

# DNA鑑定における仮説と事象の空間

検察側の主張 vs. 弁護側の主張を  
伝統的な統計検定の枠組みから眺めてみる

法数学勉強会

2010/08/28

京大(医)統計遺伝学

山田 亮

# DNA鑑定

検察側の主張

弁護側の主張

データ

裁判員・裁判官側の主張・判断

# 論文発表

著者側の主張

レフリー側の主張

データ

エディタの主張・判断

# 薬開発

開発部門の主張

経営部門の主張

データ

トップの主張・判断

# 特許申請

発明側の主張

?側の主張

データ

特許庁の主張・判断

# それぞれの判断

- 裁判
  - 疑わしきは、罰せず？

# それぞれの判断

- 論文

- 統計的に有意であれば掲載する？
- 読んで面白ければ掲載する？

# それぞれの判断

- 薬の開発
  - 社全体でペイするならば続行する？
- 特許
  - データがどうあれ、儲かる種なら押える



# それぞれの判断

- 特許
  - データがどうあれ、儲かる種なら押える

# それぞれの判断

- 裁判
  - 疑わしきは、罰せず？
- 論文
  - 統計的に有意であれば掲載する？
  - 読んで面白ければ掲載する？
- 薬の開発
  - 社全体でペイするならば続行する？
- 特許
  - データがどうあれ、儲かる種なら押える

# それぞれの判断

- データ→主張・判断を数値化
  - やり方は色々？
  - 納得がいくかどうか
  - 万人に共通な言葉かどうか
    - 論理学
    - 数学
- 確率・尤度・事前確率・事後確率・ベイズの定理

# いくつかの視点

- 尤度と尤度比
- 珍しさの尺度 P値
- (最尤)推定と推定の信頼区間

作ったスライドのどこまで話せるかは不明・・・

時間が余ったら、DNA鑑定の現実に即した例について検討したいです。

時間が足りなかったら、またの機会に続きを検討できればと思います。

ゆっくり行きましょう。

# 尤度と尤度比

# DNA鑑定

検察側の主張

弁護側の主張

データによれば  
『Xが犯人である』  
と信じる？

データ

裁判員・裁判官側の主張・判断

# DNA鑑定

検察側の主張

弁護側の主張

データによれば  
『Xが犯人である』  
と信じる？

『Xが犯人である』とすると  
データの説明が容易なので  
『Xが犯人である』と言えます。

データ

『Xが犯人でない』とすると  
データの説明が容易なので  
『Xが犯人でない』と言えます。

裁判員・裁判官側の主張・判断

『Xが犯人である』とすると  
データの説明が容易なので  
『Xが犯人である』と言えます。

# 説明が容易

『Xが犯人でない』とすると  
データの説明が容易なので  
『Xが犯人でない』と言えます。

- ある仮説で説明が容易だからって、その仮説が正しいとは限らないでしょう
- ある仮説Aが、別の仮説Bよりも説明が容易なら、AをBより信用しましょう
- 尤度
  - 仮説のもっともらしさ
  - 説明が容易な程度
- 尤度比
  - 仮説Aの尤度は仮説Bの尤度の●倍



「観察データを得る」  
と  
「Xが犯人であること」

	観察データ	観察データ以外のデータ	確率の和
犯人である	A	B	$A+B=1$
犯人ではない	C	D	$C+D=1$
	$A+C$	$B+D$	$A+B+C+D$

「観察データを得る」  
と  
「Xが犯人であること」

仮説	観察データ	観察データ以外のデータ	確率の和
犯人である	A	B	$A+B=1$
犯人ではない	C	D	$C+D=1$
	$A+C$	$B+D$	$A+B+C+D$

「観察データを得る」  
と  
「Xが犯人であること」

事象

	観察データ	観察データ以外のデータ	確率の和
犯人である	A	B	$A+B=1$
犯人ではない	C	D	$C+D=1$
	$A+C$	$B+D$	$A+B+C+D$

# 仮説の軸と事象の軸

			事象
仮説	観察データ	観察データ以外のデータ	確率の和
犯人である	A	B	$A+B=1$
犯人ではない	C	D	$C+D=1$
	$A+C$	$B+D$	$A+B+C+D$

# 仮説の軸について

「Xが犯人である」仮説  
と  
「Xが犯人でない」仮説

	観察データ	観察データ以外のデータ	確率の和
犯人である	A	B	$A+B=1$
犯人ではない	C	D	$C+D=1$
	$A+C$	$B+D$	$A+B+C+D$

『Xが犯人である』  
ときに観察データを  
得る確率

# 「観察データを得る」 と 「Xが犯人であること」

	観察データ	観察データ以外 のデータ	確率の和
犯人	<div data-bbox="233 579 645 851" style="border: 1px solid black; background-color: #f9cb9c; padding: 5px;">                     観察データを得たときに Xが犯人である尤度                 </div> A	<div data-bbox="780 579 1192 851" style="border: 1px solid black; background-color: #c8e6c9; padding: 5px;">                     『Xが犯人である』 ときに観察データを得る確率                 </div> B	<div data-bbox="1481 714 1908 899" style="border: 1px solid black; background-color: #e8f5e9; padding: 10px;"> <math>A+B=1</math> </div>
犯人ではない	C	D	<div data-bbox="1481 899 1908 1071" style="border: 1px solid black; background-color: #e8f5e9; padding: 10px;"> <math>C+D=1</math> </div>
尤度の和	A+C	B+D	<div data-bbox="48 1129 1360 1282" style="border: 1px solid black; background-color: #fce4d6; padding: 10px;"> <math>A+B+C+D</math> </div>

# 4つの数字A, B, C, D 解釈は1つ

	観察データ	観察データ 以外のデー タ	確率の和
犯人である	A	B	$A+B=1$
犯人ではな い	C	D	$C+D=1$
	$A+C$	$B+D$	$A+B+C+D$



# 4つの数字A, B, C, D

## 解釈は1つ

『Xが犯人である』とすると  
データの説明が容易なので  
『Xが犯人である』と言えます。  
「Aが大きい」

『Xが犯人でない』とすると  
データの説明が容易なので  
『Xが犯人でない』と言えます。  
「Cが大きい」

- AとCとを較べて大きい方はどちらか
- 大差がついているか

$$\frac{A}{A+C}, \quad \frac{C}{A+C}, \quad \frac{A}{C}$$

	観察データ	観察データ以外のデータ	確率の和
犯人である	A	B	A+B=1
犯人ではない	C	D	C+D=1
	A+C	B+D	A+B+C+D

# ある仮説Aが、別の仮説Bよりも説明が容易なら、AをBより信用しましょう

『Xが犯人である』とするとデータの説明が容易なので『Xが犯人である』と言えます。  
「Aが大きい」

『Xが犯人でない』とするとデータの説明が容易なので『Xが犯人でない』と言えます。  
「Cが大きい」

- AとCとを較べて大きい方はどちらか
- 大差がついているか

$$\frac{A}{A+C}, \quad \frac{C}{A+C}$$

$$\frac{A}{C}$$

尤度比

	観察データ	観察データ以外のデータ	確率の和
犯人である	A	B	A+B=1
犯人ではない	C	D	C+D=1
	A+C	B+D	A+B+C+D

- AとCとを較べて大きい方はどちらか
- 大差がついているか

$$\frac{A}{A+C}, \quad \frac{C}{A+C}, \quad \frac{A}{C}$$

尤度比

どうして割り算をする？

どうして比をとる？

	観察データ	観察データ以外のデータ	確率の和
犯人である	A	B	A+B=1
犯人ではない	C	D	C+D=1
	A+C	B+D	A+B+C+D

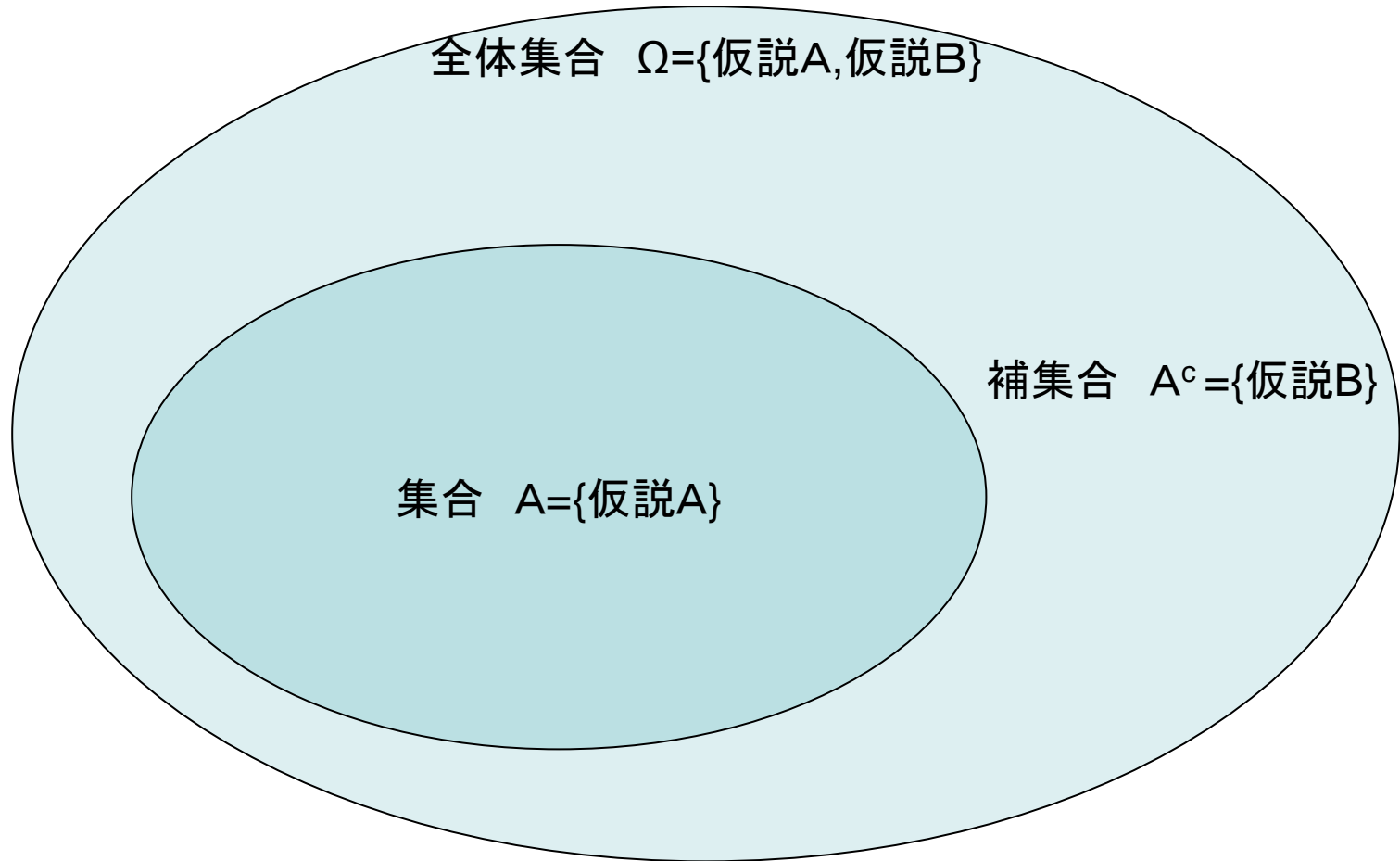
# 集合としての仮説と仮説空間

- 仮説A『Xが犯人である』
- 仮説B『Xが犯人でない』
- AかBか。それ以外はない

$$\Omega = \{A, B\}$$

$$\Omega = \{A\} \cup \{B\}$$

# 全体集合 集合と補集合



# 犯人はXか●か

検察側の主張

弁護側の主張

データによれば  
『Xが犯人である』  
と信じる？

『Xが犯人である』とすると  
『X以外の●が犯人である』とするより  
データの説明が容易なので  
『Xが犯人である』と言えます。

データ

『Xが犯人でない』とすると  
『X以外の●が犯人である』ことになり  
『Xが犯人である』とするより  
データの説明が容易なので  
『Xが犯人でない』と言えます。

裁判員・裁判官側の主張・判断

# 4つの数字A, B, C, D

## 解釈は1つ

『Xが犯人である』とすると  
『X以外の●が犯人である』とするより  
データの説明が容易なので  
『Xが犯人である』と言えます。  
「Aが大きい」

『Xが犯人でない』とすると  
『X以外の●が犯人である』ことになり  
『Xが犯人である』とするより  
データの説明が容易なので  
『Xが犯人でない』と言えます。  
「Cが大きい」

- AとCとを較べて大きいほうはどちらか
- 大差がついているか

$$\frac{A}{A+C}, \quad \frac{C}{A+C}, \quad \frac{A}{C}$$

	観察データ	観察データ以外のデータ	確率の和
Xが犯人である	A	B	A+B=1
●が犯人である	C	D	C+D=1
	A+C	B+D	A+B+C+D

「Xは犯人である」 対 「Xは犯人ではない」  
「Xは犯人である」 対 「●は犯人である」

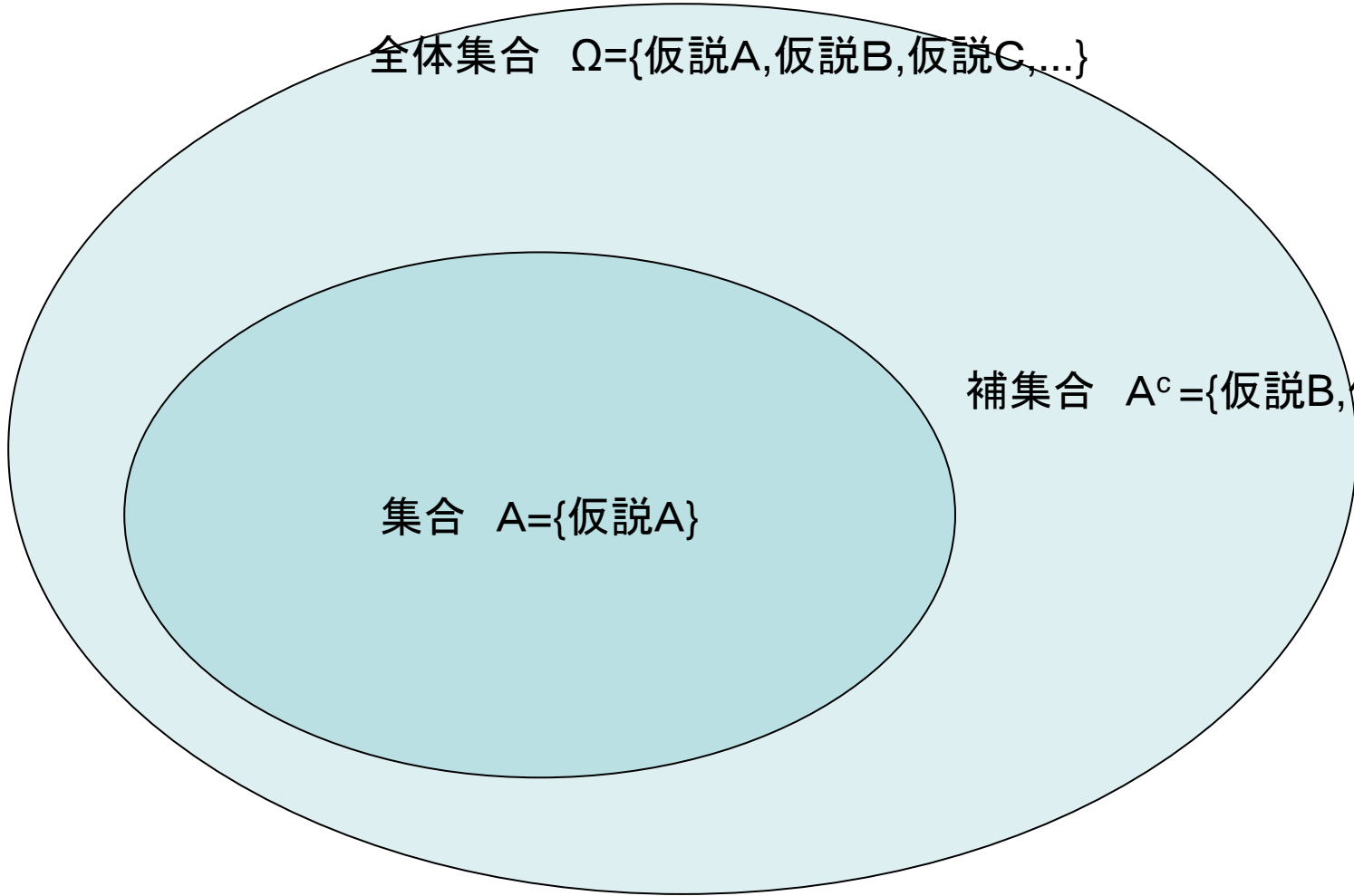
- 「Xは犯人ではない」とき、「犯人は誰か？」  
– 「犯人は●」「犯人は▲」「犯人は■」・・・
- 「犯人の候補は全部で何人？」
- 「犯人の候補のリストは？」



全体集合  $\Omega = \{\text{仮説A}, \text{仮説B}, \text{仮説C}, \dots\}$

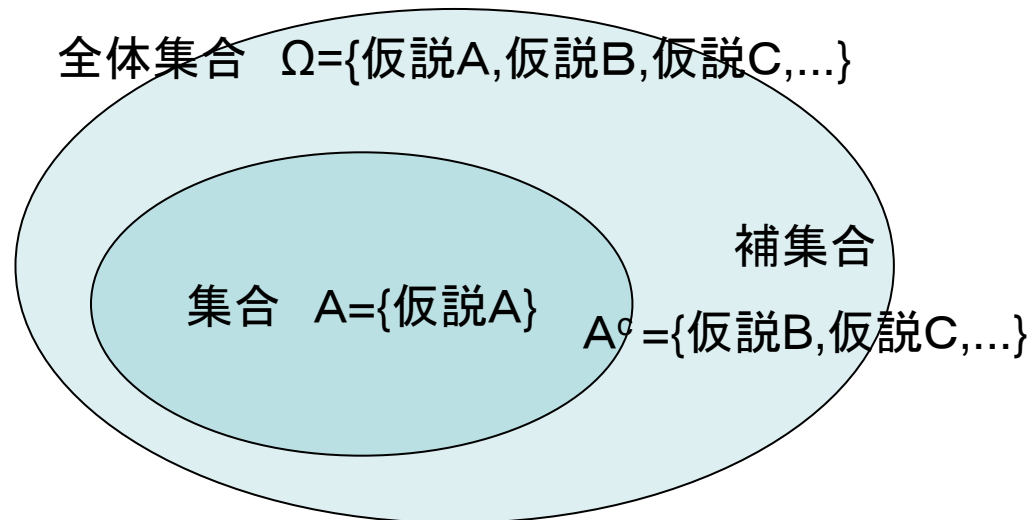
補集合  $A^c = \{\text{仮説B}, \text{仮説C}, \dots\}$

集合  $A = \{\text{仮説A}\}$



# 仮説空間と仮説

- 仮説空間は全体集合
- 仮説は仮説空間の要素
- 要素は排他的
- 要素の和は全体



# もう少しつつこんで

- 「Xは犯人ではない」とき、「犯人は誰か？」
- 「Xは犯人ではない」とき、「犯人はいるか？」
- DNA試料で考えると
  - 「試料DNAはXのDNAである」
  - 「試料DNAは●のDNAである」
  - 「試料DNAは誰かのDNAである」
  - 「試料DNAは誰のDNAでもない」

一番大事なのは、数え落としがないこと  
数え上げた上で、必要なら消去しよう

# 仮説の数だけ尤度

犯人は誰？	尤度	
A	$L_a$	
B	$L_b$	
...	...	
X	$L_x$	
犯人はいない	$L_0$	

# 仮説の数だけ尤度

犯人は誰？	尤度	
A	$L_a$	
B	$L_b$	
...	...	
X	$L_x$	
犯人はいない	$L_0$	

$$\frac{A}{A+C}, \quad \frac{C}{A+C}, \quad \frac{A}{C}$$

# 仮説の数だけ尤度

犯人は誰？	尤度	
A	$L_a$	
B	$L_b$	
...	...	
X	$L_x$	
犯人はいない	$L_0$	

$$\frac{A}{A+C}, \quad \frac{C}{A+C}, \quad \frac{A}{C}$$

$L_x/L_a$

$L_x/L_b$

...

$L_x/L_0$

のすべてが十分に大きい

$L_x/(L_a+L_b+\dots L_x+L_0)$

が十分に大きい

# 仮説の数だけ尤度

犯人は誰？	尤度	
A	$L_a$	
B	$L_b$	
...	...	
X	$L_x$	
犯人はいない	$L_0$	

$$\frac{A}{A+C}, \quad \frac{C}{A+C}, \quad \frac{A}{C}$$

$$L_x/L_a > t$$

$$L_x/L_b > t$$

...

$$L_x/L_0 > t$$

のすべてが十分に大きい



$$L_x/(L_a+L_b+\dots+L_x+L_0) > t/N_h$$

が十分に大きい

$N_h$ : X以外の仮説数

# 仮説の数だけ尤度

犯人は誰？	尤度	
A	$L_a$	
B	$L_b$	
...	...	
X	$L_x$	
犯人はいない	$L_0$	

$$\frac{A}{A+C}, \quad \frac{C}{A+C}, \quad \frac{A}{C}$$

$$L_x/L_a > L_x/(L_a+L_b+\dots L_x+L_0) > t$$

$$L_x/L_b > L_x/(L_a+L_b+\dots L_x+L_0) > t$$

...

$$L_x/L_0 > L_x/(L_a+L_b+\dots L_x+L_0) > t$$

のすべてが十分に大きい



$$L_x/(L_a+L_b+\dots L_x+L_0) > t$$

が十分に大きい



# 仮説の数だけ尤度

$$\frac{A}{A+C}, \quad \frac{C}{A+C}, \quad \frac{A}{C}$$

$(L_a+L_b+\dots L_x+L_0)$

を知るためには、仮説空間を決めなくては

$$L_x/L_a > L_x/(L_a+L_b+\dots L_x+L_0) > t$$

$$L_x/L_b > L_x/(L_a+L_b+\dots L_x+L_0) > t$$

...

$$L_x/L_0 > L_x/(L_a+L_b+\dots L_x+L_0) > t$$

のすべてが十分に大きい



$$L_x/(L_a+L_b+\dots L_x+L_0) > t$$

が十分に大きい

# 仮説空間の広さ

- 「犯人は誰だ？」
  - 仮説空間を広げる
    - 「AもBもCもDも...、みんな犯人かもしれない」
    - 「アリバイのない人はみんな」
    - 「容疑者以外もみんな」

Lx/La

Lx/Lb

...

Lx/L0

のすべてが十分に大きい

# すべての人を仮説空間にとる(?)

- 代表的な人ではなく
- すべての人

一番大事なのは、数え落としがないこと  
数え上げた上で、必要なら消去しよう

# すべての人を仮説空間にとる(?)

- 代表的な人ではなく
- すべての人

一番大事なのは、数え落としがないこと  
数え上げた上で、必要なら消去しよう

- 代表的な人
  - 集団の平均値(など)を利用
  - マーカーごとの平均値
  - マーカー組み合わせの平均値・・・出身地を考慮???
- すべての人
  - すべてが無理でも、『十分に多くの標本』があれば、標本から「すべての人」の分布のばらつき(の可能性)を考慮して、『すべての人』を仮説空間に取れる(か?)

# 仮説の重み付け

- 「犯人らしいのは誰だ？」
  - 仮説空間は変えずに、重み付け( $W_i$ )を変える
    - 現場にいなかったことが確かな人が遺留品を残す可能性は「非常に低い(ゼロではない)」～無視する??
    - その他の証拠により、「疑わしい容疑者」と「疑わしくない容疑者」がいるなら、それで重み付けする

$L_x/L_a$

$L_x/L_b$

...

$L_x/L_0$

のすべてが十分に大きい



$L_x \cdot W_x / (L_a \cdot W_a)$

$L_x \cdot W_x / (L_b \cdot W_b)$

...

$L_x \cdot W_x / (L_0 \cdot W_0)$

のすべてが十分に大きい

# 仮説の重み付け

- 「犯人らしいのは誰だ？」
  - 仮説空間は変えずに、重み付け( $W_i$ )を変える
    - 現場にいなかったことが確かな人が遺留品を残す可能性は「非常に低い(ゼロではない)」～無視する??
    - その他の証拠により、「疑わしい容疑者」と「疑わしくない容疑者」がいるなら、それで重み付けする

重み付け $W_i$ はどれくらい正確？

心象の影響を受ける???

人によって異なる???

ならば、 $W_i$ なしでデータ提示するしかないのか・・・

$$L_x * W_x / (L_a * W_a)$$

$$L_x * W_x / (L_b * W_a)$$

...

$$L_x * W_x / (L_0 * W_0)$$

のすべてが十分に大きい

# 仮説の重み付け

- 「犯人らしいのは誰だ？」
  - 仮説空間は変えずに、重み付け( $W_i$ )を変える
    - 現場にいなかったことが確かな人が遺留品を残す可能性は「非常に低い(ゼロではない)」～無視する??
    - その他の証拠により、「疑わしい容疑者」と「疑わしくない容疑者」がいるなら、それで重み付けする

$L_a \leq 1$  なので

$$L_x \cdot W_x / (L_a \cdot W_a) \geq L_x \cdot W_x / W_a$$

$L_x \cdot W_x / W_a$  が十分に大きい仮説Aは  
 $L_a$  が計算できなくても無視してよい

$$L_x \cdot W_x / (L_a \cdot W_a)$$

$$L_x \cdot W_x / (L_b \cdot W_a)$$

...

$$L_x \cdot W_x / (L_0 \cdot W_0)$$

のすべてが十分に大きい

# 仮説の省略と不等号の向き

## 検察の仕事

$L_a \leq 1$  なので

$$L_x * W_x / (L_a * W_a) \geq L_x * W_x / W_a$$

この不等号の向きを確かめるためならば

$L_x * W_x / W_a$  が十分に大きい仮説Aは  
 $L_a$  が計算できなくても無視してよい

$$L_x * W_x / (L_a * W_a)$$

$$L_x * W_x / (L_b * W_a)$$

...

$$L_x * W_x / (L_0 * W_0)$$

のすべてが十分に大きい

## 弁護の仕事

$L_x \leq 1$  なので

$$L_x * W_x / (L_a * W_a) \leq W_x / (L_a * W_a)$$

この不等号の向きを確かめるためならば

$W_x / (L_a * W_a)$  が十分に小さい仮説Aを  
提示すればよい

$$L_x * W_x / (L_a * W_a)$$

$$L_x * W_x / (L_b * W_a)$$

...

$$L_x * W_x / (L_0 * W_0)$$

のどれか一つが十分に小さい



# 仮説の省略と不等号の向き

$Lx \cdot Wx / Wa$ が十分に大きい

$Wx / (La \cdot Wa)$ が十分に小さい

左右で用いるWが異なれば、両方が同時に成立する

重み付け $W_i$ はどれくらい正確？

心象の影響を受ける???

裁判員・裁判官によって異なる???

ならば、 $W_i$ なしでデータ提示するしかないのか...

$Lx \cdot Wx / (La \cdot Wa)$

$Lx \cdot Wx / (Lb \cdot Wa)$

...

$Lx \cdot Wx / (L0 \cdot W0)$

の**すべてが**十分に大きい

$Lx \cdot Wx / (La \cdot Wa)$

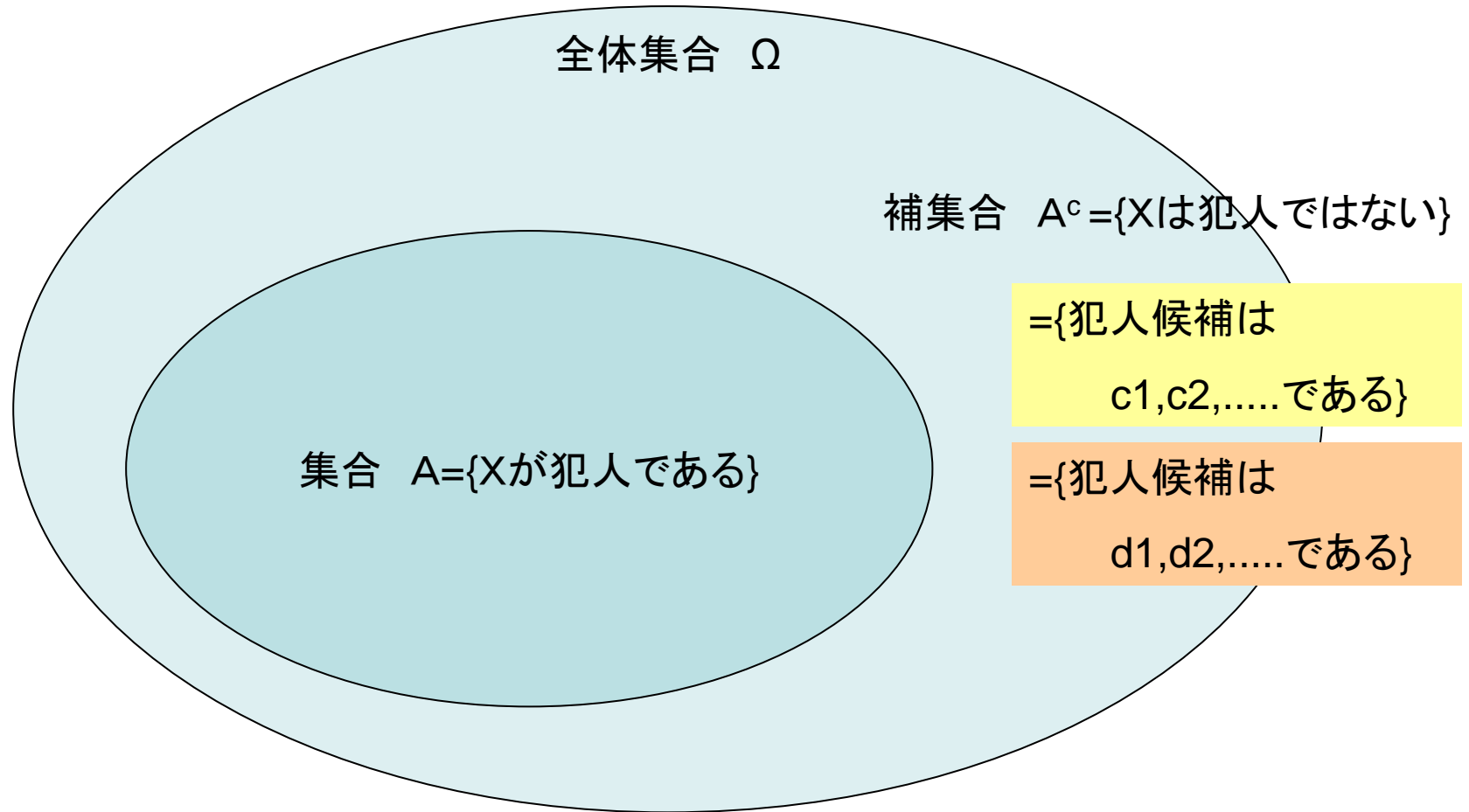
$Lx \cdot Wx / (Lb \cdot Wa)$

...

$Lx \cdot Wx / (L0 \cdot W0)$

の**どれか一つが**十分に小さい

# 異なる仮説空間は較べられない



# 複合的な仮説

- 試料DNAはXのものである
  - 試料DNAはXと●のものである
  - 試料DNAはXと●と▲のものである
  - 試料DNAはXと●と▲と...
- 
- 試料DNAは●のものである
  - 試料DNAは●と▲のものである
  - ...

# 複合的な仮説

- 複合的な仮説のそれぞれは、仮説空間の個々の要素
  - 相互に排他的
  - 全部を合わせると全体集合になる
- では、「Xは犯人である」という仮説は？
  - 試料DNAはXのものである
  - 試料DNAはXと●のものである
  - 試料DNAはXと●と▲のものである
  - 試料DNAはXと●と▲と...  
の和集合

# 複合的な仮説の一部を省略する

- その省略は本当に全体の解釈に影響を与えないほど小さいのか
  - その確認はどうやって計算するのか・・・

# 多すぎる仮説

分子・分母の項を減らして不等式が満足できる？

- 「非常に低い(ゼロではない)」～無視する？
  - $L1=L(X)+L(Xと●)+L(Xと●と▲)+L(Xと●と▲と…)+…$
  - $L2=L(●)+L(●と▲)+L(●と▲と…)+…$
  - $L1/L2$
  
  - $L1w=L(X)*W(x)+L(Xと●)*W(Xと●)+…$
  - $L2w=L(●)*W(●)+L(●と▲)*W(●+▲)+…$
  - $L1w/L2w$
  
  - $L1/L2 > t, L1/L2 < t', L1w/L2w > t, L1w/L2w < t'$ の不等式を説明するのに、分子と分母のどちらかの仮説のみを省略することは可能だが、分子・分母を省略するとき、簡単に不等式を保証できない……

实例？

# まとめ1

## 仮説空間のこと

- 仮説はいくつ？
- 仮説は排他的？
- 仮説を集めると全体になっている？
- 補集合となっている仮説の構成は？
- 興味のある仮説は単独か複合か？



## まとめ2

# 尤度の計算は仮説ごとの確率計算

- 仮説ごとに確率は計算できる？
  - 事象の場合わけは済んでいる？
  - 事象の確率計算は一意に決まっている？
  - 事象の確率計算が一意でなければ、尤度の計算も変わってくる

# まとめ3

## 仮説の重み付け(事前確率)のこと

- 重み付けの根拠は？
  - DNA試料データ以外のデータ
    - 心象も入る...
    - 裁判員・裁判官によって変わるもの
      - もちろん、検察・弁護は異なる「心象」に立っている
- 重み付けは一意に決まる？
  - 決まらなければ、コンセンサスを取らないと先に進めない
- 重み付けは複数ある？
- 新たな重みが出るたびに計算しなおすくらいなら、重み付け計算は「つどつど」にするしかない？

# 事象の空間

珍しさの尺度 P値

尤度の比較



仮説の棄却

# DNA鑑定

検察側の主張

弁護側の主張

データによれば  
『Xが犯人である』  
と信じる？

『Xが犯人である』とすると  
データの説明が容易なので  
『Xが犯人である』と言えます。

データ

『Xが犯人でない』とすると  
データの説明が容易なので  
『Xが犯人でない』と言えます。

裁判員・裁判官側の主張・判断

# 仮説の棄却

検察側の主張

弁護側の主張

データによれば  
『Xが犯人である』  
と信じる？

『Xが犯人で**ない**』とすると  
データの説明が難しいので  
『Xが犯人で**ある**』と言えます。

データ

『Xが犯人で**ある**』とすると  
データの説明が難しいので  
『Xが犯人で**ない**』と言えます。

裁判員・裁判官側の主張・判断

# 尤度の比較 仮説の棄却

検察側の主張

弁護側の主張

データによれば  
『Xが犯人である』  
と信じる？

『Xが犯人で**ない**』とすると  
データの説明が難しいので  
『Xが犯人で**ある**』と言えます。

データ

『Xが犯人で**ある**』とすると  
データの説明が難しいので  
『Xが犯人で**ない**』と言えます。

『Xが犯人で**ある**』とすると  
データの説明が容易なので  
『Xが犯人で**ある**』と言えます。

官側の

『Xが犯人で**ない**』とすると  
データの説明が容易なので  
『Xが犯人で**ない**』と言えます。

# 仮説の棄却

『Xが犯人で**ない**』とすると  
データの説明が難しいので  
『Xが犯人で**ある**』と言えます。

『Xが犯人で**ある**』とすると  
データの説明が難しいので  
『Xが犯人で**ない**』と言えます。

- データをもたらしにくい仮説は「真でない」と棄却する
- 「『棄却した仮説』とは異なる仮説」を信じてもいいかな



# 正確P値

- ある仮説のもとでは、いろいろなデータを得る可能性がある
- それぞれのデータを取得する確率が計算できる
- その確率の和は1
- 今、あるデータが得られたときに、そのデータと同程度かそれ未満の確率であるようなデータのすべての確率を足し合わせたものが正確P値
- データの珍しさが0-1で表される
  - 最も確率の高いデータが得られたときのP値は1
  - 最も確率の低いデータが得られたときのP値はそのデータを得る確率そのもの

# 4つの数字A, B, C, D

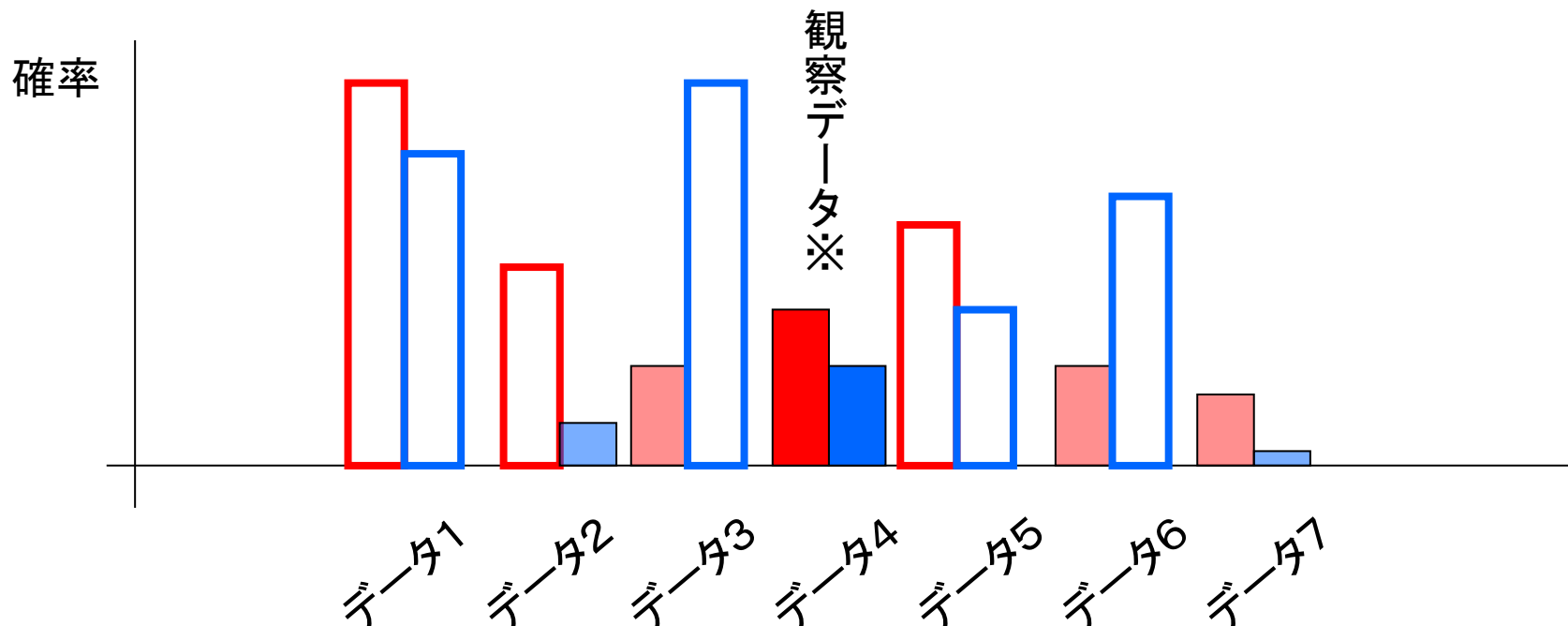
## 解釈は1つ

『Xが犯人で**ない**』とすると  
データの説明が難しいので  
『Xが犯人で**ある**』と言えます。

データ4の確率が低いです！

『Xが犯人で**ある**』とすると  
データの説明が難しいので  
『Xが犯人で**ない**』と言えます。

データ4の確率が小さいです！



# 観察データ(事象)の軸について

- ある観察をしたときに、起こりうる現象の集合
  - 個々の要素が事象
  - 個々の要素は排他的
  - 全部の事象を足し合わせると、現象全体になる
  - 個々の事象の起きる確率を足し合わせると1になる
  - 仮説空間と同じく事象空間がある
  - 確率的に起きること
  - 実験エラーを考慮すると、起こりえない事象はない
  - 観察データは、複数の「真実」に対応しているかもしれない
    - 分解が必要
  - こちらは裁判官・裁判員の心象を排除できるはずの領域・・・
  - 実験精度・エラーデータなどについて、実験者の心象は排除できない、か(?)

# (最尤)推定と推定の信頼区間

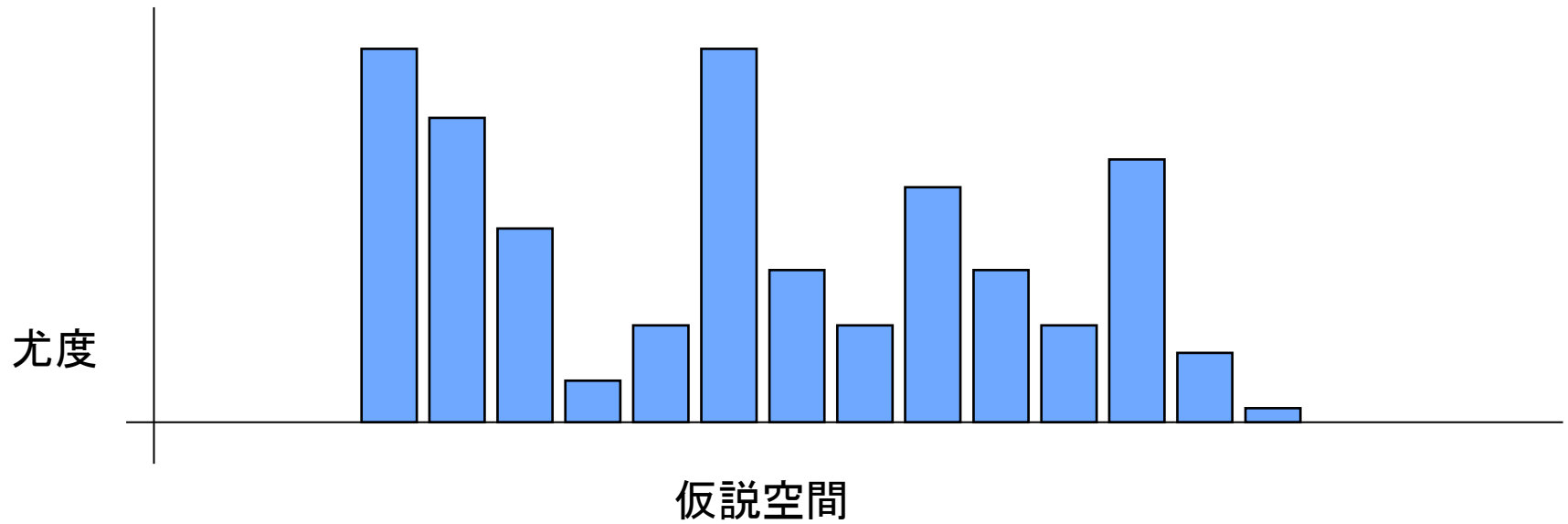
観察データのみに着目して、  
尤度に話を戻します

# 最尤推定

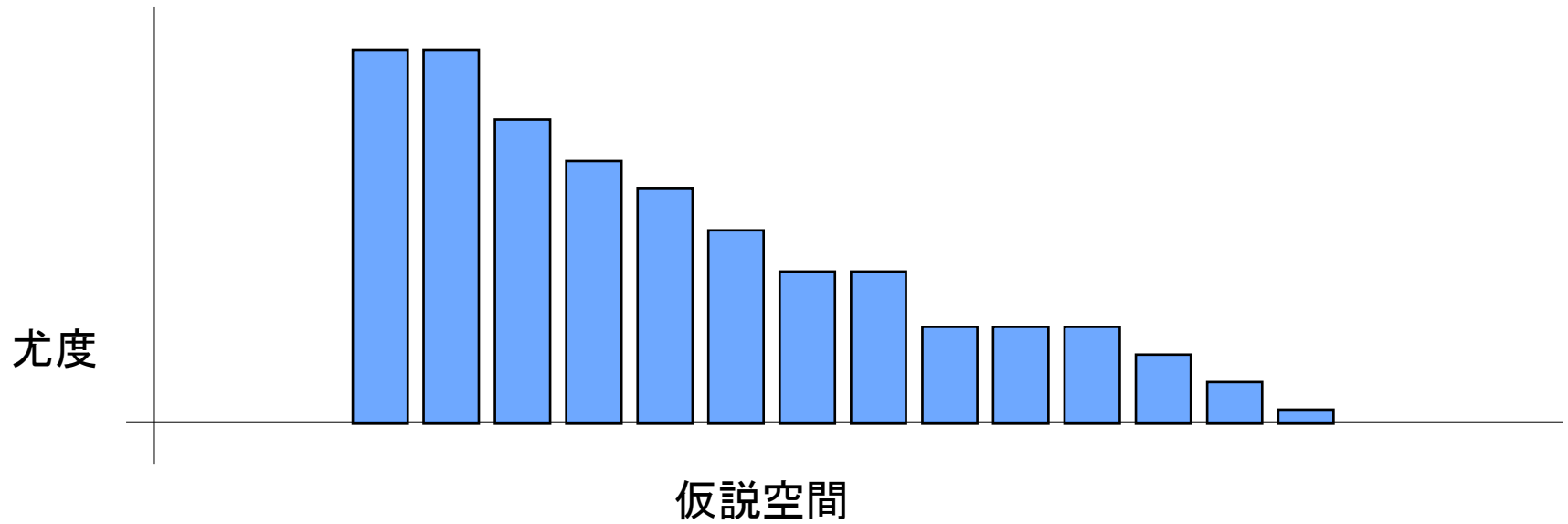
犯人は誰？	尤度
A	$L_a$
B	$L_b$
...	...
X	$L_x$
犯人はいない	$L_0$

- 仮説最尤推定  
仮説  
– 数ある仮説の中で尤度が最大の

# すべての仮説について尤度を求める



# 尤度の高い順に並べて

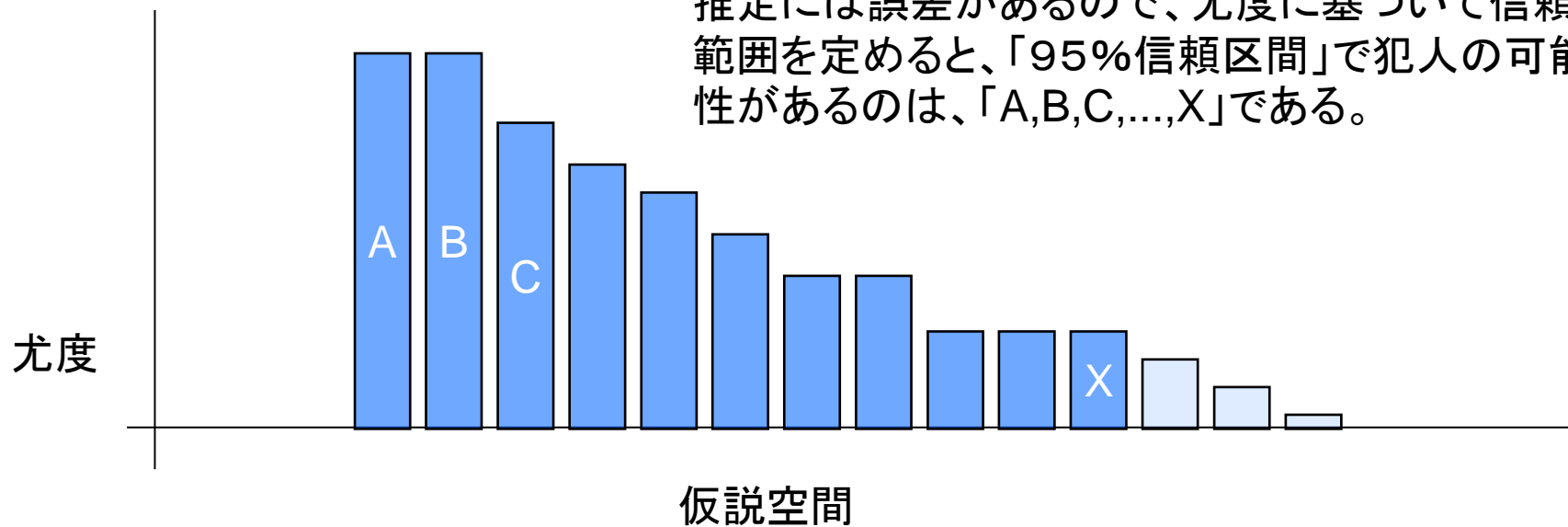




# 全体を1とみなして、上位から95%を占める仮説の範囲を定める

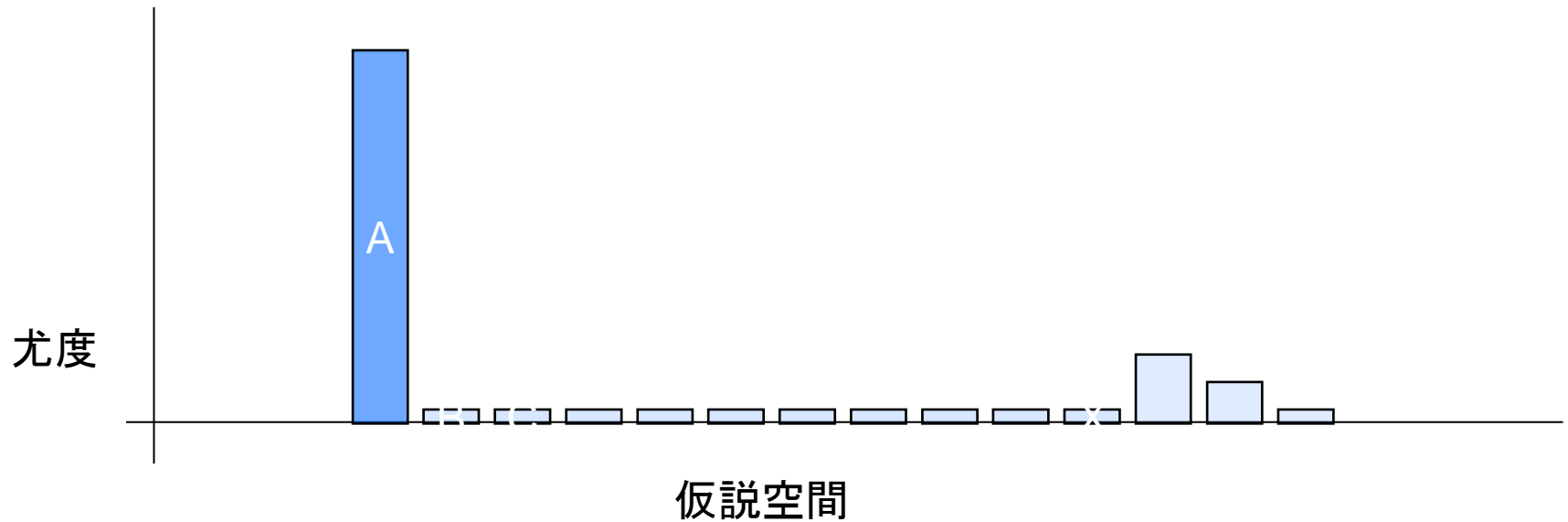
最も尤度の高い犯人(最尤推定の犯人)はAである。

推定には誤差があるので、尤度に基づいて信頼範囲を定めると、「95%信頼区間」で犯人の可能性があるのは、「A,B,C,...,X」である。

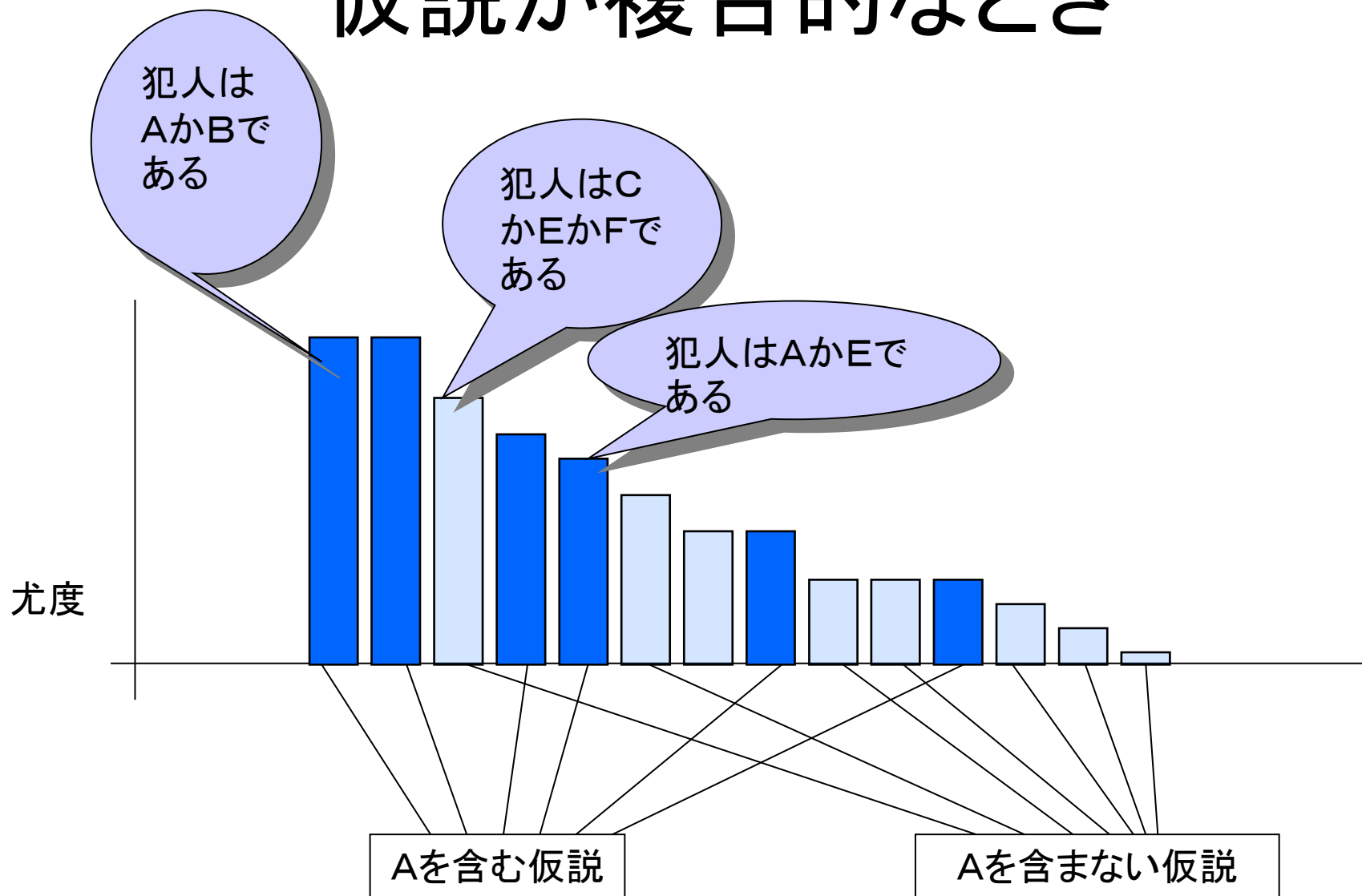


# 全体を1とみなして、上位から $\alpha\%$ を占める仮説の範囲を定める

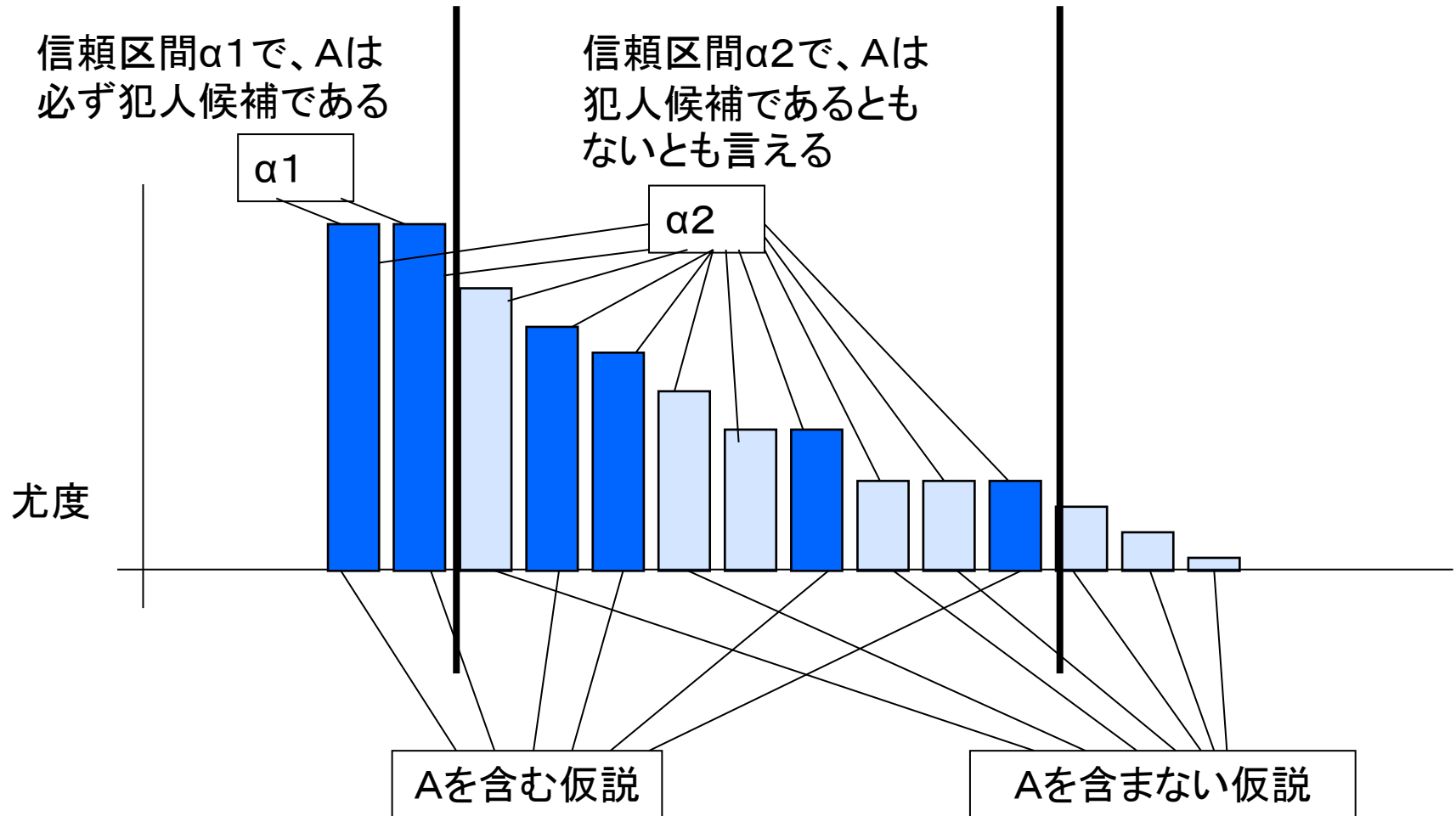
「 $\alpha\%$ 信頼区間」で犯人の可能性のあるのは、「Aだけ」である。



# 仮説が複合的なとき



# 仮説が複合的なとき



实例？

# その他に気づくことなど(1)

- DNA鑑定の有効数字のこと
  - DNA鑑定は証拠の1つ
  - 他の証拠が「犯人はXである」「犯人はXではない」という仮説に関して、おおまかな尤度しか与えないとき、DNA鑑定が細かすぎる尤度を与えているのではないか？

# その他に気づくことなど(2)

- DNA鑑定も証拠の1つ
- その他の証拠
  - 証言「この人が現場で被害者を刺していました」
  - 証言「この人が●時頃、×を走っていました」
  - 指紋が検出されました。その指紋は「容疑者のものと『断定』されました」・・・『断定』とは？
- これらと同じ次元で考えなくては使えないのでは・・・
- これらは証拠としてどう使われている？
- その目で見直すとDNA鑑定はどう見えてきて、その他の「証拠」はどう見えてくるのか？

# その他に気づくことなど(3)

- 条件・立場の違いでの違い
  - DNA鑑定の解釈は、条件によって変わる
  - 最も検察寄りの条件での解釈と最も弁護寄りの条件での解釈とが別の意味での「信頼区間」であろうか



# その他気づくことなど(4)

- パブリケーション・バイアス
  - 論文が掲載される時、その論文は「意義深い」ものに限られる。これをパブリケーション・バイアスと言う。
  - 「ぱっとしない研究成果」はそれが事実であっても陽の目を見ない
  - 1つのテーマに関して発表された複数の論文の結果を統合して評価する(メタアナリシス)では、この影響をパブリケーション・バイアスと呼んで問題視する。
- (誰かが)主張を正当化するのに適した証拠は提示され、そうでない証拠は提示されないのではないか。
- 使いやすそうな証拠は提示され、迷いを誘う証拠は提示されていないことはないか。
- 活用データの操作は、科学研究にあっては『不正義』の代表なのだが...