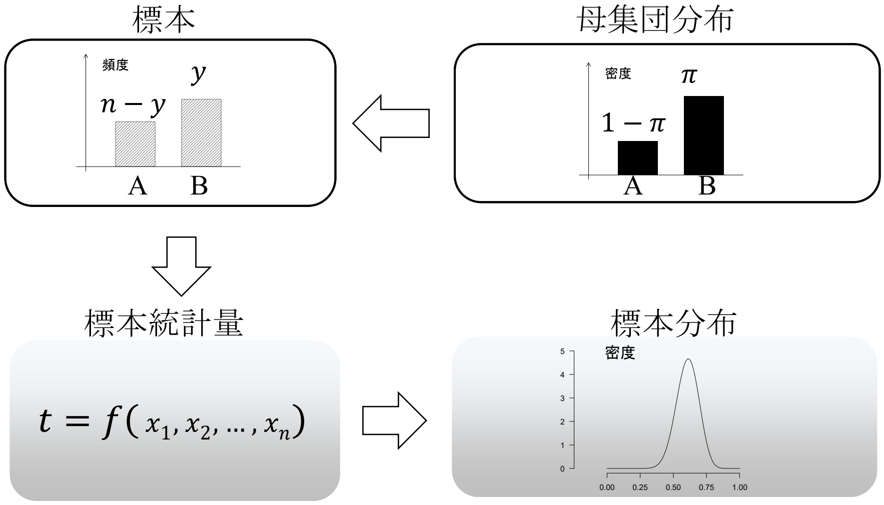
第5講　ベイズ統計学

学びのポイント

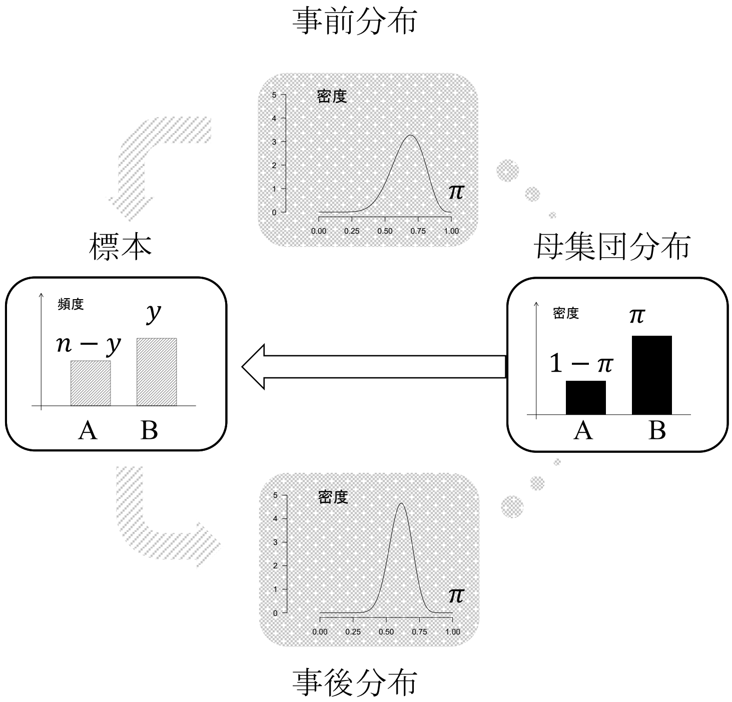
- ☐ 推測統計学のパラダイムに、これまで習ってきた**頻度主義統計学**とは異なる**ベイズ統計学**と呼ばれる枠組みがあることが分かる。
- ☐ ベイズ統計学では、研究者が母集団パラメータに対して抱く確信度を確率分布で表現することが理解できる。
- ☐ データを観測する前の研究者の信念を表した確率分布を**事前分布**と呼び、データを採った後に研究者が抱くパラメータへの確信度を表したものを**事後分布**と呼ぶことが分かる。
- ☐ ベイズの定理を用いて、事前分布と事後分布の関係が数学的に定義され、**事後分布の密度関数**は**事前分布**と**尤度**の積に比例することが理解できる。
- ☐ 事後分布の要約には、**事後期待値**・**事後中央値**・**事後確率最大値**、**事後分散**・**事後標準偏差**といった点的な要約と、**確信区間**・**最高事後密度区間**といった区間的な要約があること理解できる。
- ☐ 事前分布の選択によって、到達する事後分布の形が変化することが分かり、複数の事前分布を試して結果の移動を検討することを**感度分析**と呼ぶことが分かる。
- ☐ 事前分布を選ぶ標準的な方法に、**情報事前分布**を用いる方法、**弱情報事前分布**を用いる方法、**無情報事前分布**を用いる方法、計算が簡便な**共役事前分布**を用いる方法などがあることが分かり、その具体的な例について説明ができる。

見取り図

【頻度主義統計学】



【ベイズ統計学】



■ 目標

これまでの講義では、様々な統計モデルを構築し、その母集団パラメータに対し推定・推論を行ってきました。その推定・推論はすべて、**頻度主義統計学**というパラダイムの確率観に立脚した考え方で進められていたのですが、実は推測統計学のパラダイムには、ベイズ統計学と呼ばれる別の立場も存在しています。この講義の目標は、この**ベイズ統計学**の基本的な考え方を学ぶことにあります。

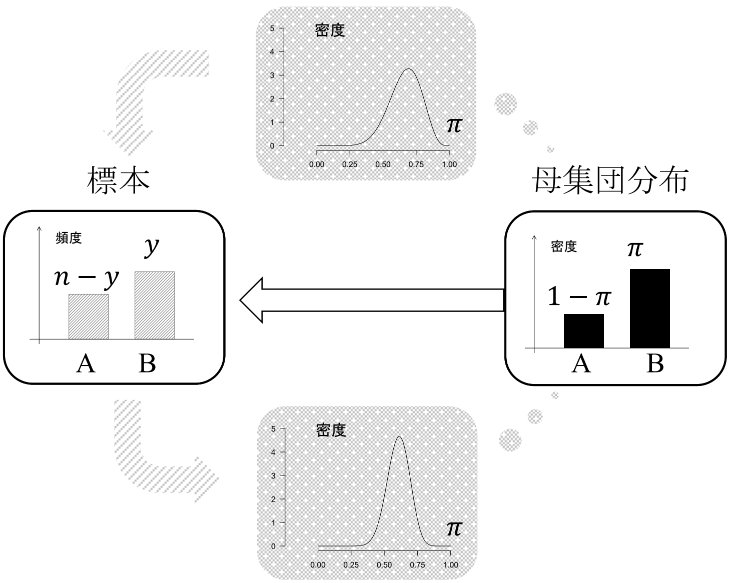
頻度主義統計学とベイズ統計学の違いは、母集団のパラメータに固定的な値を想定するか（頻度主義）、確率的な挙動を仮定するか（ベイズ統計）という点にあり、母集団にどのようなモデルを立てるのか／立てられるのかという点はパラダイムの違いに関係しません。つまり、ベイズ統計学でも、重回帰モデル、ロジステック回帰、ポワソン回帰…といったこれまで習ってきたモデルを母集団に想定することができ、そこまではこれまでの講義の内容と変わりません。しかし、頻度主義とは異なり、推定された偏回帰係数が確率的な挙動を示すので、それに応じて今まで習ってきたのとは異なる推論・解釈を行います。

ベイズ統計学って、頻度主義統計学と何が違うの？

（１）ベイズ統計学

これは、① 研究者が母集団パラメータに対して持つ信念（確信度）を確率分布で表現し、② ベイズの定理に基づき、この確率分布の更新という形でパラメータへの推論を展開する推測統計学の枠組み。

（２）パート１：母集団パラメータへの信念を確率分布で表現



① 事前分布　$p(\theta)$

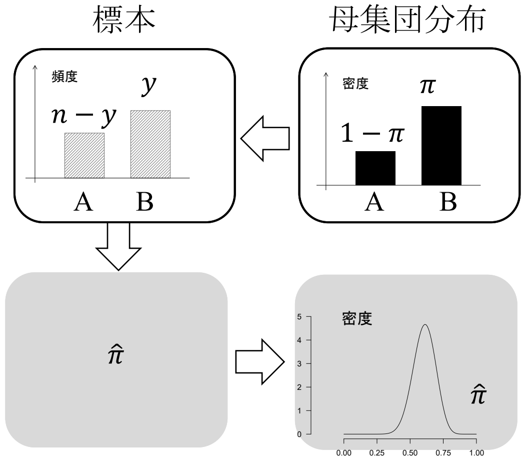
これは、データyを採る前に研究者が持っている母集団パラメータθに対する確信度を表現した確率分布。

② 事後分布　$p(\theta|y)$

これは、データyを採った後に研究者が持っている母集団パラメータθに対する確信度を表現した確率分布。

頻度主義の立場

確率は、自然／神が設定しているのであり、人間が設定するものではない。



ベイズ統計学の立場：人間の捉え方を反映した確率

確率を、人間の認識の不確かさをモデル化する道具立てとして採用する。

ポイント　人間は真実がどうであるかに関係なく、自分の視点から確率的な推論を行っている。

[1] 犯人は A さんかもしれない。

① 全知全能の視点：「A さんが犯人である」という命題に確率的な言明は不要。しいて言うなら、100%犯人か 100%犯人ではないか、のどちらか。

② 人間の視点：手元にあるデータからどのくらいその命題が確からしいのか、量化する。

[2] 明日雨が降る確率は 20～30%です。

① 全知全能の視点：「明日雨が降る」という命題に確率的な言明は不要。しいて言うなら、100%降るか 100%降らないか、決まっている。

② 人間の視点：手元にあるデータからどのくらいその命題が確からしいのか、量化する。

例題　**現代日本語の丁寧語構文の交替：信念のモデル化**

現代日本語では否定丁寧形で「走りません」「走らないです」のように「ません／ないです」という形が確率的に交替を起こすことが知られている。

そこで「ないです」という新規形を取る確率をπとして、i番目の文に登場する否定丁寧形が「ないです」を取るかどうかは、次のようなベルヌーイ分布に従うと考えた。

$$y_i \sim Bern(\pi)$$

ここでは、話を単純化するために、二つの構文の選択に特に独立変数は関与していない切片モデルを考え、πがどのような値を取るのかについてベイズ推論を行う。

○ ベイズ統計学におけるモデル式

（パート１）事前分布

$$p(\pi)$$

（パート２）尤度

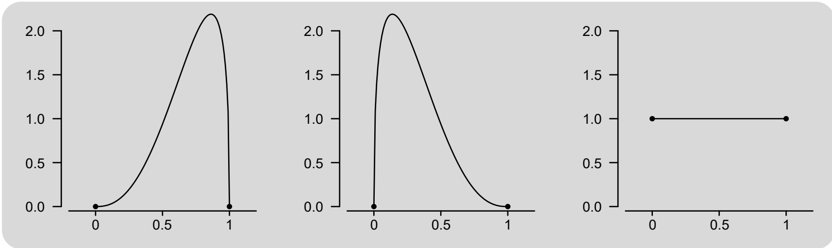
$$L(y_i|\pi) = {}_{n_i}C_{y_i}\pi^{y_i}(1-\pi)^{n_i-y_i}$$

（パート３）事後分布

$$p(\pi|y_i) \propto p(\pi)L(y_i|\pi)$$

○ 信念のモデル化

（例１）πのあたりをつける



① πの中心を１に近い値に想定する場合

先行研究や研究者自身の言語直感から「ないです」の方がよく生産されるのではないかと思った場合。

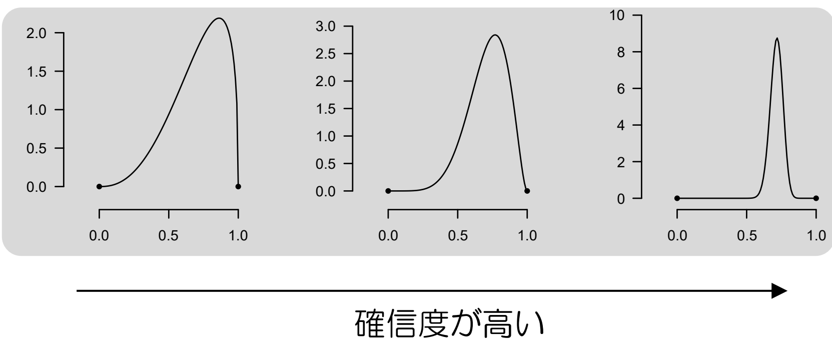
② πの中心を０に近い値に想定する場合

先行研究や研究者自身の言語直感から「ません」の方がよく生産されるのではないかと思った場合。

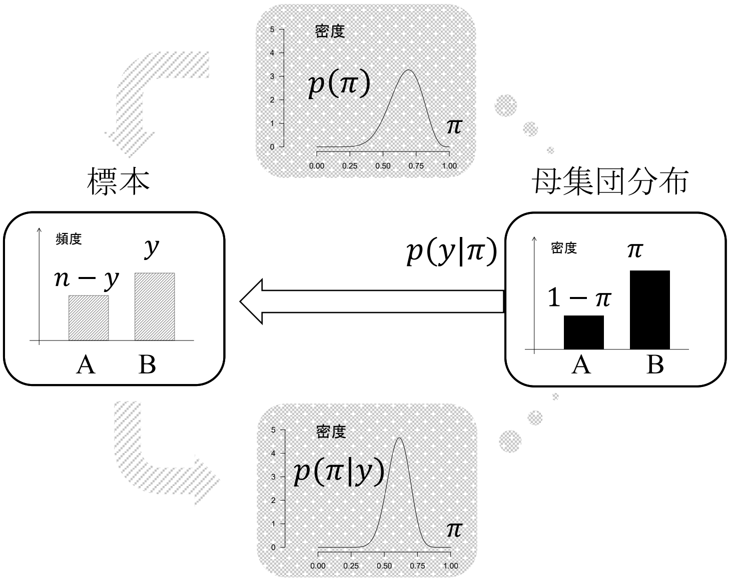
③ πに対しどこにも中心を作らない場合

先行研究や研究者自身の言語直感を議論に持ち込まず、どの値も同じくらいありうると考える場合。

（例２）πへの確信度を考える



（3） パート２：ベイズの定理で事前分布と事後分布をつなげる



分割表とベイズの定理

（1） 同時確率

$$p(x_0, y_0) = p(y_0|x_0)p(x_0)$$

$$= p(x_0|y_0)p(y_0)$$

X/Y	y_0	y_1	合計
x_0	0.4	0.1	0.5
x_1	0	0.5	0.5
合計	0.4	0.6	1

（2） ベイズの定理

$$p(x|y)p(y) = p(y|x)p(x)$$

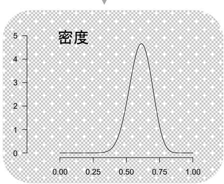
$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

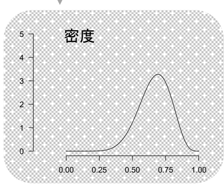
ベイズ統計学の根幹をなす三要素

① 事前分布： $p(\pi)$
② 尤度： $p(y|\pi)$
③ 事後分布： $p(\pi|y)$

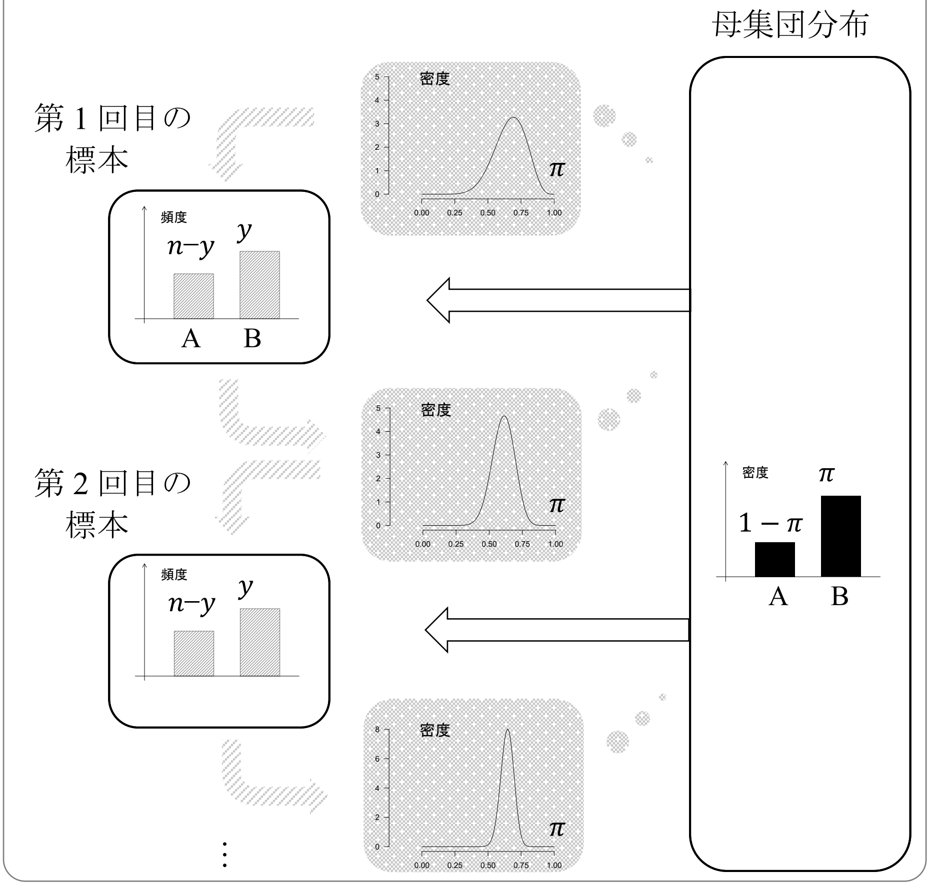
$$p(\pi|y) = \frac{p(y|\pi)p(\pi)}{p(y)}$$

$$\boxed{p(\pi|y)} \propto p(y|\pi)\boxed{p(\pi)}$$

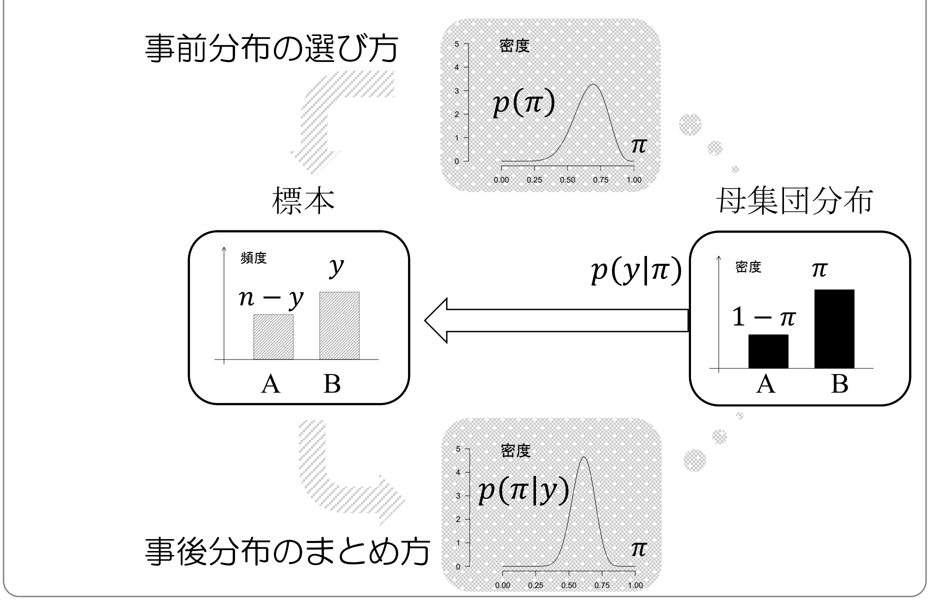




動的な推論過程としてのベイズ推論



ベイズ推論でしっかり議論をしないといけないところ



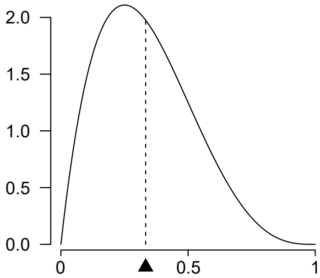
📖 ノート１　事後分布

事後分布の性質をどうやってまとめればいいの？

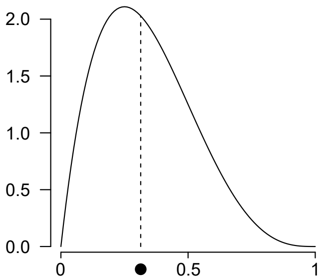
■ 点的な要約値

（１）分布の中心に関する指標

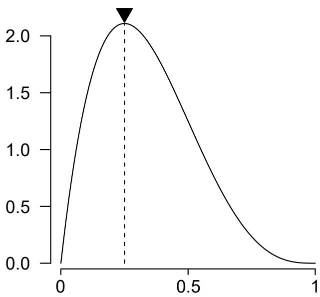
① 事後期待値 Expected a Posteriori (EAP)



② 事後中央値 Posterior Median (MED)



③ 事後確率最大値 Maximum a Posteriori (MAP)



（２）分布のばらつきに関する指標

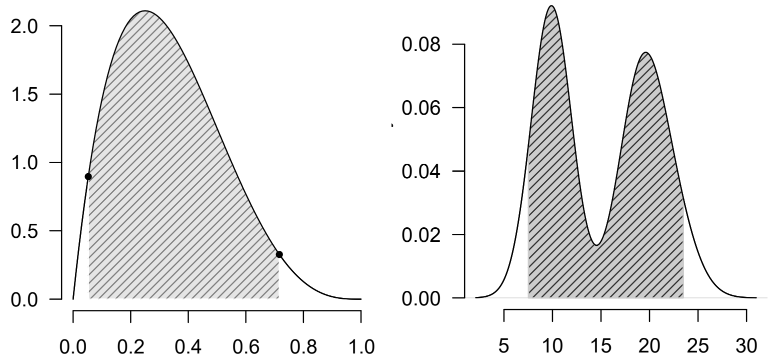
① 事後分散 Posterior Variance

② 事後標準偏差 Posterior Standard Deviation

■ 区間的な要約値

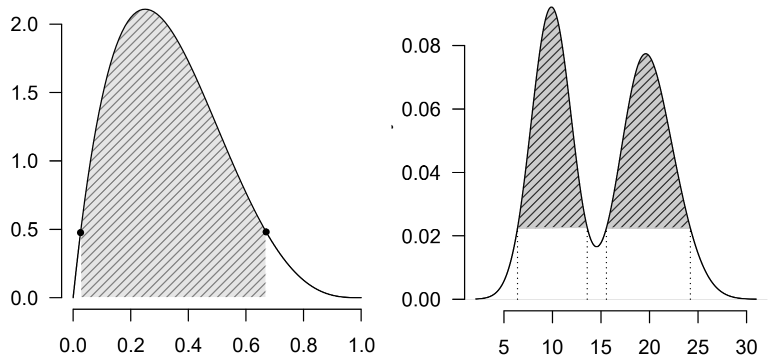
（１）確信区間 Credible Interval（ベイズ信頼区間）

これは、確率密度（事後分布の密度）の両端から$\alpha/2\%$の区間を切り取って残った中央部の$(100-\alpha)\%$の区間のこと。



（２）最高事後密度区間 Highest Posterior Density Interval (HPDI)

これは、確率密度（事後分布の密度）において、$(100-\alpha)\%$の密度の高い領域を含む区間のこと。



R における実装

```
library(coda)
#事後分布からデータを生成する
theta = as.mcmc(rbeta(1e6, 2, 40))
theta.hpd = HPDinterval(theta, prob = 0.95)
```

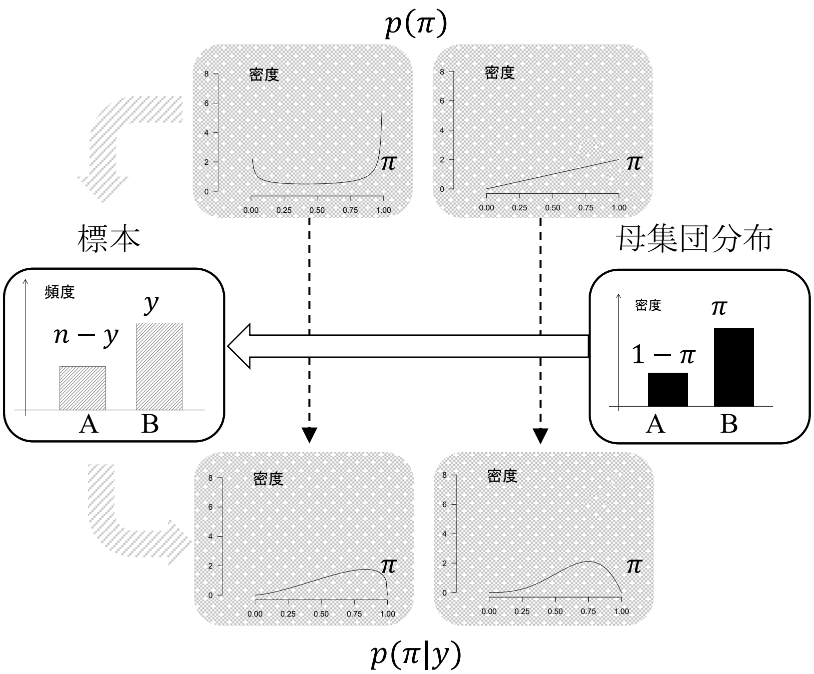
📖 ノート２　事前分布

事前分布って、どうやって選べばいいの？

（１）感度分析 Sensitivity Analysis

同じデータを採取したとしても、データを採る前の研究者の信念が異なると解釈対象の事後分布の形が変わる。

⇒そこで、複数の事前分布で試して結果の異同を検討する。



事前分布の選び方の四つのストラテジー

① 情報事前分布（主観事前分布）
⇒　先行研究から推測する

② 弱情報事前分布
⇒　「常識的」な範囲内で無情報にする

③ 無情報事前分布（客観事前分布）
⇒　可能な限り主観を排する

④ 計算が簡便な事前分布
⇒　可能な限り計算が楽なようにする
（数学的に美しい関係のあるものを使う）

（２）事前分布１：情報事前分布 Informative Prior

これは、過去の先行研究の事後分布を自分の研究の事前分布として採用するというもの。主観事前分布とも。

$$p(\theta|y) \propto L(y|\theta)\ \boxed{p(\theta)}$$

過去の知見（研究者の知識）を
反映するものを入れる

⇒ 実際の研究ではあまり多くはない。

母集団パラメータの定義域

① 正規分布$N(\mu,\sigma^2)$の　**位置パラメータμ**　は、2.1 や 3.3 のような小数点を取れるだけではなく、$-\infty$から∞までの実数の値の全てが候補となる。

② 正規分布$N(\mu,\sigma^2)$の　**尺度パラメータσ^2**　は、2.1 や 3.333…のような小数点は取れるけれど、二乗したものなので必ず正の値を取るという制約がある。

③ 二項分布$Binom(n,\pi)$の　**位置パラメータπ**　は、0.1 や 0.3 のような小数点は取れるけれど、2.1 や-0.11 のような 0 と 1 の間から飛び出した値は取ることができない。

このように、一口に母集団パラメータといっても、とりえる値にバリエーションがある。事前分布を設定するときにもこの定義域には気を付ける必要があり、πには 0 から 1 を定義域にとる確率分布を用いるべきだが、それをμ用に転用することには合理性がない。

（３）事前分布２：弱情報事前分布 Weakly Informative Prior

これは、無情報事前分布ほどではないが、なるべく特定の値に不条理な偏りがでないようにされた事前分布

$$p(\theta|y) \propto L(y|\theta)\ \boxed{p(\theta)}$$

そこまで強い研究者のバイアスを入れないことで大まかに情報がない状態を表現

弱情報事前分布の使いどころ

（１） 複雑なモデルへの対応

モデルが複雑になると、パソコンの計算が収束せず、この対処法として、常識や経験からある程度の値にゆるく絞った弱情報事前分布が用いられる。

（２） 過学習防止

過学習の恐れがある場合、絶対値の大きい回帰係数を避けるなどの目的でも弱情報事前分布は用いられる。

→　実は事前分布はリッジ回帰等の罰則に相当する。

弱情報事前分布の例

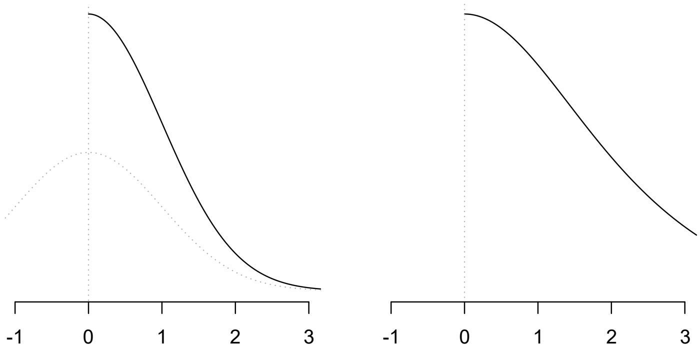
（１） $(-\infty, \infty)$の値を取るパラメータの場合

① 正規分布（情報がやや強め）

② t 分布（情報がやや弱め）

（２） $(0, \infty)$の値を取るパラメータの場合

① 半正規分布（情報がやや強め）　② 半 t 分布（情報がやや弱め）



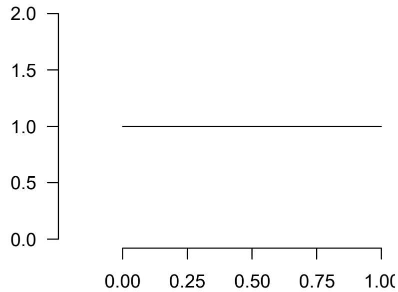
（４）事前分布３：無情報事前分布 Non-informative Prior

① 一様事前分布 Uniform Prior/Flat Prior

これは、シンプルにどの値が出るか、何の事前知識もない状態を表したもの。

$$p(\theta|y) \propto L(y|\theta)\ \boxed{p(\theta)}$$

研究者の主観に
関係なく
平坦なものを入れる



一様事前分布の哲学

「無情報」を確率分布が「平坦であること」として定義。
⇒　「平坦でない事前分布」は「無情報ではない」と考える。

※「無情報」の定義を「平坦であること」に限る必要はないので、この定義は、あり得る定義の中の一つに過ぎない。

一様分布 $Unif(\alpha,\beta)$

$$y_i \sim Unif(\alpha,\beta)$$

① 密度関数

$$p(\theta) = \frac{1}{\beta-\alpha}$$

② 期待値

$$E[y|\alpha,\beta] = \frac{\alpha+\beta}{2}$$

③ 分散

$$E[y|\alpha,\beta] = \frac{(\beta-\alpha)^2}{12}$$