二項分布モデルとベータ二項分布モデル

早稲田大学大学院　文学研究科　吉上諒

発表の流れ

・心理学領域における有用性
・母比率の異質性について
・異質性を考慮した二項分布モデル
・待機児童データへの適応
・ベータ二項分布モデル
・ネズミの胎児死亡データへの適応

心理学領域における有用性

・観測対象全体に対する特定のカテゴリに属する観測対象の比率

（例）日本人全体に対するうつ病患者の比率

＋

・観測対象の属性

（例）日本人全体に対する職業別うつ病患者の比率

観測対象の属性を考慮したモデリングが可能に。

母比率の異質性

図１　フリースローのゴール数のデータ　※人工データ

	1	2	3	4	5	...	96	97	98	99	100
ゴール成功回数	1	0	2	9	9	...	5	8	3	0	5

人数：100名（i=100）　フリースロー回数：10回（n=10）

・二項分布を仮定した際の成功回数の期待値と分散

$$E\left[X\right] = n\theta, \quad V\left[X\right] = n\theta(1-\theta)$$

・データから計算される期待値と分散の予測値

$$\hat{\theta} = \frac{1}{1000}\sum_{i=1}^{N} x_i \qquad V\left[X\right] = 10\hat{\theta}(1-\hat{\theta})$$

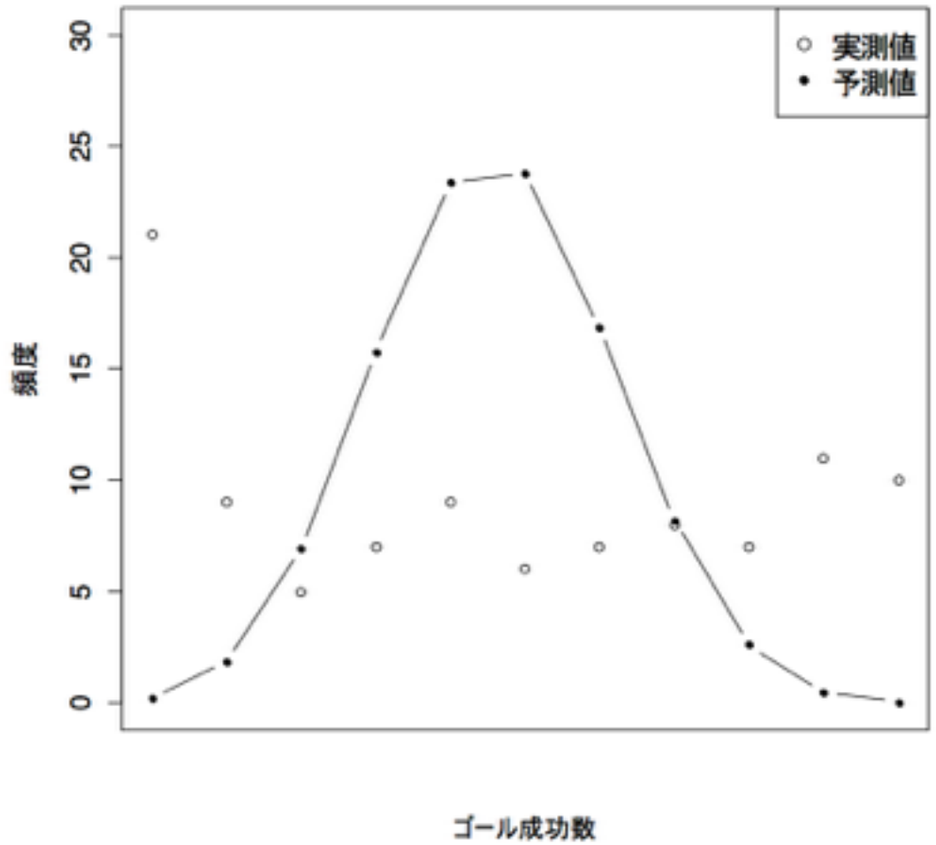
$$\hat{\theta} \simeq 0.46 \qquad = 2.48$$

母比率の異質性

分散の予測値　　約2.48
実際の標本分散　約12.75

・二項分布を仮定した場合の分散の予測値と比べて、実際のデータから求められる標本分散は約5.14倍も大きい。これを過分散という。

図２　予測値と実測値の比較



成功確率（母比率）が観測対象ごとに変わらないと仮定する二項分布モデルではデータを適切に表現できない（図２）

異質性を考慮した二項分布モデル

階層ベイズモデルを利用し、二項分布の母比率に
ベータ分布を仮定したモデル

ベータ分布の確率密度関数

$$f(x \mid \alpha_0, \beta_0) = B(\alpha_0, \beta_0)^{-1} x^{\alpha_0 - 1}(1-x)^{\beta_0 - 1}$$

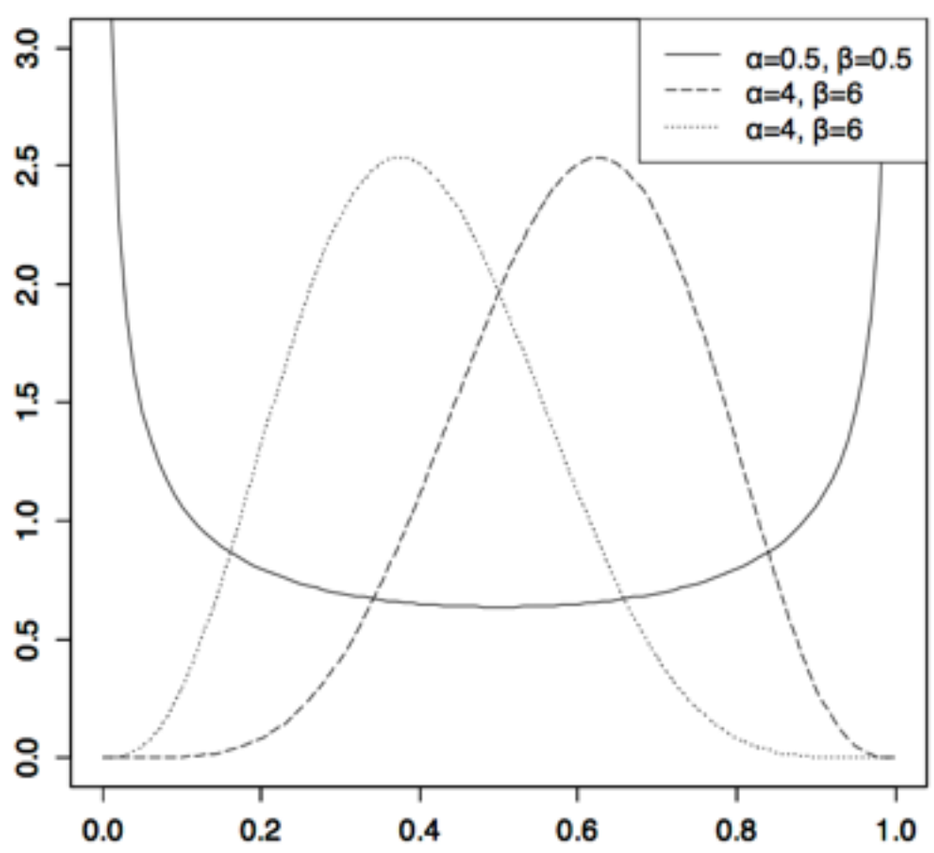
$$B(\alpha_0, \beta_0) = \int_0^1 x^{\alpha_0 - 1}(1-x)^{\beta_0 - 1} dx$$

ベータ分布に従う確率変数の期待値と分散

$$E[\Theta] = \frac{\alpha}{\alpha + \beta} = \mu \quad 0 \leq \mu \leq 1$$

$$V[\Theta] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

図３　母数の変化させた場合の
ベータ分布の密度関数



異質性を考慮した二項分布モデル

サンプリングの安定性を考慮し、Gelman, Carlin, Stern& Rubin(2015,110-111) ベータ分布の母数の再母数化を行い、κ に母数(0.1 1.5)のパレート分布を仮定する。

$$\alpha = \left(\frac{\alpha}{\alpha+\beta}\right)(\alpha+\beta) = \mu\kappa$$

$$\beta = (\alpha+\beta) - \left(\frac{\alpha}{\alpha+\beta}\right) = \kappa - \mu\kappa = \kappa(1-\mu)$$

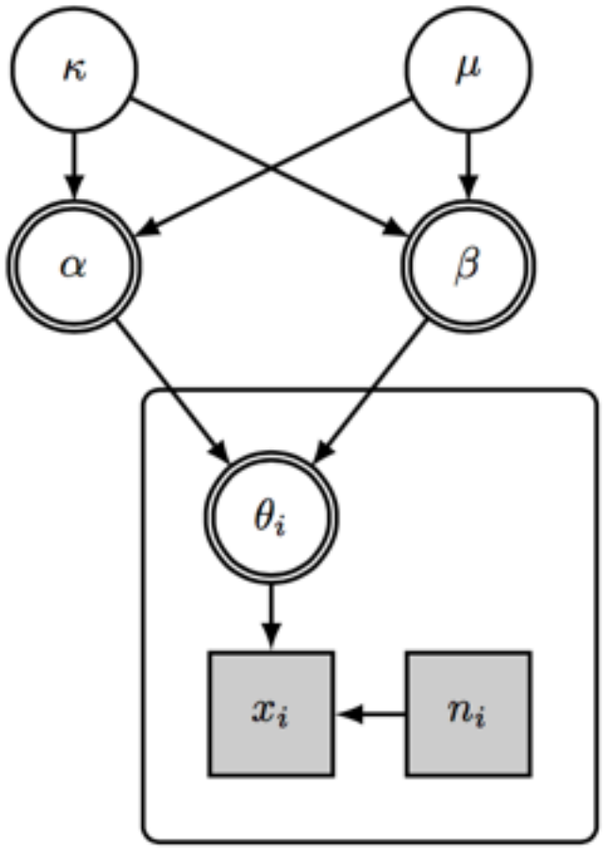
$$\kappa = \alpha + \beta$$

待機児童データへの適応

図４　各都道府県の待機児童データ

都道府県名	北海道	青森	岩手	...	宮崎	鹿児島	沖縄
全児童数	40383	26285	21664	...	18571	24963	30959
待機児童数	64	0	139	...	0	185	1721

図５　プレート図　　都道府県全体の平均待機児童率と各都道府県の待機児童率を求め、双方の大きさを比較する。



i 都道府県

$$\kappa \sim \text{Pareto}(0.1, 1.5)$$
$$\alpha = \mu\kappa, \quad \beta = \kappa(1-\mu)$$
$$\mu \sim \text{Uniform}(0, 1)$$
$$\theta_i \sim \text{Beta}(\alpha, \beta)$$
$$x_i \sim \text{Binomial}(\theta_i, n_i)$$

待機児童データへの適応

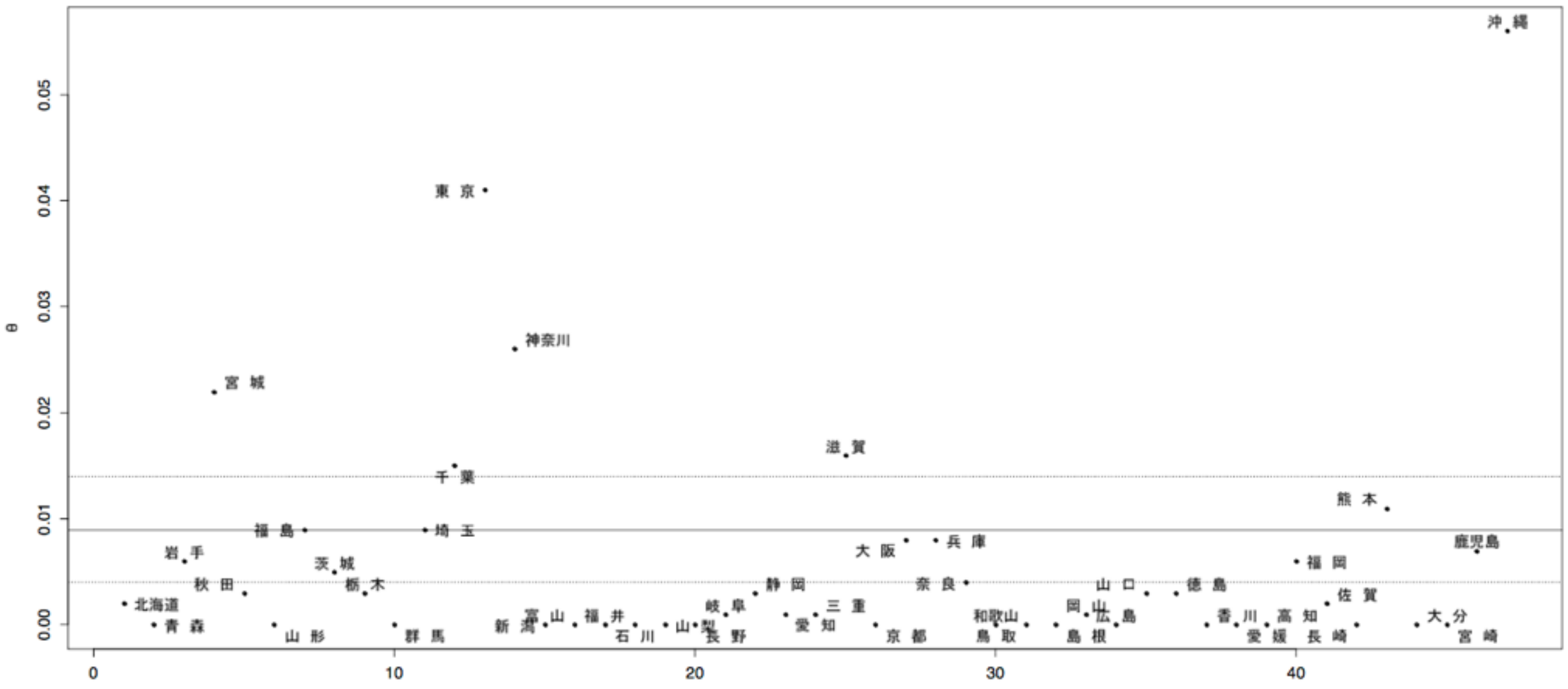
図６　分析結果

	EAP	post.sd	2.5%	50%	97.5%
μ	0.009	0.004	0.004	0.008	0.021
σ	0.001	0.001	0.000	0.000	0.002
α	0.197	0.039	0.129	0.194	0.282
β	25.135	11.506	7.536	23.563	51.859

北海道	0.002	東京	0.041	滋賀	0.016	香川	0.000
青森	0.000	神奈川	0.026	京都	0.000	愛媛	0.000
岩手	0.006	新潟	0.000	大阪	0.008	高知	0.000
宮城	0.022	富山	0.000	兵庫	0.008	福岡	0.006
秋田	0.003	石川	0.000	奈良	0.004	佐賀	0.002
山形	0.000	福井	0.000	和歌山	0.000	長崎	0.000
福島	0.009	山梨	0.000	鳥取	0.000	熊本	0.011
茨城	0.005	長野	0.000	島根	0.000	大分	0.000
栃木	0.003	岐阜	0.001	岡山	0.001	宮崎	0.000
群馬	0.000	静岡	0.003	広島	0.000	鹿児島	0.007
埼玉	0.009	愛知	0.001	山口	0.003	沖縄	0.056
千葉	0.015	三重	0.001	徳島	0.003		

待機児童データへの適応

図７　都道府県全体の平均と各都道府県の待機児童率の比較



都道府県ごとの待機児童率の異質性を考慮できている。

ベータ二項分布モデル

ベータ二項分布を利用し、二項分布の母比率の異質性を考慮したモデル

ベータ二項分布の確率密度関数

$$f(x_i \mid n, \boldsymbol{\theta}, \alpha_0, \beta_0) = f(x_i \mid n, \theta_i) \times p(\theta_i \mid \alpha_0, \beta_0)$$

$$f(x_i \mid n, \alpha_0, \beta_0) = \binom{n}{x_i} \frac{B(x_i + \alpha_0, n - x_i + \beta_0)}{B(\alpha_0, \beta_0)}$$

先のモデルと異なり、ベータ二項分布においては
二項分布の母比率を母数としない。

ネズミの胎児死亡データへの適応

図８　ネズミの胎児死亡データ

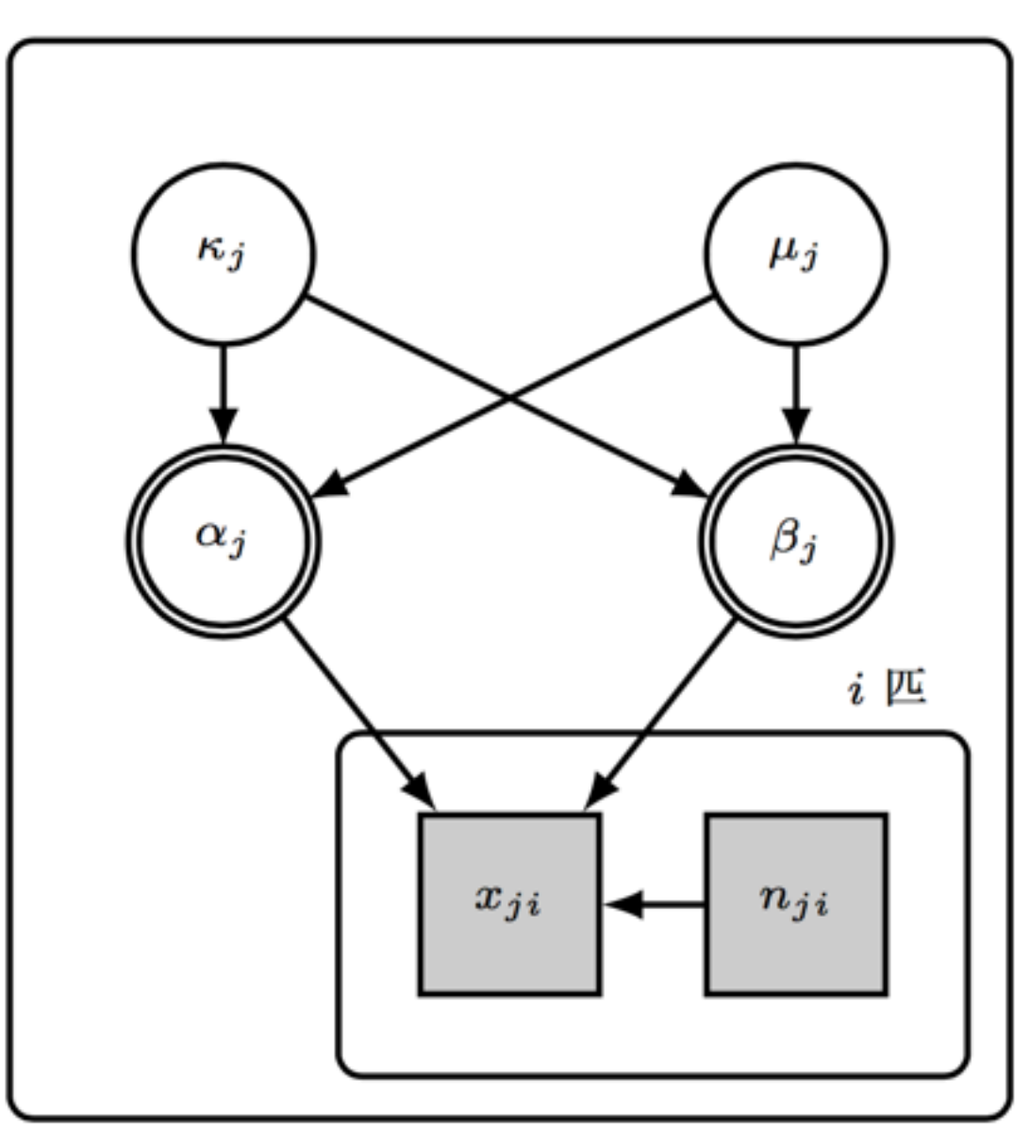
第 1 群	(4.1,10,1)	(3.2,11,4)	(4.7,12,9)	(3.5,4,4)	(3.2,10,10)
			⋮		
	(4.4,13,13)	(5.2,4,3)	(3.9,8,8)	(7.7,13,5)	(5.0,12,12)
第 2 群	(11.2,8,0)	(11.5,11,1)	(12.6,14,0)	(9.5,14,1)	(9.8,11,0)
第 3 群	(16.6,3,0)	(14.5,13,0)	(15.4,9,2)	(14.5,17,2)	(14.6,15,0)
	(16.5,2,0)	(14.8,14,1)	(13.6,8,0)	(14.5,6,0)	(12.4,17,0)

鉄欠乏のネズミに関して、（ヘモグロビンレベル、妊娠数、胎児死亡数）をまとめたデータ。ヘモグロビンレベルは個体によって異なる。
またネズミは以下の３群に分けられ、それぞれ３週間投薬を受けている。

・第１群　何も投薬をしない（31匹）
・第２群　初日と７日目に投薬を受ける（5匹）
・第３群　毎週投薬を受ける（10匹）

ネズミの胎児死亡データへの適応

図９　プレート図



$$\alpha_j = \mu_j \kappa_j$$
$$\beta_j = \kappa_j (1 - \mu_j)$$
$$\mu_j \sim \text{Uniform}(0, 1)$$
$$\kappa_j \sim \text{Pareto}(0.1, 1.5)$$
$$x_{ji} \sim \text{Beta\_Binomial}(\alpha_j, \beta_j, n_{ji})$$

ネズミの胎児死亡データへの適応

・各群の平均死亡率
・「第１群は第２群に比べて胎児死亡率が大きい」という仮説が正しい確率
・「第２群は第３群に比べて胎児死亡率が大きい」という仮説が正しい確率

なお仮説の正しい確率を算出するために、以下を定義する。

$$u_{\mu_1>\mu_2}^{(t)} = g(\mu_1^{(t)}, \mu_2^{(t)}) = \begin{cases} 1 & \mu_1^{(t)} - \mu_2^{(t)} > 0 \\ 0 & \text{それ以外の場合} \end{cases}$$

$$u_{\mu_2>\mu_3}^{(t)} = g(\mu_2^{(t)}, \mu_3^{(t)}) = \begin{cases} 1 & \mu_2^{(t)} - \mu_3^{(t)} > 0 \\ 0 & \text{それ以外の場合} \end{cases}$$

ネズミの胎児死亡データへの適応

図10　分析結果

	EAP	post.sd	2.5%	50%	97.5%
μ_1	0.763	0.049	0.658	0.765	0.849
μ_2	0.239	0.132	0.044	0.219	0.534
μ_3	0.165	0.090	0.038	0.150	0.378
$U_{\mu_1>\mu_2}$	1.000	0.020	1.000	1.000	1.000
$U_{\mu_2>\mu_3}$	0.671	0.470	0.000	1.000	1.000

- 各群の平均死亡率は第１群が0.763、第２群が0.239、第３群が0.165と推定された
- 「第１群は第２群に比べて胎児死亡率が大きい」という仮説は1.00の確率で正しいと言える。
- 「第２群は第３群に比べて胎児死亡率が大きい」という仮説は0.671の確率で正しいと言える。

参考文献

・Stan Development Team (2015).

Stan Modeling Language User's Guide and Reference Manual Stan Version 2.9.0.

・豊田秀樹 (2015)
基礎からのベイズ統計学-ハミルトニアンモンテカルロ法による
実践的入門-朝倉書店.

・厚生労働省 (2014)　保育所関連状況取りまとめ.

・李政元 (2012)　二項-ベータ階層ベイズモデルによる児童虐待相談対応率
の地域差に関する研究:都道府県政令指定都市別による多重比較
総合政策研究,41,29-36.

付録：stanコード①

```
data{
 int<lower=0> N;
 int<lower=0> S[N];
 int<lower=0> x[N];
}
parameters{
 real<lower=0,upper=1> mu;
 real<lower=0.01> kap;
 real<lower=0,upper=1-0.0001> theta[N];
}
transformed parameters{
 real<lower=0> a;
 real<lower=0> b;
 a = mu * kap;
 b = kap - a;
}
```

```
model{
 kap ~ pareto(0.1,1.5);
 for(i in 1:N){
   theta[i] ~ beta(a, b);
   x[i] ~ binomial(S[i],theta[i]);
 }
}
generated quantities{
 real sig;
 sig = a*b/((a + b)^2*(a + b +1));
}
```

付録：stanコード

```
data{
 int<lower=0> N1;
 int<lower=0> S1[N1];
 int<lower=0> x1[N1];
 int<lower=0> N2;
 int<lower=0> S2[N2];
 int<lower=0> x2[N2];
 int<lower=0> N3;
 int<lower=0> S3[N3];
 int<lower=0> x3[N3];
}
transformed data{
 real S1_m;
 real S2_m;
 real S3_m;
 S1_m = sum(S1)/N1;
 S2_m = sum(S2)/N2;
 S3_m = sum(S3)/N3;
}
```

```
parameters{
 real<lower=0,upper=1> mu1;
 real<lower=0.01> kap1;
 real<lower=0,upper=1> mu2;
 real<lower=0.01> kap2;
 real<lower=0,upper=1> mu3;
 real<lower=0.01> kap3;
}
transformed parameters{
 real<lower=0> a1;
 real<lower=0> b1;
 real<lower=0> a2;
 real<lower=0> b2;
 real<lower=0> a3;
 real<lower=0> b3;
 a1 = mu1 * kap1;
 b1 = kap1 - a1;
 a2 = mu2 * kap2;
 b2 = kap2 - a2;
 a3 = mu3 * kap3;
 b3 = kap3 - a3;
}
```

```
model{
 kap1 ~ pareto(0.1,1.5);
 kap2 ~ pareto(0.1,1.5);
 kap3 ~ pareto(0.1,1.5);
 for(i in 1:N1){
    x1[i] ~ beta_binomial(S1[i],a1,b1);
 }
 for(i in 1:N2){
    x2[i] ~ beta_binomial(S2[i],a2,b2);
 }
 for(i in 1:N3){
    x3[i] ~ beta_binomial(S3[i],a3,b3);
 }
}
generated quantities{
 real eta1;
 real eta2;
 real U1;
 real U2;
 real U3;
 real gamma;
 real omega1;
 real omega2;
 eta1 = mu1-mu2;
 eta2 = mu2-mu3;
 U1= int_step(eta1);
 U2= int_step(eta2);
 U3= U1*U2;
 gamma=(mu1*(1-mu3))/(mu3*(1-mu1));
 omega1 = S1_m*(a1*b1*(a1+b1+S1_m))/
                  ((a1+b1)^2*(a1+b1+1));
 omega2 = S1_m*(a1)/(a1+b1);
}
```

ご静聴ありがとうございました。