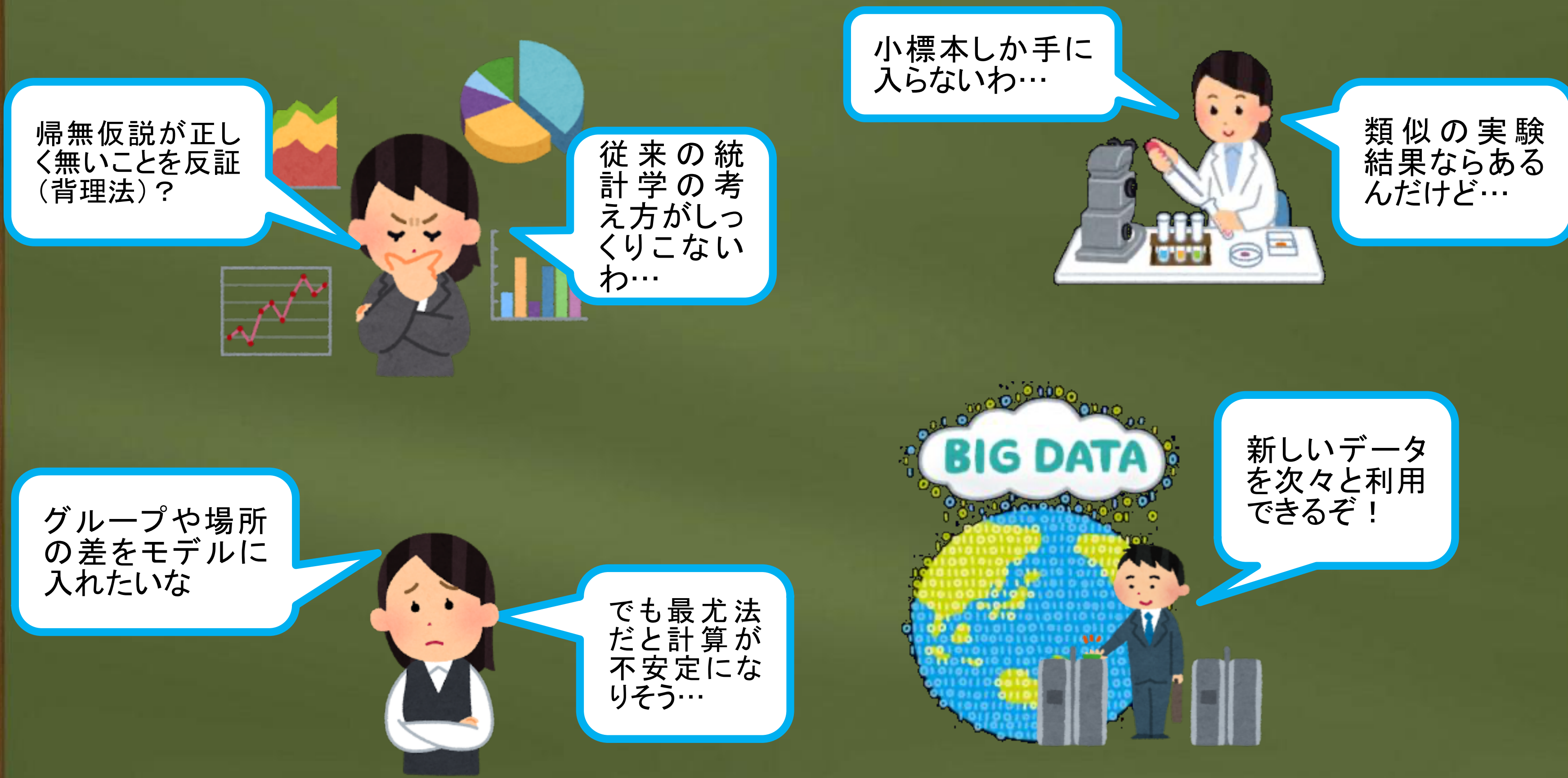
入門 統計学　第15章

ベイズ統計学

『入門　統計学　第2版　―検定から多変量解析・実験計画法・ベイズ統計学まで―』（オーム社）

※注：本書を購入された方へのサービスですので，教科書指定（参考図書は不可）していない授業での使用はお控えください。

ベイズ統計学を使うべき人



15.1　ベイズ統計学とは

ベイズの定理を土台とした推測統計学

（右が定理の生みの親であるトーマズ・ベイズ。ただし，体系的に確率的推論に本定理を使い出したのはラプラス）

T. Bayes
(1702〜1761)

統計学

- 記述統計学
  データの整理
- 推測統計学
  母数の推定
  伝統的統計学
  - 頻度論（頻度主義）
  - ベイズ統計学

考え方が異なるだけで，どちらが優れているわけではない

ブームの到来！

なぜ今，ベイズ統計学なのか？

- 理由①：情報分野の発達

→ビッグデータや機械学習との相性がよい

- 理由②：パソコン能力と計算アルゴリズムの飛躍的な向上

→課題であった複雑な積分が可能になった

- 理由③：使いやすい無料ソフト（Stan）が現れ，メジャーなソフト（SPSS，STATA）にも実装

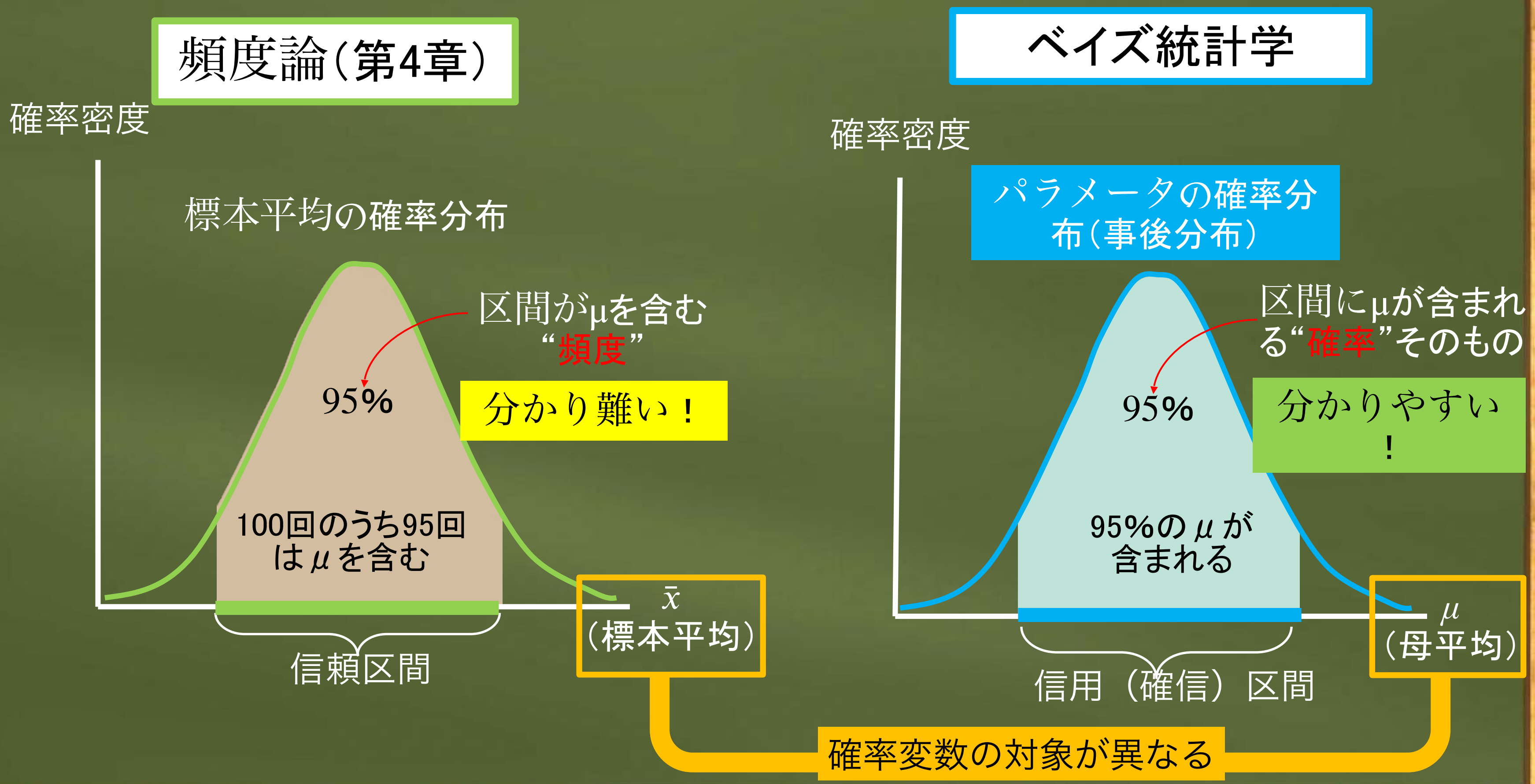
→いよいよ実践での利用が始まった

ベイズ統計学の特徴
伝統的統計学との違い

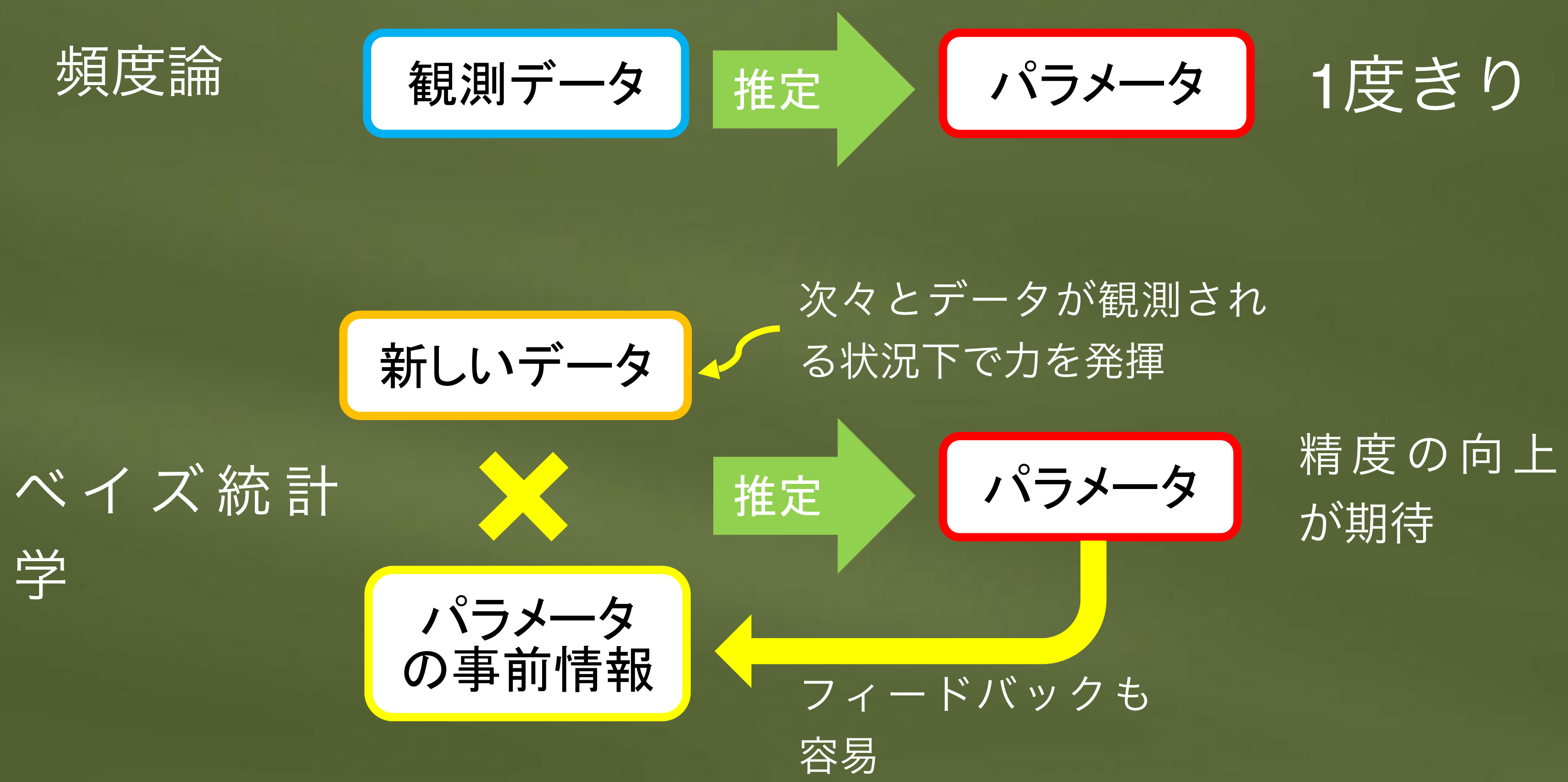
- 特徴①：パラメータ自体の分布を推定できる
- 特徴②：事前情報や新しいデータを柔軟に取り込んで分析の正確さを向上できる
- 特徴③：複雑な統計モデルを推定できる

それぞれ簡単に解説しておきましょう…

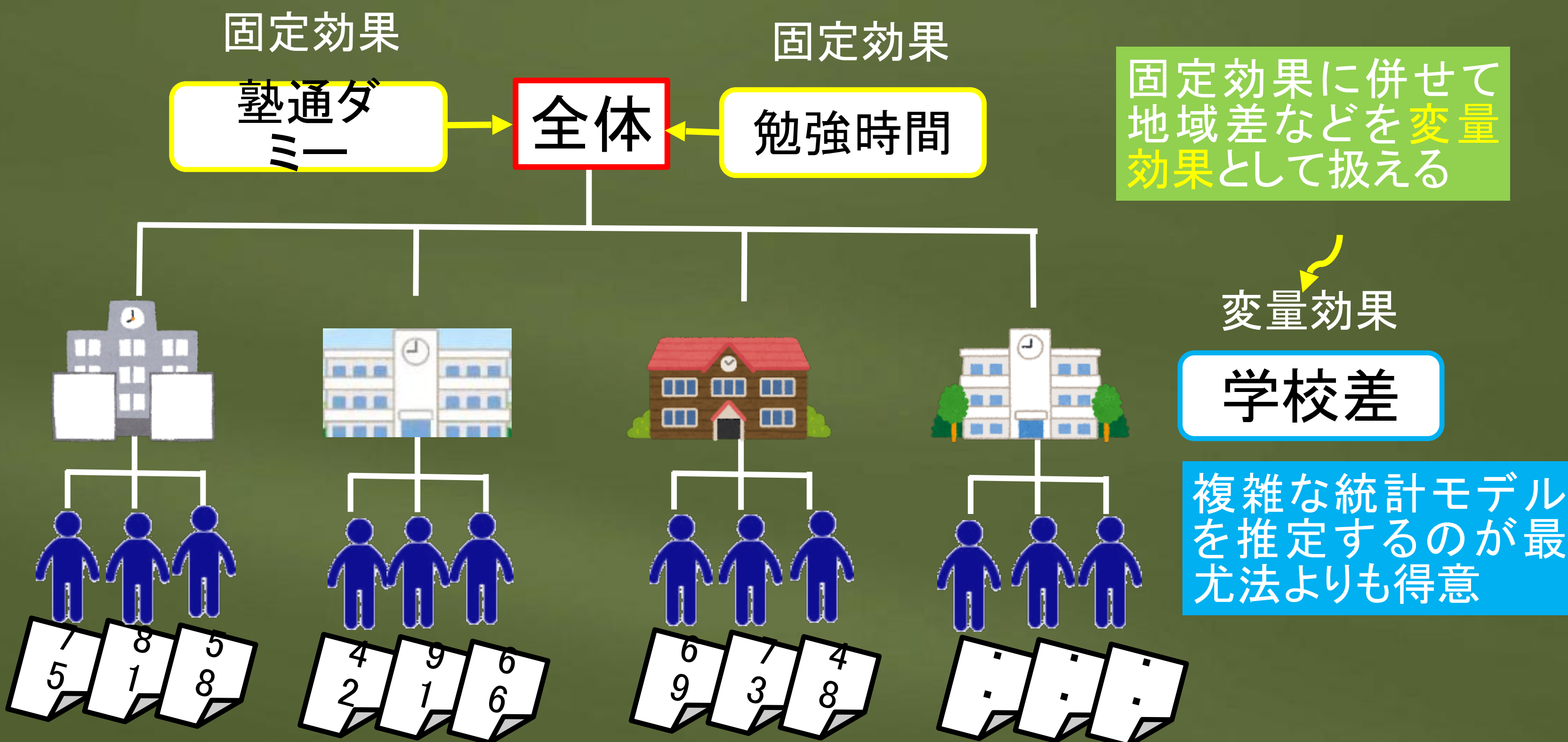
特徴①：パラメータ自体の分布を推定できる
（母平均の区間推定の例）



特徴②：事前情報や新しいデータを柔軟に取り込んで精度向上

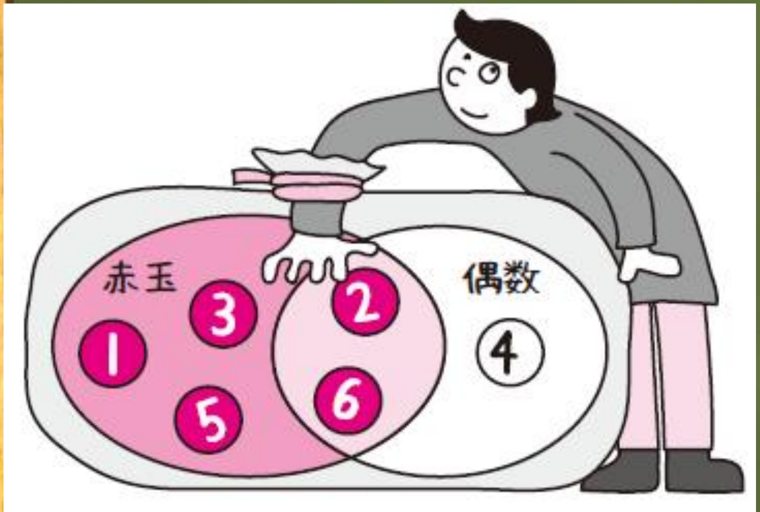


特徴③：複雑な統計モデルを扱える
（一般化線形混合モデル）



観測データ（成績）が入れ子構造の例

ベイズの定理は確率に関する式なので…
まずは確率の計算（表現法）を復習



赤玉5個と白玉1個に①～⑥の番号を振って袋に入れた状態で，赤玉が出る事象をY，偶数が出る事象をXとします。

赤玉が出る確率 $P(Y)=\frac{5}{6}$

偶数玉が出る確率 $P(X)=\frac{3}{6}$

偶数の赤玉が出る同時確率 $P(Y\cap X)=\frac{2}{6}$

同時を表す記号（キャップ）

赤玉であるときに偶数である条件付き確率 $P(X\mid Y)=\frac{P(X\cap Y)}{P(Y)}=\frac{2}{5}$

条件を表す記号（ギブン）　条件

15.2　ベイズの定理

逆の条件付き確率で求める

ベイズの定理　$\boxed{P(X|Y)} = \dfrac{\boxed{P(Y|X)}P(X)}{P(Y)}$

定理を導き出しておきましょう…

代入すればベイズの定理

条件付き確率①（先掲の式）　$P(X|Y) = \dfrac{\boxed{P(X \cap Y)}}{P(Y)}$

条件付き確率②（XとYが逆）　$P(Y|X) = \dfrac{P(X \cap Y)}{P(X)} \rightarrow P(X \cap Y) = \boxed{P(Y|X)P(X)}$

結果から原因の確率を知る

Yを結果事象（事後），Xを原因事象（事前）と考えると…

時間の流れ順

時間の流れと逆

ベイズの定理　P(

$$\frac{\mathrm{P}(}{\mathrm{P}(}$$

結果Yとしてyが観測されたとき，その原因Xがxである確率を求めることができる

↑時間の流れと逆の確率を，時間の流れ順の確率などから求めることができる点が優れている

定理を構成する確率

すでにYはyであると確定しているので，Yが起こる“確率”は不自然

観測した結果Y（データ）から見て，それを生み出した原因がXである事の“尤もらしさ”

原因Xが起こる確率（＝結果が観測される前の確率）

$$\text{事後確率}P（\text{原因}X\,|\,\text{結果}Y）=\frac{\text{尤度}P（\text{結果}Y\,|\,\text{原因}X）\times \text{事前確率}P（\text{原因}X）}{\text{全確率}P（\text{結果}Y）}$$

（逆確率とも）

結果Yとしてyが観測される確率：

原因Xがxの場合とxでない場合の2つある場合，尤度の部分のP(Y|X=x)P(X=x)とP(Y|X≠x)P(X≠x)を足し合わせる

足し合わせる（周辺化）と原因Xは消えて，結果Yだけの確率となるため，周辺尤度とも呼ぶ（後掲のサイコロのベン図を参照）

ベイズの定理を使った例題

新型コロナウィルスのPCR検査を市中でランダムに実施した場合，陽性の反応が出た人が本当に感染している確率はどれぐらいでしょうか？ただし，市中感染率は0.1%で，偽陰性は30%，偽陽性は1%であると仮定します。

感染していることは原因（時間的に前），陽性反応が出ることは結果（時間的に後）ですので，陽性反応が出た人が感染している確率は，時間の流れと逆の事後確率（逆確率）です。←ベイズの出番

P 感(染 |　　) = P 陽(性| 感) 染P　(/ P 陽（性　）

例題の解答

結果Y

原因X		陽性	陰性
	感染（0.1%）	70%	30%
	非感染（99.9%）	1%	99%

偽陰性

偽陽性

P 感(染 |) = P 陽(性| 感) 染P ( ) / P 陽 (性)

P 陽(性| 感) 染P ( ) / P 陽(性) | 感 染) + P ( ) P (

0.7×0.001 / 0.7×0.001+0.01×0.999 0.7×0.001 / 0.01069

陽性が出ても本当に感染している確率は意外と低い

個々の事象についての確率から分布に一般化

離散型確率分布の場合

原因XをパラメータΘ（シータ）（実現値はθ_i），観測データをY（実現値はy）と考えると…

結果である観測データを生み出した原因こそが統計モデルのパラメータ

飛び飛びの値が複数あるのでiを付ける

結果は2値の場合などもあるのでiは付けない

全θiのPをまとめれば事前分布（パラメータの分布）

尤度

パラメータ　データ

離散型分布のベイズの定理

$$P(\Theta=\theta_i \mid Y=y)=\frac{P(Y=y \mid \Theta=\theta_i)P(\Theta=\theta_i)}{P(Y=y)}$$

全θiのPをまとめれば事後分布

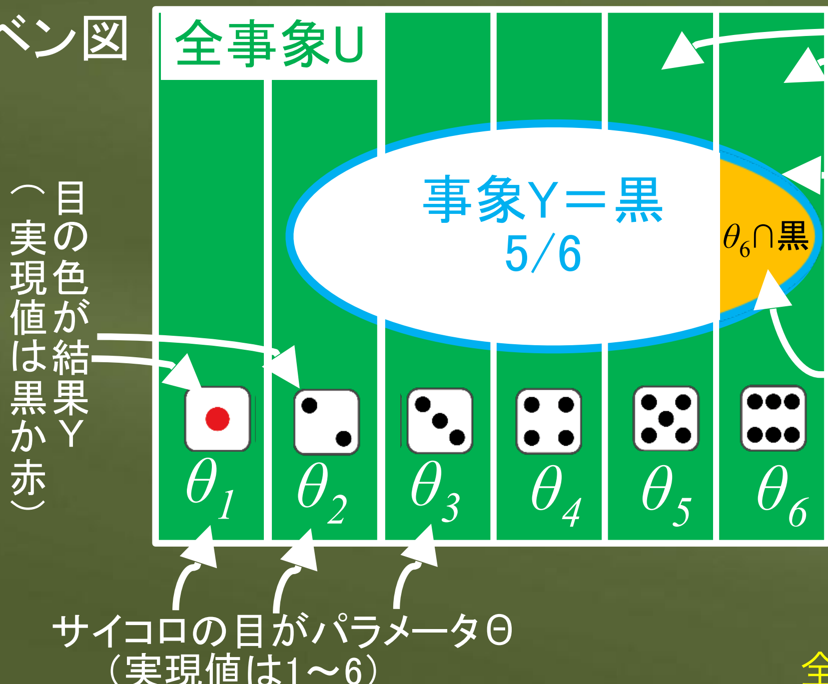
全確率（周辺尤度）

$$\Sigma\ P(Y \mid \Theta=\theta_i)P(\Theta=\theta_i)$$

離散型分布にベイズの定理を用いた事例①
（1個のサイコロ振りにおける事後分布）

サイコロの目を原因（パラメータ），目の色（黒/赤）を結果（データ）と考える

ベン図



事前確率P（Θ＝$\theta_{1\sim6}$）
いずれの目も1/6なので，事前分布は一様分布

全確率P（Y=y）
例えばP（Y=黒）＝5/6

事後確率P（Θ＝$\theta_{1\sim6}$ | Y=y）
黒目とわかったとき，6の目が出る確率は
P（Θ=6の目 | Y=黒）は次で求める

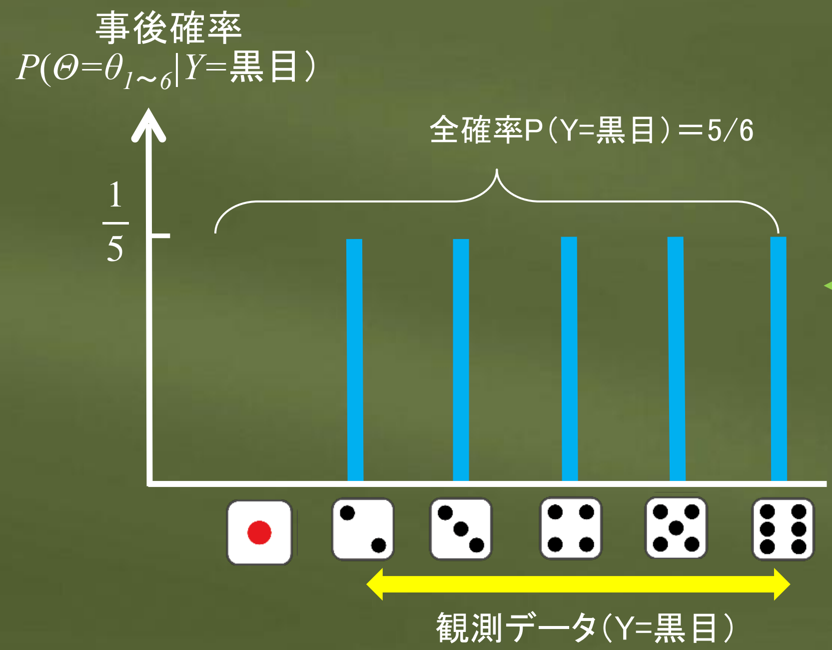
$$P(\Theta|Y)=\frac{P(\Theta\cap Y)}{P(Y)}$$

全てのサイの目の事後確率をまとめれば事後分布

離散型分布にベイズの定理を用いた事例②
（1個のサイコロ振りにおける事後分布）

黒い目が観測されたとき，6の目（パラメータ）である事後確率

$$P(\Theta=\theta_6|Y=黒)=\frac{P(Y=黒|\Theta=\theta_6)P(\Theta=\theta_6)}{P(黒)}=\frac{1\times\frac{1}{6}}{\left(\frac{0}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}\right)+\frac{1}{6}+\frac{1}{6}}=\frac{\frac{1}{6}}{\frac{5}{6}}$$



全てのPをまとめて考える

データとして黒い目（Y＝黒）が観測されたときの事後分布

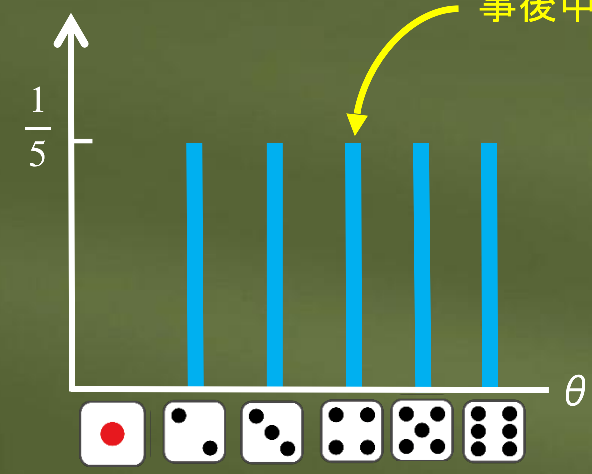
パラメータ$\theta_{1\sim6}$
（確率変数＝目）

離散型分布にベイズの定理を用いた事例③

事後分布の評価（パラメータの推定）

データから獲得

パラメータの分布（事後分布）が得られれば色々な方法でパラメータを評価できる



黒が観測されたとき，60%の確率で真のサイコロの目が含まれる信用区間[3, 5]

事後中央値MED（事後分布全体の確率が半分になる点）＝4

事後期待値（平均値）EAP （第2章）

$$E(\quad \sum_{i=1}^{6} \quad i \quad i \quad \frac{1}{5} \quad \frac{1}{5} \quad \frac{1}{5} \quad \frac{1}{5} \quad \frac{1}{5}$$

事後分散（第2章）

$$V(\quad \sum_{i=1}^{6} \quad i \quad 2 \quad i \quad 2 \quad 2 \quad \frac{1}{5}$$

$$+(3-4)^2\times\frac{1}{5}+(4-4)^2\times\frac{1}{5}+(5-4)^2\times\frac{1}{5}+(6-4)^2\times\frac{1}{5}=$$

事後確率最大値MAP：事後確率が最も大きなθが推定値
（今回は2～6が同じ事後確率なので適さない）

連続型確率分布のベイズの定理

現実の世界では重さや長さなど，変数（パラメータ）は連続している…
（注：確率Pの代わりに確率密度関数fを導入）

尤度　事前分布

事後分布

連続型分布のベイズの定理

$$f(\theta \mid y) = \frac{f(y|\theta)f(\ )}{f(y)}$$

確率密度関数　周辺尤度

$$f(y) \leftarrow \int_{-\infty}^{\infty} f(y|\theta)f(\ )$$

周辺化によってθと関係なくなり，確定しているyだけで決まる（つまり定数）

事後分布の面積を1にする役割がある

ため正規化（規格化）定数とも呼ぶ

簡略表現

プロポーション
比例

ベイズ統計学の基本式

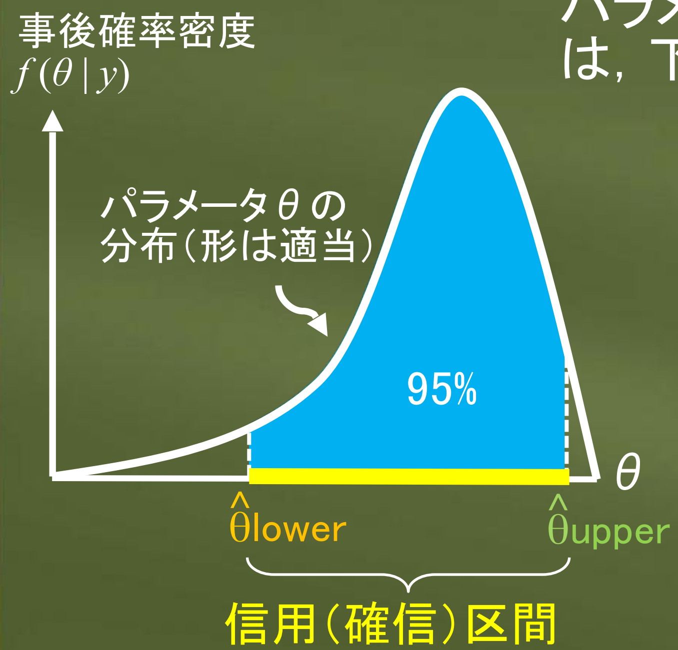
$$f(\theta \mid y) \propto f(y|\ )f(\ ) \Rightarrow$$

$\boxed{\text{事後分布∝尤度×事前分布}}$

連続型事後分布によるパラメータ評価①

（区間推定）

（離散型同様）事後分布から**パラメータの推定量** $\hat{\theta}$ を求められる



パラメータを95%の確率で含む信用（確信）区間は，下記の上下限界値で挟まれた範囲：

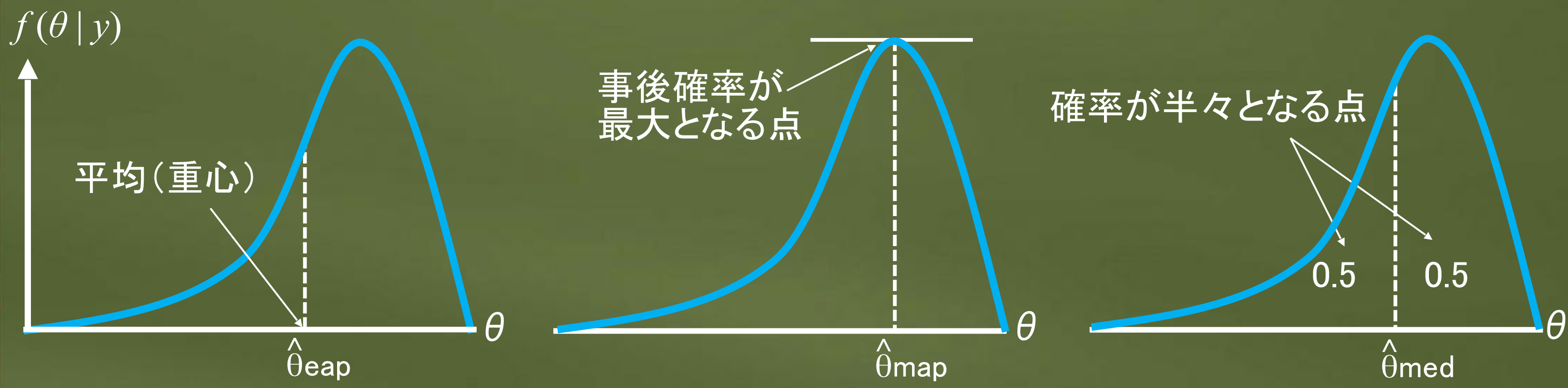
上限値 $F(\hat{\theta}_{upper} \mid y) = \int_{\hat{\theta}_{upper}}^{\infty} f(\theta \mid y)d\theta = 0.025$

下限値 $F(\hat{\theta}_{lower} \mid y) = \int_{-\infty}^{\hat{\theta}_{lower}} f(\theta \mid y)d\theta = 0.025$

下から積分して0.025となる値が下限値

連続型事後分布によるパラメータ評価②

（3種類の点推定）



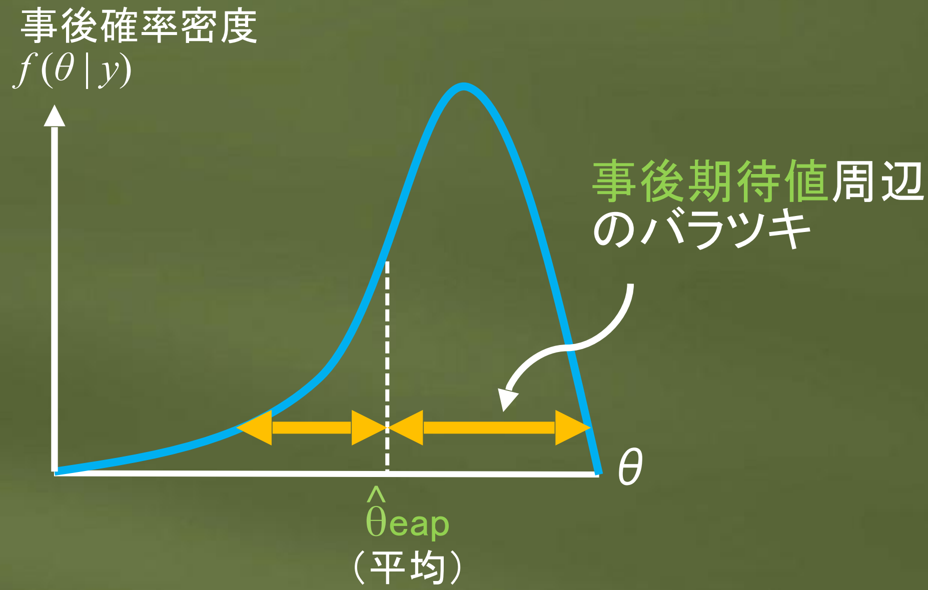
事後期待値（EAP）　　事後確率最大値（MAP）　　事後中央値（MED）

$$\hat{\theta}_{eap} = E(\Theta) = \int_{-\infty}^{\infty} \theta f(\theta | y) d\theta \qquad \hat{\theta}_{map} = \max_{\theta} f(\theta | y) \qquad F(\hat{\theta}_{med} | y) = \int_{-\infty}^{\hat{\theta}_{med}} f(\theta | y) d\theta = 0.5$$

左右対称の事後分布ならば全て同じ推定値となる

連続型事後分布によるパラメータ評価③

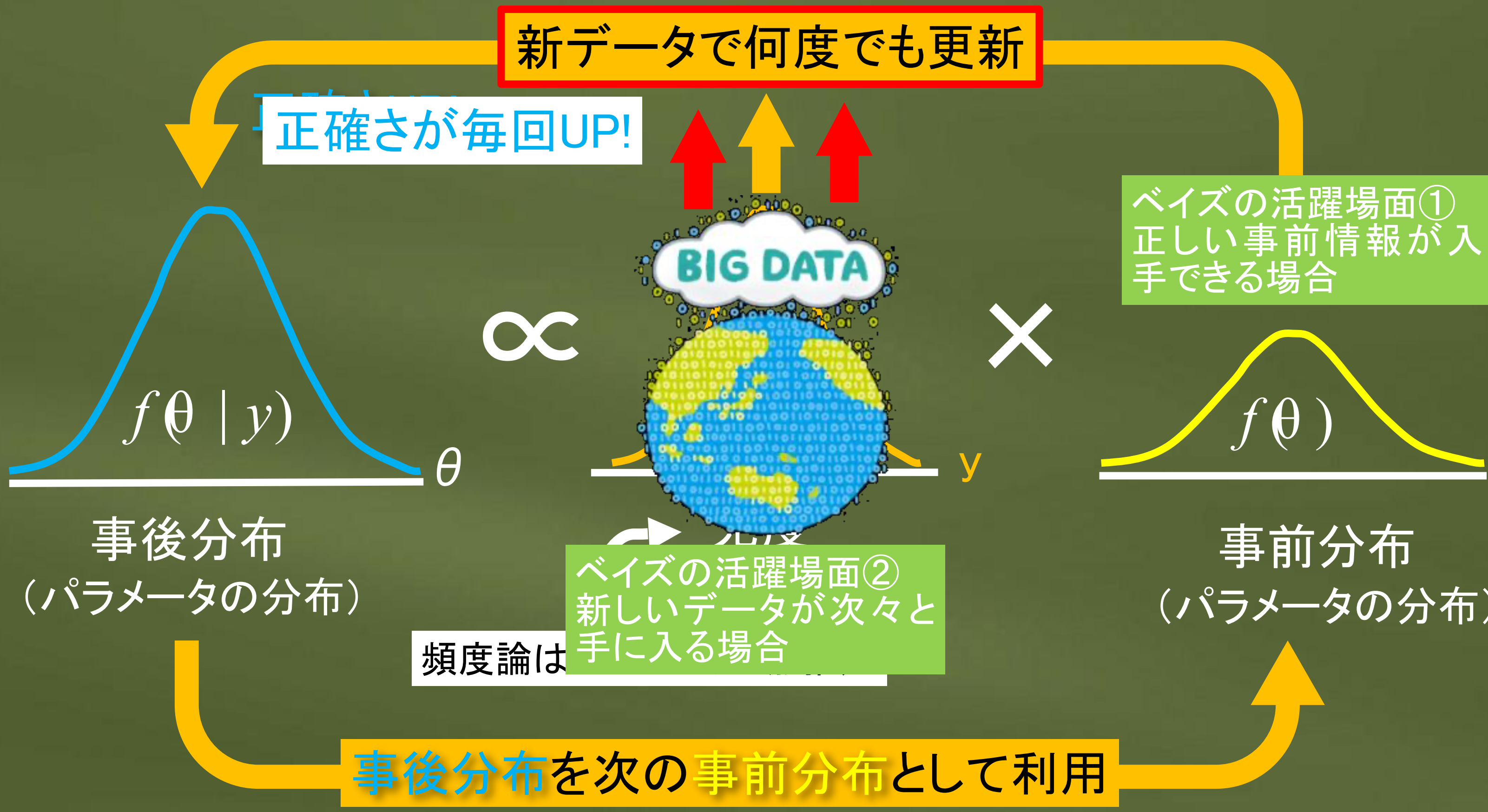
（事後分散）



事後分散 $V(\Theta) = \int_{-\infty}^{\infty} (\theta - E(\Theta))^2 f(\theta \mid y) d\theta$

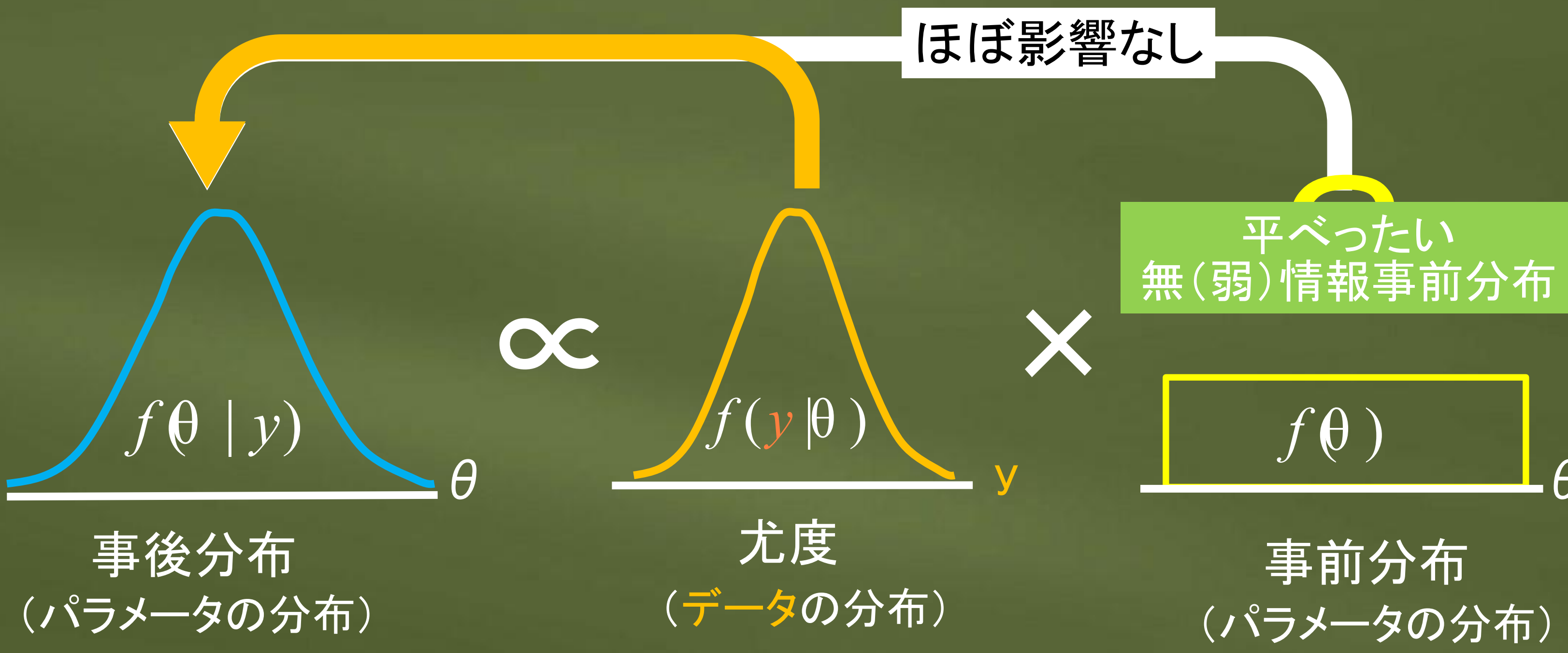
15.3　活躍場面　その1

ベイズ更新



パラメータに関する事前情報がなかったら？

事前情報を利用できなくても，パラメータの分布が得られるというベイズ統計学の利点は生かせる



回帰モデルのベイズ推定①

（線形単回帰の事後分布）

単回帰モデル　$y = \alpha + \beta x + u \qquad u \sim N(0, \sigma^2)$

yが従う分布に注目して表現し直すと…

yは平均$\alpha+\beta x$，分散σ^2の正規分布に従っている

$$\boxed{y \sim N(\mu, \sigma^2) \qquad \mu = \alpha + \beta x}$$

データ（yとx）は，αとβとσ^2をパラメータとした統計モデルから生成（次掲図参照）

単回帰モデルのパラメータの事後分布

$$\underbrace{f(\alpha, \beta, \sigma^2 \mid y, x)}_{\text{事後分布}} \propto \underbrace{f(y, x \mid \alpha, \beta, \sigma^2)}_{\text{尤度}} \times \underbrace{\boxed{f(\alpha)} \times \boxed{f(\beta)} \times \boxed{f(\sigma^2)}}_{\text{事前分布}}$$

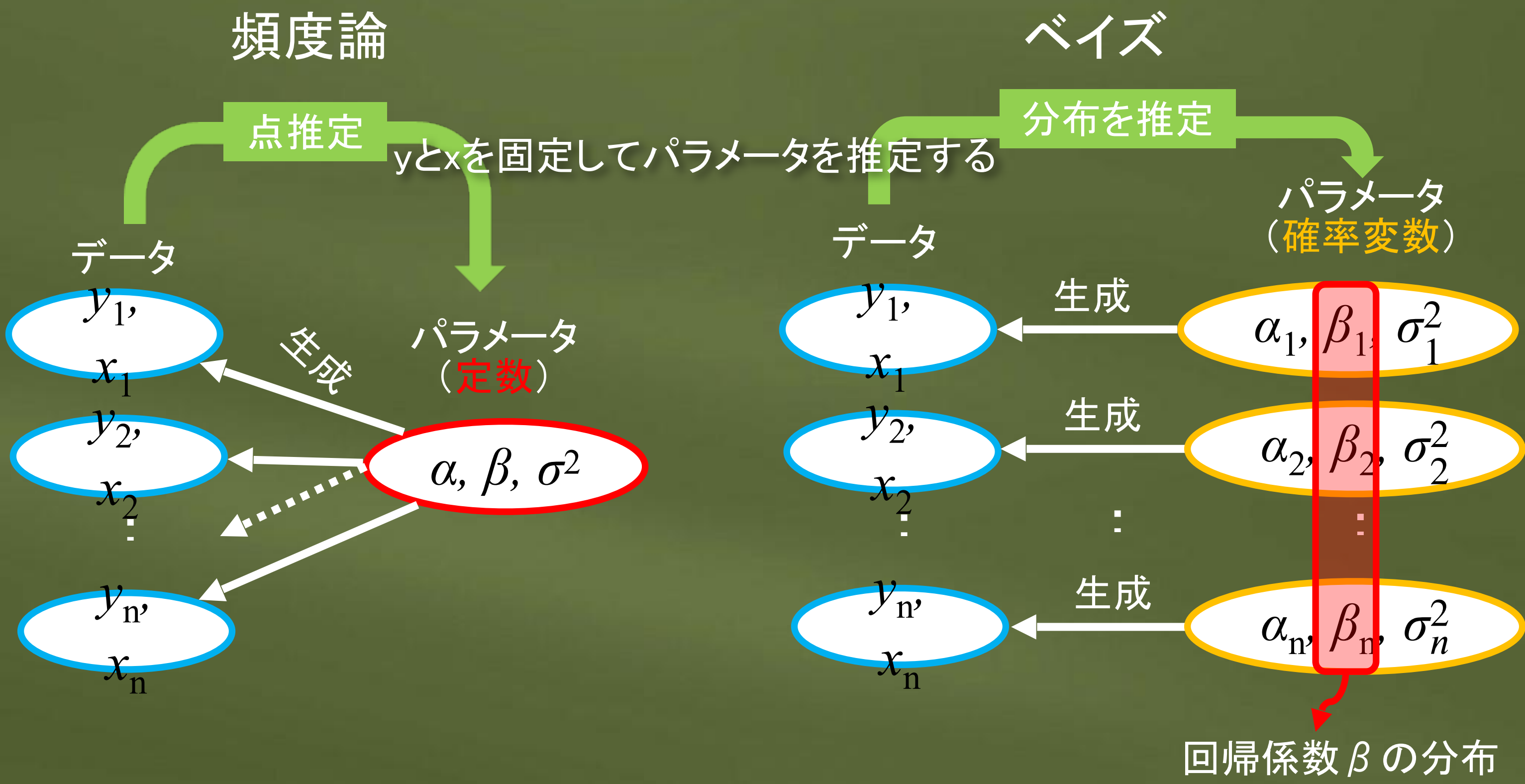
多変量の分布だが，周辺化していけば単純になる（例えばσで積分すればαとβの事後分布になる）

$-\infty \sim \infty$なので$\sigma=100$の正規分布など

$0 \sim \infty$なので正値の一様分布など

回帰モデルのベイズ推定②

（頻度論との比較）



15.4　活躍場面　その2

入れ子構造のデータ（線形単回帰の例）

学校j	生徒i	成績yi	学習時間xi
1	1	75	4
1	2	81	6
1	3	58	1
2	4	42	0
2	5	91	5
2	6	66	3
3	7	69	2
3	8	73	2
3	9	48	0
…	…	…	…

グループ差　？　結果　固定効果

これまでは固定効果※のみのモデルを考えてきた…

固定効果のみ

生徒iの学習時間xi

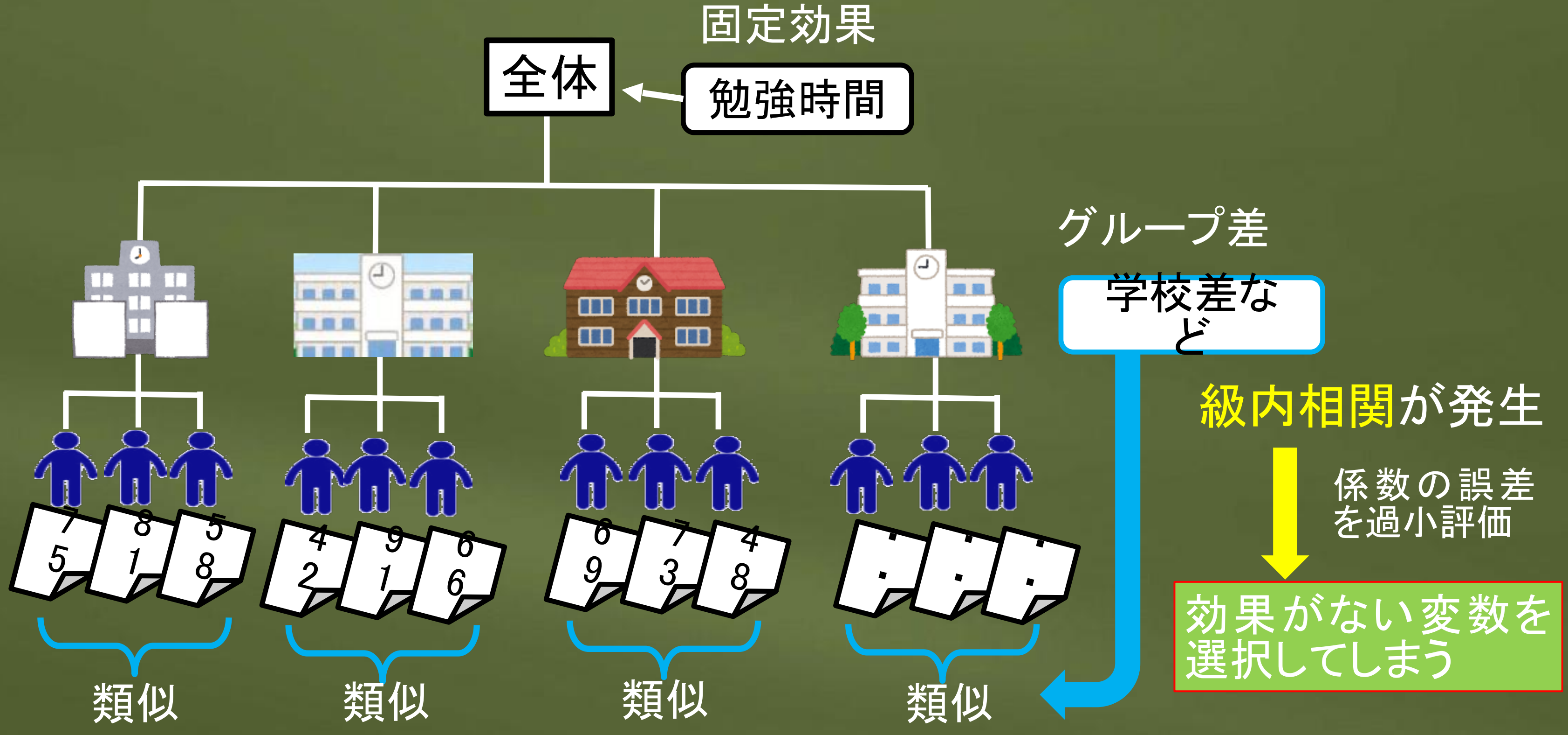
$$y_i = \alpha + \beta x_i + u_i \qquad u_i \sim N(0, \sigma^2)$$

未知のパラメータは3つ

※固定効果(fixed effect)：変数の1つ1つの特定の水準が結果に与える影響（定数項αも固定効果に含める）

学校差などグループ差がある場合，モデルに取り込まないと，どうなる？

入れ子構造のデータの課題



グループ差の処理法

- グループごとに別々のモデルを推定

  →モデルごとの標本サイズが小さくなる

- グループごとにダミー変数を作成して，固定効果として1つのモデルを推定

  →グループ数が多いと煩雑になり，階層間で多重共線性が発生する可能性

- 変量効果（ramdom effect)として1つのモデルに取り込み推定

  →混合モデル（この後解説）

固定効果と混合効果の
混合（効果）モデル

jはグループ，黄色は推定すべきパラメータ

$$y_{ij} = \alpha_j + \beta_j x_{ij} + u_{ij}$$

定数項　係数

$$\beta_j = \mu_\beta + r_{\beta j}$$

$$\alpha_j = \mu_\alpha + r_{\alpha j}$$

固定効果　変量効果

グループ差によるバラツキ
の効果（確率分布に従う）

$$\left.\begin{aligned} u_{ij} &\sim N(0,\sigma_y^2) \\ r_{\beta j} &\sim N(0,\sigma_\beta^2) \\ r_{\alpha j} &\sim N(0,\sigma_\alpha^2) \end{aligned}\right\} \text{正規分布}$$

平均0，分散σ²の正規
分布とする場合が多い

注：混合モデルは本来，計画行列を使って表記しますが，難しいので，本書では『StanとRでベイズ統計モデリング（松浦）』や『基礎から学ぶマルチレベルモデル（Ita Kreft)』の表記法を採用しています。

複雑な混合モデルに最尤法は適さない

混合モデルに最尤法を用いた場合の問題点：

①各グループが平均よりも正や負にどのぐらい大きいのか（r_jの個別値）は計算できない（知りたくない場合はOK）

→変量効果の分散σ^2を積分で求めるときに周辺化されて消えるため

混合効果　固定効果 変量効果

$$\beta_j = \mu_\beta + r_{\beta j}$$

グループjの全体平均μ_βからのズレなので，パラメータはj個ある

②そもそも最尤法はパラメータの多い複雑なモデルには適さない

→複雑なモデルでは漸近性（大標本ならば最尤推定量が正規分布に従う性質）が保証されないため，真の値から離れたり検出力が弱くなる

→パラメータがそれぞれ異なる確率分布に従う場合に確率の積を計算するのは大変なので不安定になる

階層ベイズモデル①

ベイズならば，尤度に対して，データ構造に応じたパラメータの事前分布を次々と乗ずることで，幾多のパラメータをも混ぜこぜにした1つの事後分布として扱える（グループ別の変量効果も計算できるし，漸近性も不要）

データを生成する統計モデルのパラメータ　事前分布を設定

データyが従う分布に注目した表現

$$y_{ij} \sim N((\mu_\alpha + r_{\alpha j}) + (\mu_\beta + r_{\beta j})x_{ij}, \sigma_y^2)$$

$$r_{\beta j} \sim N(0, \sigma_\beta^2)$$

$$r_{\alpha j} \sim N(0, \sigma_\alpha^2)$$

パラメータrを生成するパラメータ（ハイパーパラメータ）

事前分布に階層性を持たせる

超事前分布を設定

階層ベイズモデル②

多次元の事後分布から，いずれかのパラメータを推定するには多
重の積分計算∫∫∫∫が必要…どうする？

事後分布　　　　　　　　　　　　　　尤度

$$\boxed{f(\mu_\alpha,\mu_\beta,\sigma_y^2,r_{\alpha j},r_{\beta j},\sigma_\alpha^2,\sigma_\beta^2|y_{ij},x_{ij})}\propto\boxed{f(y_{ij},x_{ij}|\mu_\alpha,\mu_\beta,\sigma_y^2,r_{\alpha j},r_{\beta j})}$$

$$\times\boxed{f(\mu_\alpha)\times f(\mu_\beta)\times f(\sigma_y^2)\times f(r_{\alpha j}|\sigma_\alpha^2)\times f(r_{\beta j}|\sigma_\beta^2)}\times\boxed{f(\sigma_\alpha^2)\times f(\sigma_\beta^2)}$$

事前分布　　　　　　　　　　　　　　超事前分布

パラメータ　ハイパーパラメータ

生成

事前分布が階層構造

（階層ベイズは，データが持つ階層構造の事ではない）

15.4　事後分布（によるパラメータ）の評価

マルコフ連鎖モンテカルロ法（MCMC）

直接的な積分はあきらめて，事後分布から取ってきた乱数を使って推定

手順①：事後分布からマルコフ連鎖を用いてサンプリングする

多変数の確率分布から乱数（パラメータと見なす）を取り出すこと

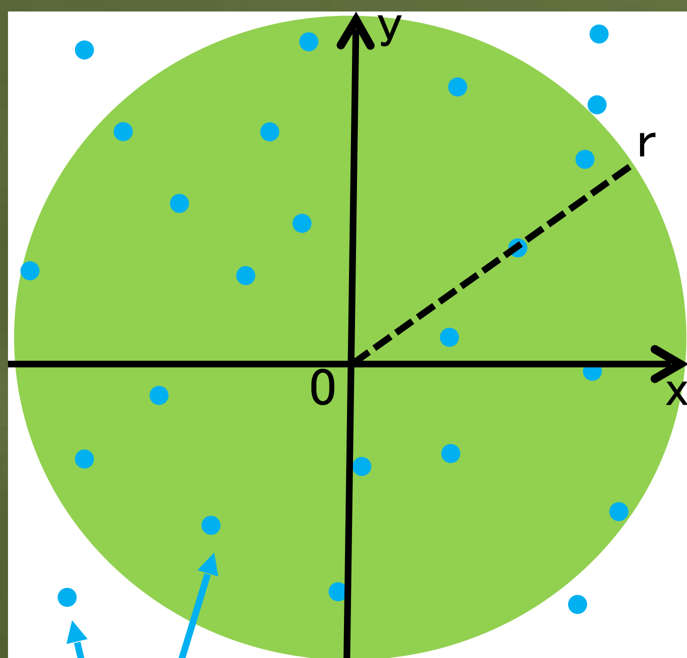
MCMC
（数種類ある）

モンテカルロ法

手順②：サンプル（乱数の集合）に対して平均や分散を計算して積分の代わりとする

モンテカルロ法（積分の近似計算）

円の面積の計算例



ランダムに点(乱数)をばらまく

本来は $\int_{-r}^{r}\sqrt{r^2-x^2}\,dx$×2という積分計算が必要（$x^2+y^2=r^2$より)

確率p

円の中に入った個数
÷全体の個数

∴円の面積＝外枠(四角形)の面積×確率p

球の体積（3次元）：箱の中でランダムに点をばらまき，球に入った数から求める

↑パラメータが増えて次元が上がっても容易に対応

マルコフ連鎖（乱数の生成ルール）

高次元空間で完全にランダムに乱数を発生させるのは不効率

マルコフ連鎖で解決

未来の確率変数（乱数）X_{n+1}が現在の値X_nだけから決定されるという考え方（それより過去は無関係）

1期前の状態に依存

（マルコフ連鎖を規定する条件付き確率）
遷移核

$$\mathrm{P}(X_{n+1}=x \mid X_n=x_n,\ldots,X_1=x_1)=\boxed{\mathrm{P}(X_{n+1}=x \mid X_n=x_n)}$$

過去の値が全部わかっているという条件の下での未来の確率

現在の値だけがわかっているという条件の下での未来の確率

メリット①：1期前のみを参考にするため，効率よく乱数を取り出せる
メリット②：p_1からp_2，p_2からp_3…と次々に求めていくと値は収束する
（乱数を取ってくる事後分布が定常分布になる）

マルコフ連鎖の事例

今日X_1　明日X_2　あさってX_3　100日後X_{101}

鼻水は
かなり出ていた

鼻水の出る確率：
今日出てたから…
$p_2=0.9$

鼻水の出る確率：
前日も少し出てたから…
$p_3=0.6$

鼻水の出る確率
前日結構出てたから…
$p_{101}=0.8$

上がったり下がったりしながら徐々に収束する

一番基本的なMCMC

メトロポリス・ヘイスティング法（M-Hアルゴリズム)①

MCMC：遷移核を上手く設定して，事後分布が定常分布になるマルコフ連鎖を作り，そこから乱数を生成

　→マルコフ連鎖の各要素X_{n+1}を，事後分布から取ってきた乱数（パラメータ）θ_{n+1}とみなす

次の乱数として提案されたθ

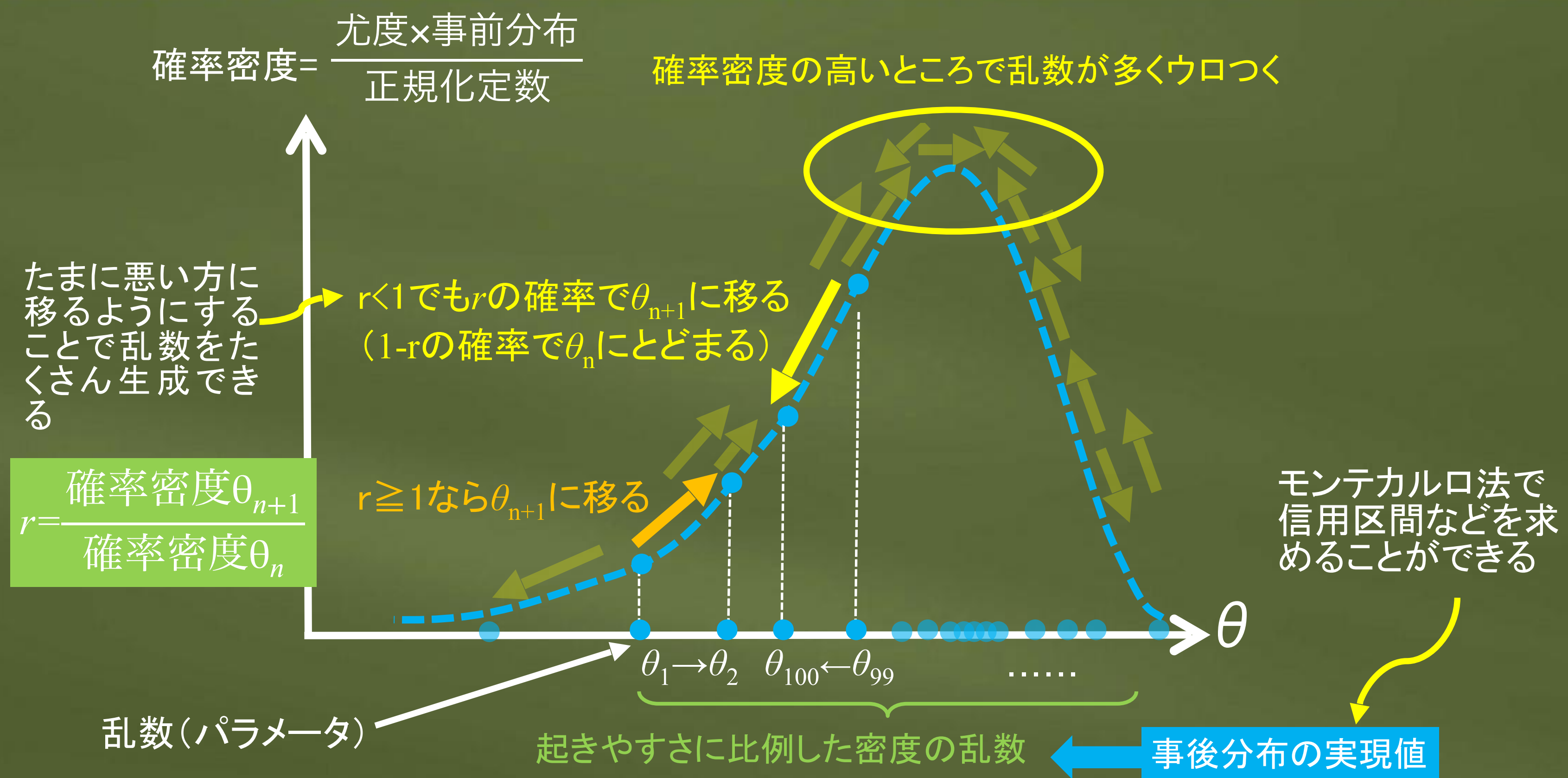
M−Hアルゴリズムの遷移核

$$r=\frac{事\qquad\theta_{n+1}}{事\qquad\theta_{n}}=\frac{\dfrac{f(y|\theta_{n+1})f(\quad_{n+1})}{\cancel{f(y)}}}{\dfrac{f(y|\theta_{n})f(\quad_{n})}{\cancel{f(y)}}}$$

1より大きければθ_nよりもθ_{n+1}の方が確率密度が大きい（パラメータとして適している）

積分が面倒な正規化定数が消えるというメリット

メトロポリス・ヘイスティング法②



15.5　まとめ

ベイズの注意点とソフトウェア

- あまり複雑なモデル（多階層の入れ子構造など）を想定しない

　→MCMCでパラメータの推定値が収束し難くなる

- 恣意的な事前分布を設定しない

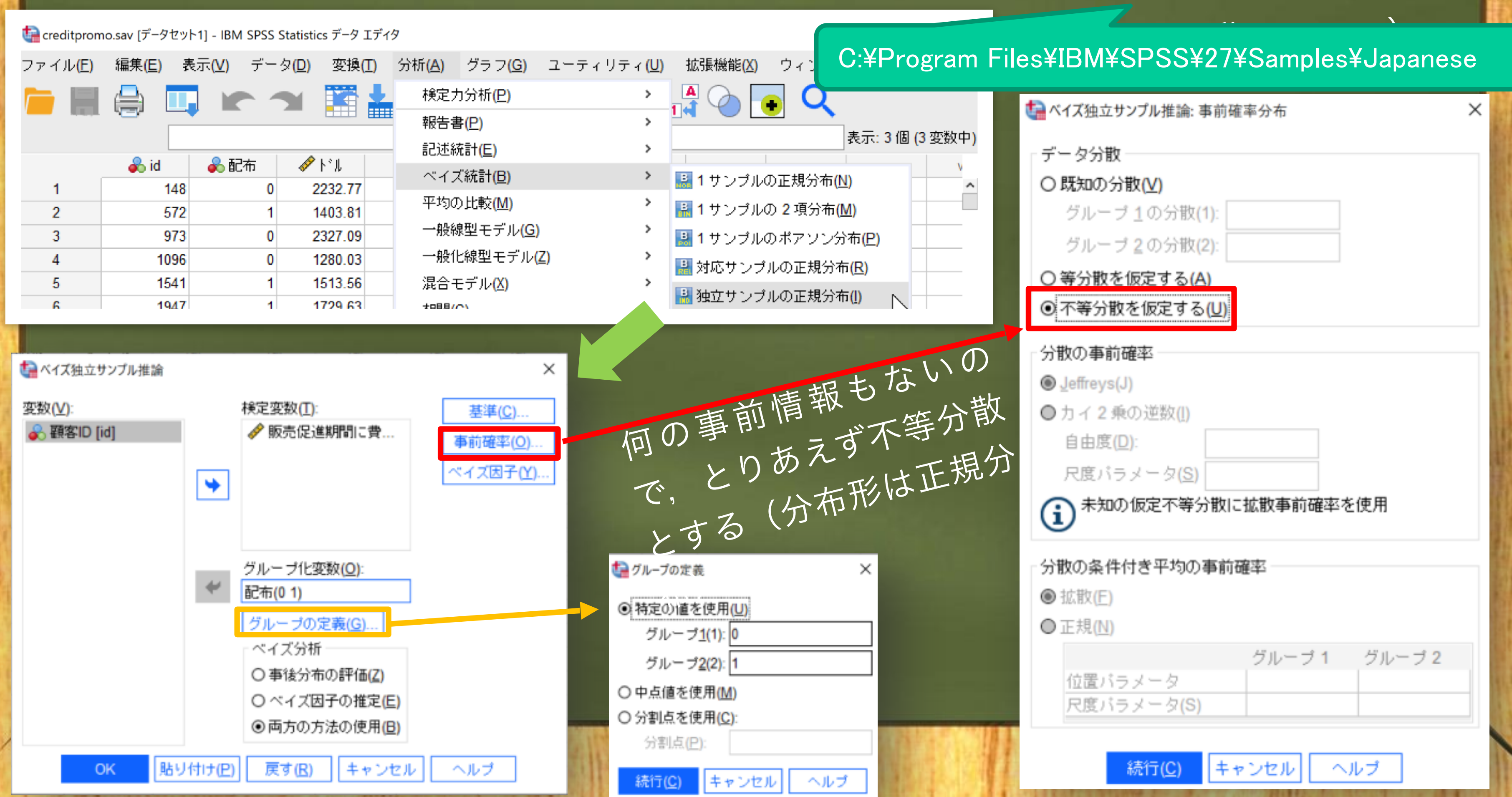
　→データ取得が1回きりで，客観性や公平性が重視される分野なら頻度論にしておく

（参考）

SPSSによるベイズ統計学

独立した2群の平均の差の推定①

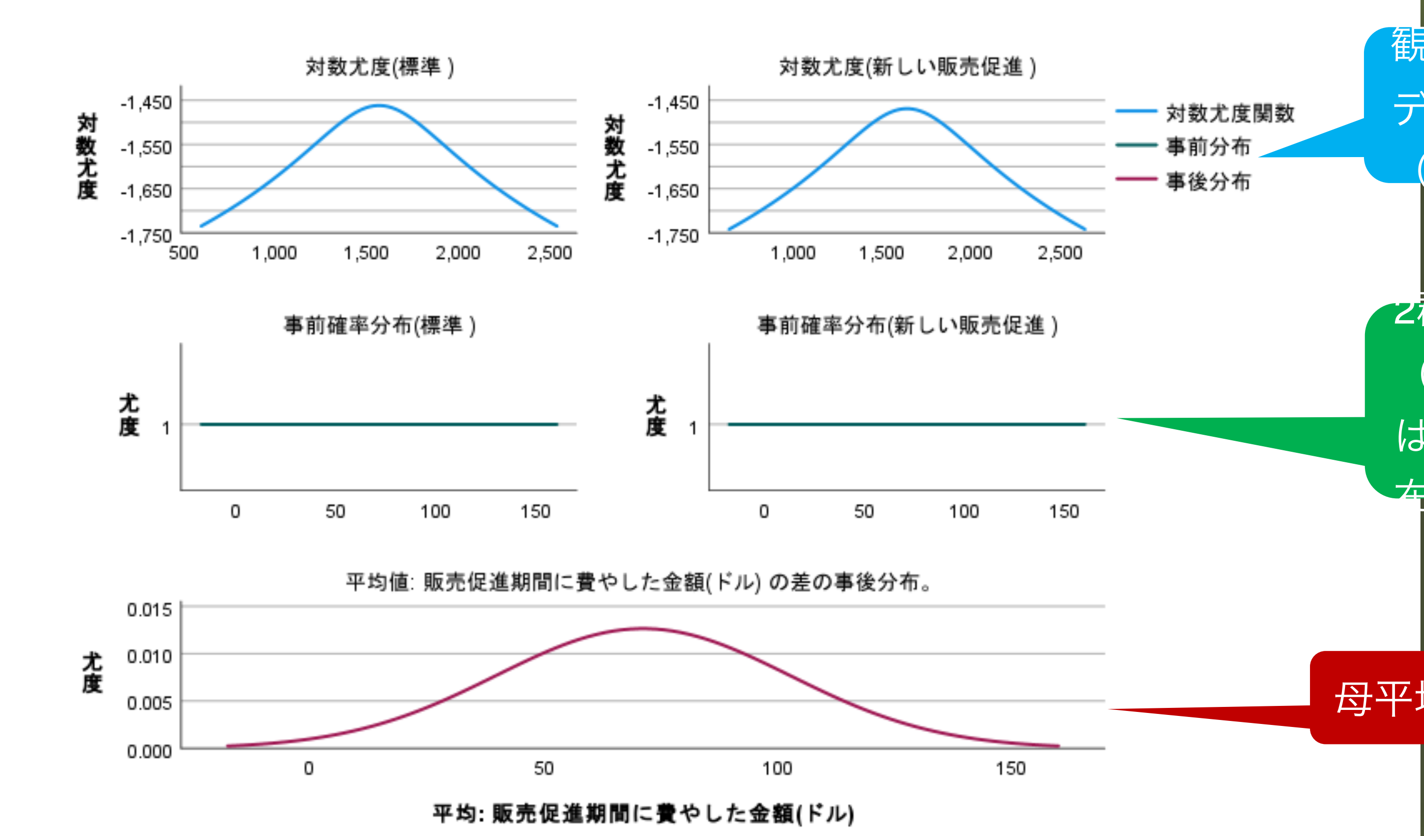
新しい販売促進がクレジットカード使用額に与えた影響分析（サンプルファイル：



（参考）

SPSSによるベイズ統計学

独立した2群の平均の差の推定②



観測された2群のデータの分布（対数尤度）

×

2群の事前分布（今回は事前情報はないので一様分布）

||

母平均の差の確率分布

（参考）

SPSSによるベイズ統計学

独立した2群の平均の差の推定③

独立サンプル平均値の事後分布評価[a]

	事後分布			95% 信用区間	
	最頻値	平均値	分散	下限	上限
販売促進期間に費やした金額(ドル)	71.1110	71.1110	997.691	9.1767	133.0452

a. 分散の事前確率: Diffuse。平均の事前確率: Diffuse。

真値（母平均の差）が95％の確率で含まれると信用できる区間

検定のようなこともできる（パラメータの分布が求まるので必要性は低い）↓

ベイズ因子独立サンプルの検定 (方法 = Rouder)[a]

	平均値の差	プールされた差の標準誤差	ベイズ因子[b]	t 値	自由度	有意確率 (両側)
販売促進期間に費やした金額(ドル)	71.1110	31.45914	1.150	2.260	498	.024

a. グループ間で不等分散を仮定します。

b. ベイズ因子: 帰無仮説 対 対立仮説。

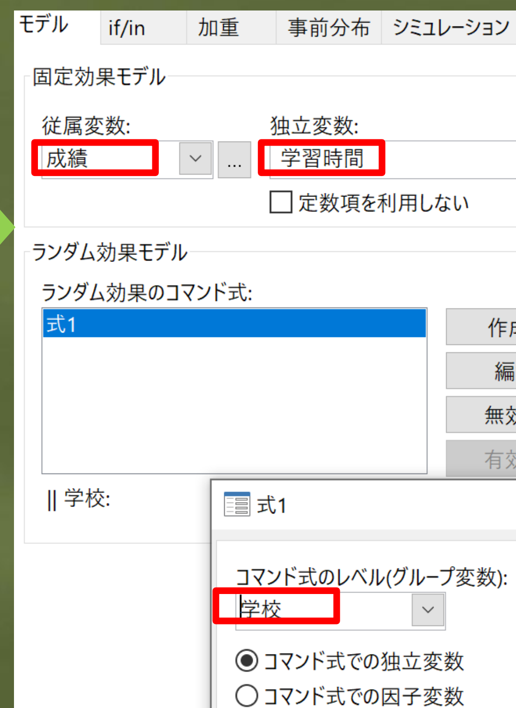
グループ2（新しい販売促進）のモデルの方がグループ1（標準）のモデルよりも尤もらしい。

（参考）

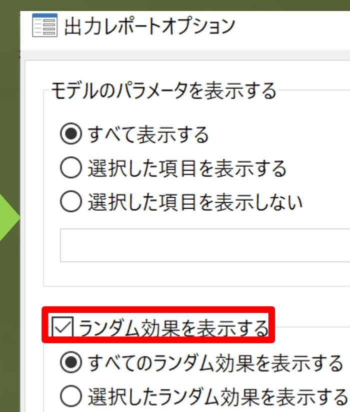
STATAによるベイズ統計学
線形混合モデル①

表15.1の成績のデータ（線形混合モデル）

固定効果に学習時間，
変量効果に学校を設定



学校別の変量効果
の値を表示する



（参考）

STATAによるベイズ統計学
線形混合モデル②

```
Bayesian multilevel regression                  MCMC iterations  =     12,500
Metropolis-Hastings and Gibbs sampling          Burn-in          =      2,500
                                                MCMC sample size =     10,000
Group variable: 学校                             Number of groups =          3

                                                Obs per group:
                                                             min =          3
                                                             avg =        3.0
                                                             max =          3

                                                Number of obs    =          9
                                                Acceptance rate  =      .8121
                                                Efficiency:  min =      .0318
                                                             avg =      .2317
Log marginal-likelihood                                      max =      .6318
```

事後分布の評価

	平均 Mean	標準偏差 Std. Dev.	標準誤差 MCSE	中央値 Median	信用区間 Equal-tailed [95% Cred.	Interval]
成績						
学習時間	6.812599	1.714658	.021572	6.828337	3.405889	10.23423
_cons	49.85974	6.470371	.137288	49.771	36.81623	63.04366
U0[学校]						
1	-1.012968	3.438104	.192809	-.1212398	-11.00991	3.773118
2	-.5387752	3.044413	.144845	-.0163995	-8.70026	4.795246
3	.7505432	3.20943	.144893	.1334103	-4.996328	9.346132
学校						
U0:sigma2	31.59324	192.6665	4.77957	.9246839	.0073184	242.2062
e.成績						
sigma2	84.94981	76.05439	1.09737	62.85645	21.48052	283.0922

乱数の数

M-H法で提案した θ_{n+1}が受容された比率（0.1以下は問題あり）

回帰係数

学校別の変量効果

学校差の大きさ（変量効果の分散）

成績（残差）の分散

以上で第15章は終了です。