2017.09.22

日本心理学会第81回大会TWS

ベイジアンになると何がどう変わるのか

ベイジアンデータ解析入門

話す人：小杉考司（山口大学教育学部）

ベイジアンになると？

- ベイズ統計を使うと今までの「仮説検定」「実験計画」の考え方がどのように変わるのか，を解説します。
- ポイントは次の二点だけです。
  - 「**点**」から「**幅**」へ
  - 「**ないない**」から「**あるある**」へ

お話の流れ

- 頻度主義の考え方とベイジアンの考え方
- ベイズ流の帰無仮説検定はどのように行うか
- 従来の仮説検定の大問題とベイズ流のやり方による克服

求めるものが確率変数に

（C）岡田先生

	頻度主義	ベイズ主義
母数θ	定数	確率変数
データx,y	確率変数	定数

頻度主義では，たった一つの真値を求めて議論する

→仮説は真か偽のどちらかである

ベイズ主義では，データから考えられる母数の分布を考える

→確信できる程度を見定める

頻度主義とベイジアン

- いずれも「標本の特徴」から真実を見たいと考えるが
  - 「真実は常に一つ！」とするのが頻度主義者
  - 「真実はわからないけれども，どの辺りにあるか，主張の強さで表現することはできる」とするのがベイジアン

信頼区間
Confidential Intervals

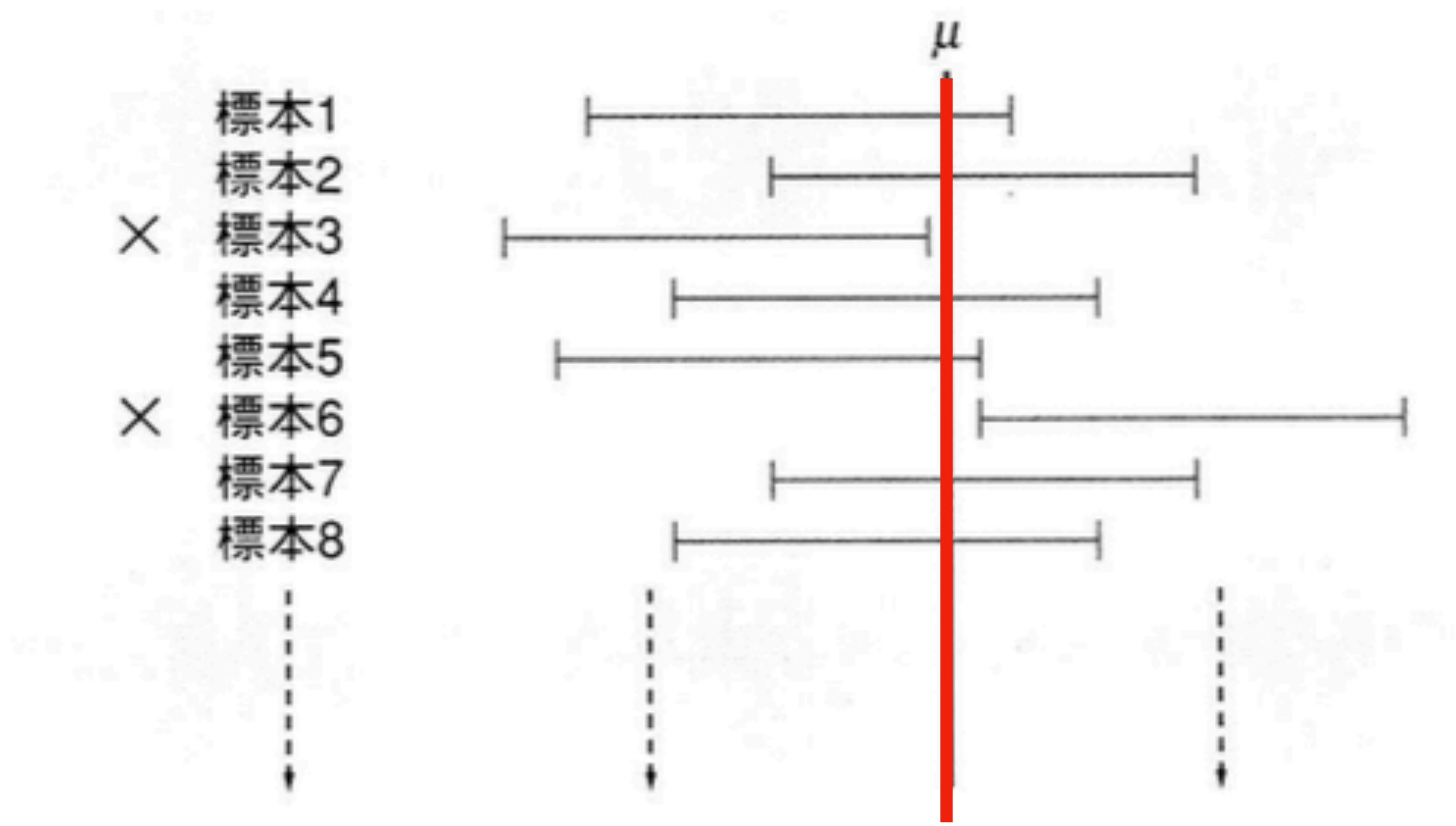


図 8.5 区間の標本による変動

出典；山内光哉(著)心理教育のための統計法第二版.サイエンス社

信頼「区間」というが，分布の情報を持っていない

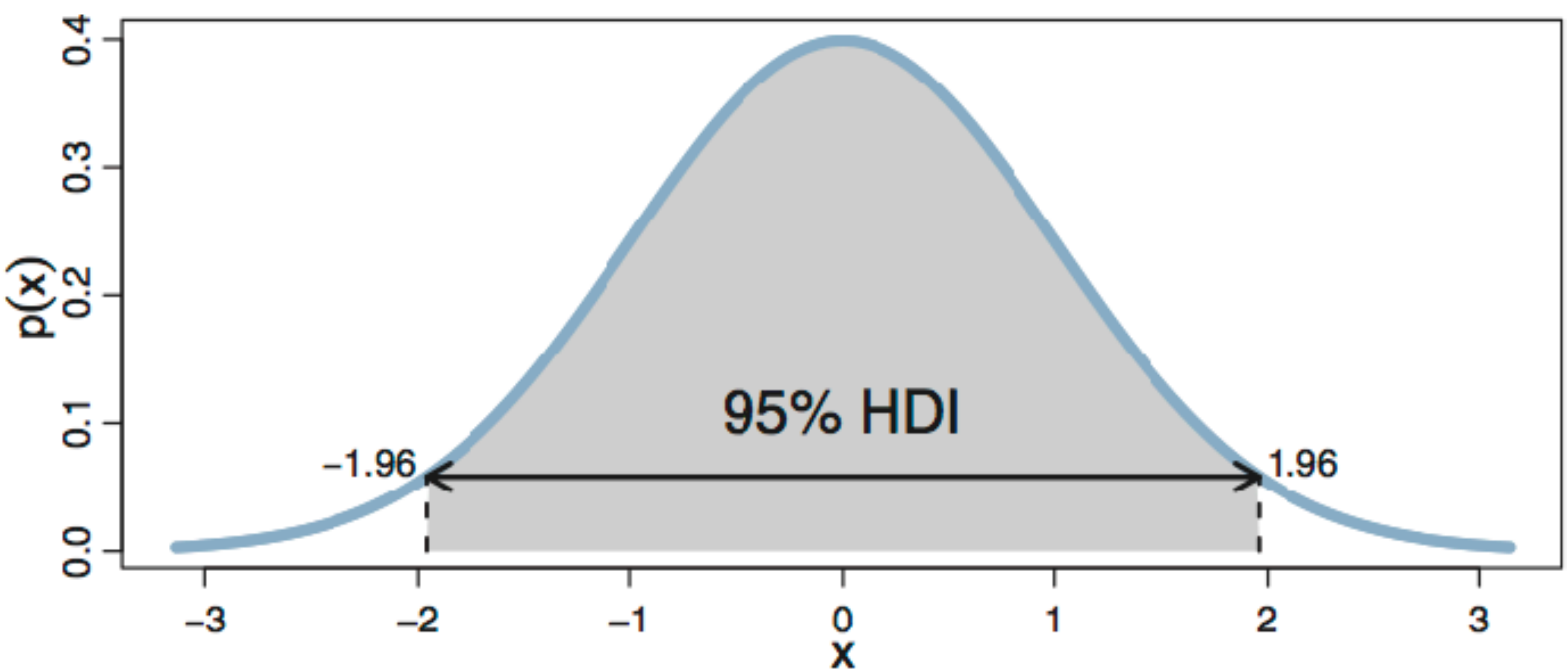
＝幅ではないことに注意

確信区間

Credible Intervals

あるいは最高密度区間

Highest Density Intervals



確信できる「区間」でμの取りうる可能性の広さ

幅が狭い＝自信がある／幅が広い＝自信がない

「点」から「幅」へ

- これまでは「**点**」の表現だったので，仮説は「真か偽か」の二択，一点張り
  - 母平均を標本平均から推定しても，点推定なのでほぼ確実に外れている。信用区間で表現しても結果は「当たるか外れるか」
- ベイジアンは母数が確率分布＝**幅**をもっているので，仮説は「どれぐらい確からしいか」であり，幅の広さで自信の強さを表現するようになる

「**ないない**」から「**あるある**」へ

- 帰無仮説検定は「ないない」づくし
  - 帰無仮説は差が「**ない**」とする
    - t検定の場合，muA=muBの一点張り
  - それに従って統計量を算出しp値を計算
  - 差が「**ない**」とは「いえ**ない**」，が結論

「**ないない**」から「**あるある**」へ

- ベイジアンは「あるある」で語る
  - 手元のデータから確率変数としての母数を推定する
  - 母数はこの辺りに「**ある**」という
  - 群間の差が0である可能性がこれぐらい「**ある**」。もちろん差が**ある**可能性がこれぐらい「**ある**」ともいう。
    - 一点張りでないので，幅を持って自信の強さを表現

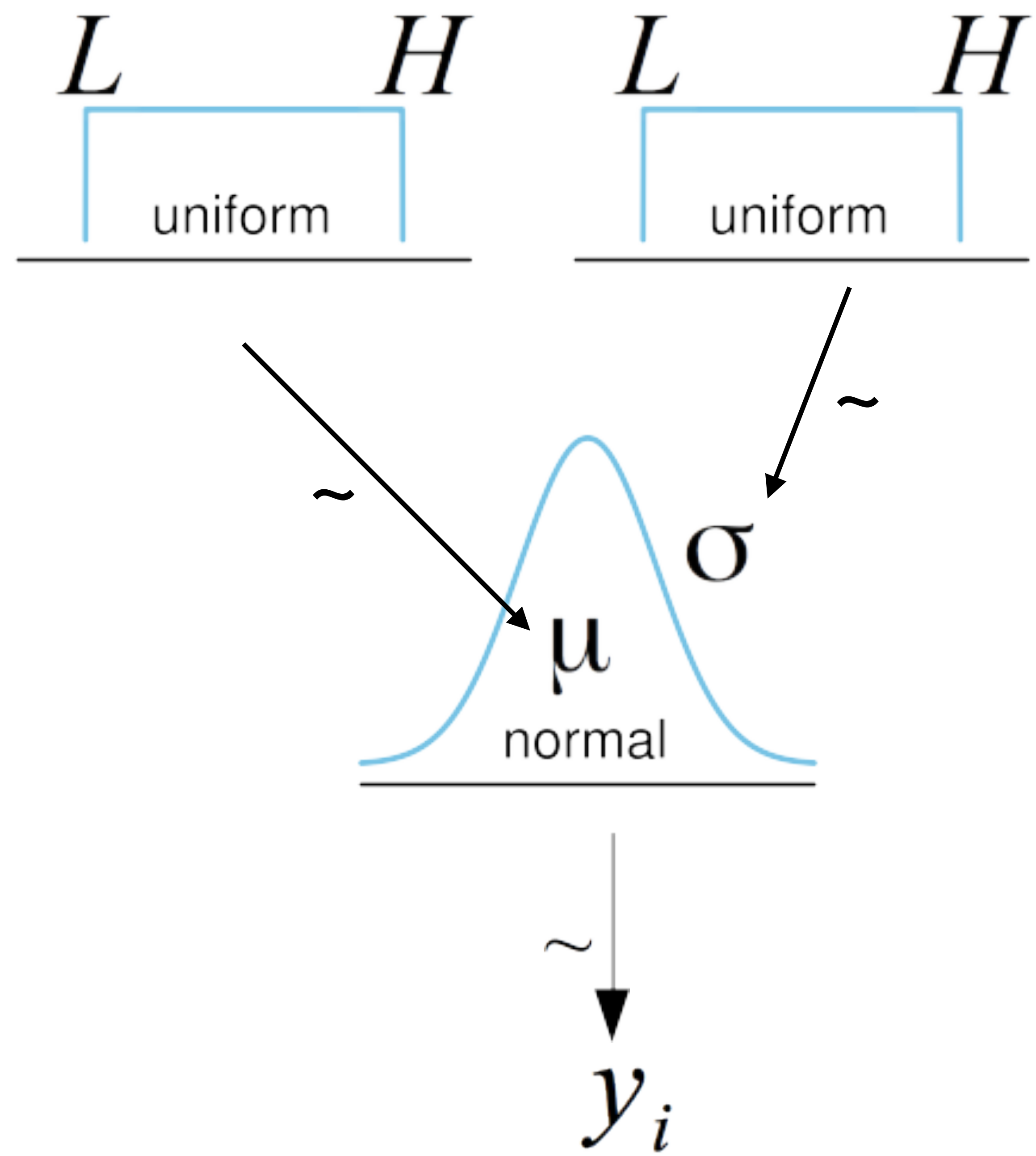
具体例1；2群の差の検定

- 独立した2群の平均値の差の検定ーt検定の場合
  - 変数に正規分布を仮定
  - 分散が等しいかどうかで自由度を調整
  - 2群の平均値に差はないという帰無仮説を，有意水準（5%）で棄却できるかどうかが勝負

ベイジアン分析の場合

各変数に確率分布を考える

- 変数に正規分布を仮定
- 正規分布は二つのパラメータ, μとσを持っている
- μとσがどういう分布をしているかわからないので，一様分布を仮定しようと思う。
- σは0より大きいとしておく



ベイジアン分析の場合

各変数に確率分布を考える

- 変数に正規分布を仮定

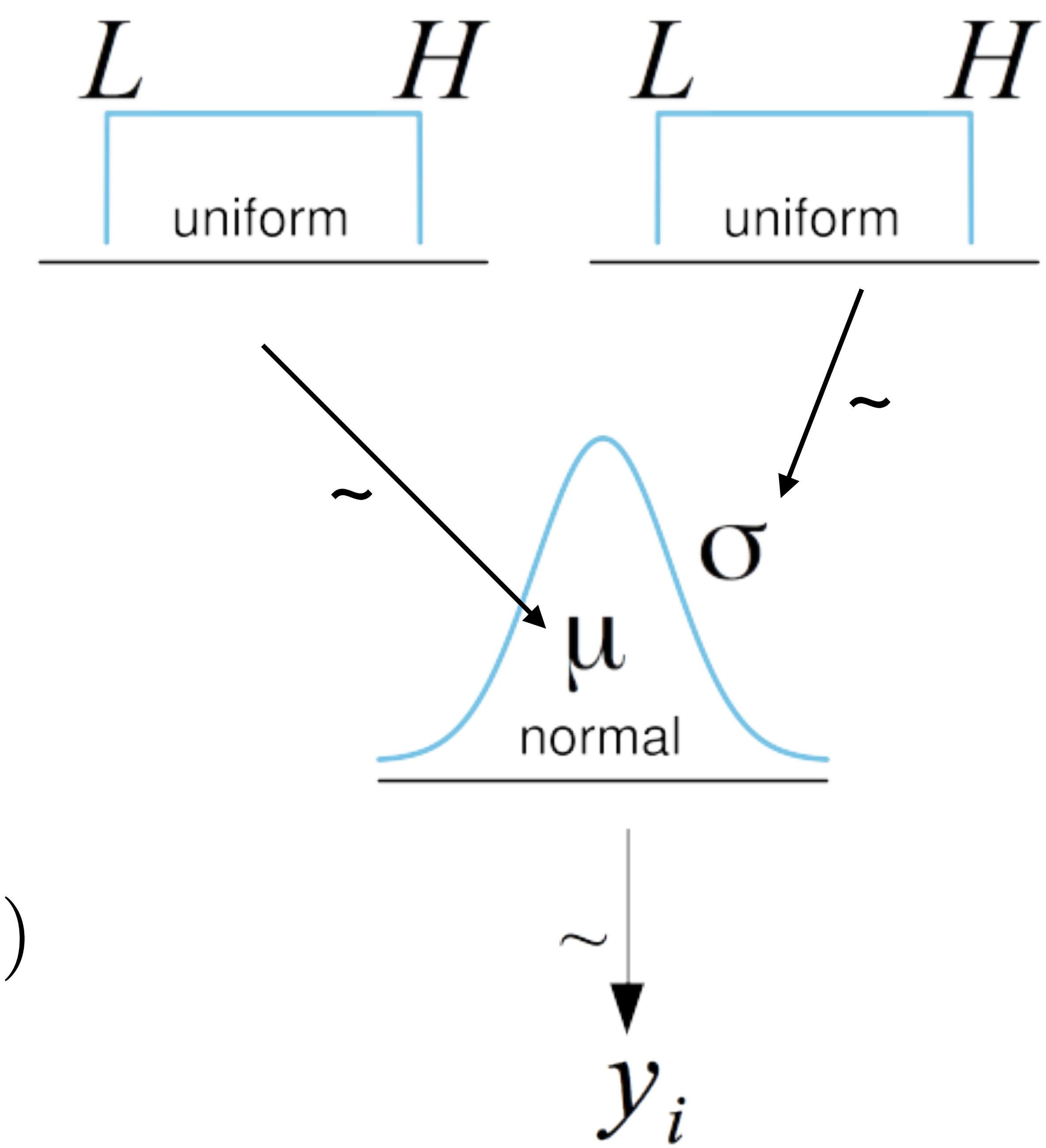
$$Y_i \sim N(\mu, \sigma)$$

- μ と σ がどういう分布をしているかわからないので，一様分布を仮定しようと思う。

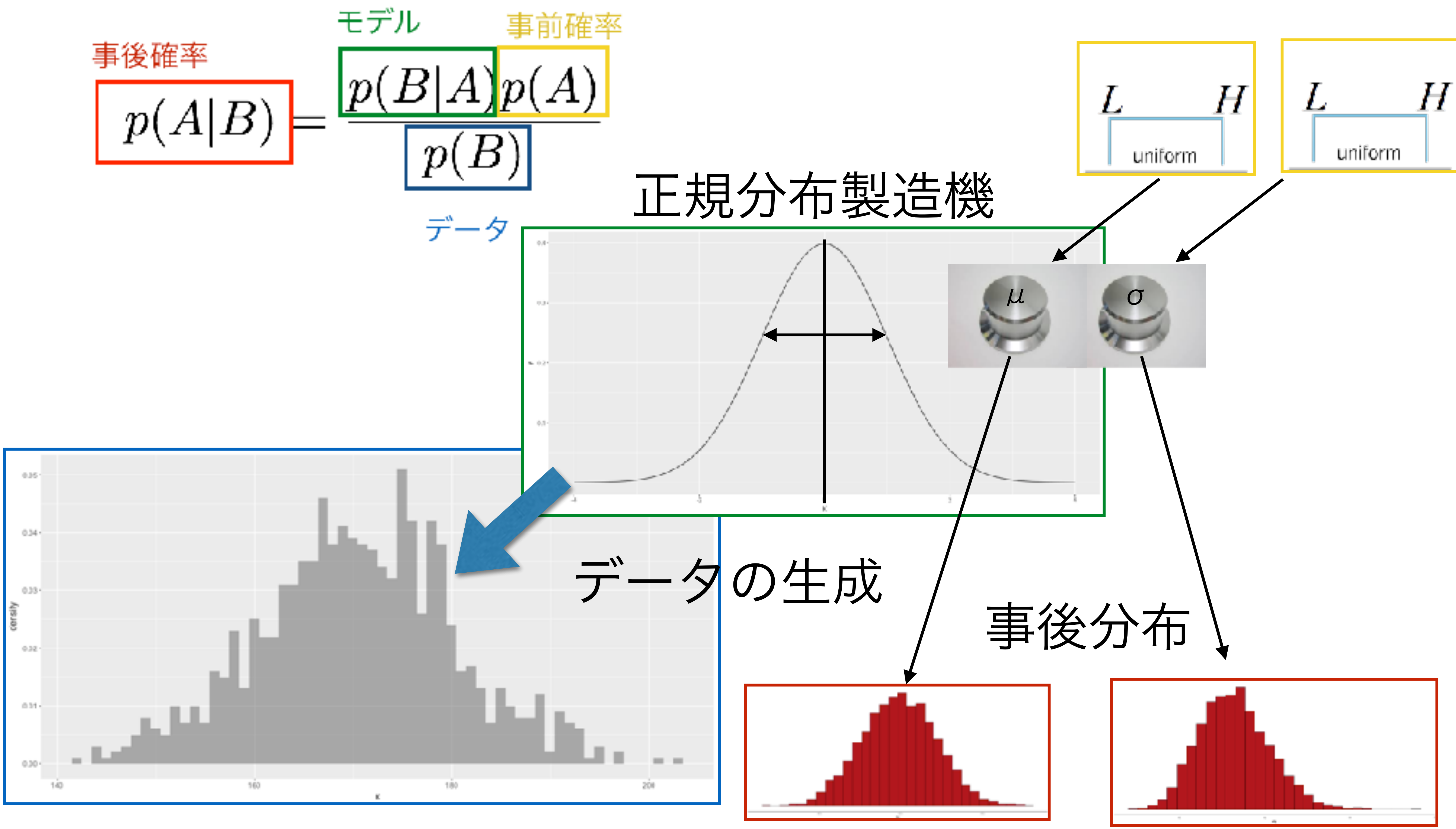
- σ は0より大きいとしておく

$$\mu \sim Uniform(-100, 100)$$

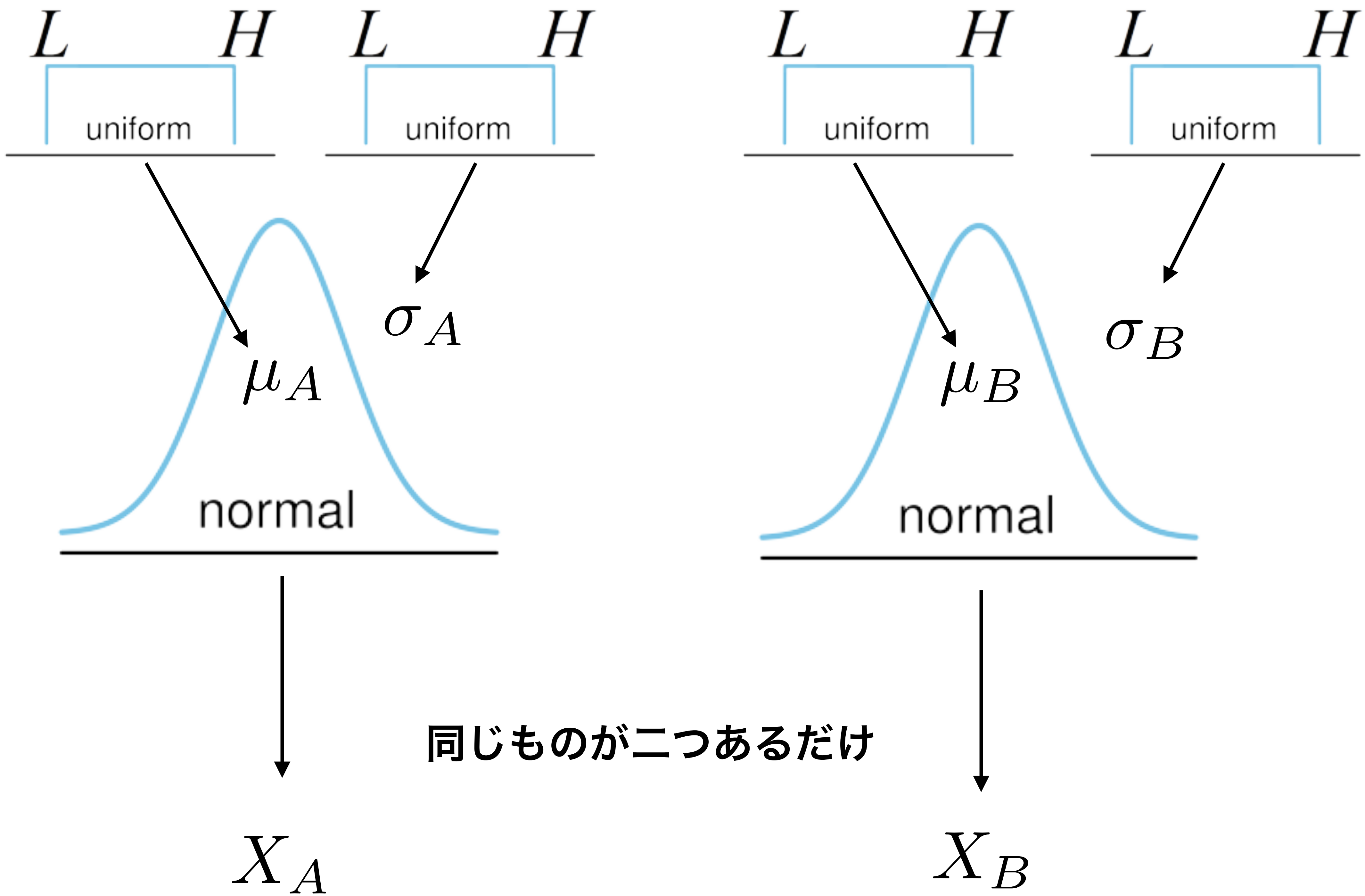
$$\sigma \sim Uniform(0, 100)$$



ベイズの法則はこう使われる



対応のない2群なので



L
H
uniform
L
H
uniform
L
H
uniform
L
H
uniform
σ_A
μ_A
normal
σ_B
μ_B
normal
同じものが二つあるだけ
X_A
X_B

ベイジアンソフトウェア

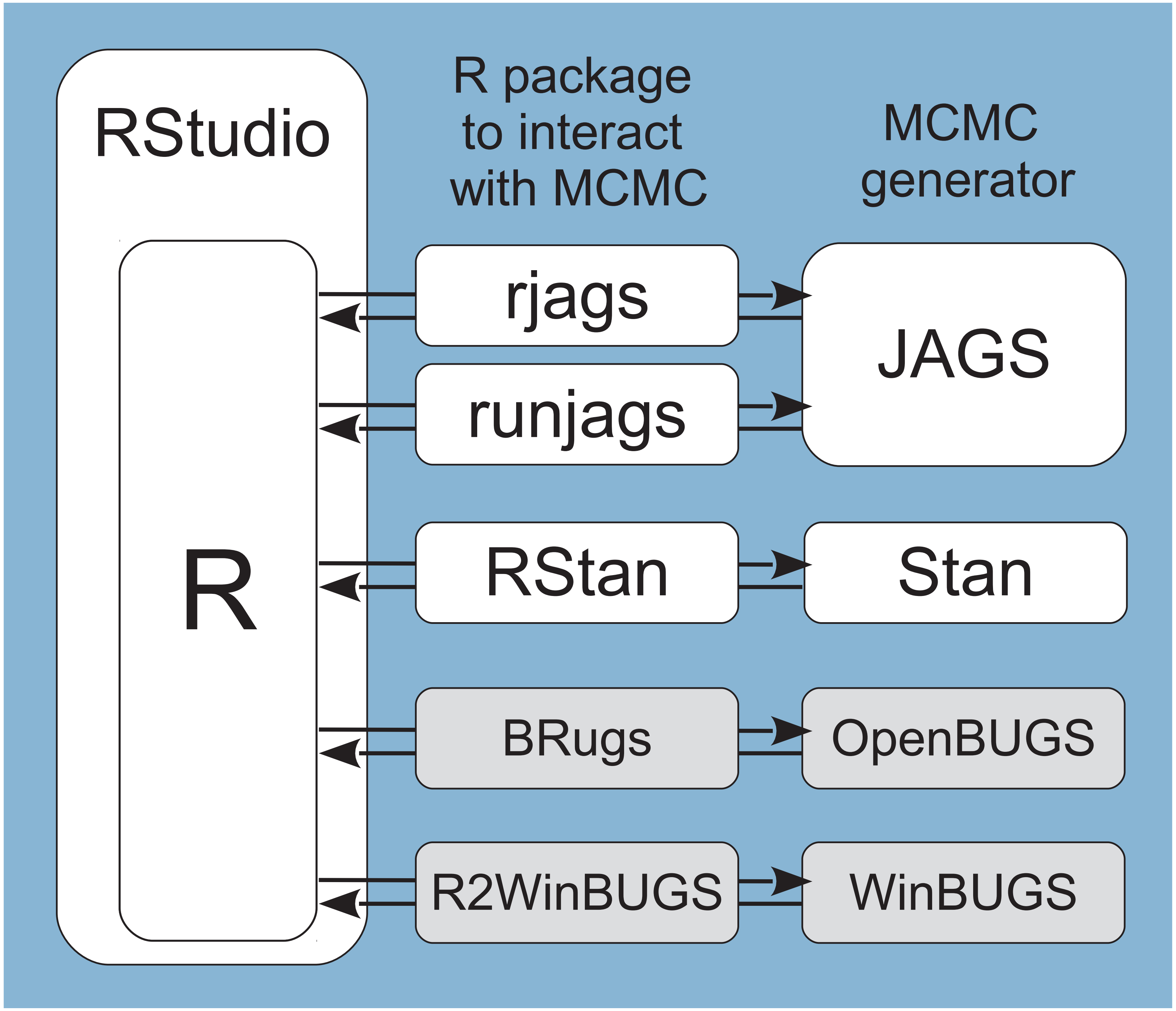


図8.1

実際の推定するコード

```
model{
  # likelihood
  for(i in 1:N){
    X1[i] ~ dnorm(muA,sig)
    X2[i] ~ dnorm(muB,sig)
  }
  # prior
  muA ~ dnorm(0,1.0e-4)
  muB ~ dnorm(0,1.0e-4)
  sig <- pow(sigma,-2)
  sigma ~ dt(0,1,1)T(0,)
}
```

example1_jags.txt

↑JAGS

Stan→

```
data{
  int<lower=0> N;
  real X1[N];
  real X2[N];
}

parameters{
  real muA;
  real muB;
  real<lower=0> sig;
}

model{
  // likelihood
  for(i in 1:N){
    X1[i] ~ normal(muA,sig);
    X2[i] ~ normal(muB,sig);
  }
  // prior
  sig ~ student_t(4,0,5);
  muA ~ normal(0,1000);
  muB ~ normal(0,1000);
}
```

example1.stan

キックするコード

```r
# ベイズ推定に使うデータセット（共通）
dataSet <- list(N=N,X1=X1,X2=X2)
# JAGSによる推定
library(runjags) # ライブラリの読み込み
runJagsOut <- run.jags(method="parallel",
                       model="JPA2017/kosugitti/example1_jags.txt",
                       monitor=c("muA","muB","sigma"),
                       data=dataSet,
                       n.chains=4,
                       adapt=1000,
                       burnin=1000,
                       sample=10000,
                       summarise=TRUE)
# 結果の出力
runJagsOut

# Stanによる推定
library(rstan)
rstan_options(auto_write = TRUE)
model1 <- stan_model("JPA2017/kosugitti/example1.stan")
stanModelFit <- sampling(model1,data=dataSet,chains=4,warmup=1000,iter=10000)

# 結果の出力
stanModelFit
```

得られる結果は標本

- 設定した条件で取り得るmuA,muB,sigのパラメータの組み合わせがたくさん得られており，それのヒストグラムは確率分布とみなせる，というのがMCMCアプローチ。

- 結果は標本なので標本の記述統計，グラフ化がそのまま結果の解釈に用いることができる

結果が示すもの

- 「点」から「幅」へ

一つの代表値（点）で考えても良いけれども・・・

パラメータの取りうる可能性の幅

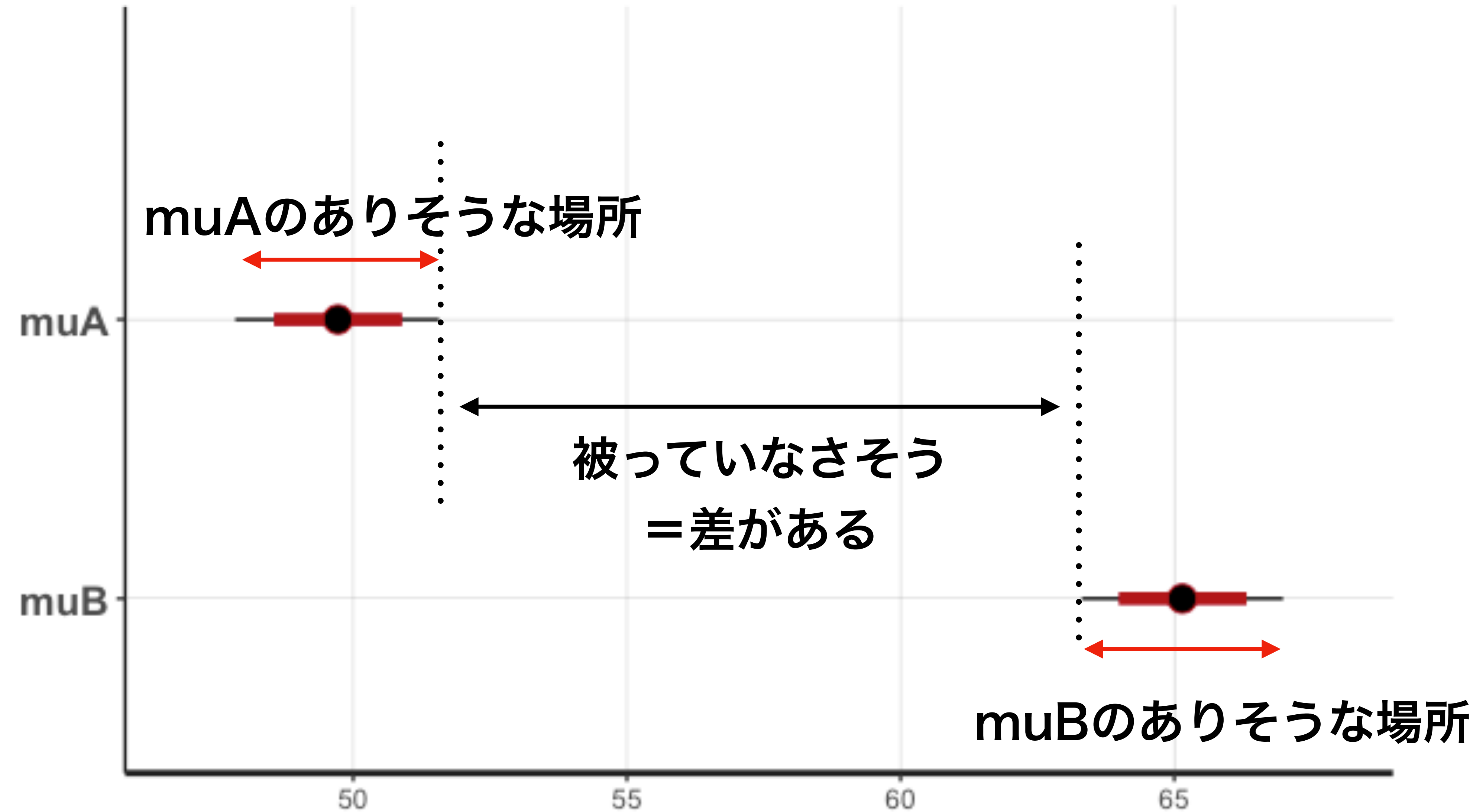
```
> # 結果の出力
> stanModelFit
Inference for Stan model: example1.
4 chains, each with iter=10000; warmup=1000; thin=1;
post-warmup draws per chain=9000, total post-warmup draws=36000.

       mean se_mean   sd   2.5%    25%    50%    75%  97.5% n_eff Rhat
muA   49.72    0.01 0.94  47.84  49.12  49.72  50.33  51.57 28844    1
muB   65.15    0.01 0.93  63.31  64.54  65.15  65.75  66.99 30539    1
sig    2.90    0.00 0.52   2.10   2.53   2.83   3.20   4.11 22964    1
lp__ -29.89    0.01 1.33 -33.32 -30.50 -29.55 -28.91 -28.38 13347    1

Samples were drawn using NUTS(diag_e) at Fri Sep  1 17:24:33 2017.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

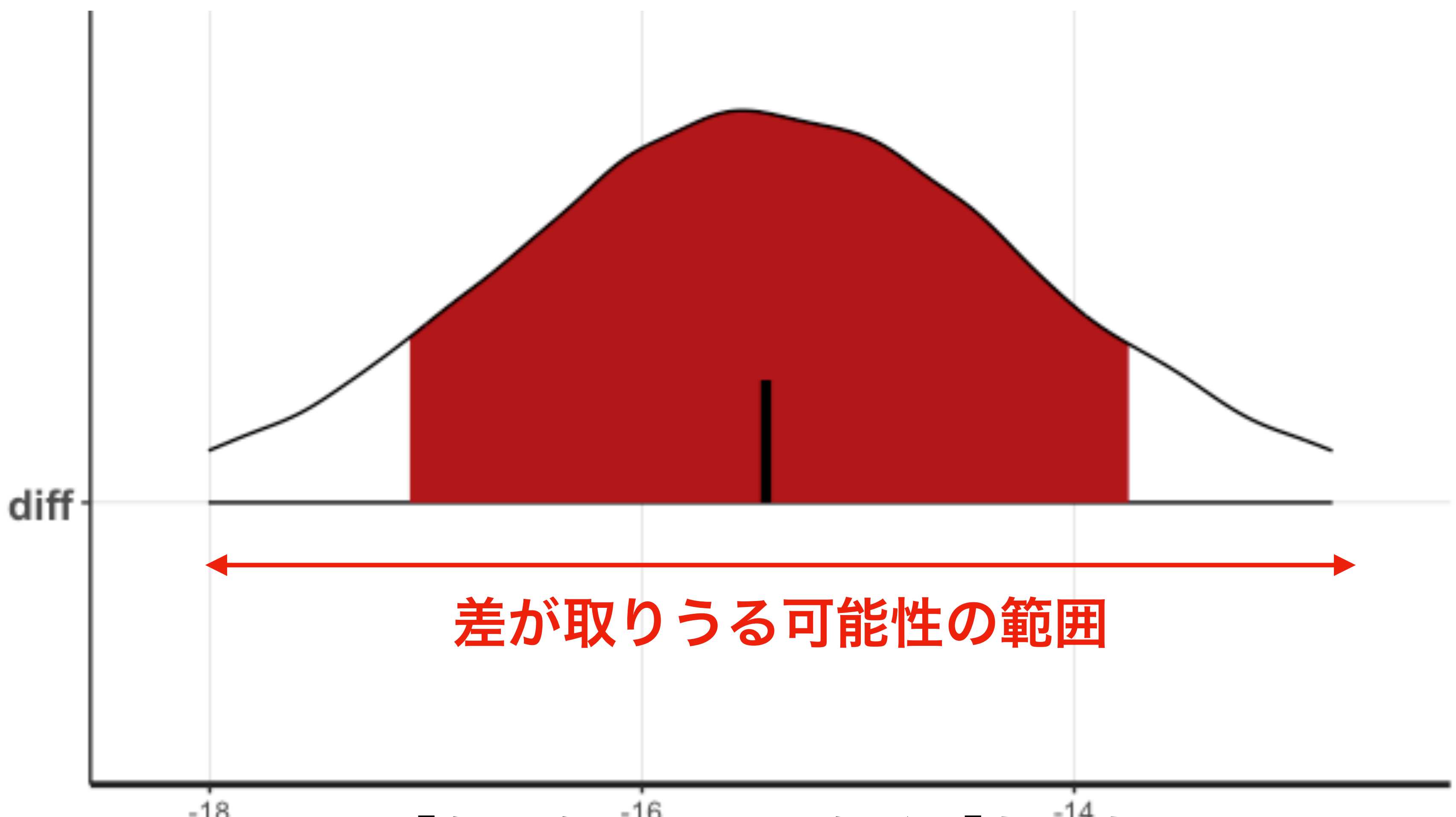
結果が示すもの

```
> plot(stanModelFit,pars=c("muA","muB"))
ci_level: 0.8 (80% intervals)
outer_level: 0.95 (95% intervals)
```



差の分布

```
# 差の分布
## JAGSの場合
JagsSample <-  as.data.frame(as.matrix(runJagsOut$mcmc))
JagsSample$diff <- JagsSample$muA - JagsSample$muB
## Stanの場合
model1a <- stan_model("JPA2017/kosugitti/example1a.stan")
stanModelFit1a <- sampling(model1a,data=dataSet,chains=4,warmup=1000,iter=10000)
plot(stanModelFit1a,show_density=T,pars="diff")
```



「ないない」ではなく「あるある」

ROPE;実質的に等価な範囲

- Region of practical equivalenceの略。
- 帰無仮説が「差がない；muA=muB」という一点張りだったのに対し，「実質的にこの程度の差であれば意味がないよね」という領域。
  - ex)身長の差が0〜3cmぐらいであれば，両群の体の大きさは実質的に同じと見なそう，など。
- HDIがROPEに含まれるかどうかが重要

t検定のときに気になること

- 群のサイズがアンバランスだと計算が面倒に
  - このモデルなら簡単に対応できます
- 分散の等質性で補正しなくていいの？
  - 分散を個別に推定する＝違うものと仮定するだけで対応できるようになる。

異なる分散のコード

JAGSのばあい

```
model{
  # likelihood
  for(i in 1:N){
    X1[i] ~ dnorm(muA,sig)
    X2[i] ~ dnorm(muB,sig)
  }
  # prior
  muA ~ dnorm(0,1.0e-4)
  muB ~ dnorm(0,1.0e-4)
  sig <- pow(sigma,-2)
  sigma ~ dt(0,1,1)T(0,)
}
```

example1_jags.txt

```
model{
  # likelihood
  for(i in 1:N){
    X1[i] ~ dnorm(muA,sig1)
    X2[i] ~ dnorm(muB,sig2)
  }
  # prior
  muA ~ dnorm(0,1.0e-4)
  muB ~ dnorm(0,1.0e-4)
  sig1 <- pow(sigma1,-2)
  sigma1 ~ dt(0,1,1)T(0,)
  sig2 <- pow(sigma2,-2)
  sigma2 ~ dt(0,1,1)T(0,)
}
```

example2_jags.txt

異なる分散のコード

Stanのばあい

```
data{
  int<lower=0> N;
  real X1[N];
  real X2[N];
}

parameters{
  real muA;
  real muB;
  real<lower=0> sig;
}

model{
  // likelihood
  for(i in 1:N){
    X1[i] ~ normal(muA,sig);
    X2[i] ~ normal(muB,sig);
  }
  // prior
  sig ~ student_t(4,0,5);
  muA ~ normal(0,1000);
  muB ~ normal(0,1000);
}
```

example1.stan

```
data{
  int<lower=0> N;
  real X1[N];
  real X2[N];
}

parameters{
  real muA;
  real muB;
  real<lower=0> sig1;
  real<lower=0> sig2;
}

model{
  // likelihood
  for(i in 1:N){
    X1[i] ~ normal(muA,sig1);
    X2[i] ~ normal(muB,sig2);
  }
  // prior
  sig1 ~ student_t(4,0,5);
  sig2 ~ student_t(4,0,5);
  muA ~ normal(0,1000);
  muB ~ normal(0,1000);
}
```

example2.stan

異なる分散の結果

```
> stanModelFit2
Inference for Stan model: example2.
4 chains, each with iter=10000; warmup=1000; thin=1;
post-warmup draws per chain=9000, total post-warmup draws=36000.

        mean se_mean   sd  2.5%    25%    50%    75%  97.5% n_eff Rhat
muA    52.75    0.02 2.70  47.37  51.04  52.76  54.47  58.12 30337    1
muB    66.69    0.01 1.37  63.95  65.84  66.69  67.54  69.40 28564    1
sig1    8.30    0.01 1.94   5.48   6.93   7.99   9.31  12.94 27171    1
sig2    4.22    0.01 1.09   2.68   3.47   4.03   4.75   6.87 24795    1
lp__  -43.65    0.01 1.54 -47.48 -44.40 -43.29 -42.52 -41.75 13166    1
```

- 違うものとして推定しているだけで，解釈に違いがない

柔軟な仮説が組める

- 「差が○cm以上ありそうな可能性」が表現できる
- 「どちらの標準偏差が大きいか」なども考えられる
  - 分散が心理変数に対応する例なども考えられる
- パラメータの大小，あるいはパラメータを数式で表現することもできる＝**ベイジアンモデリング**

それでもやはり
踏ん切りがつかない人へ

コインを24回投げて7回表が出たとき，このコインはイカサマか？

- 「コインが表になる確率が0.5である」という帰無仮説を棄却できるかどうか，と考えると思います。

- どうやってデータが得られたか，を考えて見ましょう
  - 24回投げよう，と最初から決めていた場合
  - 7回表が出たらやめよう，と最初から決めていた場合
  - 5分間投げ続けよう，と最初から決めていた場合

帰無仮説検定的には全て結果が違う！

帰無仮説検定の危うさ

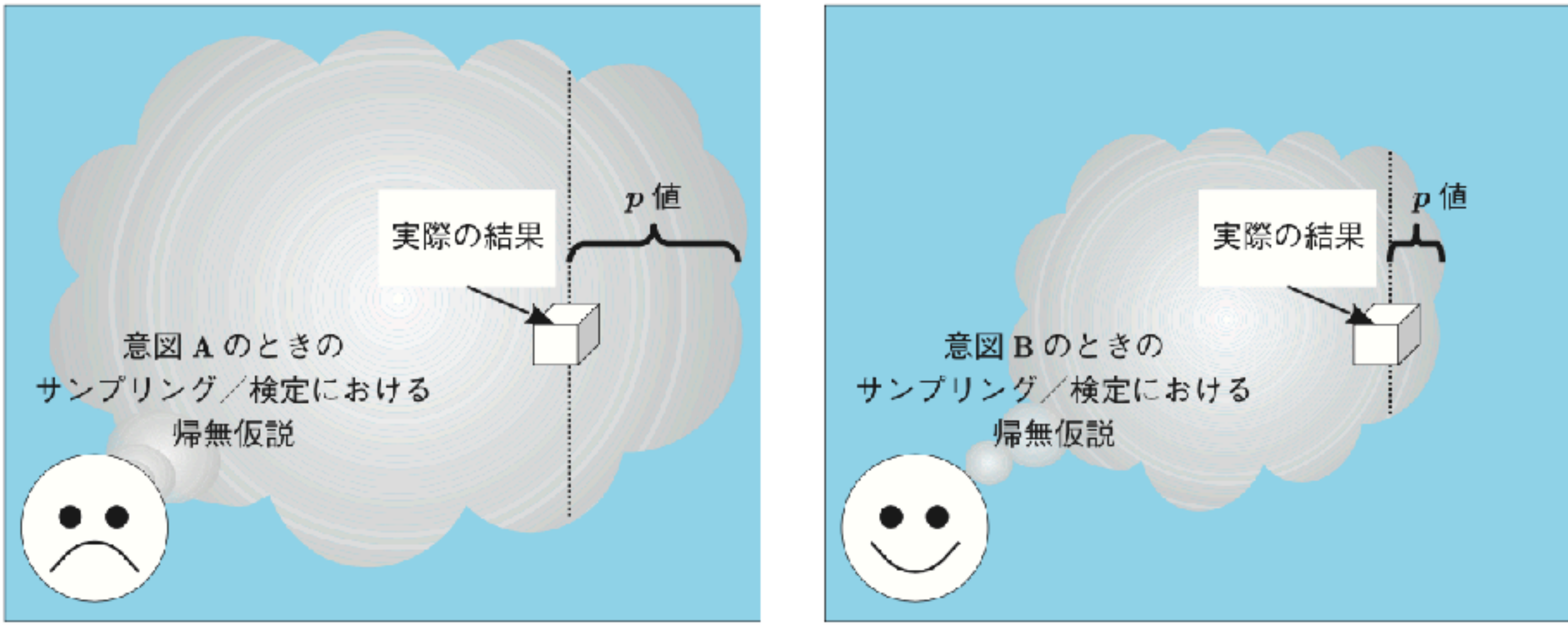


図 11.1　帰無仮説は仮想的な結果の集合（雲）を生成する．この結果のほとんどは集合の中央にあるが，一部はブロックで印が付けられた実際の結果の向こうにある．p 値は実際の結果以上に極端な雲の比率である．左：サンプリング意図 A のもとでは，仮想的な可能性の集合には実際の結果を超える部分が多くある．それゆえ，p 値は大きい．右：サンプリング意図 B のもとでは，仮想的な可能性の集合には実際の結果を超える部分が少ない．それゆえ，p 値は小さい．

コインを24回投げて7回表が出たとき，このコインはイカサマか？

		N								
		1	2	3	4	5	6	7	8	...
	0									...
	1									...
	2	-								...
	3	-	-							...
z	4	-	-	-						...
	5	-	-	-	-					...
	6	-	-	-	-	-				...
	7	-	-	-	-	-	-			...
	8	-	-	-	-	-	-	-		...
	...	-	-	-	-	-	-	-	-	...

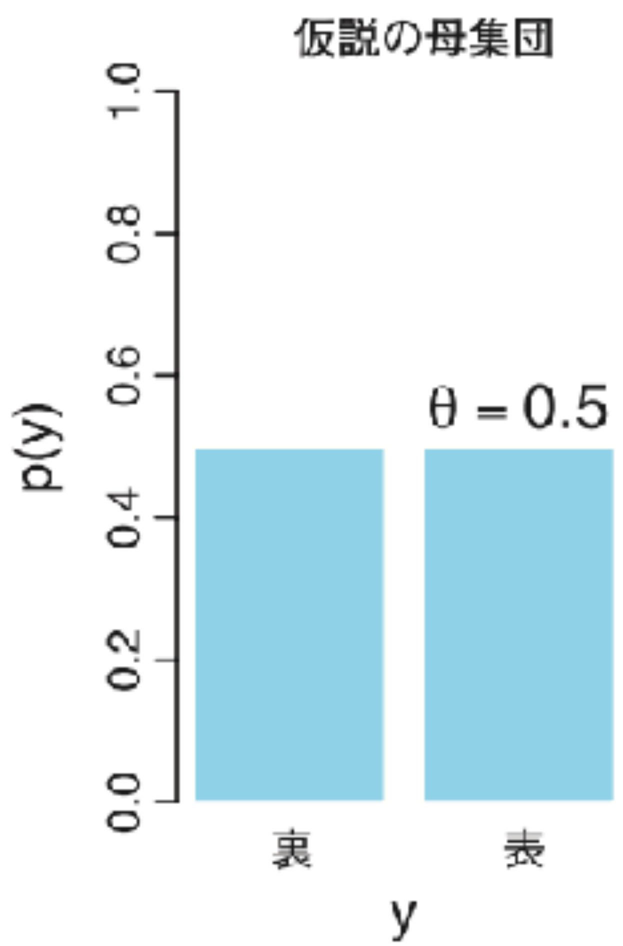
		N								
		1	2	3	4	5	6	7	8	...
	0									...
	1									...
	2	-								...
	3	-	-							...
z	4	-	-	-						...
	5	-	-	-	-					...
	6	-	-	-	-	-				...
	7	-	-	-	-	-	-			...
	8	-	-	-	-	-	-	-		...
	...	-	-	-	-	-	-	-	-	...

		N								
		1	2	3	4	5	6	7	8	...
	0									...
	1									...
	2	-								...
	3	-	-							...
z	4	-	-	-						...
	5	-	-	-	-					...
	6	-	-	-	-	-				...
	7	-	-	-	-	-	-			...
	8	-	-	-	-	-	-	-		...
	...	-	-	-	-	-	-	-	-	...

図 11.2　コイン投げのサンプル空間．列は N の候補となる値であり，行は z の候補となる値である．左の表：N が決められていると考えられるときの可能性の空間であり，$N = 5$ の列が陰影で強調されている．中央の表：z が決められていると考えられるときの可能性の空間であり，$z = 4$ の行が陰影で強調されている．右の表：時間が決められていると考えられるときの可能性の空間であり，サンプルサイズの確率が異なる列の陰影で示されている．

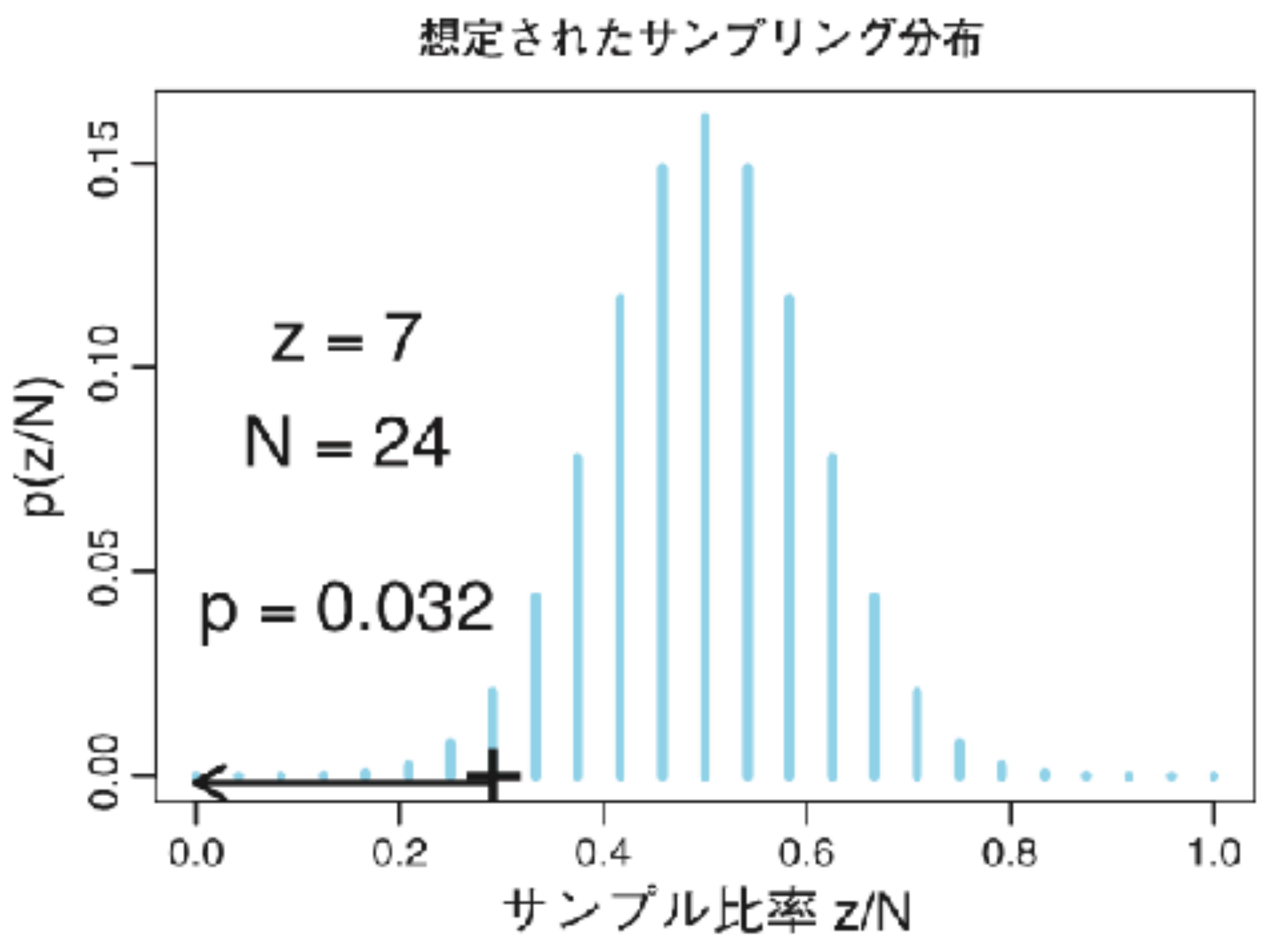
		N								
		1	2	3	4	5	6	7	8	...
z	0									...
	1									...
	2	-								...
	3	-	-							...
	4	-	-	-						...
	5	-	-	-	-					...
	6	-	-	-	-	-				...
	7	-	-	-	-	-	-			...
	8	-	-	-	-	-	-	-		...
	...	-	-	-	-	-	-	-	-	...

24回投げよう と決めていた場合



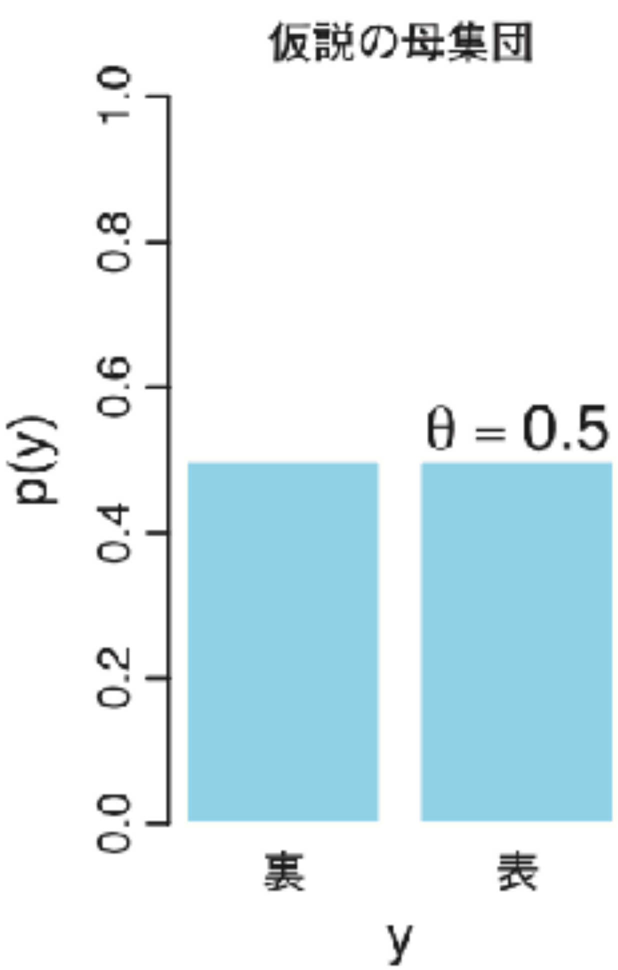
N を固定

= 24



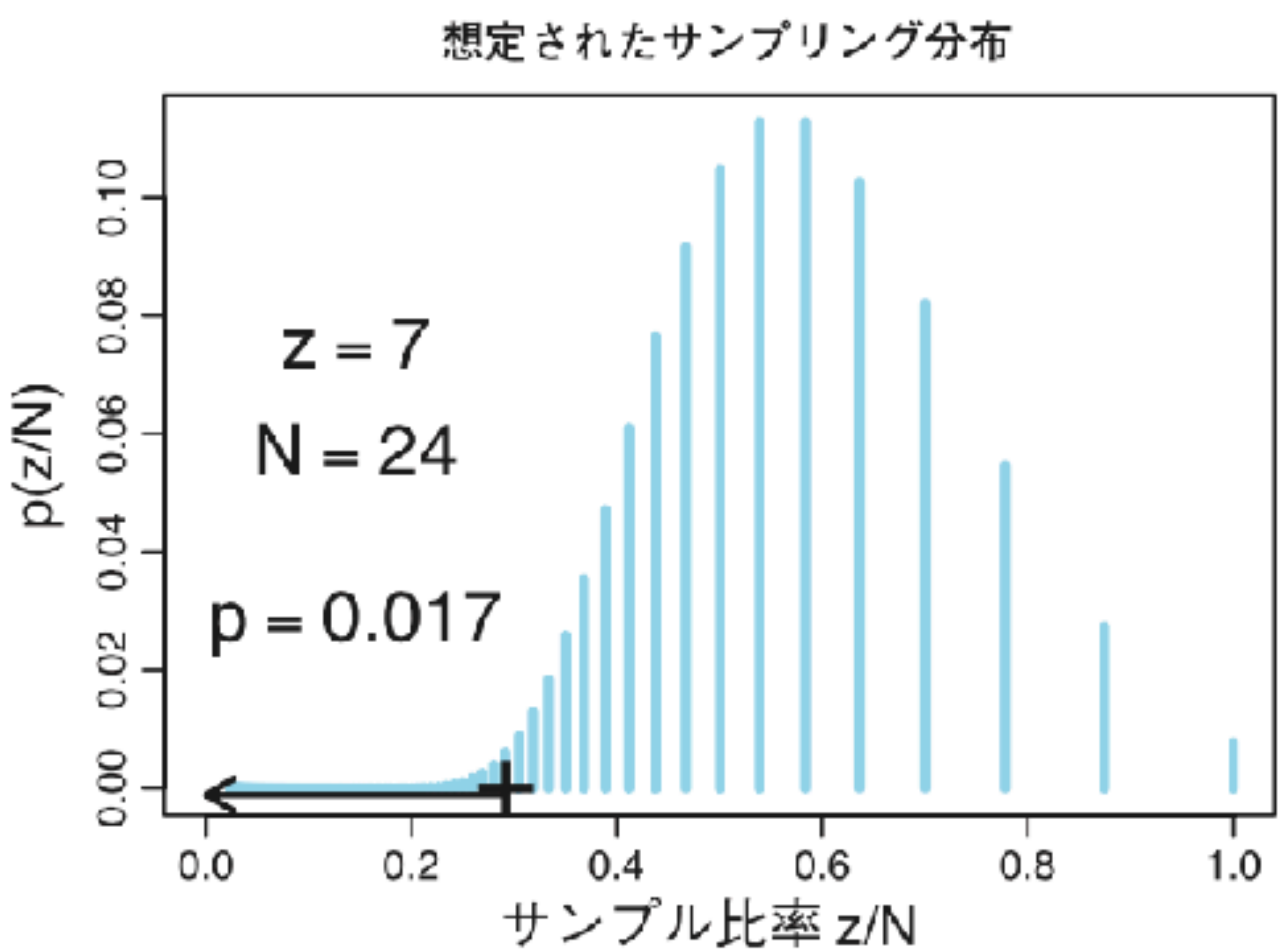
7回表を見よう と決めていた場合

z \ N	1	2	3	4	5	6	7	8	...
0									...
1									...
2	-								...
3	-	-							...
4	-	-	-						...
5	-	-	-	-					...
6	-	-	-	-	-				...
7	-	-	-	-	-	-			...
8	-	-	-	-	-	-	-		...
...	-	-	-	-	-	-	-	-	...



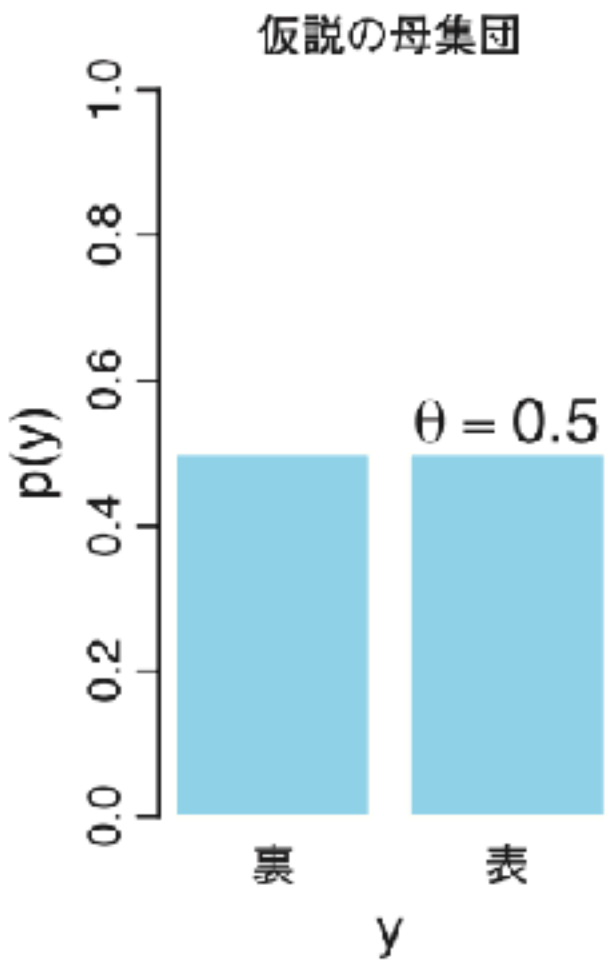
z を固定 = 7

⟹



		N								
		1	2	3	4	5	6	7	8	...
z	0									...
	1									...
	2	-								...
	3	-	-							...
	4	-	-	-						...
	5	-	-	-	-					...
	6	-	-	-	-	-				...
	7	-	-	-	-	-	-			...
	8	-	-	-	-	-	-	-		...
	...	-	-	-	-	-	-	-	-	...

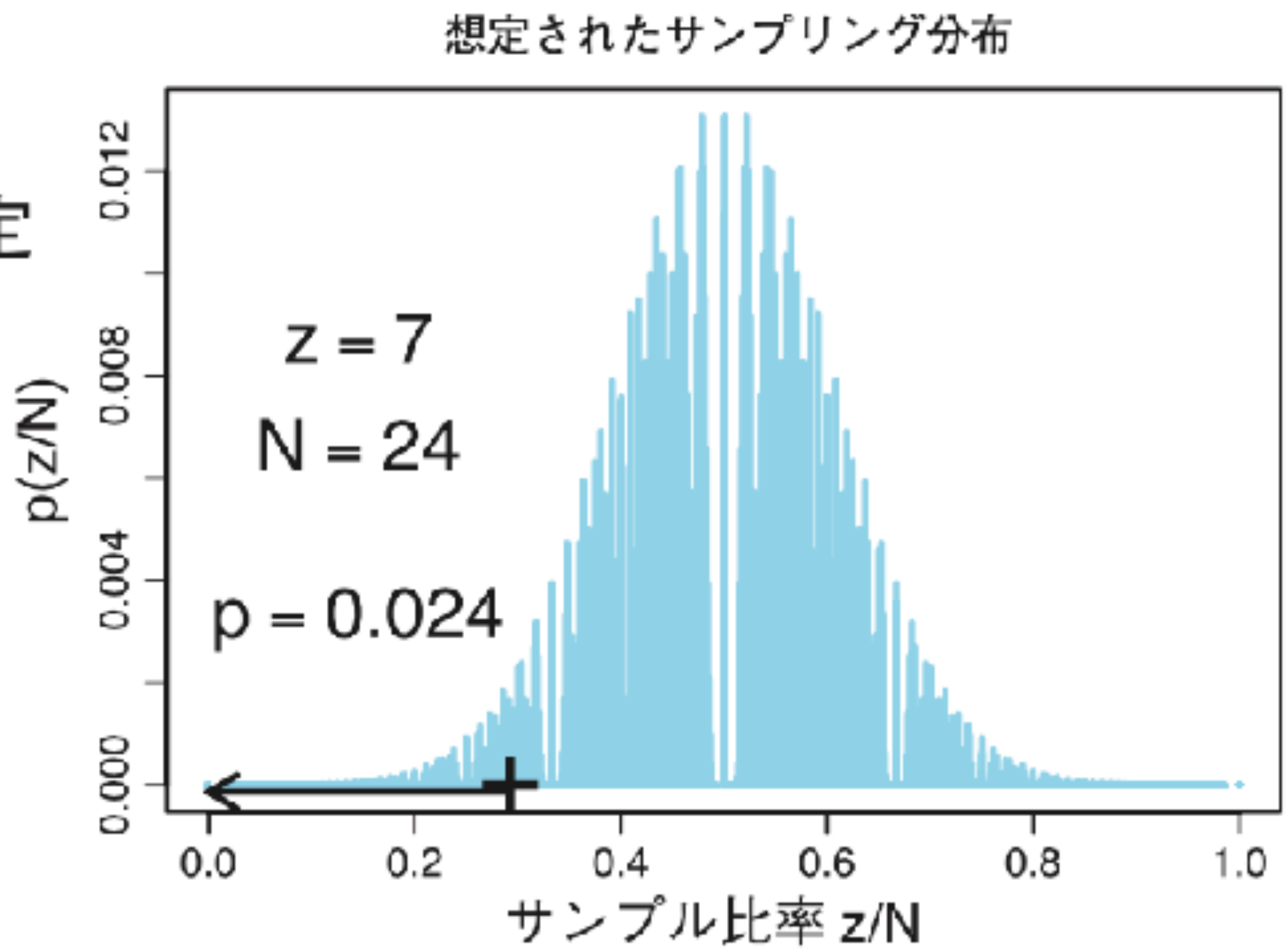
5分間実験しようと決めていた場合



時間を固定

$\lambda = 24$

$\Longrightarrow$



データは常にひとつ！

- データが同じなのに結果が違う，というのは直感的におかしいと思いませんか？

- **サンプルサイズの決め方**（例数設計），**分析計画の設計**（どことどこの差を見るか，一実験あたりの危険率補正）が重要な理由！

  - 「取れるだけ取ってごらん」「有意差が出るまで頑張って見よう」「実験なら10人，調査なら100人かな」…

事後分布は常にひとつ！

- ベイジアンにとってN＝24でz＝7から得られるパラメータの取りうる範囲，は常に一つ。
- 複雑な実験計画でもこれは同じ。ある群，あるセルから得られるパラメータの事後分布は常に一つ。
- ベイジアンにとって，色々な**群の比較は事後分布の異なる読み方に過ぎない。**分析計画を先に考える必要がない
- ベイジアンにとって，例数設計は効果量に基づいて考えればよく，極端に言えば設計しなくても問題にはならない。

効果量は「勝ち目のある勝負をしよう」という意味もあったが，
どんな勝負からも逃げずにデータだけで勝負すると言うのであればそれでもよい。

従来型手法は現実的でない

- 決まり切った手続きに正しく乗せるためには，たくさんの仮定を満たしている必要があり，それからちょっとでもはみ出すと本当は意味がないことにある。
    - （自戒を込めて）どこかで諦めていた
- 細心の注意を払って答えにたどり着いても，言えることは「ないない」であって，当たらぬ真実に想いを馳せるだけ

事前分布と事後分布

- 事後分布には尤度と事前分布が含まれている
- 事前分布＝主観確率＝思い込みはダメ！という心理学教育
- ではこんなことを考えてみましょう。

釘でコイントス？をした。
釘の頭でバランスをとって上を向いたら「表」
転がってしまったら「裏」
24回投げる計画で，7回表が出た。

事前分布は適切に

- 帰無仮説検定は（釘であれなんであれ）「表と裏が出る確率に差はない」と仮定し，それを棄却できない。

> おや? 私たちが話しているのは釘のことである.どうして帰無仮説を棄却できないのか.

- 帰無仮説検定は，手続きの厳格化によってあらゆる対象に対応できる。釘であれ，コインであれ。
- しかし，釘が「50/50の確率で表が出る」という帰無仮説は「適切」ではない。

- 事前分布が事後分布に影響を与えるのは間違いない
  - 曖昧な事前分布であればデータの特徴が事後分布に。
  - 強烈な事前分布を覆すには，大量のデータがいる。

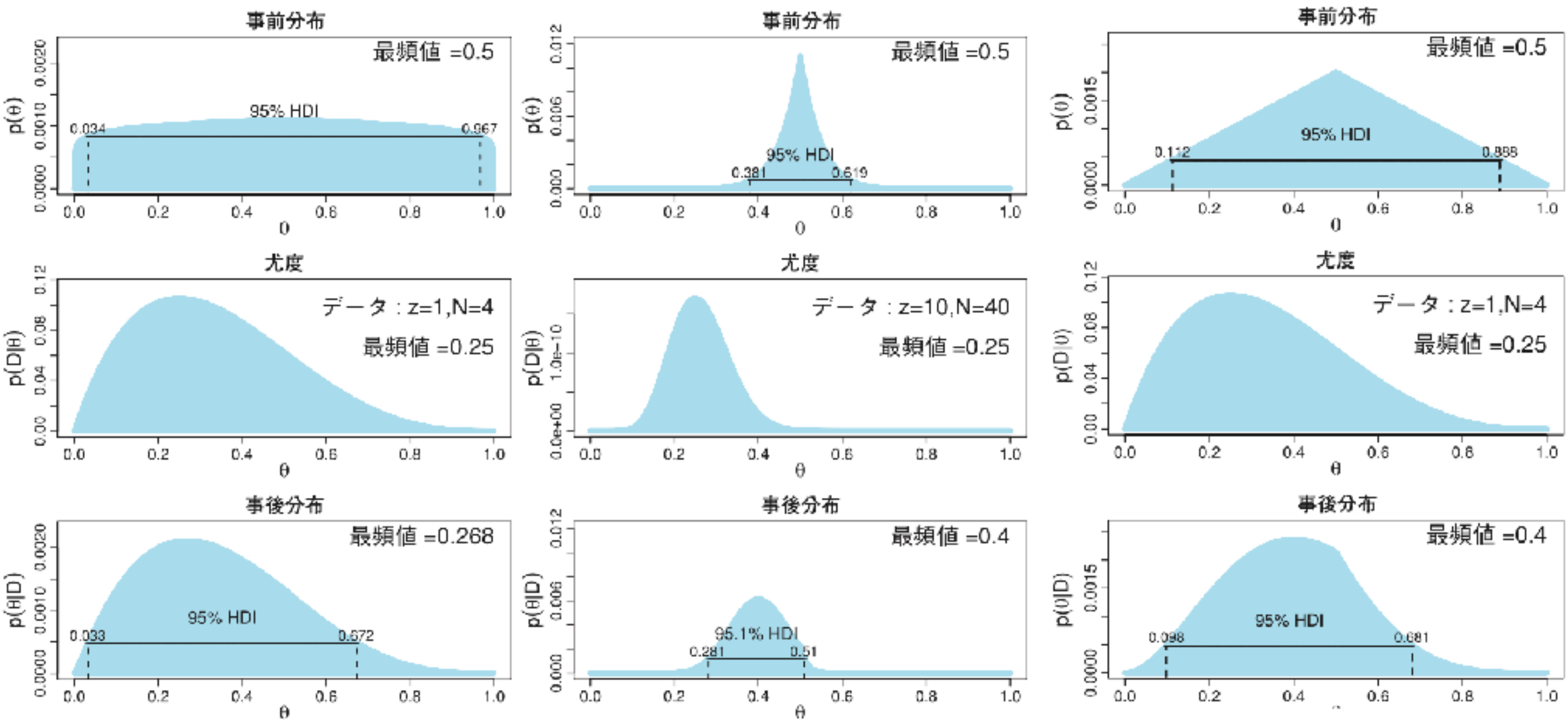


図5.2と5.3より

だからこそベイジアン

- ベイジアンモデリングとは「○○分析」という，極めて一般化された手続きに還元されない分析をすることを意味する＝**明確なモデリングの意図**が必要
- ○×分析のフォーマットに合わせてデータの性質を見るという視点をやめて，データを生成する仕組みをモデリングする
    - 平均値などの**パラメータに構造をもたせることをモデリング（する）**と言います。

2017.09.22

日本心理学会第81回大会TWS

ベイジアンになると何がどう変わるのか

ベイジアンデータ解析入門

話す人：小杉考司（山口大学教育学部）