心理学者のための
ベイズ統計入門

清水裕士
関西学院大学

自己紹介

- 清水裕士
  - 関西学院大学社会学部　准教授
- 専門
  - 社会心理学，グループ・ダイナミックス
- 趣味
  - 心理統計学
  - 統計ソフトウェアの開発

今日のお話

- いまなぜ心理学でベイズ統計なのか
  - 仮説検定を超えて
  - 最尤推定を超えて

- ベイズ統計学とは
  - ベイズの定理の理解を深める

- マルコフ連鎖モンテカルロ法
  - ベイズ推定のアルゴリズムについて

- まとめ

いまなぜ心理学でベイズ統計なのか

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*†

Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available. Replication effects were half the magnitude of original effects, representing a substantial decline. Ninety-seven percent of original studies had statistically significant results. Thirty-six percent of replications had statistically significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effect size; 39% of effects were subjectively rated to have replicated the original result; and if no bias in original results is assumed, combining original and replication results left 68% with statistically significant effects. Correlational tests suggest that replication success was better predicted by the strength of original evidence than by characteristics of the original and replication teams.

心理学の再現性問題

- 実験結果が再現されない
  - 疑わしい研究実践：QRP
    - 有意になるまでデータを足す
    - いろんな指標で検定を試す
    - データを見た後で仮説を設定する：HARKing
- 有意性検定に対する過度な信頼と圧力
  - 統計的に有意な結果なら大丈夫
  - とりあえず有意にならないと論文にならない

仮説検定からもう一歩

- *p*値「だけ」からの脱却
  - 効果量に注目しよう
  - 信頼区間に注目しよう
  - 効果量の信頼区間も
- 仮説検定からの脱却
  - ベイズ統計の可能性
  - より自由な仮説設定へ
  - 予測に使いやすい

頻度主義からベイズ統計へ

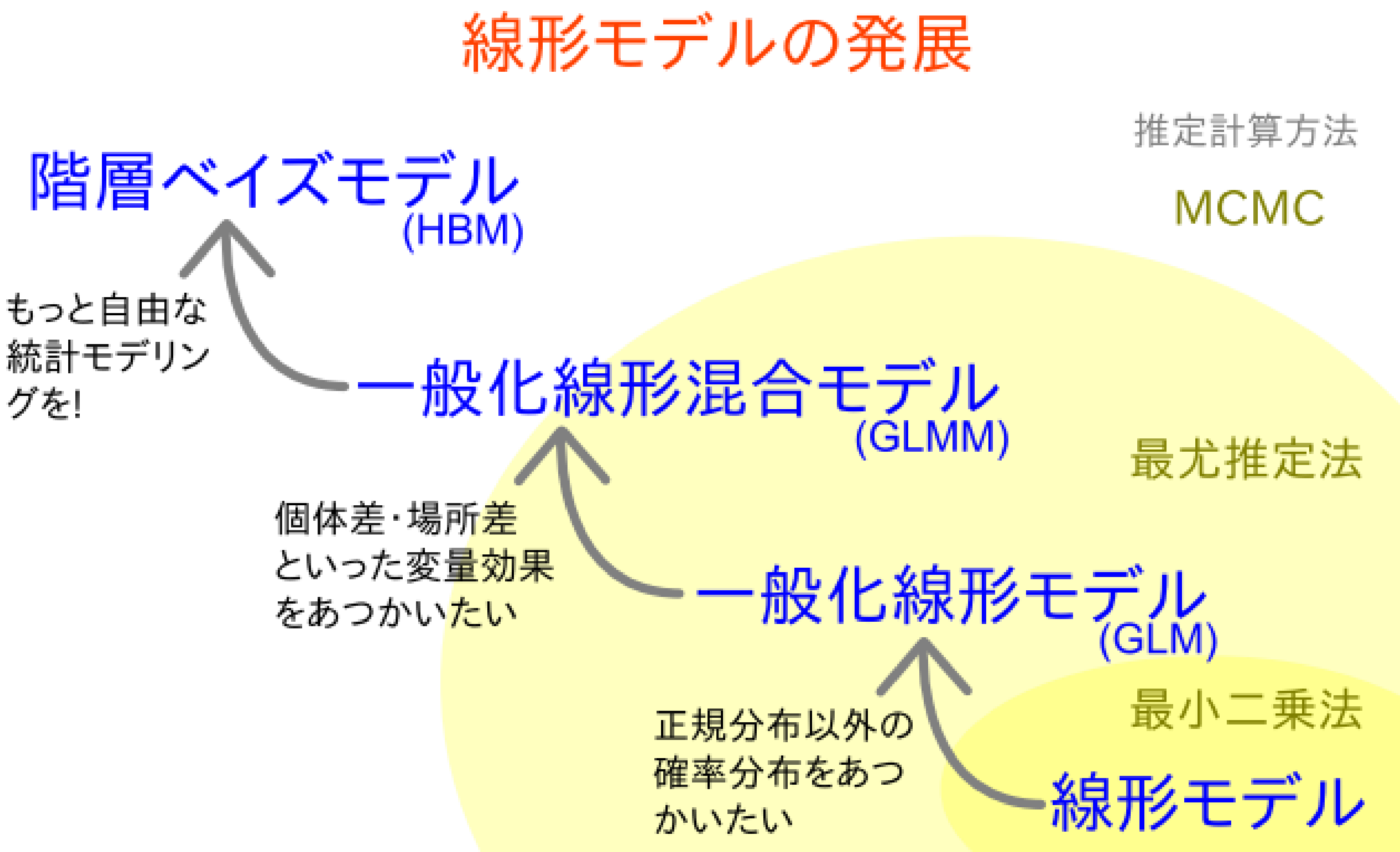
- 仮説検定がどうも変な理屈である
  - 帰無仮説を設定して，それが間違えていることをデータで示す
  - 95%信頼区間は，同じサンプルサイズのデータをたくさんとったときに，その範囲に真値が95%の確率で含まれる区間のこと
- ベイズのほうが直感に合う？
  - 研究仮説が正しい確率がわかる
  - パラメータの確率的範囲が直接わかる

心理統計とベイズ推定

- これまでの心理統計モデリング
  - 統計モデルにデータを合わせる
    - 連続量を分割してカテゴリ変数にして分散分析
    - データを対数変換して正規分布に近づける
    - 参加者効果を固定効果として推定
    - →なんとか分散分析できるようにもっていく

- 近年の心理統計モデリング
  - データに合わせて統計モデルを選択する
    - 説明変数が連続量なら回帰分析＋交互作用
    - 正規分布でないなら，指数分布族に拡張
    - 参加者効果を変量効果として推定
    - →一般化線形混合モデル

一般化線形混合モデル



久保(2012)から引用

最尤法からベイズ推定へ

- 最尤法が推定できるモデルには限りがある
  - 異なる確率分布を混ぜるような複雑なモデルは推定が難しい場合がある
    - できないわけじゃないけども近似推定となる
    - 一般化線形混合モデルや，混合分布モデルなど
    - 情報量規準などが（厳密には）正しさが保証されない

- ベイズ推定でこれらは解決する
  - MCMC法で推定することで，複雑なモデルでも正確に解くことが可能
  - 情報量規準も正しさが保証される

ベイズ統計とベイズ推定

- 似て非なる，やっぱり似た二つ
  - ベイズ（主義）統計と頻度主義統計は，思想が根本的に異なる
  - しかし，ベイズ推定と最尤推定は，基本的に同じ次元で理解することができる
    - ただベイズ推定する以上，ベイズ統計の考え方と無関係ではいられないのは確か

- 心理学からはどっちからベイズに入ればいい？
  - 個人的な感覚では，まずはベイズ推定，その後に徐々にベイズ統計の考え方が広まっていくのでは？
  - ただ実験系だとベイズ主義からのほうが入りやすい？

頻度主義仮説検定の限界

心理統計といえば仮説検定

- 帰無仮説と有意水準を設定
  - H0:μ=0と仮定する
    - 自動的に対立仮説H1：μ≠0となる
  - α=0.05
- データから検定統計量を計算
  - *t*値が自由度29で2.50だった
  - データによって計算された検定統計量が危険率αにおける棄却域に含まれていれば帰無仮説を棄却
    - つまり，H1:μ≠0が正しいと結論する
  - そうでなければH0を保留する

*p*値を求める

- *p*値が0.05より小さければ有意
  - p<.05に祈りをささげる

- *p*値とはなんだったか
  - 帰無仮説が正しいと仮定したとき，実現値より極端な値が出る確率
    - 帰無仮説が正しい確率では「ない」
    - 対立仮説が間違えている確率でも「ない」

仮説検定の限界

- 手続きがややこしい
  – 背理法を使った論法
  – 帰無仮説が正しいと仮定してそれを否定する
- *p*値が知りたいわけじゃない
  – 帰無仮説が正しいと仮定した場合の，データが得られる確率
  – 別に帰無仮説を正しいなんて思ってない
    - そもそも，帰無仮説は基本的に間違えている

知りたいのは仮説が正しい確率

- 「ボクの考えた最強の仮説」は正しいの？
  - 頻度主義統計学では教えてくれない

- なぜか
  - p値は「真値についての仮説」を所与とした，データが得られる確率
    - 真の平均=0と仮定した場合の，標本平均$\bar{x}$が得られる確率
    - 確率的に変化するのはデータ$\bar{x}$であって真値μではない
      - 真値は定数だから
  - 仮説の正しさの程度は，仮説検定ではわからない
    - データの得られる確率のみがわかる

仮説の正しさが知りたい

- ベイズ統計学で知ることができる
  - ベイズ統計では，知りたい値が確率変数であると考える
    - つまり，母平均μは確率変数
  - 逆にデータは固定的であると考える
    - 「すでに実現したもの」は確率的に変化はしない

- 仮説が正しい確率
  - データを所与とした場合の，仮説（例えばμ>0）の確率がわかる

ただし，そこまでには長い道のり

- まずは，ベイズの定理の説明
  - そしてその前に確率の話
- ベイズ推定の話
  - 統計モデリングとからめて
  - 最尤推定法の話も
- 仮説検定の話
  - ベイズファクターなんかの話も
  - しばらくお付き合いください

ベイズ統計学とは

ベイズ統計学

- ベイズさんが考えた定理を使って推論する
  - ベイズの定理

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

  - P(θ|X)：事後分布（パラメータの分布）
  - P(x|θ)：尤度
  - P(θ)：事前分布
  - θ：パラメータ
  - x：データ

なるほど，わからん

もうちょっと噛み砕いてみる

- ベイズの定理

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

  – これを理解するためには，いくつかの準備が必要

  – 確率の話
  – 確率分布の話
  – あとでやる

確率の話

P()ってなんだ

- 確率の関数
  - P(x)と書けば，事象xが起こる確率を意味する
  - サイコロを振って1の目が出る確率はP(x=1)
    - 1/6が返ってくる

- 確率は0～1の範囲しかとらない
  - 0は0%，1は100%　0.2は20%を意味する

P(θ|x)ってなんだ

- この「たて棒」なに？
  - xで条件付けられたθの確率であることを示す
    - 条件付き確率という
  - たとえばP(はい｜男性)は,「男性が『はい』と答える確率」を意味する

	男性	女性	合計
はい	8	4	12
いいえ	2	6	8
合計	10	10	20

- 条件付き確率を計算してみる
  - P(はい|男性)は8 / 10 = 0.8
  - P(はい|女性)は4 / 10 = 0.4

ベイズの定理

- P(はい|男性)の確率は，次の式で計算できる

$$P(\text{はい} \mid \text{男性}) = \frac{P(\text{男性} \mid \text{はい})P(\text{はい})}{P(\text{男性})}$$

- P(男性 | はい)の確率
  - はいと答えた人の中の男性の割合
    - 8/12 = 2/3
  - P(はい)= 12/20 = 0.6
  - P(男性)= 10/20 = 0.5

	男性	女性	合計
はい	8	4	12
いいえ	2	6	8
合計	10	10	20

- P(はい|男性)
  - (2/3 * 0.6) / 0.5 = 0.8・・・先ほどの結果と一致

ベイズの定理

- このベイズの定理は･･･

$$P(\text{はい} \mid \text{男性}) = \frac{P(\text{男性} \mid \text{はい})P(\text{はい})}{P(\text{男性})}$$

- このようにも書ける

$$P(\text{はい} \mid \text{男性}) = \frac{P(\text{男性} \mid \text{はい})P(\text{はい})}{P(\text{男性} \mid \text{はい})P(\text{はい}) + P(\text{男性} \mid \text{いいえ})P(\text{いいえ})}$$

  – つまり，{はい，いいえ}を反応iで表すと，

$$P(\text{男性}) = \sum_{i=1}^{n} P(\text{男性} \mid \text{反応}_i)P(\text{反応}_i)$$

確率分布の話

分布に拡張

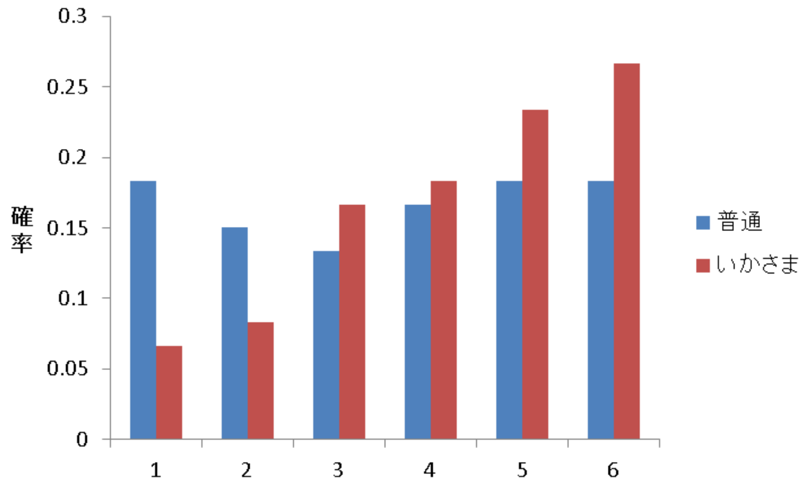
- ダイスを60回振ってみた
  - 「普通のダイス」と「いかさまのダイス」がある

	1	2	3	4	5	6	合計
普通	11	9	8	10	11	11	60
いかさま	4	5	10	11	14	16	60
合計	15	15	20	20	25	25	120

  - P(x=2|いかさま)= 5/60 = 1/12
    - ここまではさっきと同じ

分布としてみる

- 確率に関する分布・・・確率分布
  - 特に，離散変数の場合は離散分布と呼ぶ



- xが変数の場合，P(x)は確率分布になる
  - 青い分布はP(x|普通)を表す
  - 赤い分布はP(x|いかさま)を表す

確率分布の性質

- 全部足すと1になる
  - 確率の定義
    - 逆に，全部足して1にならないものは確率と呼ばない

- さっきの表を確率に直すと･･･

	1	2	3	4	5	6	合計
普通	.183	.150	.133	.167	.183	.183	1
いかさま	.067	.083	.167	.183	.233	.267	1
合計	.125	.117	.150	.175	.208	.225	1

確率分布でもベイズの定理

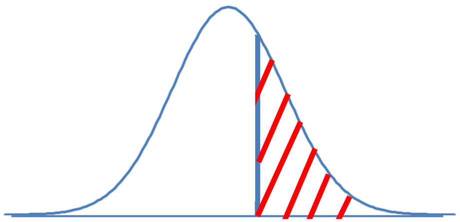
- xが変数になっても同様に成り立つ

$$P(x|\text{いかさま}) = \frac{P(\text{いかさま}|x)P(x)}{P(\text{いかさま})}$$

  - ダイスがいかさまである場合の，ダイスの目（1～6）が出る確率を表す
  - 複数の確率を同時に計算している，ともいえる

連続変数に拡張

- 正規分布を例に考える
  - 男性の身長と女性の身長がそれぞれ独立した正規分布に従うとする
    - 男性・・・平均170cm，*SD*10cmの正規分布
    - 女性・・・平均155cm，*SD*10cmの正規分布
- 条件付き確率
  - P（x>175cm|男性）は，男性で身長が175cmより高い確率を表す



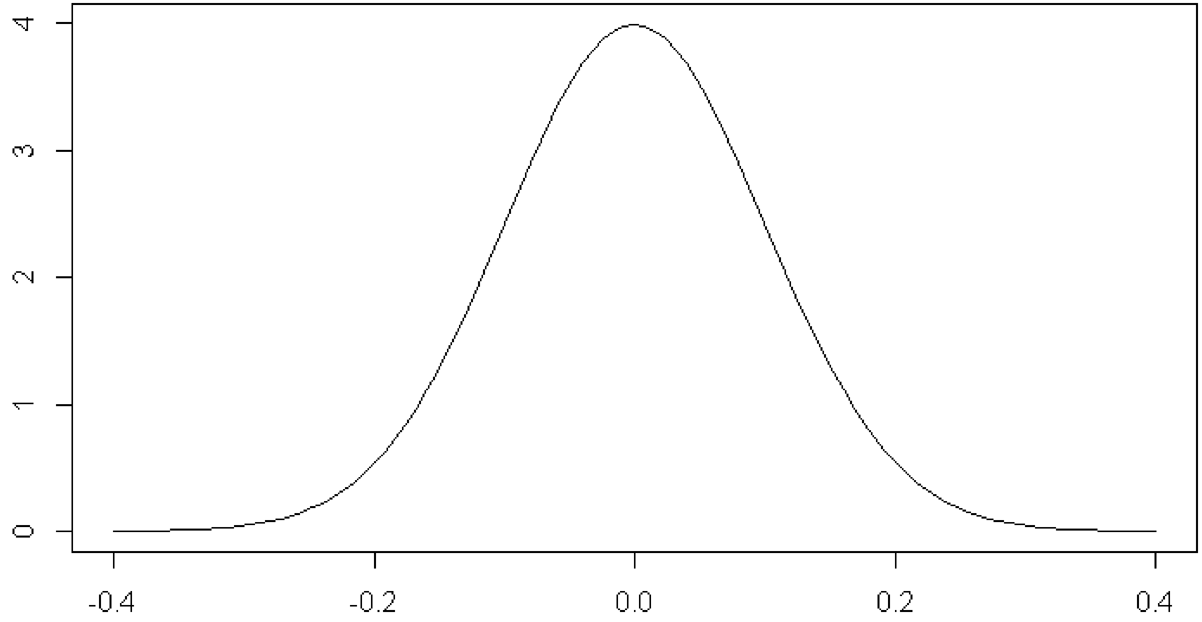
連続分布の場合の注意

- 連続値の確率分布･･･確率密度関数
  - P(x=175｜男性)は，男性で身長が170cmの確率･･･をあらわすが･･･
    - これを確率密度という　確率密度は1を超えることがある
  - 連続値の場合，ぴったり170cmの人が実現するとは考えにくい（微妙に違いがあるはず）

- 確率密度は1を超えることがある
  - 平均170，標準偏差0.1の正規分布で，170が実現する確率密度は3.98になる

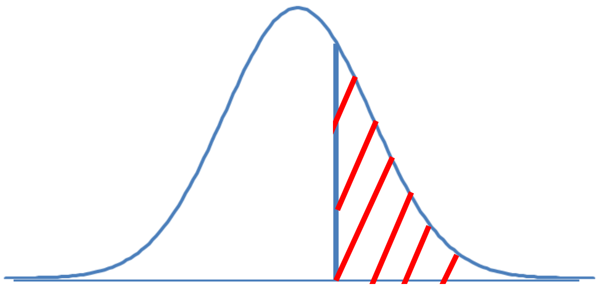
確率密度関数

- 平均=0, *SD*=0.1の正規分布
  - 確率密度は1を超えることがある
  - 面積は1を超えることはない（ぴったり1になる）



連続分布の場合の注意

- 確率・・・積分した値
  - 連続値の場合は範囲=面積で確率を考える
    - 定積分した値



  - 連続値の確率分布を-∞～∞で積分すると1
    - 離散変数では総和が1になったことと対応

確率密度関数もベイズの定理

- 175cmより高い男性の確率

$$P(x > 175|\text{男性}) = \frac{P(\text{男性}|x > 175)P(x > 175)}{P(\text{男性})}$$

- 男性の身長全体についての確率分布

$$P(x|\text{男性}) = \frac{P(\text{男性}|x)P(x)}{P(\text{男性})}$$

確率密度関数もベイズの定理

- 別の表現ではこう書くこともできる

$$P(x|\text{男性}) = \frac{P(\text{男性}|x)P(x)}{\int_{-\infty}^{\infty} P(\text{男性}|x)P(x)\,dx}$$

  - 離散変数の場合と同様，分母は分子についてxのあり得る確率の総体を表す
    - これによって，P(x|男性)が確率の定義に収まる
  - 離散だとΣ（総和），連続だと∫（積分）になる
    - さっきの離散分布の場合を思い出そう

$$P(\text{はい}\mid\text{男性}) = \frac{P(\text{男性}\mid\text{はい})P(\text{はい})}{P(\text{男性}\mid\text{はい})P(\text{はい}) + P(\text{男性}\mid\text{いいえ})P(\text{いいえ})}$$

統計モデリングの話

ベイズの定理と統計モデリング

- ベイズの定理はわかった（として）
  - 統計学とどういう関係があるの？

- 統計モデリングにおいて出てくるもの
  - データ
  - パラメータ

  - データをうまく表現できるパラメータが知りたい

統計モデリングとベイズ統計

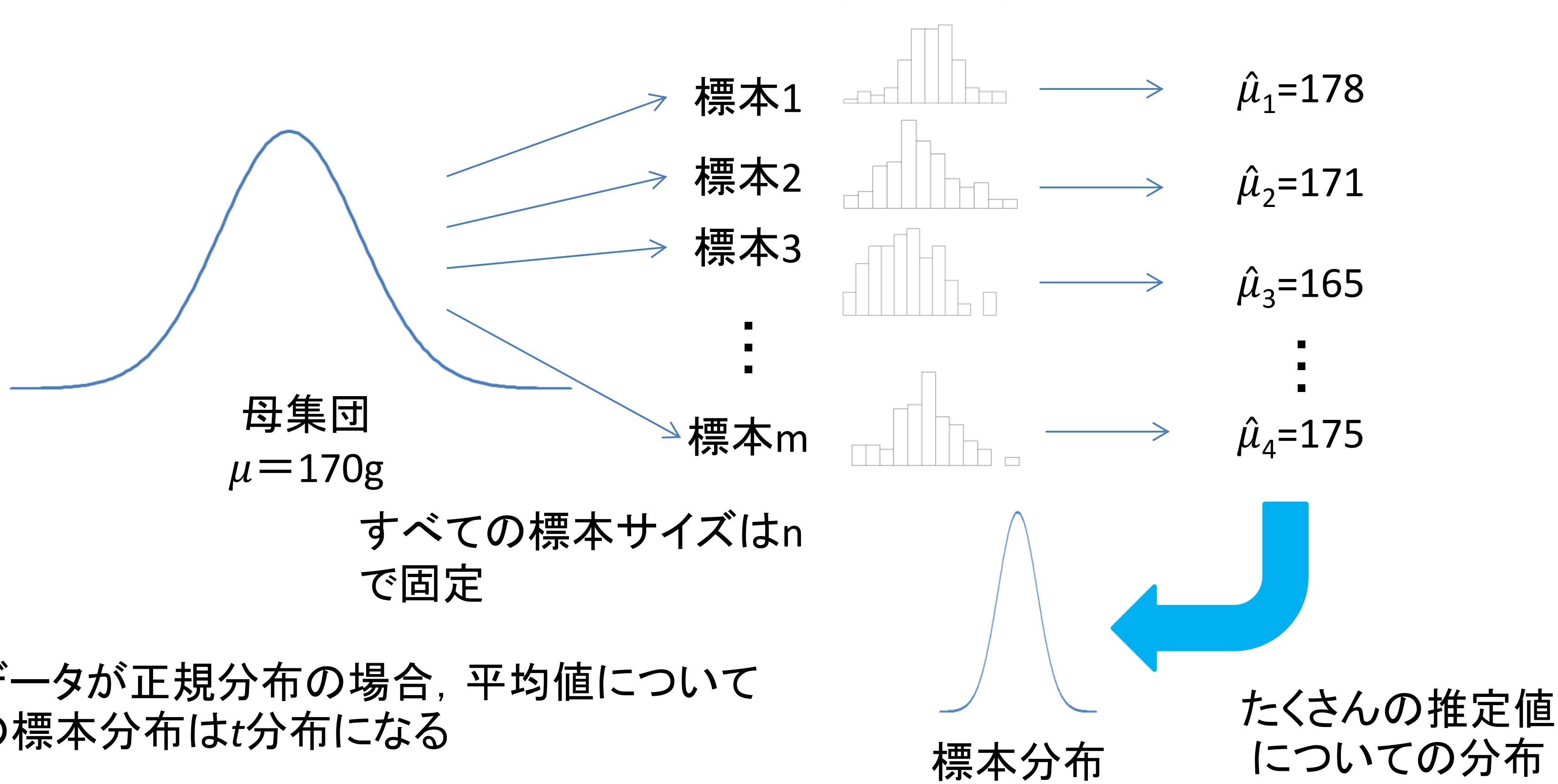
- 統計モデリングでは何の確率が知りたいのか
  - パラメータの確率が知りたい

  - パラメータの確率・・・？

- 頻度主義統計学とベイズ統計学の違い1
  - 頻度主義：パラメータは固定値
  - ベイズ：パラメータは確率分布

頻度主義の考え方：標本分布

- たくさんの標本の推定値についての分布



データが正規分布の場合，平均値についての標本分布はt分布になる

標本分布

- パラメータの分布
  - データの分布ではない

- 頻度主義･･･標本分布の理論
  - 正規分布のパラメータであるμは，固定値
    - 一つの値として決まってるはず　170gとか
  - でも，標本をとることで，ばらつく
    - 標本変動

- 標本分布の評価
  - 一つの標本からμを推定する場合，それはどれくらい標本変動の影響を受けるのか？
  - *t*分布によって推定することができる

標本分布がわかると･･･

- 信頼区間がわかる
  - t分布の2.5，97.5パーセンタイルが95%信頼区間

- 帰無仮説検定ができる
  - （帰無仮説μ=0の場合）0を中心としたt分布において，ある標本の$\hat{\mu}$（より大きい値）が生じる確率がp値

ベイズ統計では

- パラメータは固定値とは考えない
  - 真の値が決まっている，という立場に立たない
  - パラメータを推定する不確実さを表すものとして確率分布を利用

- パラメータμはだいたいこれぐらい
  - データxは，N(μ, σ)に従っている
  - ただ，μもはっきりとはわからず，*t*分布に従うと考える
    - 標本変動というより，不確実さとして分布を使う
    - 主観確率ともいう

知りたいのはパラメータの分布

- データを使って，パラメータを推論する
  - 真のパラメータはよくわからないが，わからないなりに「なんとなく推論」する

- 推測統計にベイズ定理を応用
  - パラメータの確率分布を知るために，ベイズの定理を使う
  - このように，推測統計にベイズの定理を用いる統計学をベイズ統計学と呼ぶ

統計モデリングとベイズの定理

統計モデリング

- 確率分布を何かあてはめて，データを推測
  - ほとんどの場合，心理学では正規分布を使う

- ベイズ統計学を統計モデリングにつかう

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

  - θはパラメータ
  - xはデータ

統計モデリングとしてのベイズの定理

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

- P(θ|x)は・・・
  - xであった場合のθの確率分布
  - xはデータ，θはパラメータなので・・・

得られたデータを条件とした，パラメータの確率分布

統計モデリングとしてのベイズの定理

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

- じゃあ，P(x|θ)とか，P(θ)ってなんだ？
  - データが得られる前のパラメータの分布
  - ちょっと具体的に説明して以降
- 二項分布を使って説明
  - パラメータが一つだけなので説明が簡単

二項分布を例に考える

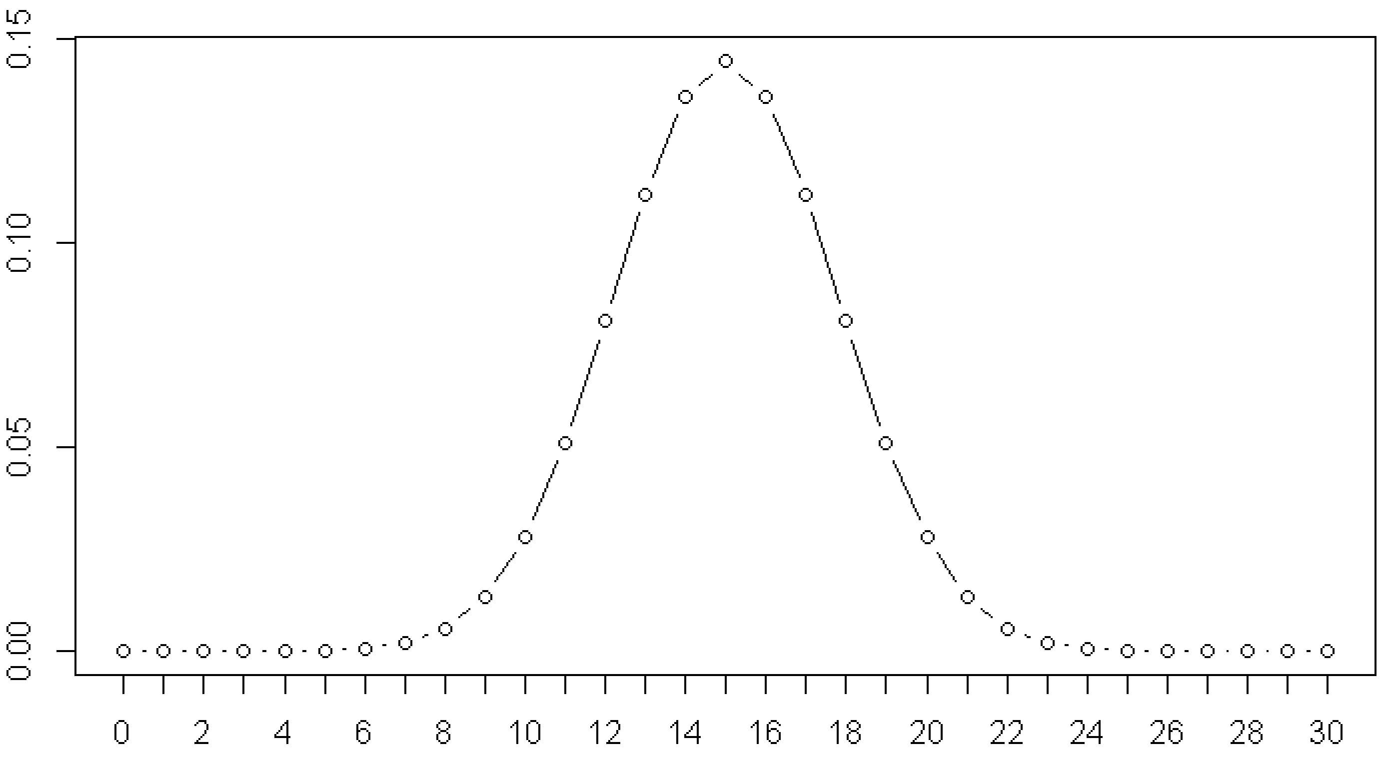
- 二項分布
  - 成功率pで成功する試行をn回行ったとき，x回成功する確率を表す分布
  - Binomial(x | p, n)と表記する
    - ただし，nはここでは固定して考えるのでpのみがパラメータ

$$Binomial(x|p,n) = {}_nC_x(p)^x(1-p)^{n-x}$$

- コイントスを30回繰り返す（表が出る確率は1/2）
  - 成功率0.5，試行数30の二項分布に従う
  - 15回成功する確率が一番高い

二項分布

- Binomial(x|0.5, 30)の分布



成功率pはそもそもわからん

- 実際は成功率pは未知
  - 成功率pを推定したい
  - 実現した値（データ）から，成功率を知る
    - たとえば，30回中15回表が出た，というデータが手に入った

- 頻度主義統計学の推測
  - 母比率を最尤推定する
    - 成功率pの最尤推定値は15/30＝0.5と，簡単に求まる
  - その95%信頼区間を計算する
    - 標本サイズが大きい場合は標準誤差は$\sqrt{\frac{p(1-p)}{n}}$で近似できる
      - 今回は30なのでギリOK

最尤推定法

- 尤度が最大になるパラメータを推定値とする
  - 尤度とは，パラメータを変数とした確率モデル

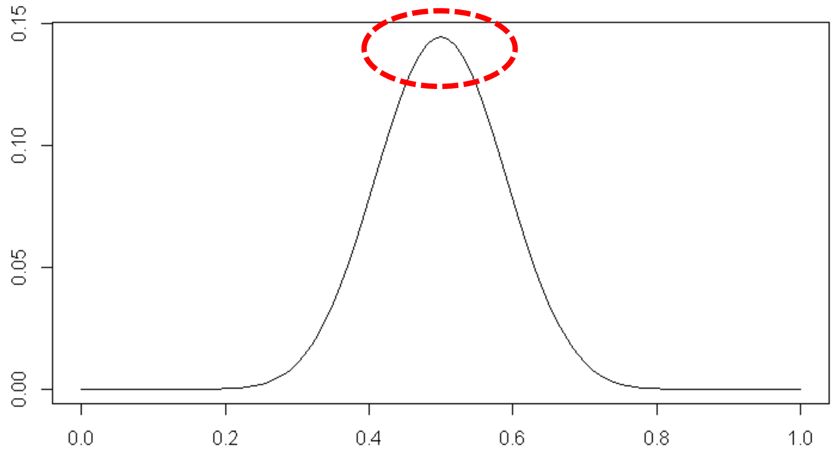
$$Binomial(x|p,n) = {}_nC_x(p)^x(1-p)^{n-x}$$

  - xではなくて，pの関数だと考えると，尤度となる
    - 一見確率分布と同じに見えるが，xが定数でpが変数である点が違う
    - 尤度は確率と違ってpについて積分しても1にならない

- 最尤推定法の考え方
  - 手元にあるデータが最も得られやすいようなパラメータが，真値の推定値としてよいだろう

最尤推定法

- 尤度関数
  - データを固定してパラメータを変化させたときの関数



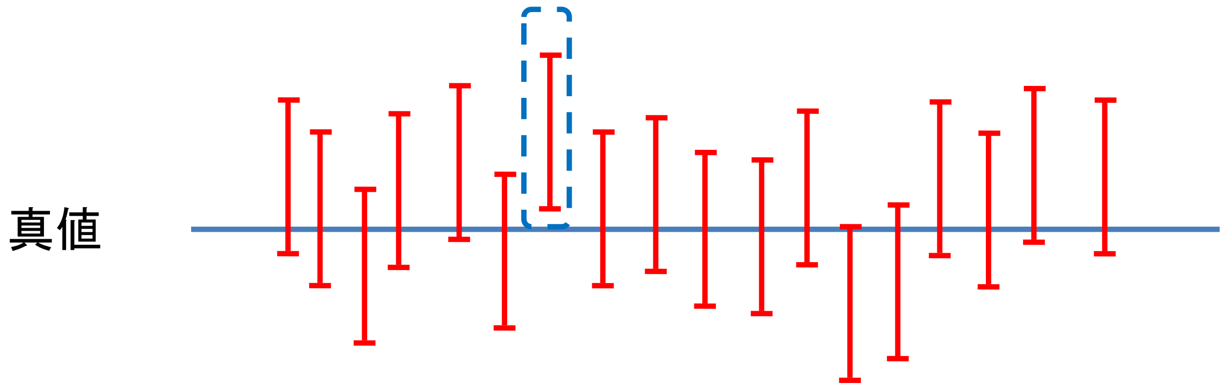
- 最尤推定値
  - 母比率の最尤推定値は，成功数/試行数で求まる
    - つまり，15/30=0.5

95%信頼区間

- 標準誤差を計算
  - 標本推定値($\hat{p}$)がどれくらい標本で変化するか
    - 標本が大きい場合，比率の推定値は正規分布に近似できる
  - 標準誤差 = $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.5(1-0.5)}{30}}$ ≒0.09
- 信頼区間
  - 0.5-0.09*1.96 ~ 0.5+0.09*1.96 ＝ 0.32~0.68

95%信頼区間の解釈

- たくさん標本をとった場合，その範囲に真値が入る確率が95%となるような区間

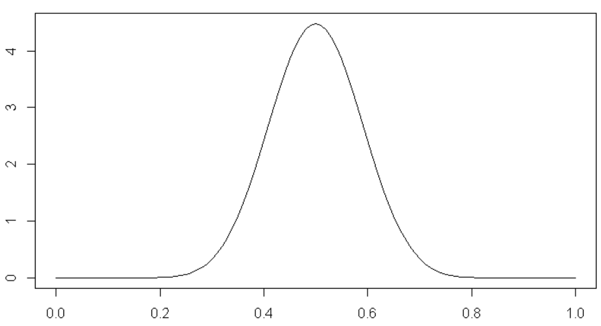


  - 真値は定数なので，「その範囲に95%の確率で真値が含まれる区間」と定義するのは間違い

ベイズ統計による推定

- 成功率pを確率分布として考える
  - たぶん0.5である確率が高そうだけど，もしかしたら0.4とか，場合によっては0.6かもしれない
  - データが有限である以上，成功率は「なんとなく」しか推論できない

- 30回中15回成功した場合の成功率pの分布
  - パラメータの不確実さを確率分布で表現



ベイズの定理で考える

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

- P(x|θ)は・・・
  - パラメータがθの場合の，xが得られる確率
  - つまり，データがどういうパラメータの確率分布に従っているかを表現している
  - つまり，尤度と対応している

- P(x|θ)として何を選択するかが重要
  - 尤度は確率モデルを表している
  - どういう確率モデルを用いるかによって予測の仕方は大きく異なる

→尤度として二項分布の尤度関数を当てはめてみる

二項分布によるベイズ統計モデリング

- 30回中15回成功した場合の成功率pの分布
  - パラメータθが，ここではpとなっているのに注意

$$P(p|x) = \frac{Binomial(x|p,30)P(p)}{P(x)}$$

- Binomial(x|p, 30)は・・・
  - xを所与とした，成功率p，試行数30の二項分布の尤度（関数）を表す
  - 成功率がpの場合の，「30回中15回成功した」というデータが得られる確からしさについての関数

P(p)はなに？

- データに条件づけられていないpの確率分布
  - つまり,「データを得る前のパラメータpの分布」
  - これをパラメータの事前分布と呼ぶ
- 事前の分布・・・？
  - データを得る前にパラメータについて持っている事前の知識を反映したもの
  - 試行前に「コインはだいたい半々で表が出るだろう」という信念があるなら，0.5がやや高めの分布となる
  - これらについては後で詳述

P(x)ってなんだっけ

- データが得られる確率（周辺尤度ともいう）
  - 次の式で計算できる

$$P(x) = \int_{-\infty}^{\infty} P(x|\theta)P(\theta)\, d\theta$$

  - 尤度と事前分布の積を，パラメータθで積分
  - つまり，パラメータがあり得る範囲全部について「尤度と事前分布の積」を足し合わせる
  - パラメータが増えると，この計算が大変になる

- ただし・・・
  - パラメータに依存しないので，定数として得られる
  - （今のところは）尤度と事前分布の積を確率分布に戻す（積分して1になるようにする）ためだけのものと考えてOK

事後分布

- P(p|x)は，
  - データを条件とした成功率pの確率分布
  - データを得た後の分布なので，事後分布と呼ぶ

- ベイズ統計では･･･
  - 尤度と事前分布を使って，事後分布を推定するのが目的

ベイズの定理と統計モデル

- 日本語で書くとこうなる

$$\text{事後分布} = \frac{\text{尤度} \times \text{事前分布}}{\text{データに関する定数}}$$

  - 事後分布∝尤度×事前分布
    - ∝は比例を意味する　分布が同じ形です，ということ
    - 事前のパラメータについての知識に対して，データを統計モデルに当てはめた結果を追加すると，データを加えた後のパラメータの推論となる

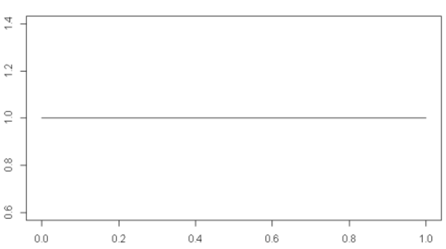
ベイズ統計で出てくるもの

- 事前分布
  - パラメータの事前知識
    - だいたいパラメータはこのあたりだろう，という予測
- 事後分布
  - データを得た後のパラメータについての推論
    - 事前知識とデータから推論した結果
- 尤度（確率モデル）
  - データの生成メカニズムを表す
  - 心理学ではほとんど正規分布が用いられる

ベイズ推定の結果

- 事後分布には尤度と事前分布が必要
  - 尤度は二項分布を選択
  - それによってパラメータは成功率pが決まる

- じゃあ事前分布どうしよう？
  - 事前知識が0の場合
    - 成功率について，あらかじめ知識がないとする
    - あらゆる成功率について一定確率
    - あとで詳しく触れる

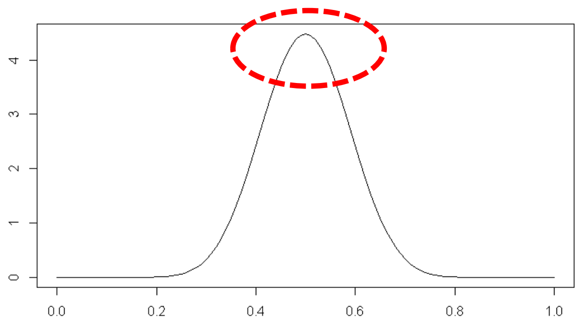


最尤推定とベイズ推定

- 事前分布が一様分布の場合
  - 最尤推定値とベイズ推定におけるMAP推定値は一致する
    - MAP推定値：Maximum a posteriori　事後確率最大値

- なぜか
  - 事前分布が一様分布ということは，尤度の形は変化しない
  - パラメータすべての場合で同じ値をかけるから

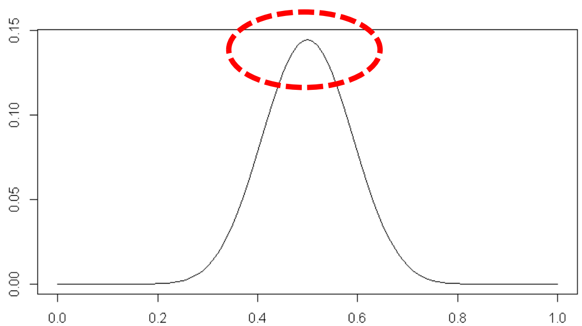
尤度に一様分布をかけても形は同じ



事後分布

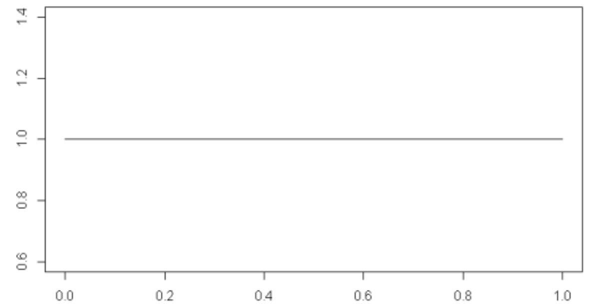
形が同じなので，尤度が最大になるパラメータと，MAPは一致する

$\propto$



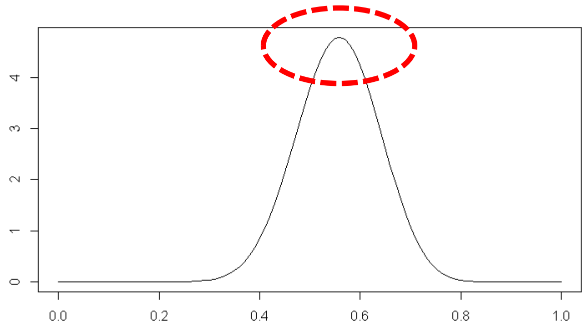
尤度

$\times$



事前分布

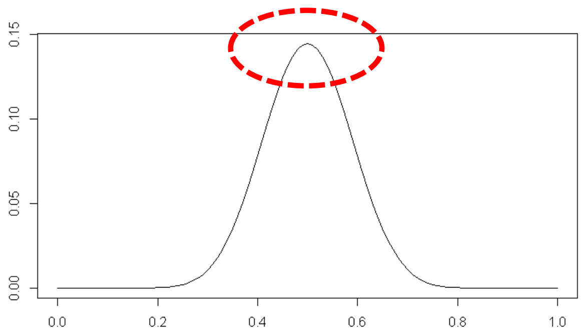
事前分布が一様分布じゃない場合



事後分布

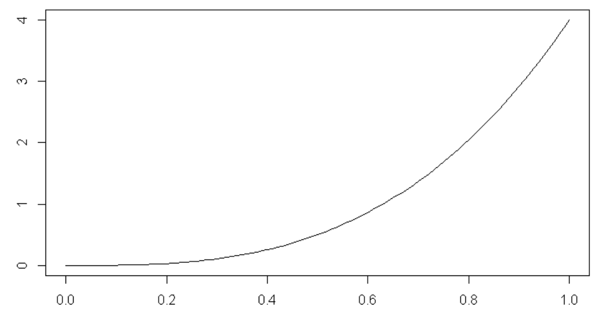
一様分布じゃないと，最尤推定値とMAPは一致しない

$\propto$



尤度

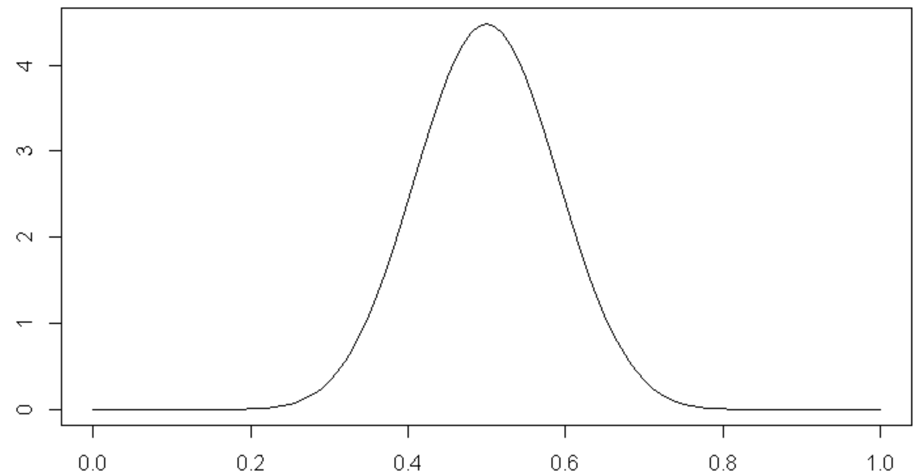
$\times$



事前分布

成功率pの事後分布

- 0.5が一番確率（密度）が高い



- 95%信用区間（credible interval）
  - 今回の成功率pはベータ分布，beta(16,16)に従うので，直接パーセンタイル点を求めることで計算できる
  - 信用区間は，0.33～0.67

最尤推定とベイズ推定

- 最尤法
  - 尤度が最大となる値を推定値として用いる
    - 固定された値
    - 信頼区間は，別の標本で得られうる推定値
      - 確率的に変動するのはあくまでデータ
- ベイズ法
  - 事後分布が推定したいもの
    - 固定値とは考えない
    - 信用区間は，推定値の確率的な範囲

事前分布の話

事前分布の話

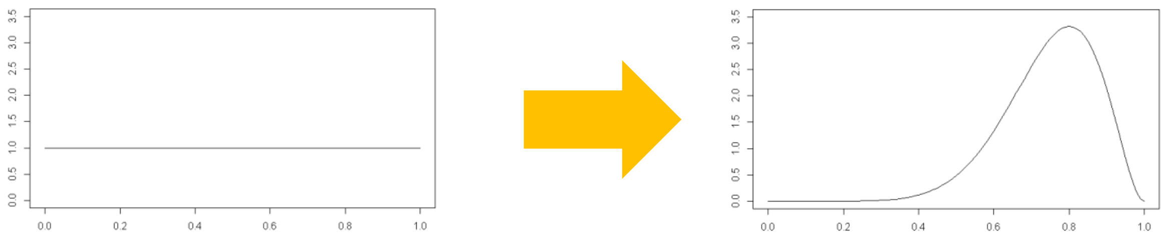
- ベイズ推定には事前分布が登場
  - データを得る前の知識を推論に活用できる

- ただし，恣意的になるという指摘も
  - 事前分布を変えると，結果は変わる
  - 恣意的な事前分布を設定すると，研究者に有利な結果を導くことも可能になる・・・？

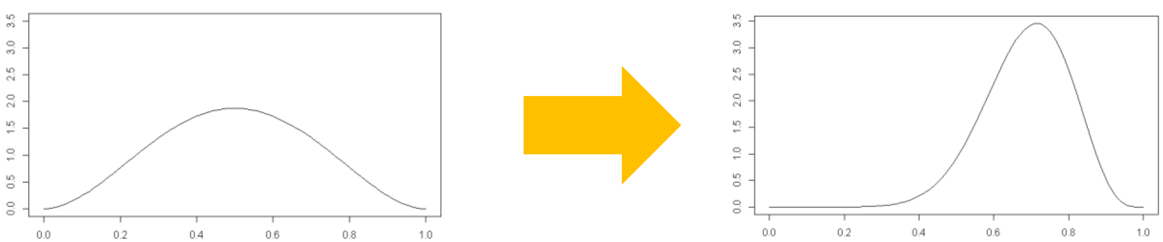
事前分布の違いによる結果の違い

- 10回投げて8回表が出た場合の推論

- 「まったく予想がつかない・・」・・・0.8



- 「コインはだいたい1/2で表になる！」・・・0.71



無情報事前分布

- 事前知識がきわめて少ないことを表す分布
  - まさに無情報な分布
- パラメータが有限区間の場合
  - 成功率のような0～1の範囲のものとか
  - 一様分布でOK
- パラメータが∞区間の場合
  - 正規分布に従う連続量とか
  - 平均0，分散1000の正規分布などを用いる
    - 分散が大きい=情報が少ない
  - パラメータが0～∞の場合
    - 逆ガンマ分布や半コーシーなどが使われる

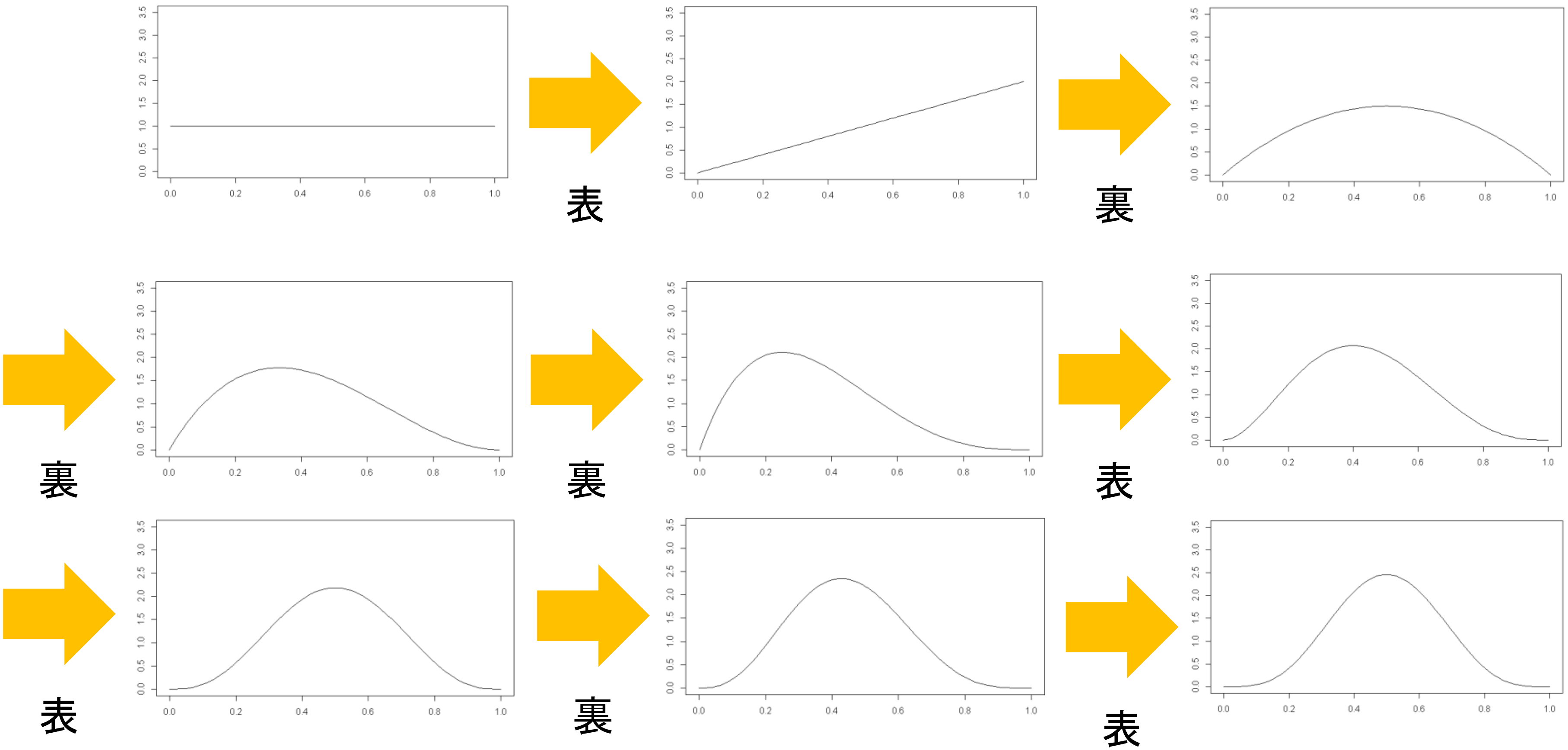
とりあえずの結論

- 特別な理由がなければ一様分布でいい
  - 豊田（2015）
- 理由
  - 最尤法は事前分布が一様分布の時のMAPと一致する
    - 従来の推定量と同じ理解が可能
  - パラメータの種類がどんなものでも使える
  - わかりやすい

ベイズ推定を更新する

- データをどんどん足していくことを考える
  - 事後分布をデータの追加に合わせてどんどん更新していくことができる
  - ベイズ推定の強みの一つ
- 昨日の事後分布は明日の事前分布
  - 事後分布は，データを得た後の分布
  - 次のデータを得る前でもあるので，次のデータから見れば事前分布となる

コイントスをするたびに更新



だんだん分散が小さくなっている → 情報量が増えている

ベイズ統計学と仮説検定

仮説検定

- 仮説を設定して，それが正しいかチェック
  - 頻度主義では，帰無仮説が間違えている＝対立仮説が正しいことを導くという論法
- ベイズ統計では仮説の正しい確率がわかる
  - いくつかの対立する仮説の正しい確率を計算して，それを比較する
  - 自分の研究仮説が正しいことを直接示せる

仮説が正しい確率

- データを所与とした場合の仮説の確率
  - P(H1|x)
    - データxを得た後の，H1が成立する確率

- 遠くにいる人の性別を当てたい
  - 男性か，女性か
  - 事前仮説として，男性であると想定
    - 対立仮説は女性

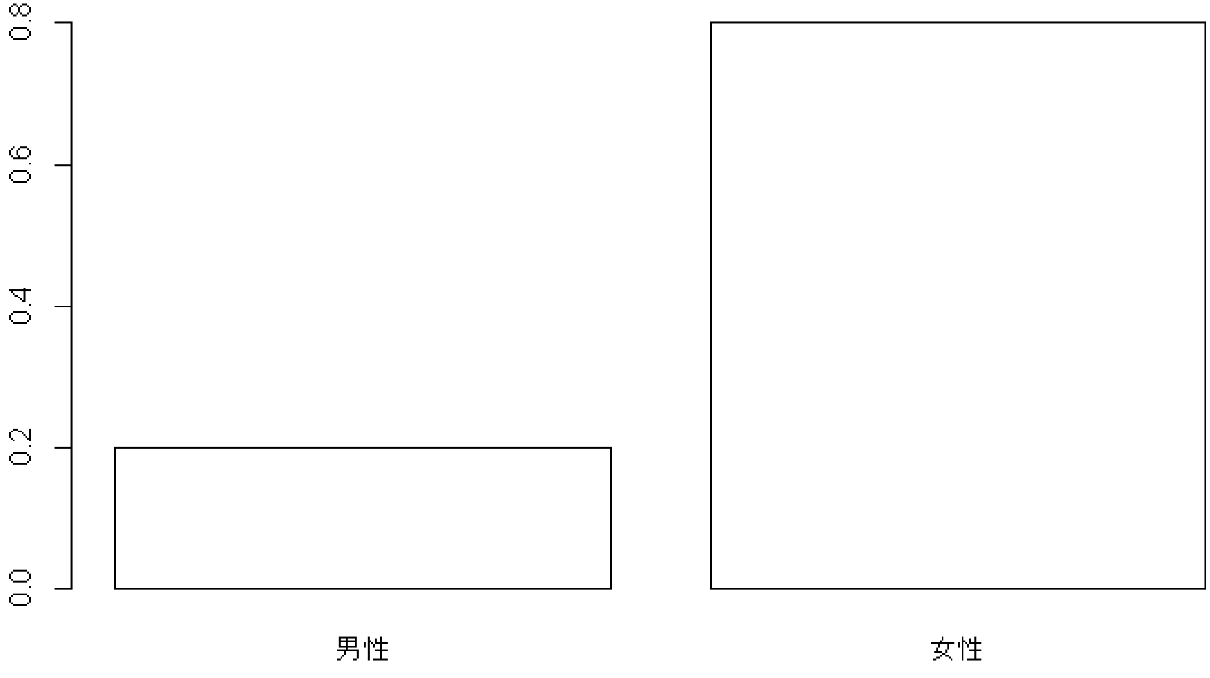
ベイズの定理で仮説の正しさを推論

- なんとなく，髪の毛が長い気がする
  - データ・・・髪の毛が長い
- スカートをはいているっぽい
  - データ・・・スカートをはいている
- ハイヒールもはいってるっぽい
  - データ・・・ハイヒールをはいている

ベイズの定理で仮説の正しさを推論

- 事後確率
  - P（男性|髪長い・スカート・ハイヒール）を求める
- 事前確率
  - P（男性）
  - 男性か女性かは最初はわからないので，0.5
- 尤度
  - P(髪長い・スカート・ハイヒール|男性）
  - 男性の場合，髪が長くてスカートはいてハイヒールを履く確率はかなり低い
- 事後確率
  - 事後確率∝尤度×事前確率
  - 仮説：男性である，の事後確率もかなり小さくなりそう

仮説の正しい確率（事後分布）



仮説を比較

- P（男性|データ）とP（女性|データ）を比較
  - 次の量を考える

$$\text{事後オッズ} = \frac{P(\text{男性} \mid \text{データ})}{P(\text{女性} \mid \text{データ})}$$

  - 男性であるという仮説と，女性であるという仮説の比をとる。
    - オッズとは成功率と失敗率の比のこと
    - 例えば男性仮説の事後確率が0.2で，女性仮説の事後確率が0.8の場合，事後オッズは0.25となる
    - これは，女性であると考えるほうが男性であると考えるより4倍あり得る，ということを意味する

ベイズファクター

- 事後オッズは事前分布に依存する
  - さっきの例では事前知識なしを仮定
  - しかし，道によっては男性のほうがそもそも通りやすいとか，事前知識を活用できる場合もある

- データだけから見た仮説の確からしさ
  - 尤度の比を用いればよい
  - これをベイズファクターという

$$\text{ベイズファクター} = \frac{P(\text{データ} \mid \text{男性})}{P(\text{データ} \mid \text{女性})}$$

一般的な表現

- 事後オッズ
  - データを得た後の仮説の確からしさ

$$\text{事後オッズ} = \frac{P(\text{仮説 }1|x)}{P(\text{仮説 }0|x)}$$

- ベイズファクター
  - どちらの仮説の場合に今回のデータが得られやすいか

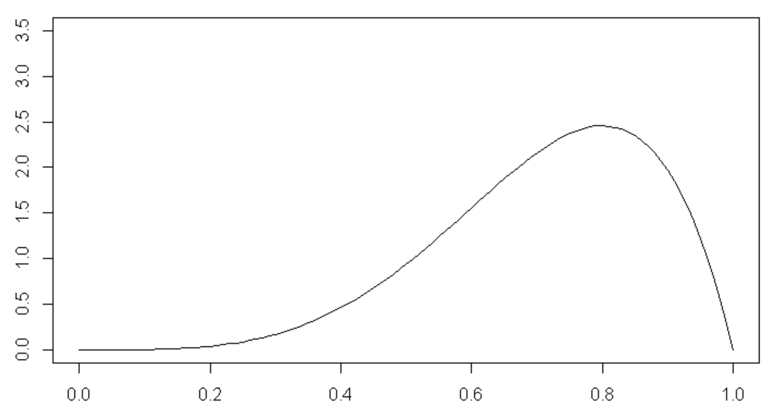
$$\text{ベイズファクター} = \frac{P(x|\text{ 仮説 }1)}{P(x|\text{ 仮説 }0)}$$

事後オッズとベイズファクター

- ベイズファクターは次式で計算できる
  - ベイズファクター ＝ 事後オッズ/事前オッズ
    - 事前オッズ･･･事前確率のオッズ
  - データを得ることで，事後オッズが事前オッズに比べてどれほど変化したかを表す量といえる
- 事前オッズが1の場合
  - データを手に入れる前の事前知識がない場合
  - ベイズファクター＝事後オッズ となる

ベイズファクターの実際

- ベイズファクターは計算が大変
  - 今回の例のように，パラメータが離散的な場合（男性，女性というように）は簡単に計算ができる
  - 連続変量の場合
    - 積分が必要
    - p >=0.5とp<0.5を比較する・・・など



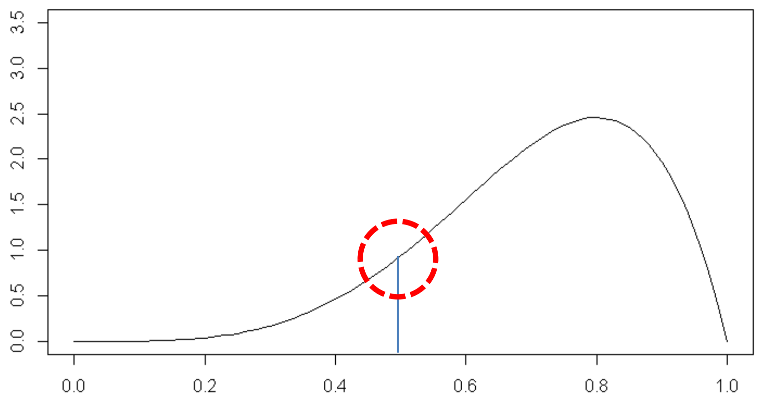
定点と範囲の仮説を比較する

- 5回コインを投げて4回表が出た
  - 表が出る確率pについての仮説を立てる
  - $p > 0.5$
    - 半々よりも表が出やすいコインだ
    - 範囲による仮説（厳密には$0.5<p<1$）
  - $p = 0.5$
    - コインは半々で表が出る
    - 定点的な仮説

- 構造的には帰無仮説検定と同じ
  - 定点的な仮説と範囲的な仮説は比較が難しい

ベイズファクターの実際

- 定点的な仮説が成立する確率
  - パラメータが特定の値の場合の確率は，直接は計算しにくい
    - 確率密度であって，確率ではない
    - 定点同士なら，確率密度の比として比較はできるが・・



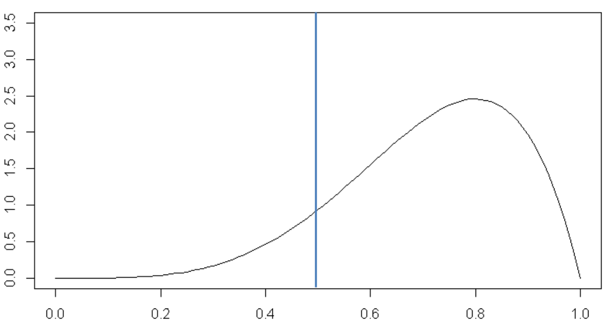
情報仮説の評価

- 情報仮説
  - 不等号によって表現された範囲的な仮説
  - p>0.5や，μ1＞μ2など

- 無情報仮説
  - パラメータについて制約を置かない仮説
  - 帰無仮説と異なり，特定の値であることを仮定しない
  - 無情報仮説に比べて，情報仮説が成立する確率がどれほど大きいかを調べることで仮説を検証する

情報仮説の評価

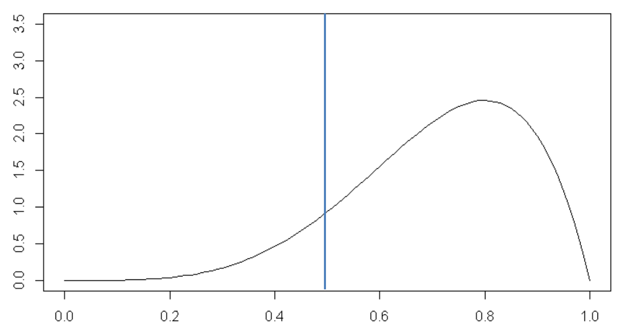
- 無情報仮説と情報仮説を比較
  - p＞0.5の仮説は，制約がない仮説に比べて何倍確からしいか
  - p＞0.5の確率は，事後分布から0.89
    - 仮説が予測する範囲は0～1の範囲のうち，半分
      - 仮説の複雑さで重みづける必要がある
    - 無情報仮説に対するBFは0.89/0.5＝1.78



情報がないなら，一様分布となり，「パラメータpは0.5以上になる確率」は0.5になる
0.5に比べて0.89は1.78倍高い

補仮説との比較

- 仮説が正しくない，という仮説の確率
  - p>0.5の確率は0.89
    - 積分で計算
  - その仮説が正しくない確率（p<=0.5) は(1-0.89)
  - BF = $\dfrac{0.89/0.5}{(1-0.89)/(1-0.5)}$ = 8.09
  - P>0.5という仮説が正しい確率は，それが間違えている確率に比べて8倍ありえる



補仮説の確率は，パラメータpが0.5より小さくなる場合の面積に対応する（1-0.89=0.11）。
0.11にくらべて0.89は8倍高い

マルコフ連鎖モンテカルロ法

ベイズ推定って実際どうやるの？

- コイントスの例
  - 二項分布を使った簡単な例
  - パラメータが一つで，分布も簡単なので解析的に事後分布が求まる

- 一般的な話
  - 実は，ほとんどのモデルは解析的には解けない
  - 数値計算でも難しい

ベイズ推定の何が難しいのか

- またもやベイズの定理

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

  – 尤度は確率モデルが決まれば計算できる
  – 事前分布も，一様分布なら特に問題はない

  – P(x)が問題

P(x)ってなんだっけ

- データが得られる確率（周辺尤度ともいう）
  - 次の式で計算できる

$$P(x) = \int_{-\infty}^{\infty} P(x|\theta)P(\theta)\,d\theta$$

  - 尤度と事前分布の積を，パラメータθで積分
  - つまり，パラメータがあり得る範囲全部について「尤度と事前分布の積」を足し合わせる

  - パラメータが増えると，この計算が不可能になる

じゃあどうしよう・・・？

- 積分を避ける方法はないものか・・・
  - シミュレーションを用いれば・・・

- マルコフ連鎖モンテカルロ法
  - MCMC法とか呼ばれる
  - マルコフ連鎖による乱数を用いたモンテカルロ法

  - 細かい原理は，今回はパス

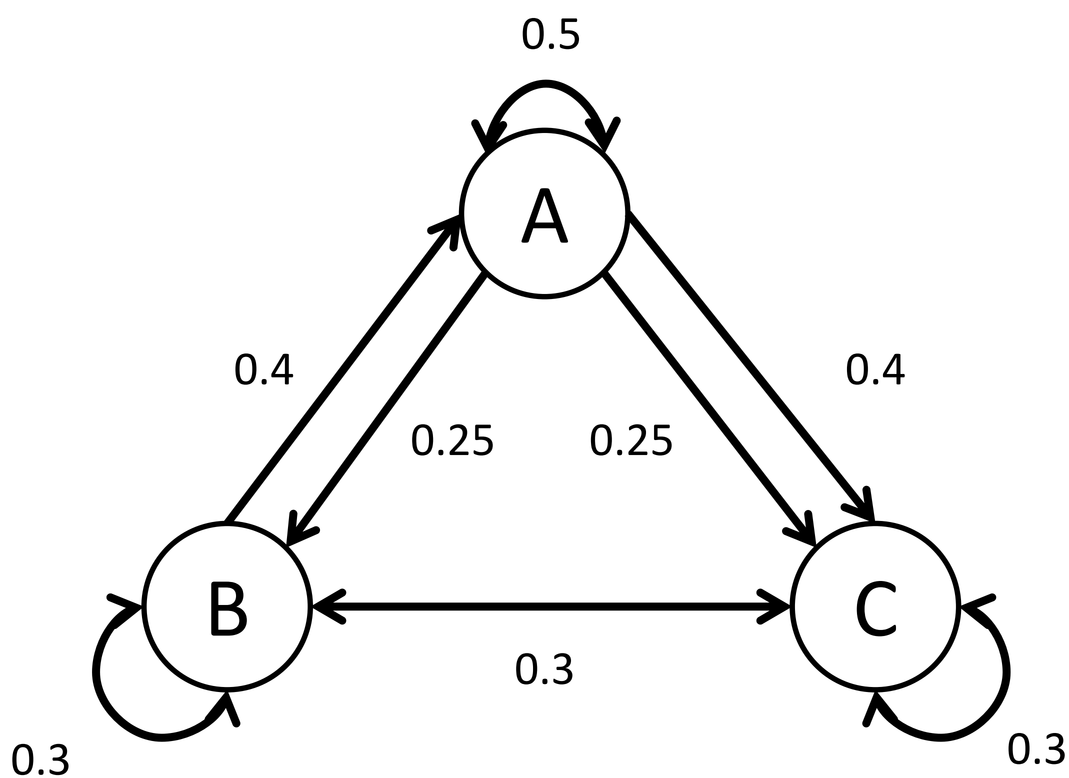
マルコフ連鎖って何よ

- ある状態から，別の状態に推移する確率だけが決まっている

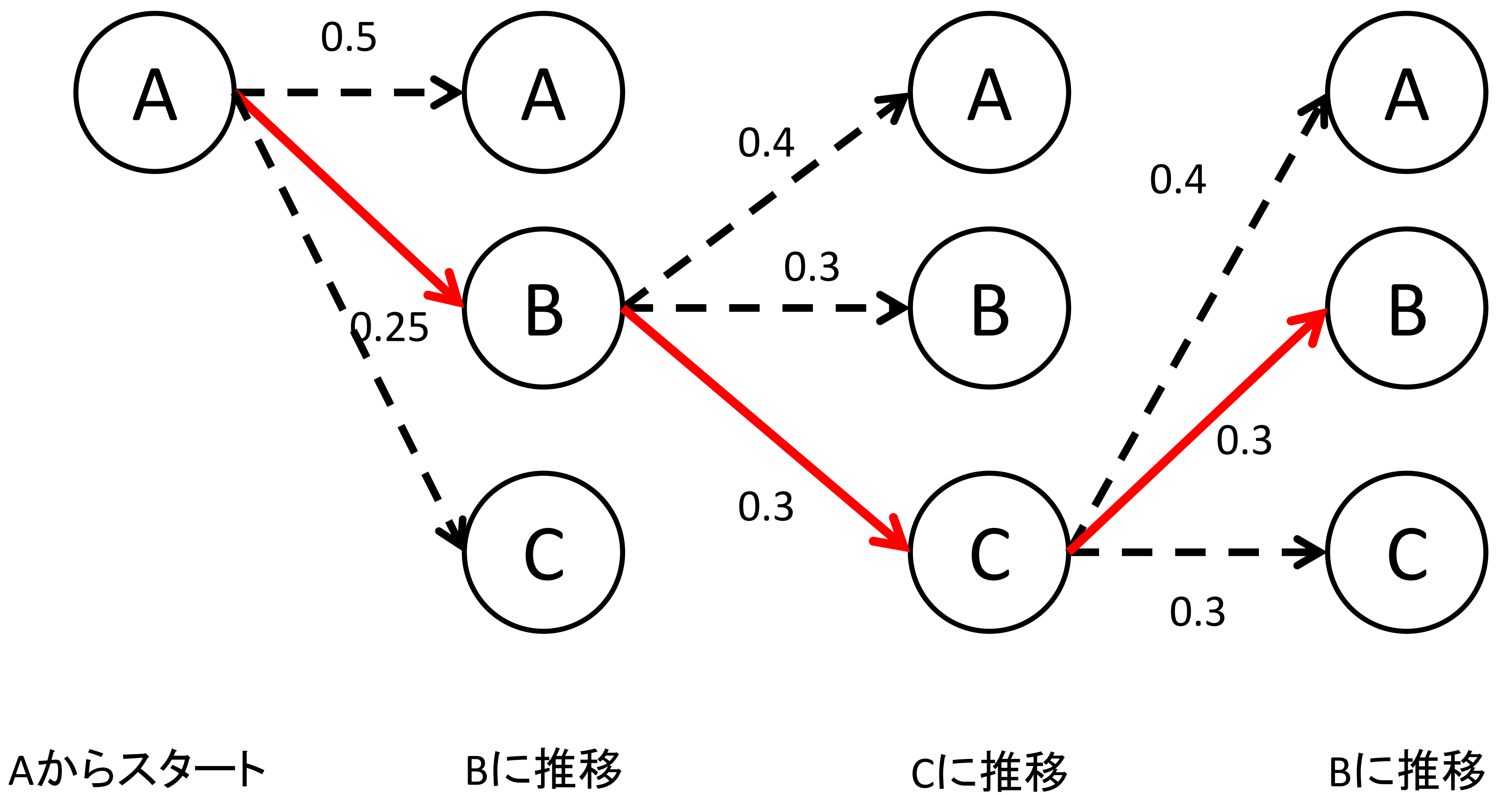
	A	B	C
A	0.5	0.25	0.25
B	0.4	0.3	0.3
C	0.4	0.3	0.3

- 推移確率行列
  - AからBに移行する確率は0.25
  - BからAは0.4
  - CからCは0.3

マルコフ連鎖



マルコフ連鎖



マルコフ連鎖

- 初期分布
  - 最初にA,B,Cのどの状態にあるかを決めるもの
  - たとえば1/3，1/3，1/3
- 推移確率に従って状態を変えていく
  - 2時点目においてAである確率は，
  - 初期分布×推移確率行列
  - というように，推移確率行列をかけていくことで計算ができる

定常分布

- マルコフ連鎖を続けていると･･･
  - 各状態にたどり着く確率は変化しなくなってくる
  - それを定常分布という

- 定常分布は推移確率行列のみから決まる
  - つまり，初期分布には依存しない
  - ついでに先ほどの推移確率行列の場合，
  - A=8/18，B＝5/18，C=5/18という定常分布となる

連続分布でも同様に

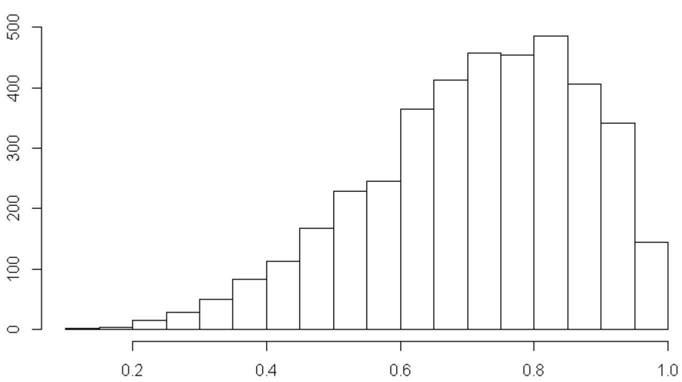
- 連続分布でも定常分布が存在する
  - いくつかの条件がいるが･･･
  - MCMCを使う場合は，この条件を満たす

- MCMCは何をやっているか
  - 事後分布が定常分布となるようなマルコフ連鎖を発生させる
  - 各パラメータが推移する確率はいずれ収束し，それが定常分布＝事後分布となる

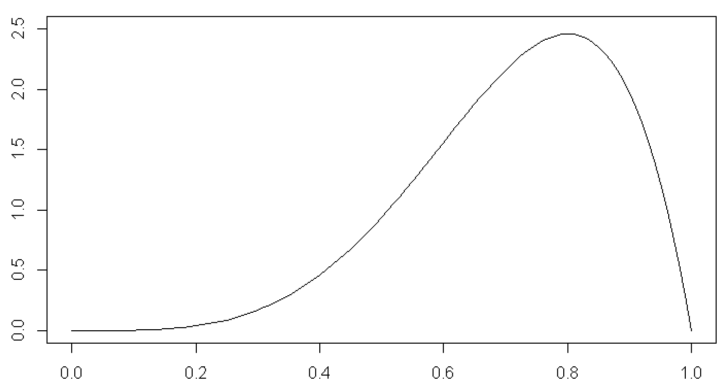
MCMC法は何をするのか

- パラメータの推定値をたくさん計算する
  - 1000とか2000とか
  - 推定値をどんどんマルコフ連鎖で生成する

- 推定値の集合体が事後分布になる



MCMCによる事後分布

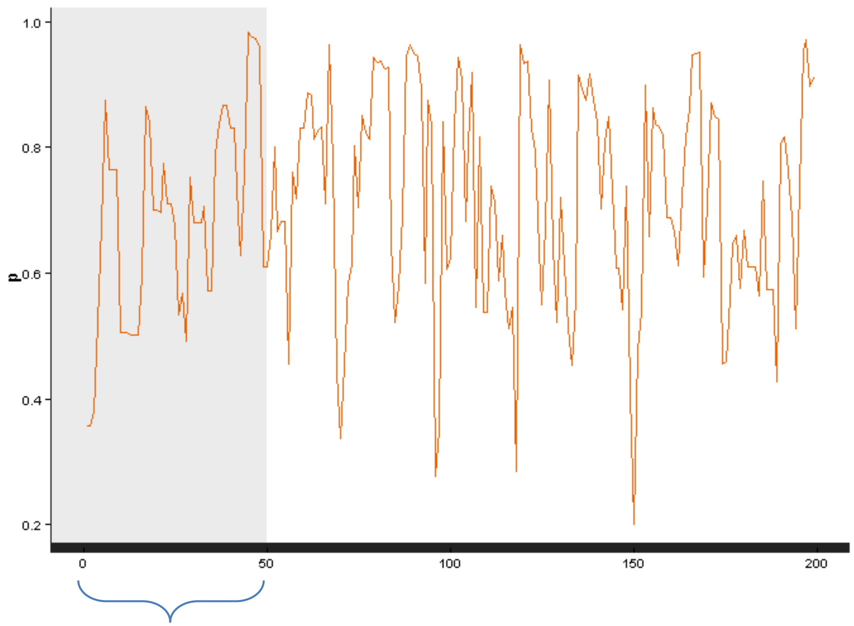


解析的に求めた事後分布

MCMCのイメージ

初期値　　　　　　　　　　マルコフ連鎖に従って乱数が生成

0.38 ⟶ 0.87 ⟶ 0.76 ⟶ 0.75 ⟶ 0.45 - →

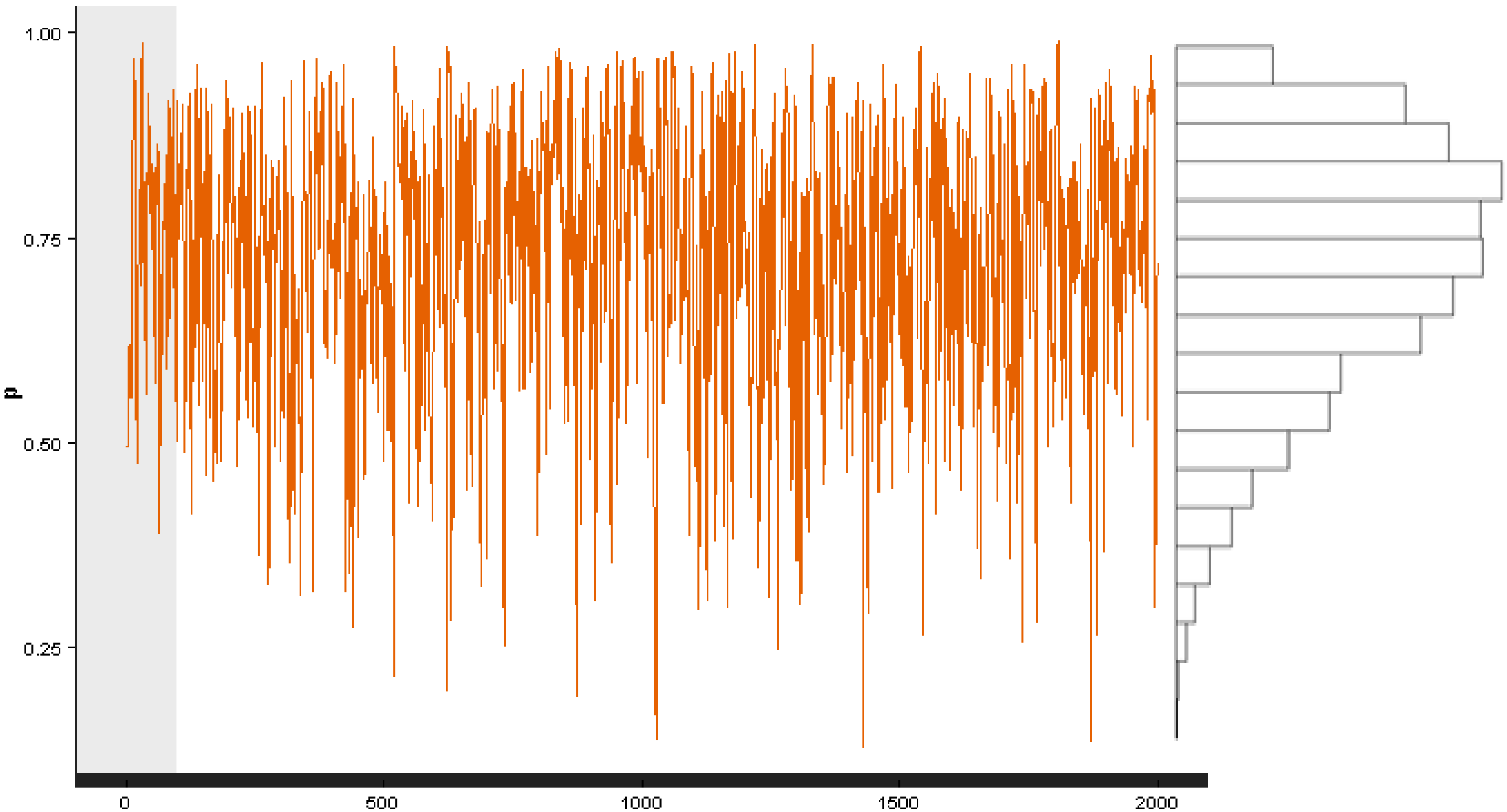


初期値から順番に，推定された値に対応したマルコフ連鎖の推移確率行列に従ってどんどん値が決まっていく

最初のほうのパラメータは初期値に依存してしまうので，
切り捨てることが多い→バーンイン期間という

2000回走らせた場合

- バーンイン期間として100個を捨てた
  - 定常分布に収束する



事後分布の要約

- 普通に平均値を計算すればよい
  - Rだとmean()を使えばすぐ出せる
  - 平均値0.712
- 標準誤差はSDを計算すればよい
  - sd()をつかう
  - SDは0.164
- 95%信頼区間
  - パーセンタイル点を計算する
  - quantile(object, probs=c(0.025,0.975))を使う
  - 0.352~0.957

いつになったら収束するのか

- 複数のマルコフ連鎖を走らせる
  - 違う初期値から始まったマルコフ連鎖を比較して，収束を判断する
- $\hat{R}$で判断（stanなどの場合）
  - 「あーるはっと」と読む
  - 1になれば収束。1.05以下ぐらいでOKと判断する

MCMCを実行するソフトウェア

MCMCを実行するフリーソフト

- WinBugs
  - けっこう昔からあるソフト
  - R上でも動くパッケージがある
- Stan
  - 比較的最近のフリーソフトで，今勢いがある
  - R上でも動く
- JAGS
  - Bugsに似た文法の，比較的最近のフリーソフト
  - R上でも動く

有償ソフト

- Mplus
  - 構造方程式モデル用のプログラム
  - ベイズ推定ができる

- Amos
  - 同じく構造方程式モデル用のプログラム
  - ベイズ推定ができる

- SAS
  - PROC　MCMCがある
  - ただ，University Editionだと無償で使える

今回使うのは・・・

- Stan
  - フリーで使える
  - Rで使える
  - 開発に勢いがある
    - 変分ベイズを使った推定もできるようになった
  - 清水が最初に使ったのがStanだった

rstanパッケージをインストール

- 2.9.0が最新バージョン
  - 2016年1月29日時点
- 使い方については今日話しません
  - それだけで1日終わる
  - もし興味あればまた呼んでください
- 今日は簡単なstanコードだけ走らせます

平均値をMCMCで推定してみる

平均値の推定

- といっても，実はいろんなやり方がある
  - まず分布を決めないといけない

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

  - P(x|θ)は確率モデルを表す（尤度の部分）
  - ここにどういう確率分布を仮定するかで，事後分布はかわってくる

とりま，正規分布で

- 正規分布

$$Normal(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

  – パラメータは平均値μと標準偏差σ

データ

- Rで正規乱数を発生
  - y <- rnorm(100,5,1) #平均5，SD1の乱数を100個
- Stanコードを書く("normal.stan")

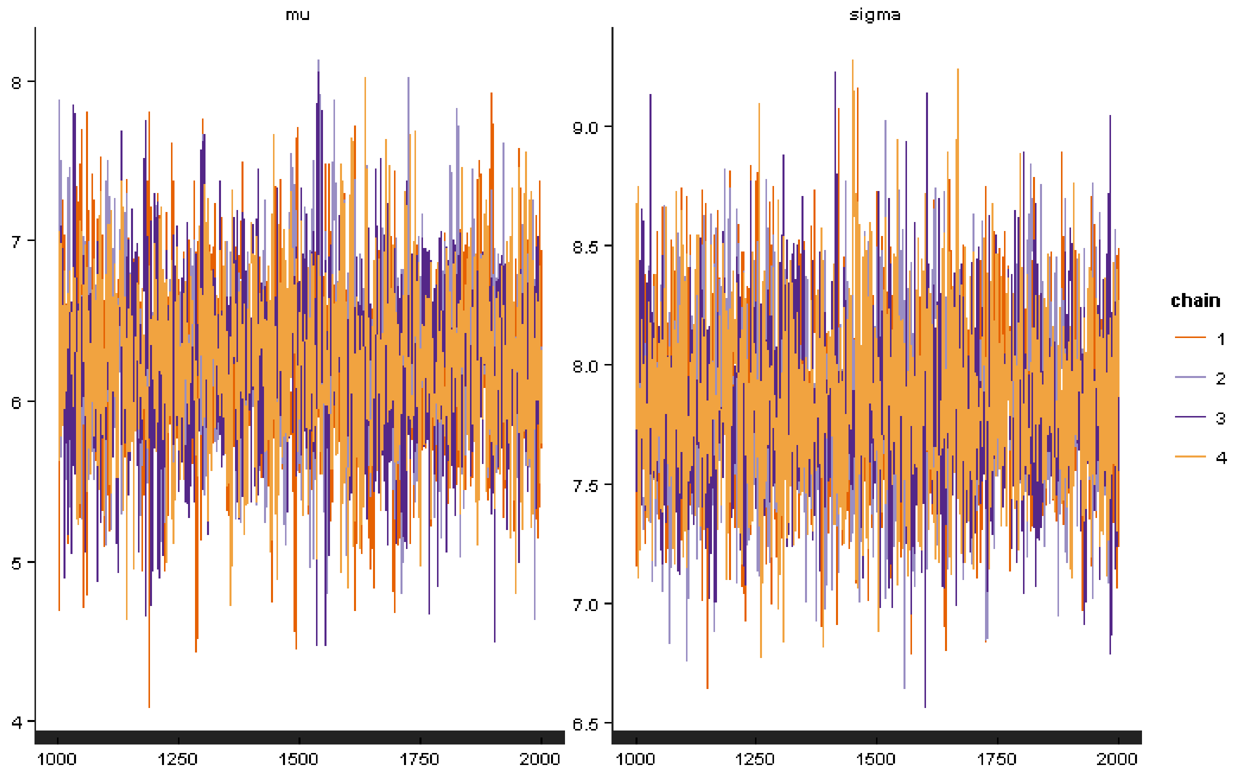
```
data{
  real y[100];
}
parameters{
  real mu;
  real<lower=0> sigma;
}
model{
  y ~ normal(mu,sigma);
}
```

- Stanを走らせる

  >fit <- stan(file="normal.stan",data=list(y=y))

MCMCで推定

>traceplot(fit)



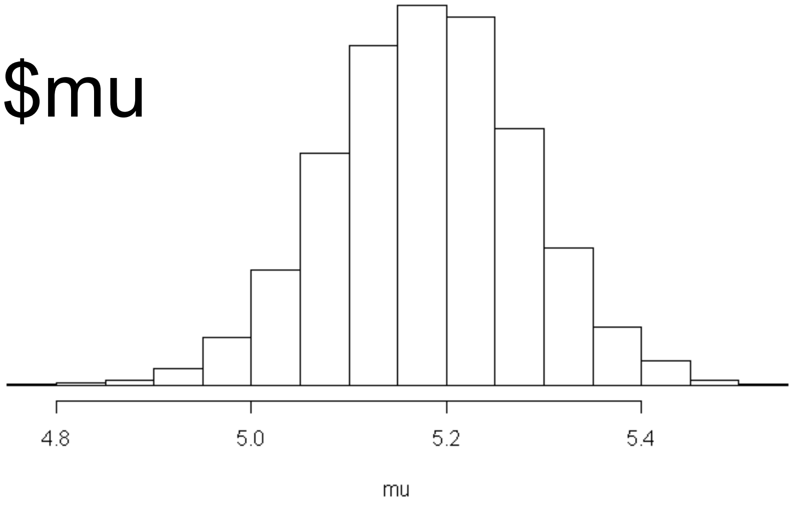
推定結果

- stanの出力

	mean	se_mean	sd	2.5%	50%	97.5%	n_eff	Rhat
mu	5.18	0	0.10	4.98	5.18	5.38	2196	1
sigma	1.03	0	0.07	0.90	1.02	1.19	2351	1

- 平均値の事後分布

```
>mu <- rstan::extract(fit)$mu
>hist(mu)
```



要約統計量

- 平均値
  >mean(mu)
  5.1802

- 95%信用区間
  >quantile(mu, probs=c(0.025,0.975))
  4.984144 5.377316

- モード（事後確率最大値）
  >density(mu)$x[which.max(density(mu)$y)]
  5.19473

平均値の差の推定

二群の平均値の差の検定

- いわゆる*t*検定
  - ただし，*t*検定は二群の分散が等しいという仮定がある
  - Welchの検定を用いて補正する必要がある

- ベイズなら，そんな仮定はいらない
  - 分散が違っても，それぞれモデリングできる
  - なんなら分布は正規分布じゃなくてもいい

データ

- Rで二つのデータを作る
  - 母数がN(7,6)のグループと，N(6,2)のグループ
  - x <- rnorm(100,7,6)
  - y <- rnorm(100,6,2)

- 平均値
  - x:平均=7.54　SD=5.95
  - y:平均=6.21　SD=2.14

普通にWelchの検定

>t.test(x,y)

t = 2.1043, df = 124.21, p-value = 0.03737

– 5%水準で有意

- 続いてMCMCでやってみる

stanコード

```
data{
  real x[100];
  real y[100];
}
parameters{
  real mu_x;
  real mu_y;
  real<lower=0> sigma_x;
  real<lower=0> sigma_y;
}
model{
  x ~ normal(mu_x,sigma_x);
  y ~ normal(mu_y,sigma_y);
}
generated quantities{
  real diff;
  diff <- mu_x - mu_y;
}
```

二つの群は異なる分散の正規分布から発生したと仮定すれば，標準誤差や自由度の補正などは不要

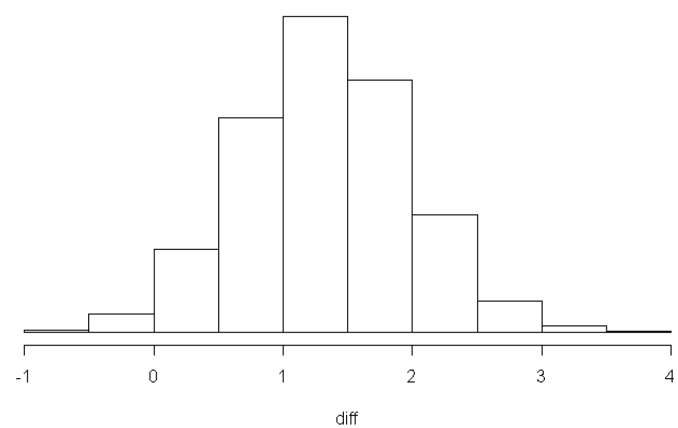
なお，sigma_xとsigma_yを同じパラメータで推定すれば，*t*検定と全く同じ結果が得られる

結果

- Stanの出力

	mean	se_mean	sd	2.5%	50%	97.5%	n_eff	Rhat
mu_x	7.54	0.01	0.60	6.36	7.53	8.70	2762	1
sigma_x	6.03	0.01	0.43	5.25	6.01	6.91	3096	1
mu_y	6.21	0.00	0.22	5.78	6.21	6.64	3266	1
sigma_y	2.17	0.00	0.16	1.89	2.16	2.52	2711	1
diff	1.33	0.01	0.65	0.07	1.33	2.60	2865	1
lp__	-353.08	0.04	1.40	-356.62	-352.77	-351.31	1462	1

- 差の事後分布
  - 平均1.33[0.07, 2.60]
  - 差があると見てよさそう



せっかくなのでベイズファクター

- ベイズファクターは･･･
  - 他の仮説よりも，自分の仮説がどれほど確からしいかを知ることができる
  - ただし，帰無仮説検定のように，定点的な仮説(μ=0のような)とは比較が難しい

- 情報仮説による評価をしてみる
  - $\mu_x > \mu_y$という仮説が正しい確率が知りたい
  - この仮説が，他の仮説と比べてどれくらい確からしいか知りたい

$\mu_x > \mu_y$の確率

- xの平均値ほうがyの平均値より大きい確率

  ```
  >diff <- rstan::extract(fit)$diff
  >length(diff[diff>0]) / length(diff)
  ```

  - Diffの数全体に対する，0より大きいdffの数の比

  0.98025

  →今回のデータから考えて，98%は正しい

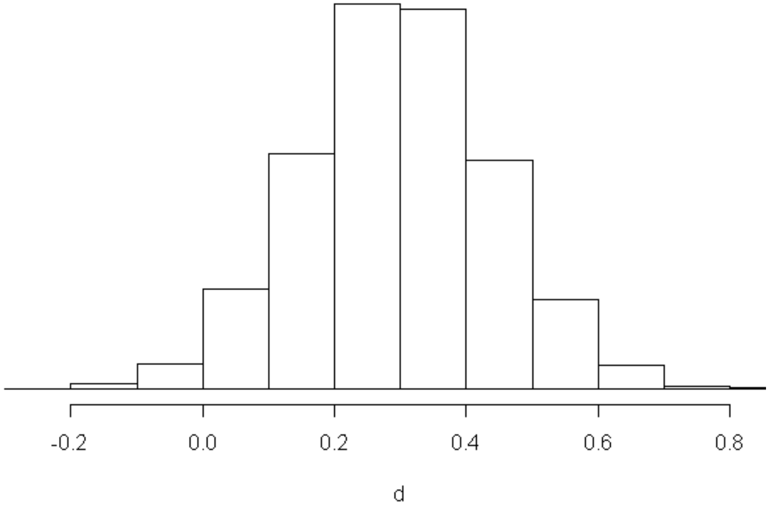
- xがyより大きいという仮説のベイズファクター
  - $\mathrm{BF} = \dfrac{0.98 / 0.5}{(1-0.98) / (1-0.5)} = 49$
  - xがyより大きい，という仮説はその逆の仮説より49倍あり得る
  - 無情報仮説に対しては，0.98/0.5 = 1.96倍

効果量dも計算してみる

- Cohenのdの計算
  >sigma_p <- (99*(var(x)+var(y))) / 198
  >d <- diff / sigma_p^0.5

- dの平均値
  >mean(d)
  0.296791

- dの事後分布
  >hist(d)



MCMCで回帰分析

一般化線形混合モデルをMCMCで

- stanコードを書かなくても，自動で走らせてくれる

- rstanarm
  - stanチームが作成したパッケージ

- brms
  - lme4のベイズ版を目指したパッケージ

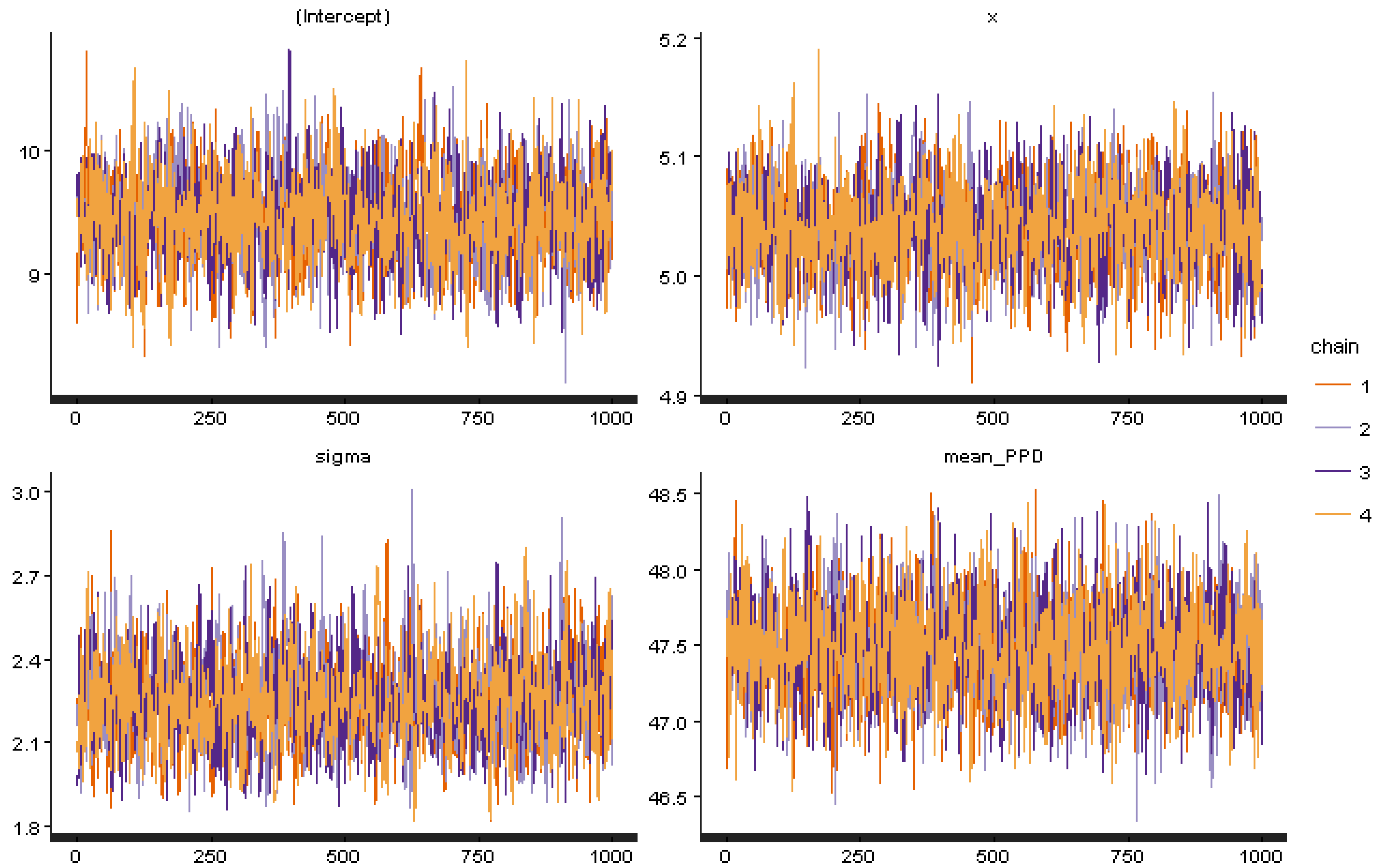
- glmmstan
  - 清水が作ったパッケージ
  - glmmstanでググってもらえると解説スライドが出てきます

rstanarmを使ってみる

```
>library(rstanarm)
>z <- 10 + x*5 + rnorm(100,0,2)
>fit.reg <- stan_glm(z~x)
```

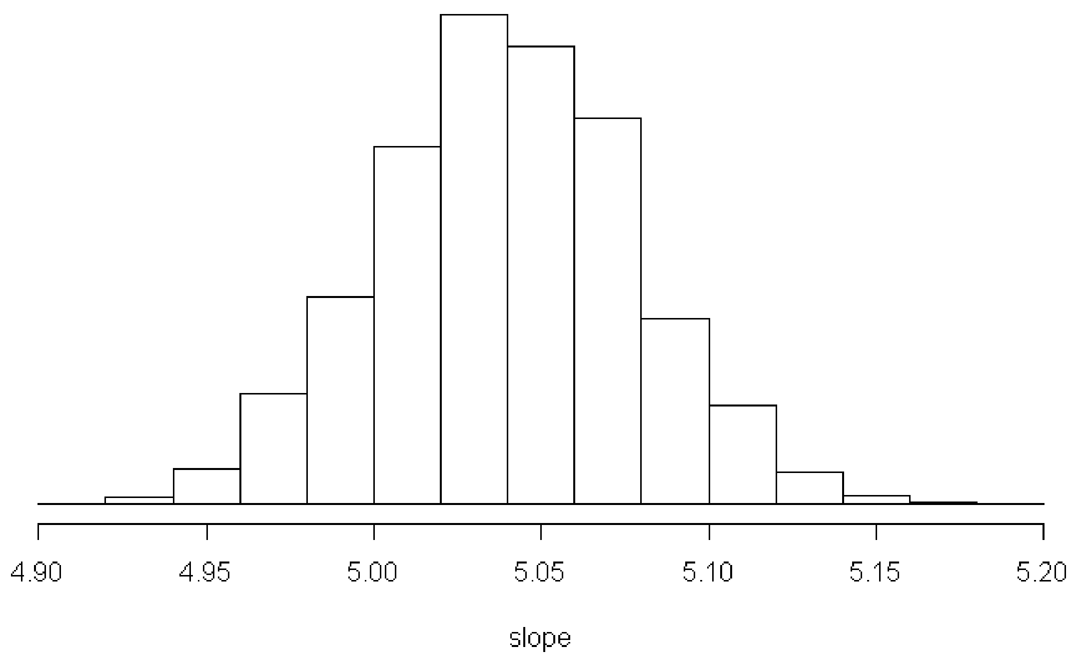
```
Estimates:
                 mean     sd       2.5%     97.5%
(Intercept)        9.463    0.372    8.735   10.209
x                  5.040    0.038    4.965    5.112
sigma              2.262    0.165    1.966    2.612
mean_PPD          47.485    0.324   46.839   48.121
log-posterior -132.002    1.314 -135.529 -130.560
```

マルコフ連鎖の様子


(Intercept)
x
sigma
mean_PPD
chain
1
2
3
4
10
9
5.2
5.1
5.0
4.9
3.0
2.7
2.4
2.1
1.8
48.5
48.0
47.5
47.0
46.5
0
250
500
750
1000

事後分布

```
>slope <- as.data.frame(fit.reg)$x
>hist(slope)
```



それ以外のモデル

- stan_glmer()を使えば・・・
  - 一般化線形混合モデルが実行可能
  - いろんな分布，いろんな変量効果を組み合わせて実行することができる

まとめ

心理学におけるベイズ統計

- 心理学でなぜベイズ統計なのか
  - これまでの心理統計の限界
  - 再現性の問題
  - 複雑なモデルを使う必要性

- ありえる，二種類のモチベーション
  - 検定ではなく，仮説を直接評価したい！
  - 最尤法では推定が難しいモデルを推定したい！

仮説評価とベイズ統計

- ベイズファクター
  - 検定に代わる新しい評価基準になりえる
  - ただし，ベイズファクターを直接計算するのは難しいので，今後の発展が待たれるところ

- 情報仮説の評価
  - 比較的簡単に計算できるが・・・
  - 複雑な仮説やモデルには対応しにくい
  - 評価基準もまだコンセンサスはできてなさそう

最尤推定からベイズ推定へ

- 調査系の心理学者は必要性があるかも
  - ただし，最尤法で解くのが難しいモデルは，心理学ではまだまだ使われていない
    - ロジスティックHLMとか，そのあたりはありえるか
- 常にベイズ推定でもいい
  - 最尤法で解けても，ベイズ推定のほうが便利なことはある
  - パラメータの差の評価や，間接効果も簡単
  - 無理な補正などが必要ない

今のところの結論

- 別に今すぐベイズじゃなくていい（
  - 明日できないと困るものではない
  - ベイズ統計じゃないとできないことは，まだまだマニアックな範囲かもしれない
- しかし，確実にその日は来る
  - 今のうちにベイズ統計に親しんでおくと，いずれ使い道が出てきたり，査読で指摘される日も
  - 技術は多く持っていて困ることはない