

法廷のための ベイジアン・ネットワーク

法数学勉強会

2013/09/28

京大(医)統計遺伝学分野

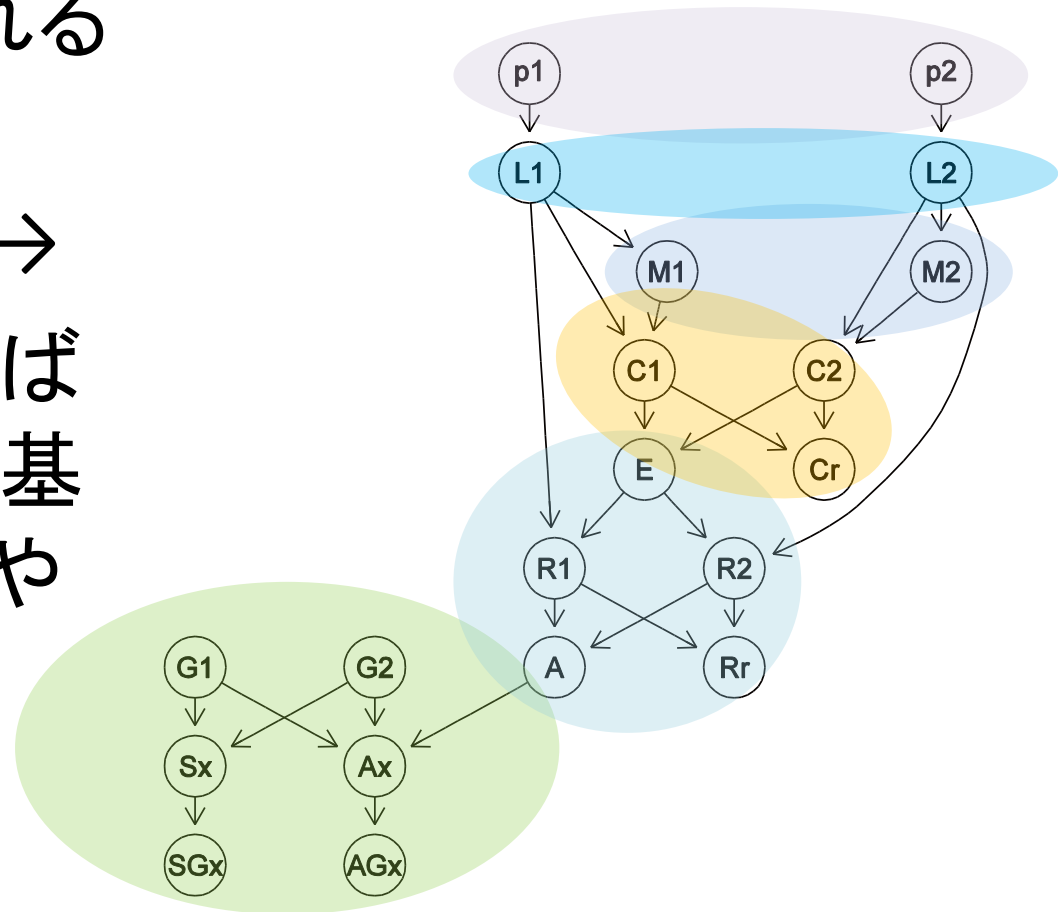
山田 亮

今日の目標

- 自力でベイジアン・ネットワークを作れるようになること

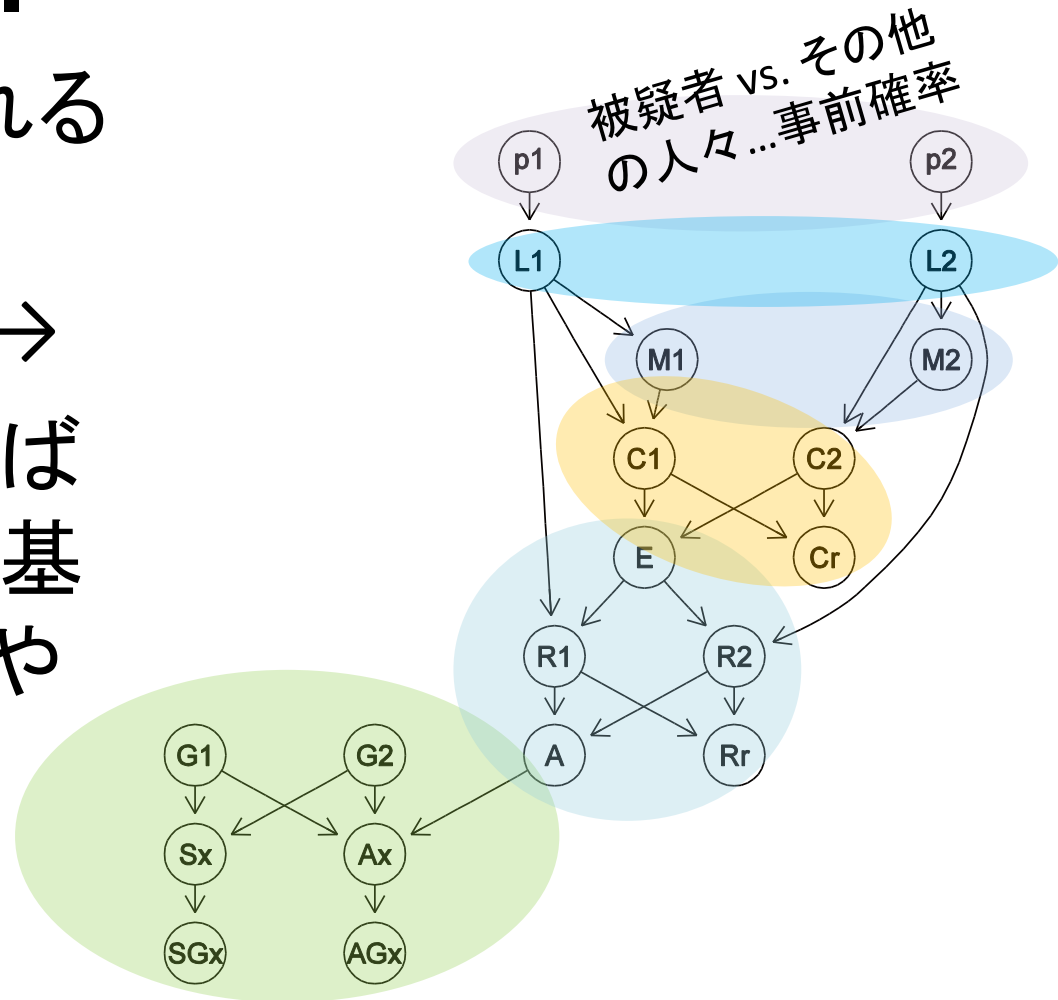
今日の目標

- 自力でベイジアン・ネットワークを作れるようになること
- (仮の)最終形は →
- 最後まで、行ければよし、行けなくても基礎だけはしっかりやりましょう



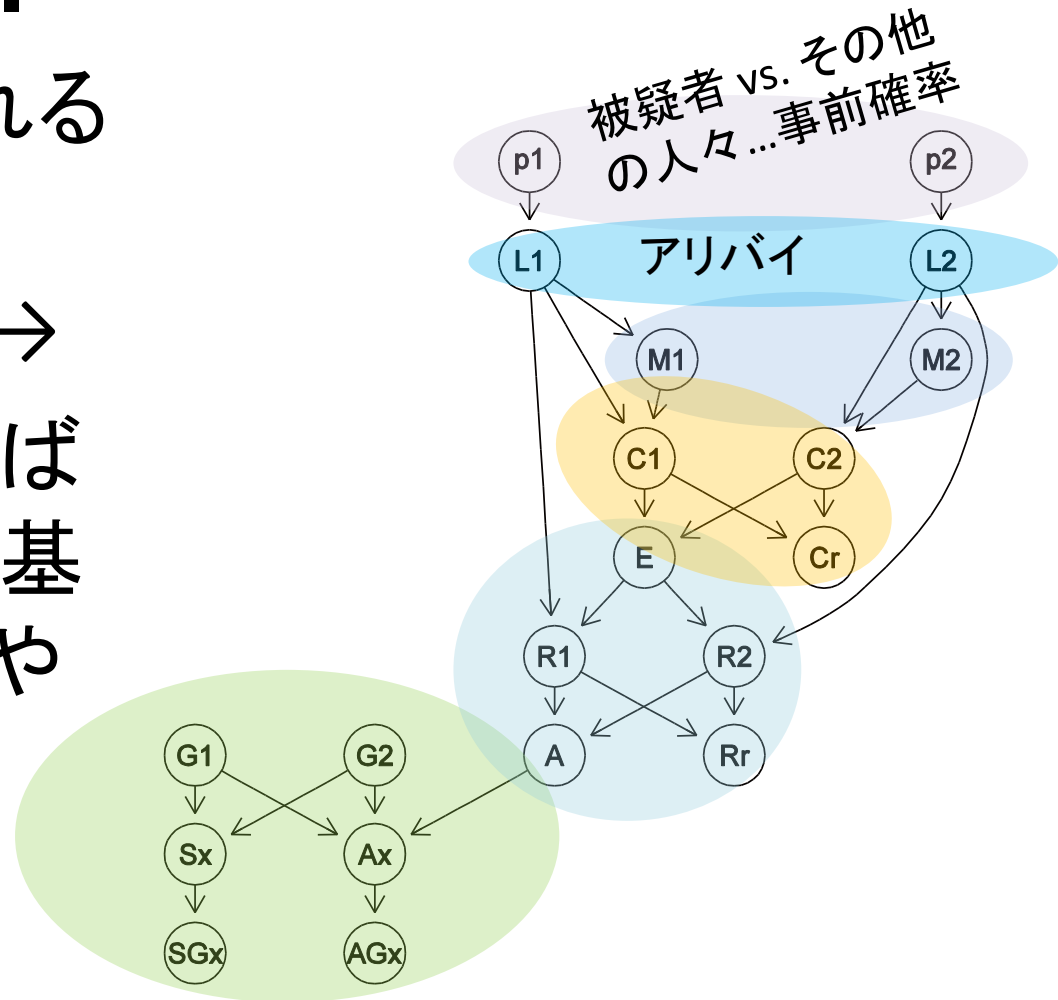
今日の目標

- 自力でベイジアン・ネットワークを作れるようになること
- (仮の)最終形は →
- 最後まで、行ければよし、行けなくても基礎だけはしっかりやりましょう



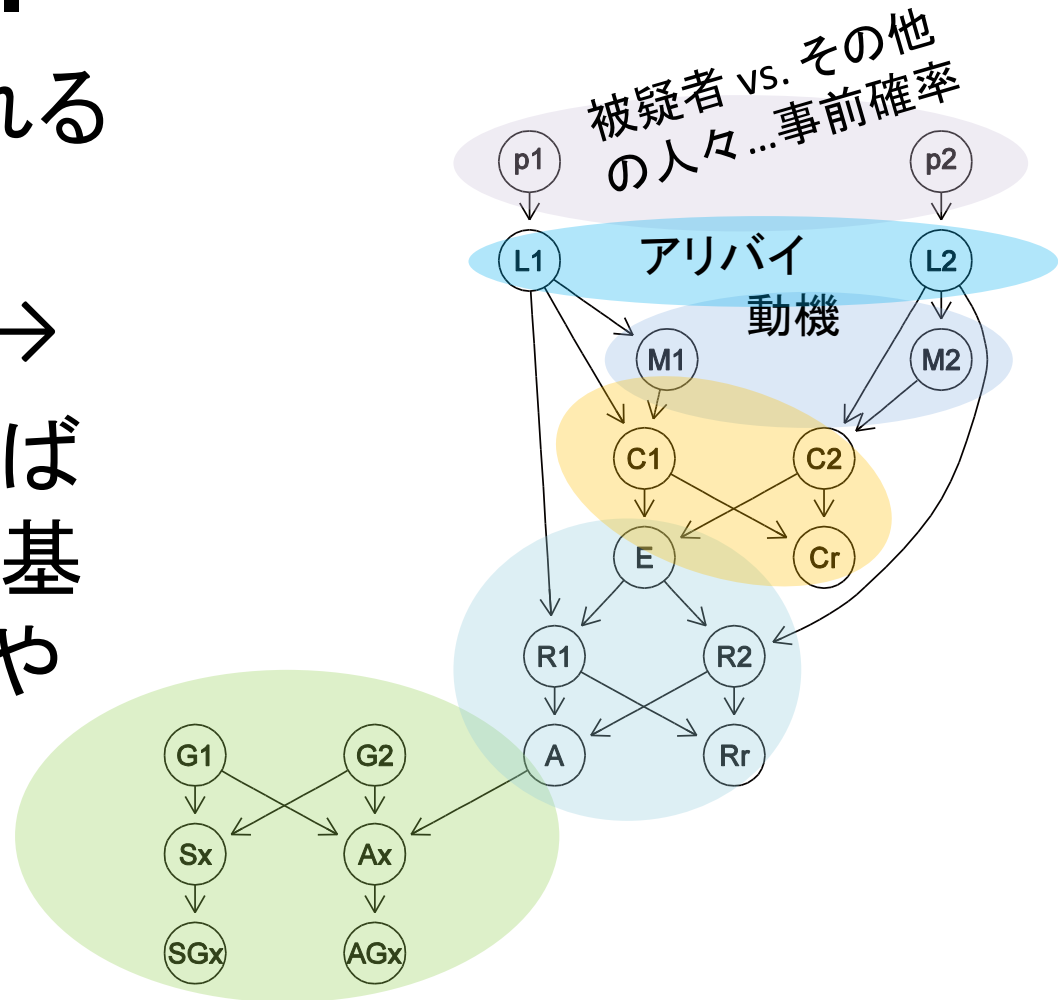
今日の目標

- 自力でベイジアン・ネットワークを作れるようになること
- (仮の)最終形は →
- 最後まで、行ければよし、行けなくても基礎だけはしっかりやりましょう



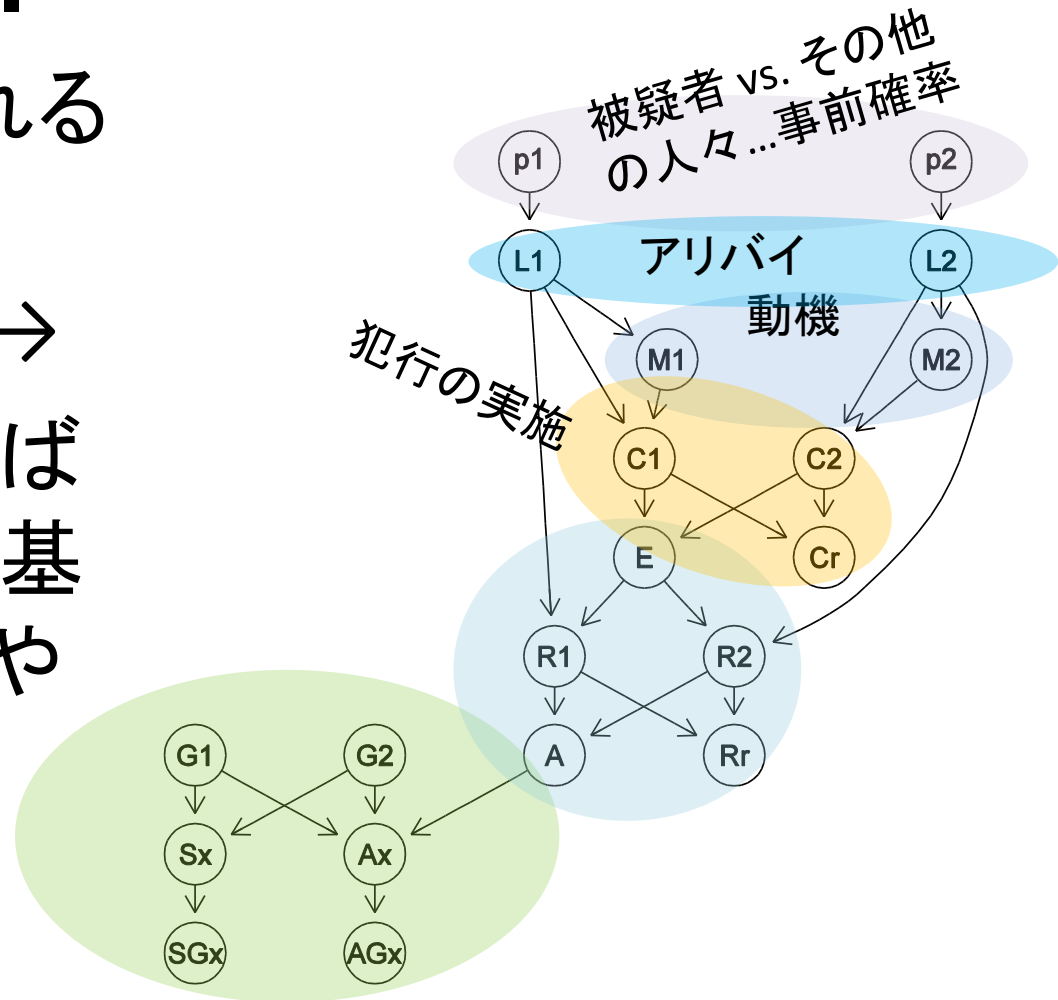
今日の目標

- 自力でベイジアン・ネットワークを作れるようになること
- (仮の)最終形は →
- 最後まで、行ければよし、行けなくても基礎だけはしっかりやりましょう



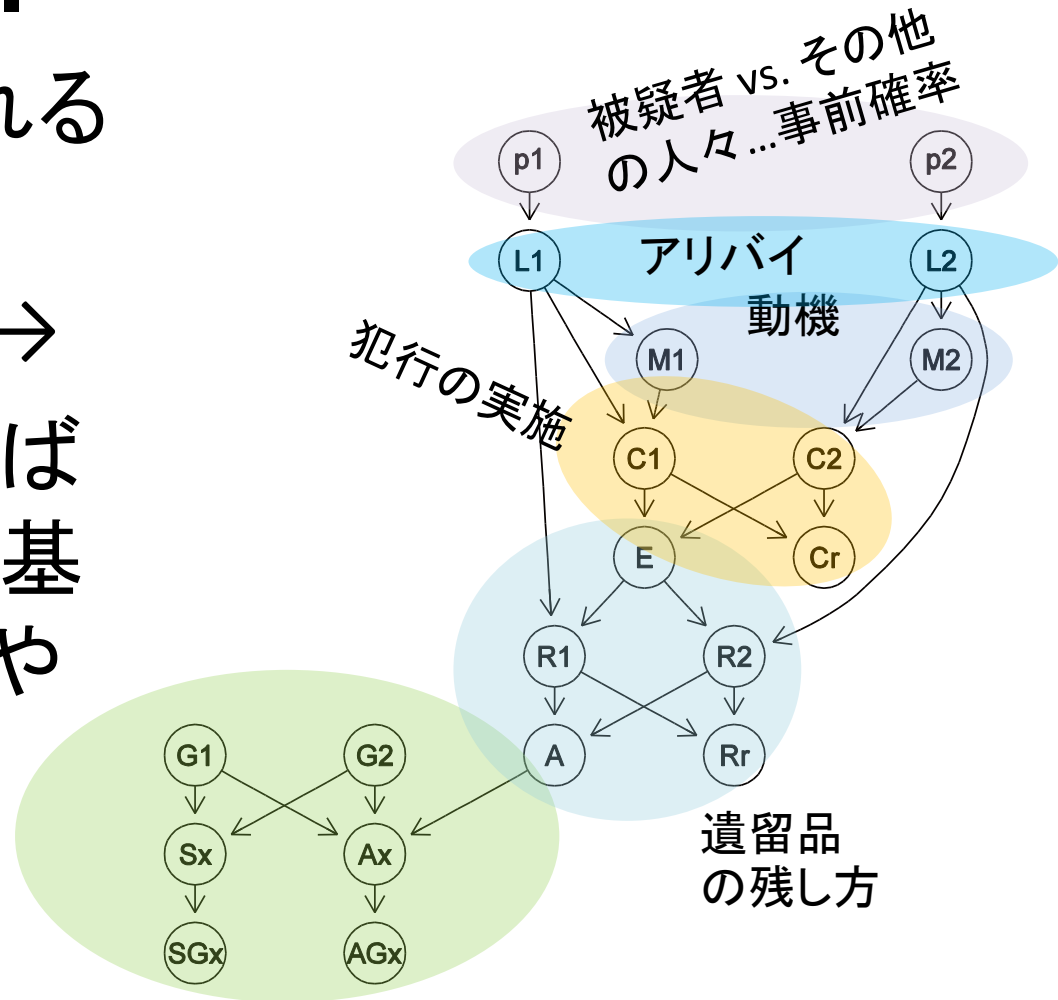
今日の目標

- 自力でベイジアン・ネットワークを作れるようになること
- (仮の)最終形は →
- 最後まで、行ければよし、行けなくても基礎だけはしっかりやりましょう



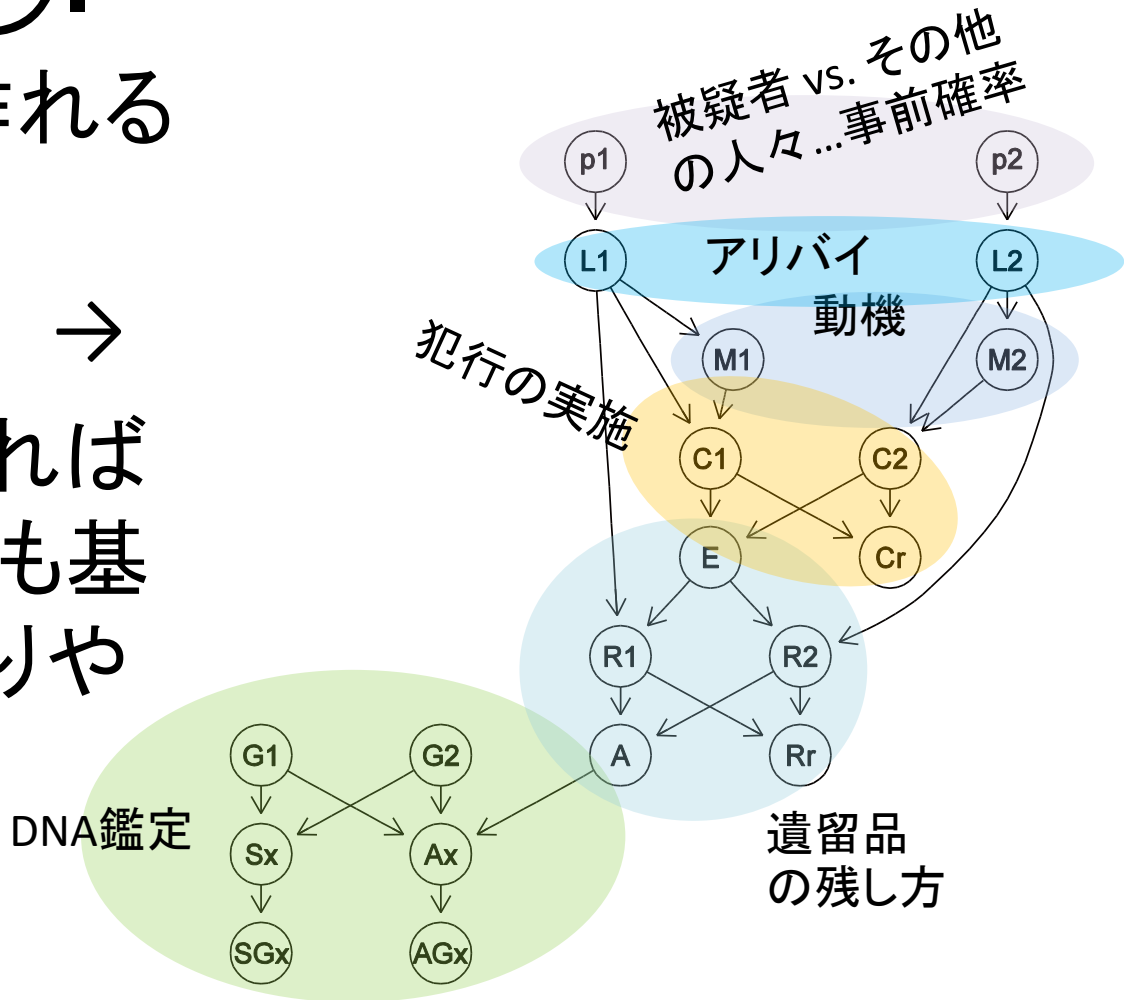
今日の目標

- 自力でベイジアン・ネットワークを作れるようになること
- (仮の)最終形は →
- 最後まで、行ければよし、行けなくても基礎だけはしっかりやりましょう

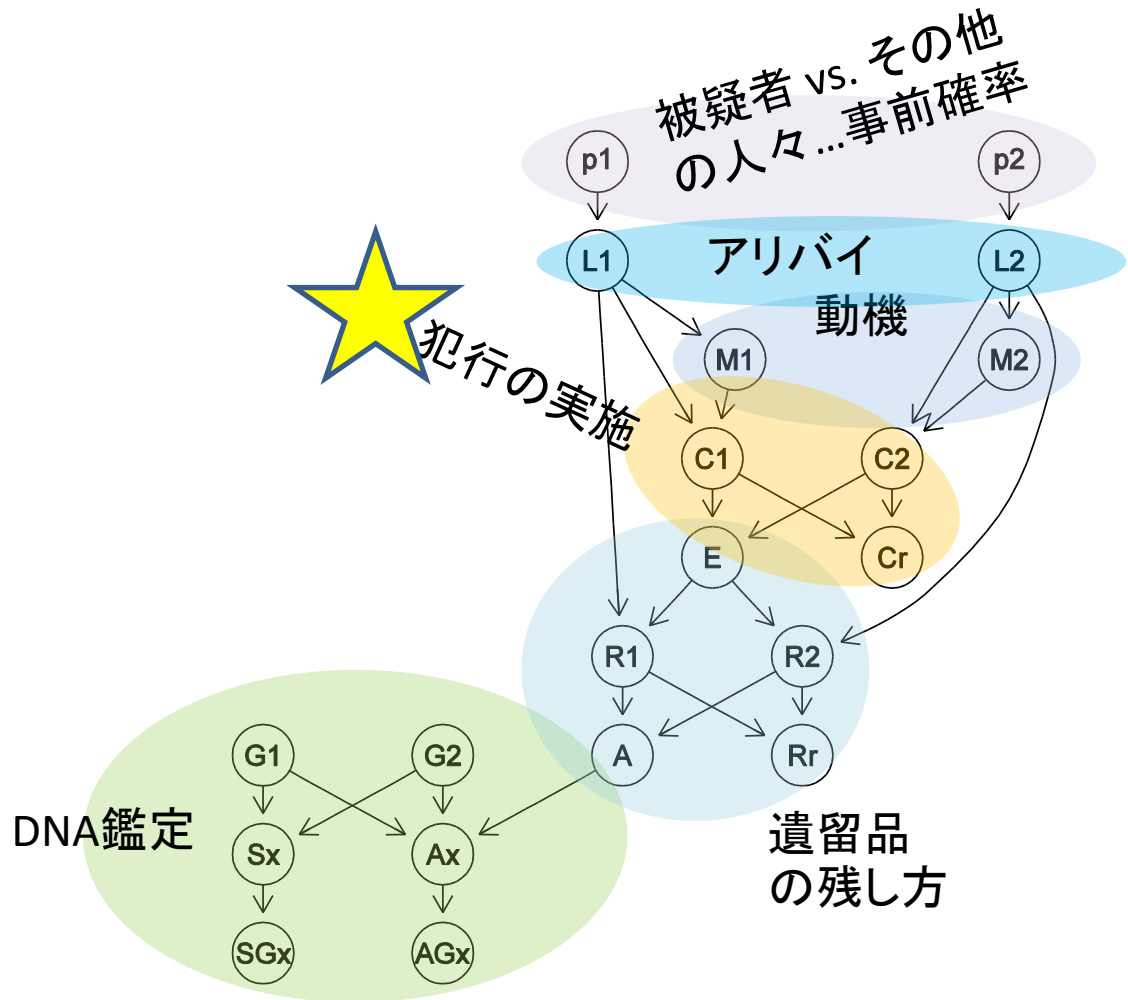


今日の目標

- 自力でベイジアン・ネットワークを作れるようになること
- (仮の)最終形は →
- 最後まで、行ければよし、行けなくても基礎だけはしっかりやりましょう

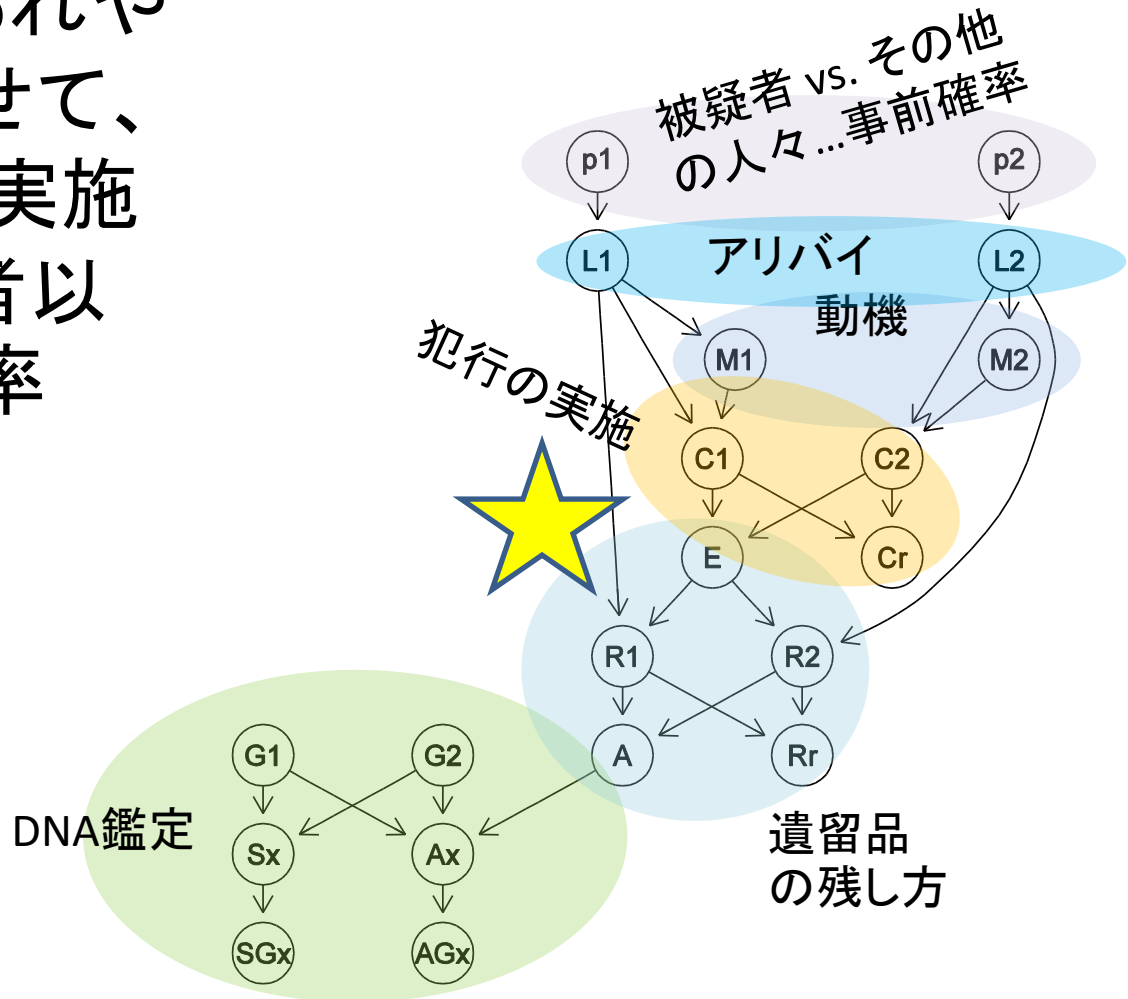


今日の目標



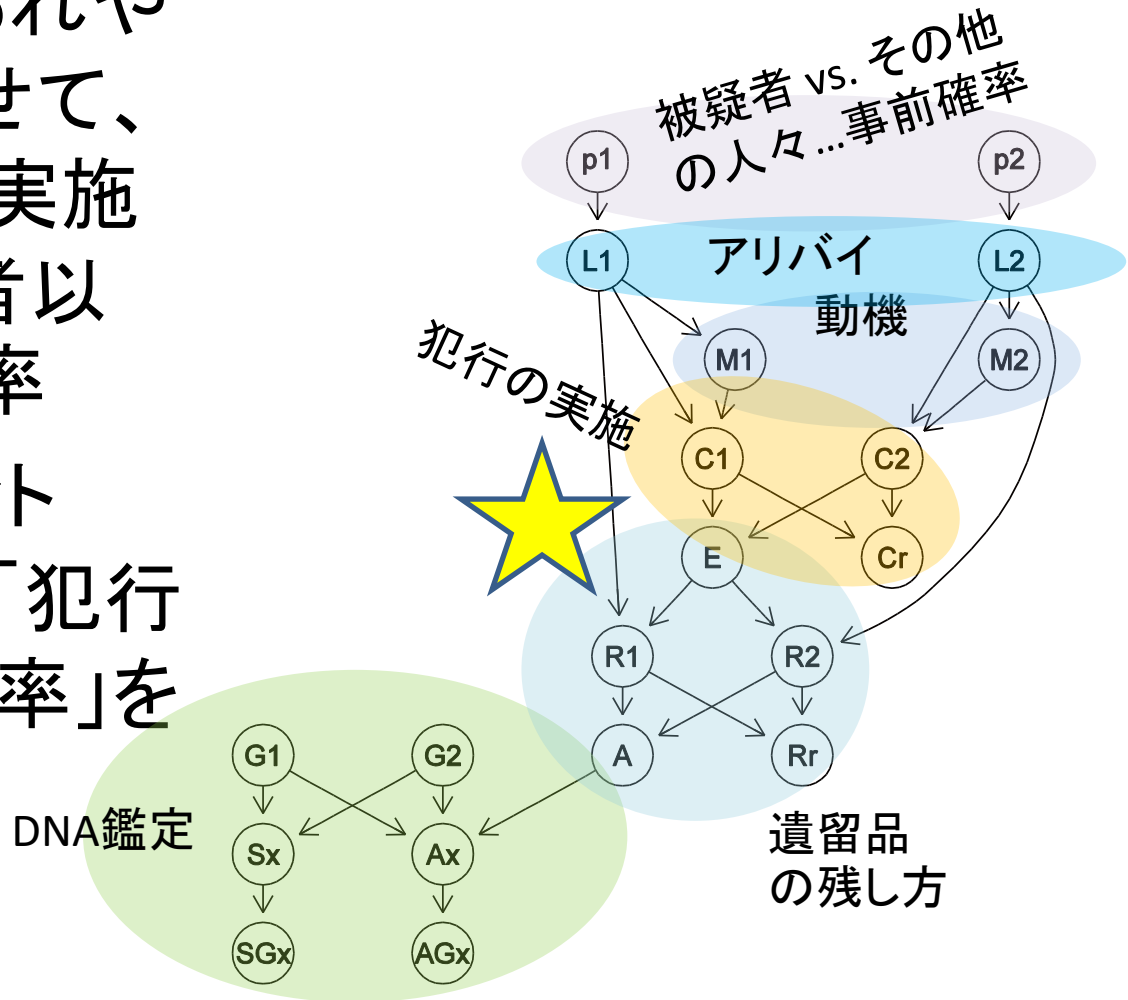
今日の目標

- 知りたいのは、あれやこれやを考え併せて、被疑者が犯行を実施した確率と被疑者以外が実施した確率



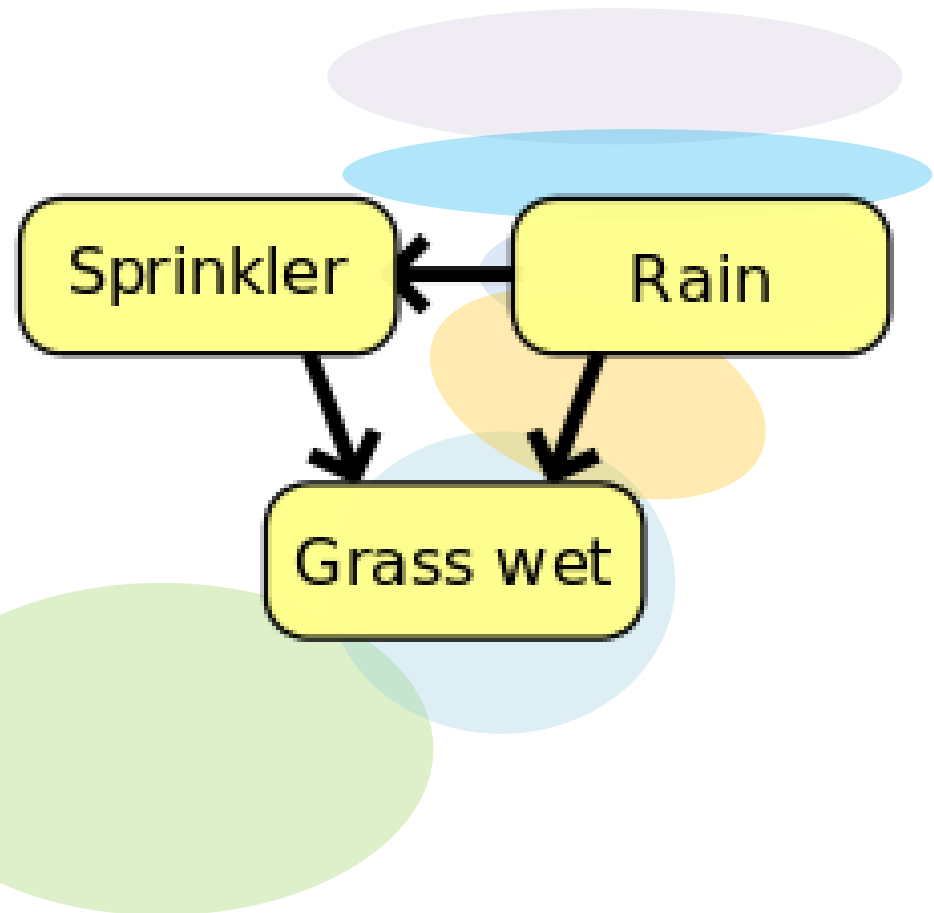
今日の目標

- 知りたいのは、あれやこれやを考え併せて、被疑者が犯行を実施した確率と被疑者以外が実施した確率
- DNA鑑定は、ネットワークを介して、「犯行実施」の「事後確率」を動かす



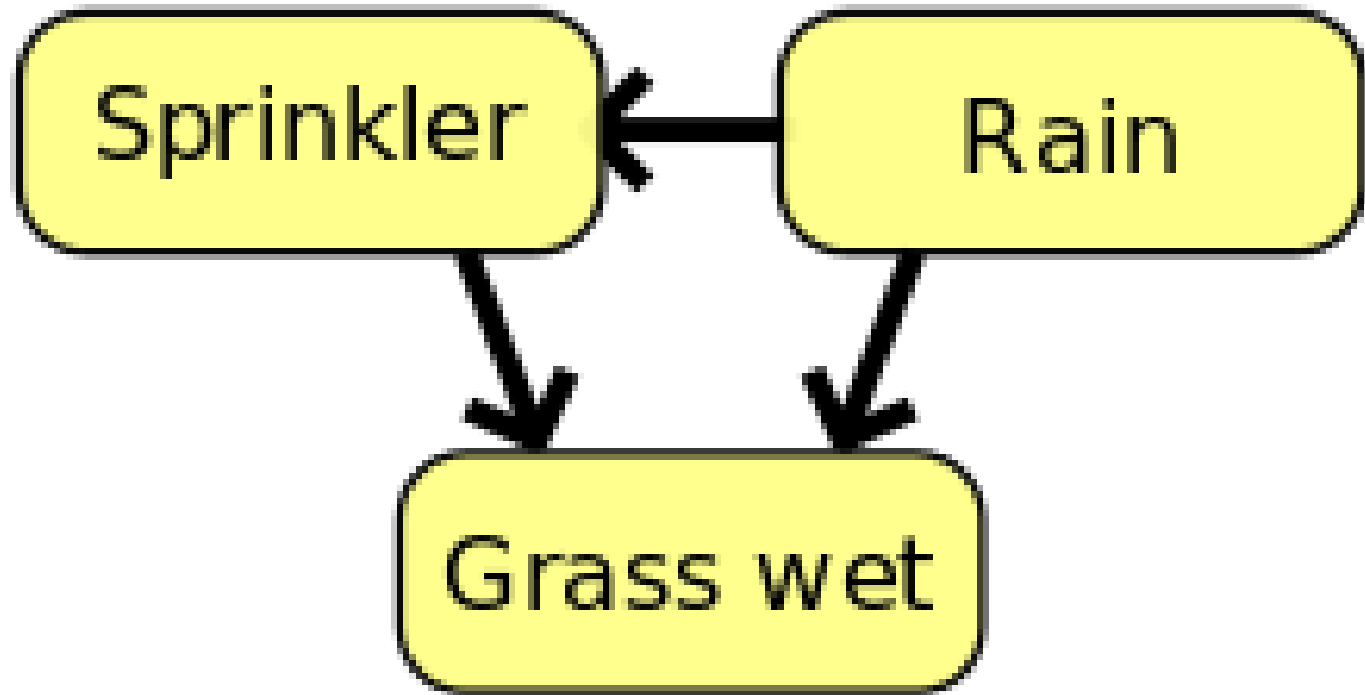
今日の目標

- 自力でベイジアン・ネットワークを作れるようになること



- 基礎だけはしっかりやりましょう

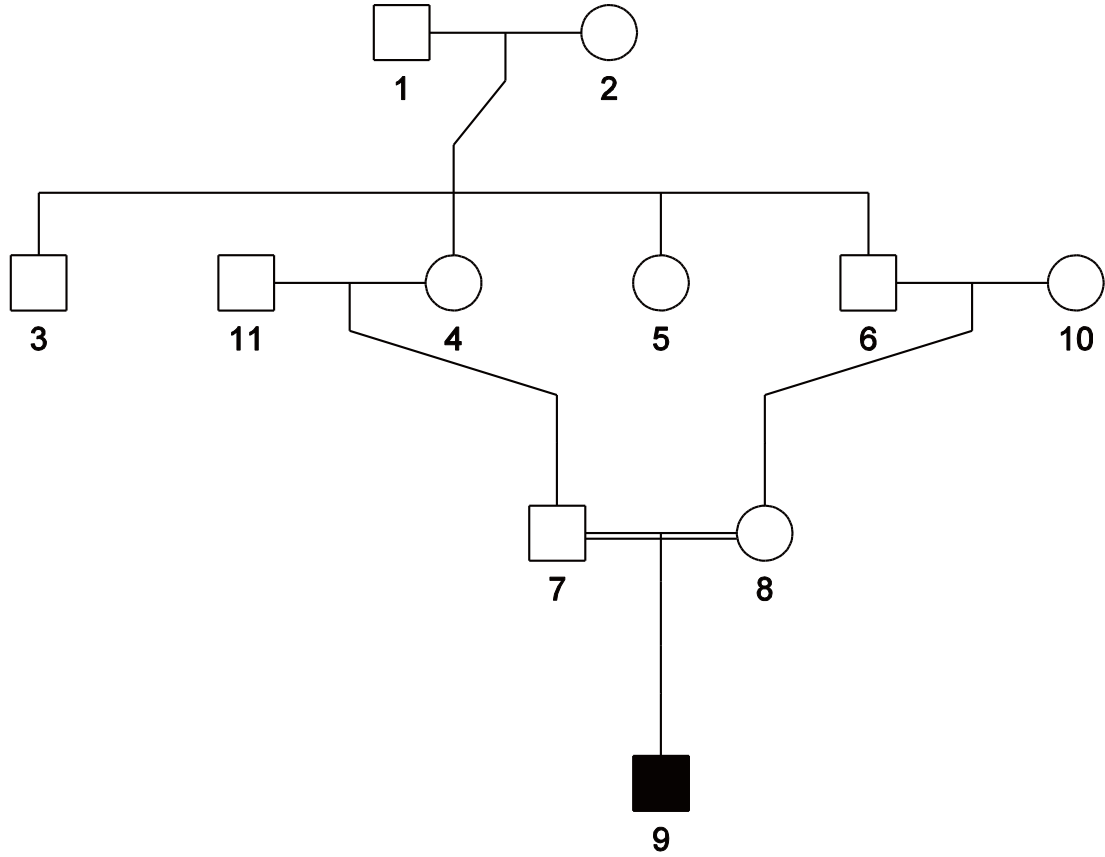
ベイジアン・ネットワーク 「・・・だったら、・・・」の連鎖



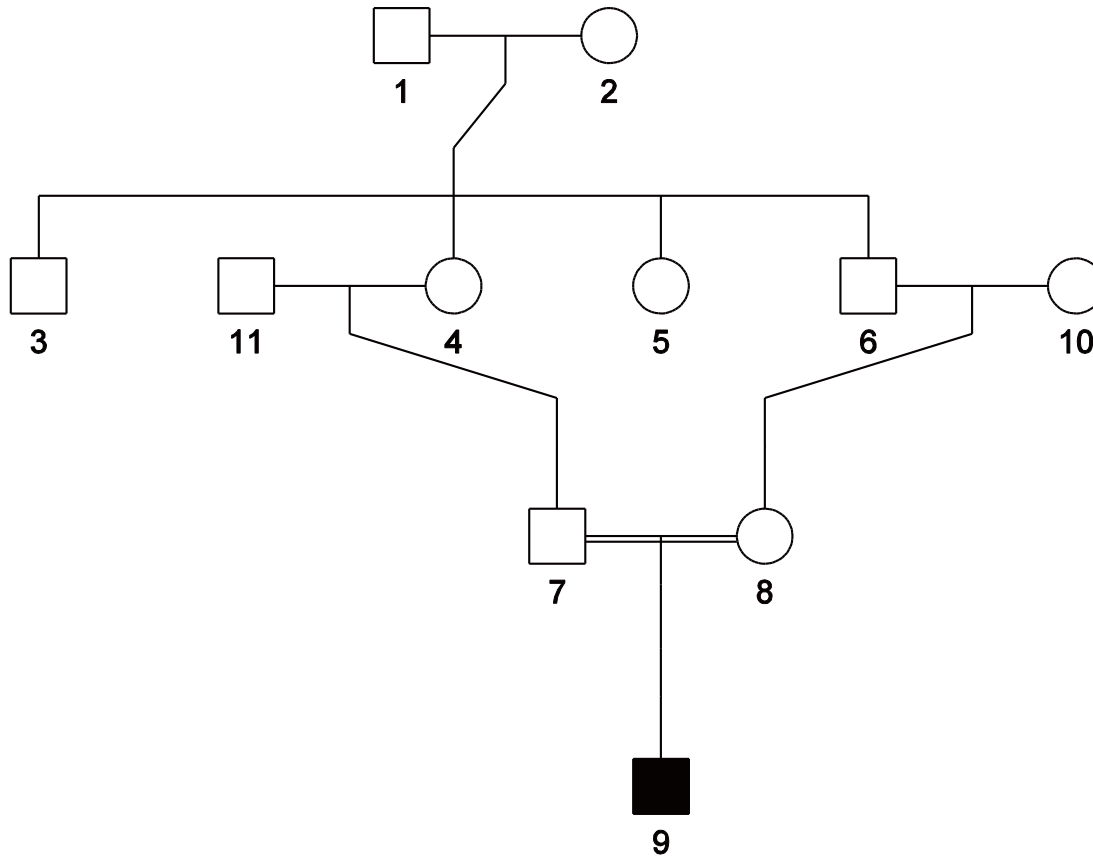
ベイジアン・ネットワーク という グラフ

- Directed (向きのある)グラフ
- A-cycle (サイクルがない) グラフ
- 併せて
 - Directed Acyclic Graph
 - 有向非循環グラフ
 - 略して DAG (頻出単語です)

家系図というDAG(有向非循環グラフ)

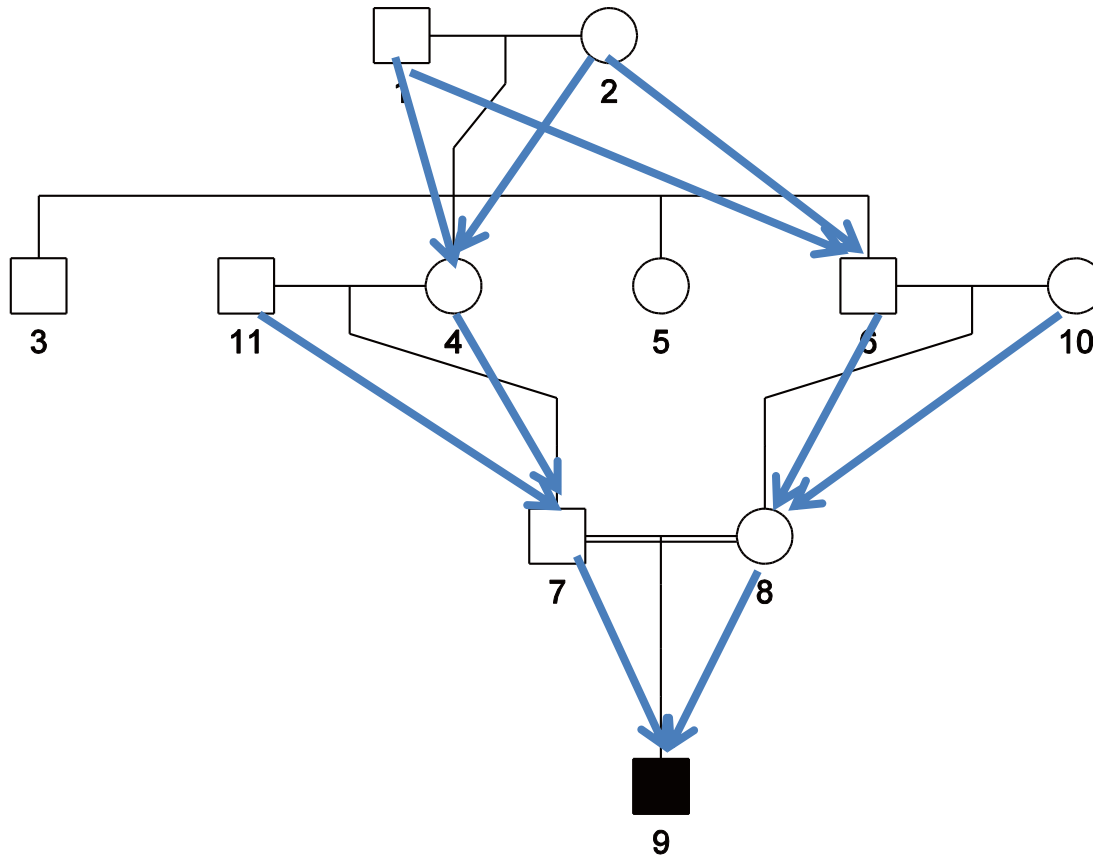


家系図というDAG(有向非循環グラフ)



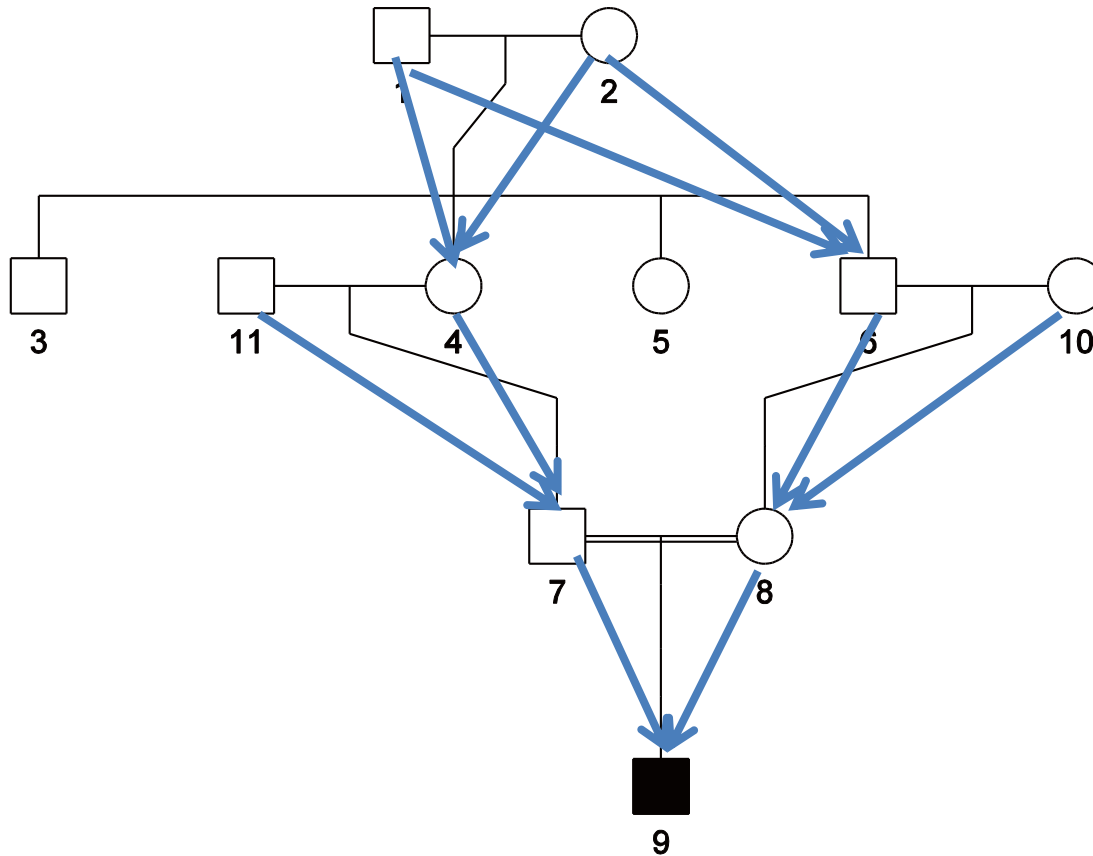
- これはDAG?
- これはグラフ？

家系図というDAG(有向非循環グラフ)



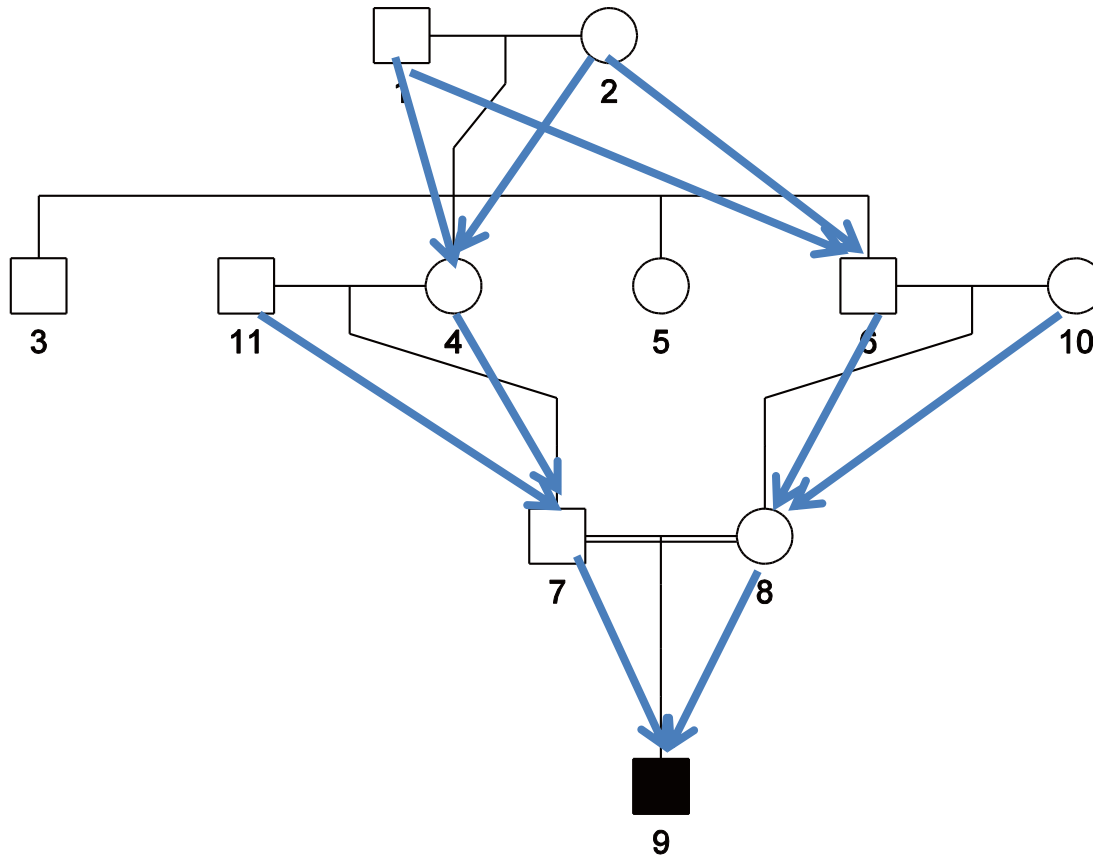
- これはDAG?
- これはグラフ？

家系図というDAG(有向非循環グラフ)



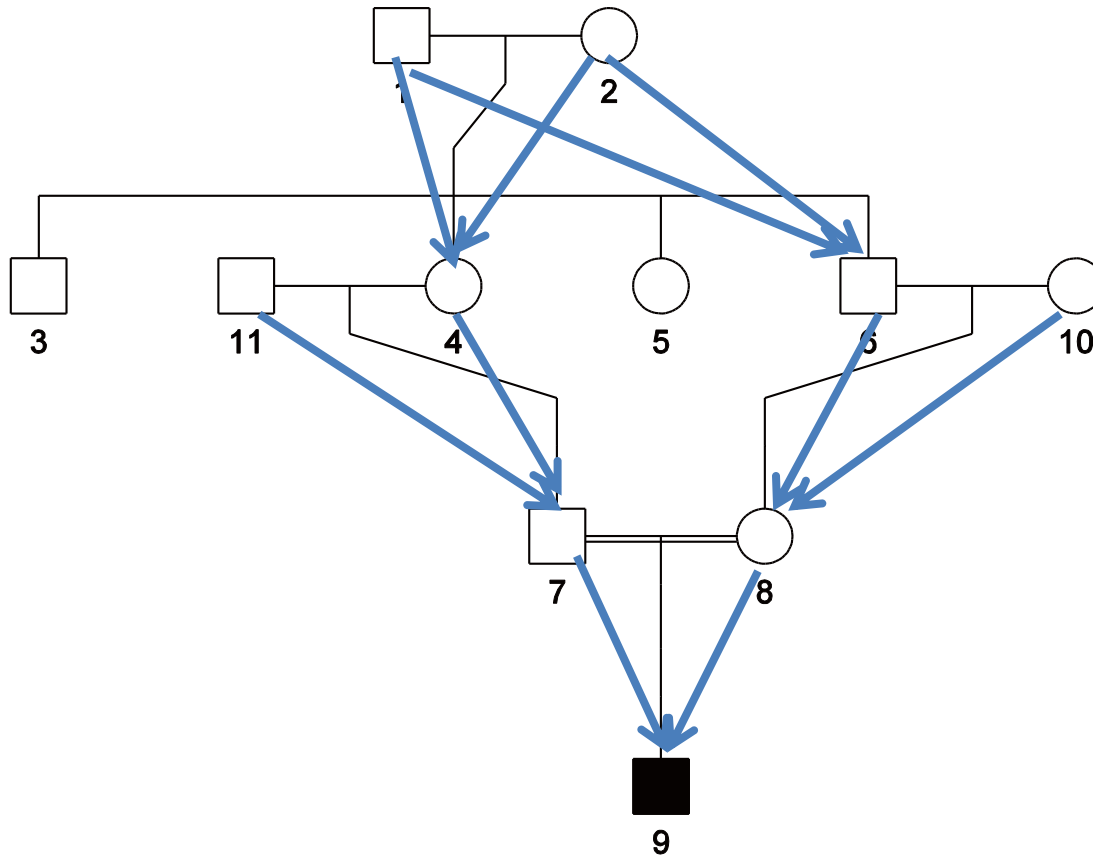
- これはDAG?
- これはグラフ?
- これは有向?

家系図というDAG(有向非循環グラフ)



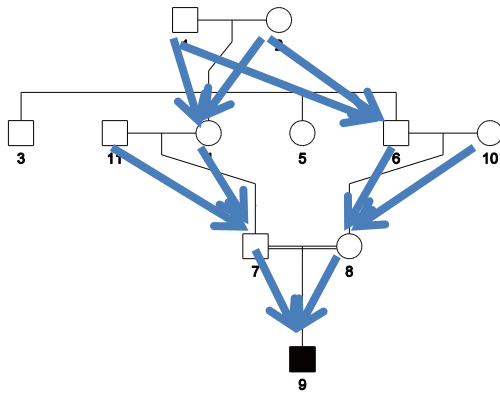
- これはDAG?
- これはグラフ?
- これは有向?

家系図というDAG(有向非循環グラフ)



- 非循環だけれど、『ぐるり』がある
- 近親婚
- ベイジアン・ネットワークでは、『独立ではない影響』

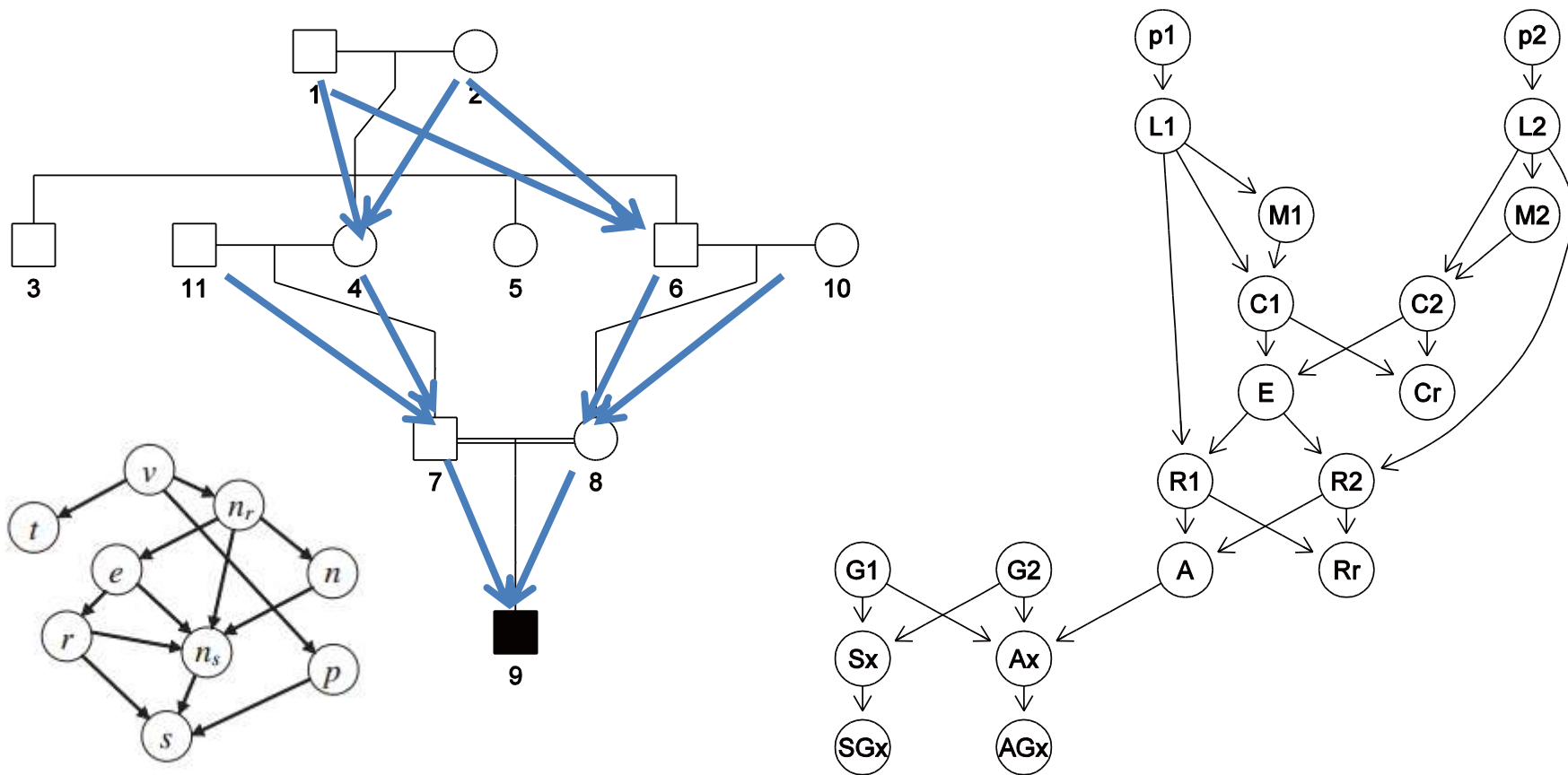
家系図というDAG(有向非循環グラフ)



- 近親婚・独立ではない影響
- ジェノタイプの計算をするときには、「親子」関係だけに着目して、 $1/2$ の確率の伝達を計算する
- ベイジアン・ネットワークも同じ。複雑なネットワークも個々のノード(子)への親の寄与がたくさん集まったものとみなす

家系図もベイジアン・ネットワーク

家系図と一般的ベイジアン・ネットワークとは何が違う？

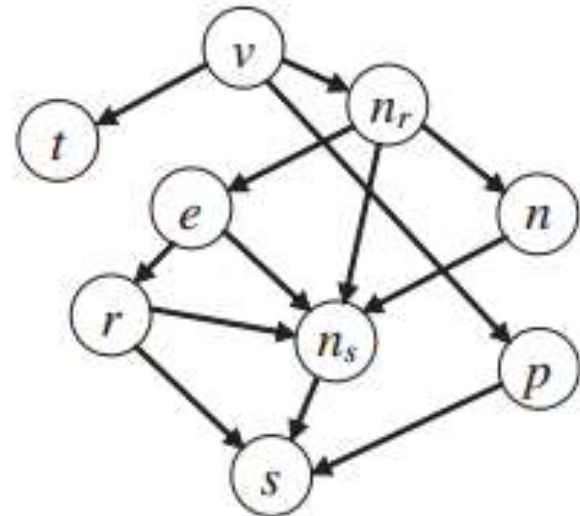


DAGは親子関係図

- ネットワーク全体は
 - 親子の関係のみからなる
 - すべての親子関係を抜き出そう

すべての親子関係を抜き出そう

- ノードは何個ある？
- 親の人数の最小値は？
- 親の人数の最大値は？
- 親の人数別にノード数を集計してみよう



ベイジアン・ネットワークと確率変数

- ベイジアン・ネットワークの
 - ノードは「確率変数」
 - 「親子関係」には「ルール」がある
 - すべての「親子関係」には「ルール」がある・必要

DAGは「親子関係」の集まりなので・・・

- 1つの「親子関係」を作れば、ベイジアン・ネットワークの出来上がり
- では「親子関係」を1つ作ってみる

確率変数と親子関係ルールの例

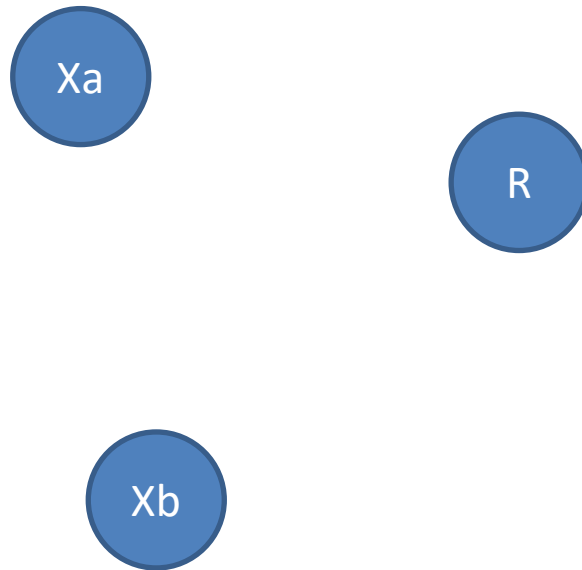
- 2人(A,B)でジャンケン
 - ジャンケンをして「手」の出し方に応じて、A,Bの相対的位置が変わる、というゲーム
 - A,Bが(グー・チョキ・パー)を出す確率は確率変数
 - X_a, X_b
 - 1回のジャンケンでAがBに対して±何歩進むかも確率変数
 - R
 - ±何歩？
 - 場合を列挙しよう

3個の確率変数 X_a, X_b, R が作る ベイジアン・ネットワーク

- 描いてみよう、ベイジアンネットワーク
 - ノードは確率変数
 - ノードの数は？

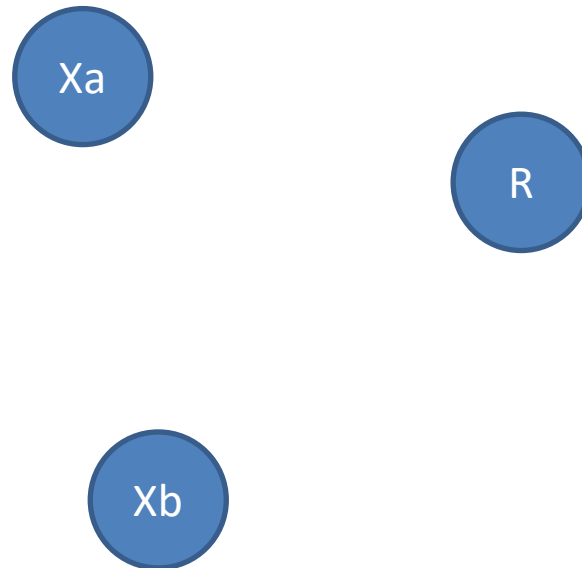
3個の確率変数 X_a , X_b , R が作る ベイジアン・ネットワーク

- 描いてみよう、ベイジアンネットワーク
 - ノードは確率変数
 - ノードの数は？



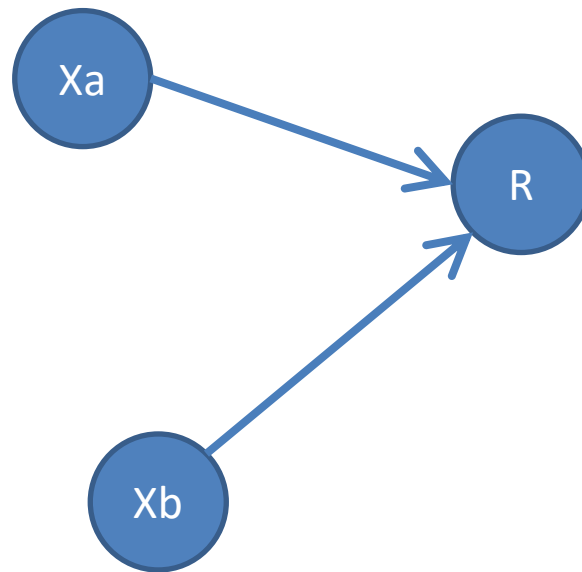
3個の確率変数 X_a , X_b , R が作る ベイジアン・ネットワーク

- 描いてみよう、ベイジアンネットワーク
 - 原因・理由は「親」～エッジの始点
 - 結果は「子」～エッジの終点



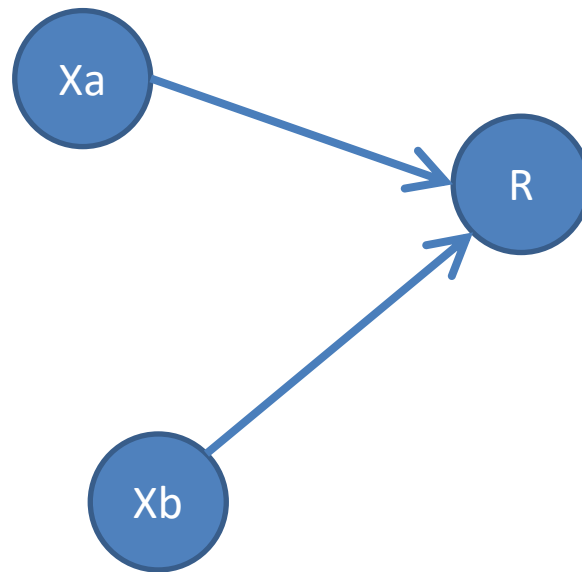
3個の確率変数 X_a , X_b , R が作る ベイジアン・ネットワーク

- 描いてみよう、ベイジアンネットワーク
 - 原因・理由は「親」～エッジの始点
 - 結果は「子」～エッジの終点



3個の確率変数 X_a , X_b , R が作る ベイジアン・ネットワーク

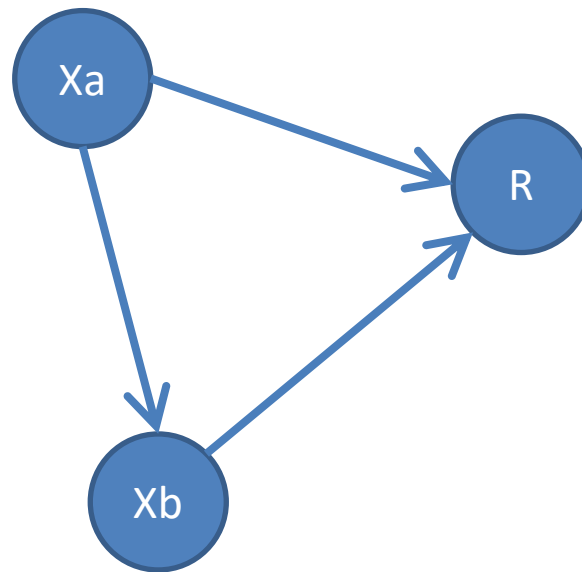
- 描いてみよう、ベイジアンネットワーク
 - 原因・理由は「親」～エッジの始点
 - 結果は「子」～エッジの終点



Bさんが「ずる」をして、
「後だし」するとして
ら・・・

3個の確率変数 X_a , X_b , R が作る ベイジアン・ネットワーク

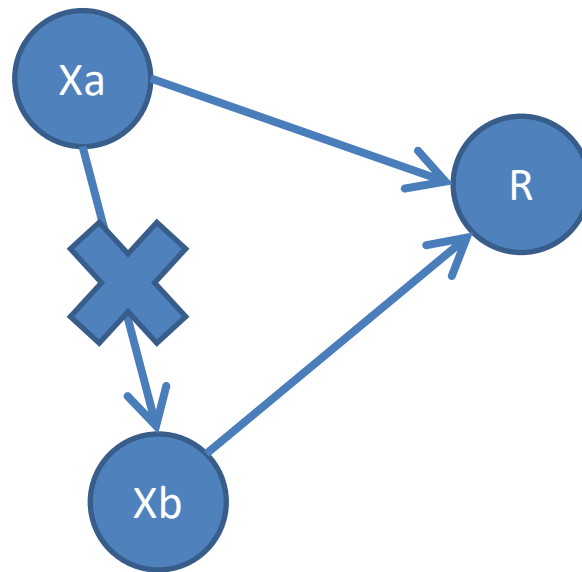
- 描いてみよう、ベイジアンネットワーク
 - 原因・理由は「親」～エッジの始点
 - 結果は「子」～エッジの終点



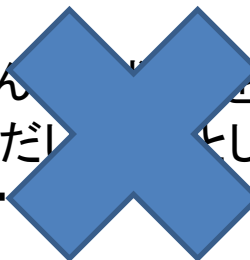
Bさんが「ずる」をして、
「後だし」するとしたら...

3個の確率変数 X_a , X_b , R が作る ベイジアン・ネットワーク

- 描いてみよう、ベイジアンネットワーク
 - 原因・理由は「親」～エッジの始点
 - 結果は「子」～エッジの終点



Bさん...をして、
「後だ...とした
ら...



確率変数の事象の数

- 2人(A,B)でジャンケン
 - X_a, X_b
 - A,Bが(グー・チョキ・パー)を出す
 - 事象はそれぞれ3通り
 - R
 - 1回のジャンケンでAがBに対して±何歩進むか

確率変数の事象の数

- 2人(A,B)でジャンケン
 - X_a, X_b
 - A,Bが(グー・チョキ・パー)を出す
 - 事象はそれぞれ3通り
 - R
 - 1回のジャンケンでAがBに対して±何歩進むか
 - 「ぐ・り・こ」 (±3歩)
 - 「ち・よ・こ・れ・ー・と」 (±6歩)
 - 「ぱ・い・な・つ・ぶ・る」 (±6歩)

確率変数の事象の数

- 2人(A,B)でジャンケン

- X_a, X_b

- A,Bが(グー・チョキ・パー)を出す
 - 事象はそれぞれ3通り

- R

- 1回のジャンケンでAがBに対して±何歩進むか
 - 「ぐ・り・こ」 (±3歩)
 - 「ち・よ・こ・れ・ー・と」 (±6歩)
 - 「ぱ・い・な・つ・ぷ・る」 (±6歩)
 - 全部で、+3,-3,+6,-6,0(アイコも忘れずに)の5通り

「親子関係」を表すテーブル

A	B	R(+6)	R(+3)	R(0)	R(-3)	R(-6)
グー	グー	0	0	1	0	0
チョキ	グー	0	0	0	1	0
パー	グー	1	0	0	0	0
グー	チョキ	0	1	0	0	0
チョキ	チョキ	0	0	1	0	0
パー	チョキ	0	0	0	0	1
グー	パー	0	0	0	0	1
チョキ	パー	1	0	0	0	0
パー	パー	0	0	1	0	0

親子関係テーブルを使う

- 確率の計算

- $X_a = (a(\text{グー}), a(\text{チョキ}), a(\text{パー})) = (0.1, 0.4, 0.5)$
- $X_b = (b(\text{グー}), b(\text{チョキ}), b(\text{パー})) = (1/3, 1/3, 1/3)$
- のとき、確率変数が $(+6, +3, 0, -3, -6)$ である確率はいくつか？

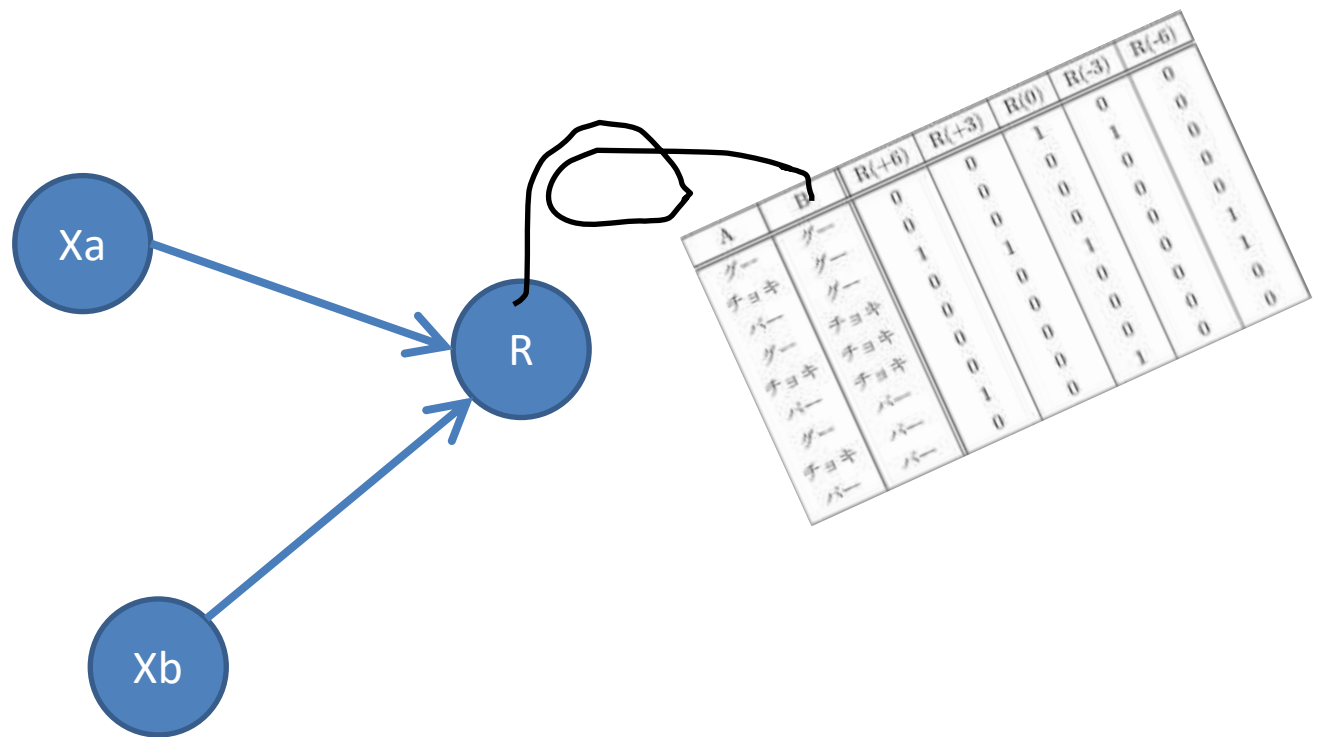
親子関係テーブルを使う

- 尤度の計算

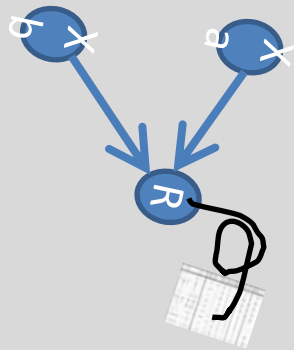
- 確率変数Rが -6 であったという
- X_a がグーであった尤度はいくつか？

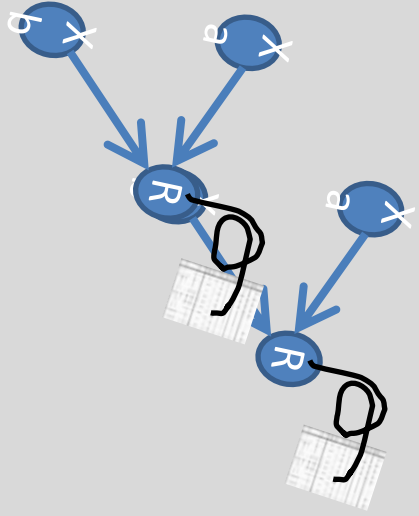
- 確率変数Rが +6 で、 X_b がパーであったという
- X_a がチヨキであった尤度はいくつか？
- X_a がパーであった尤度はいくつか？

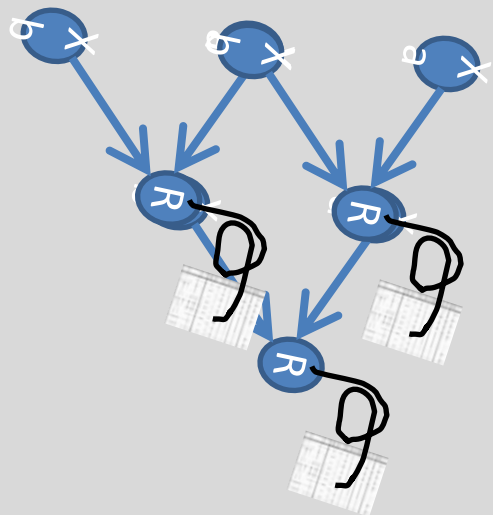
ベイジアン・ネットワーク(のピース)の 出来上がり

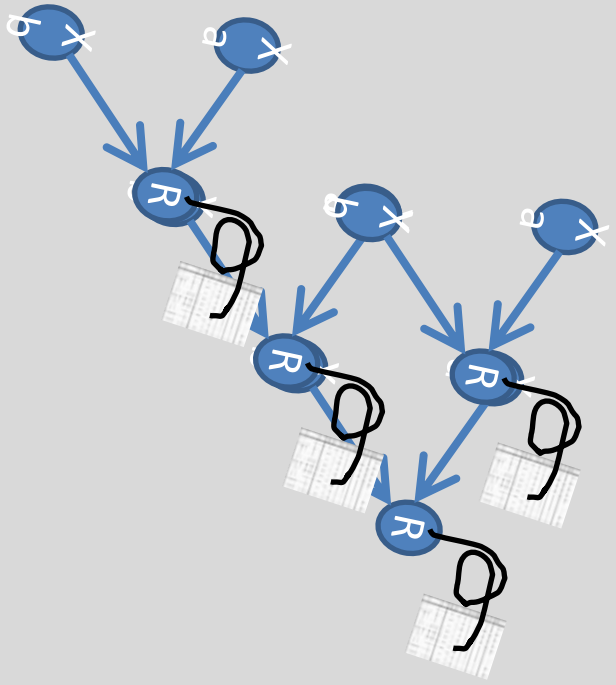


ベイジアン・ネットワークの ピースの組合せ









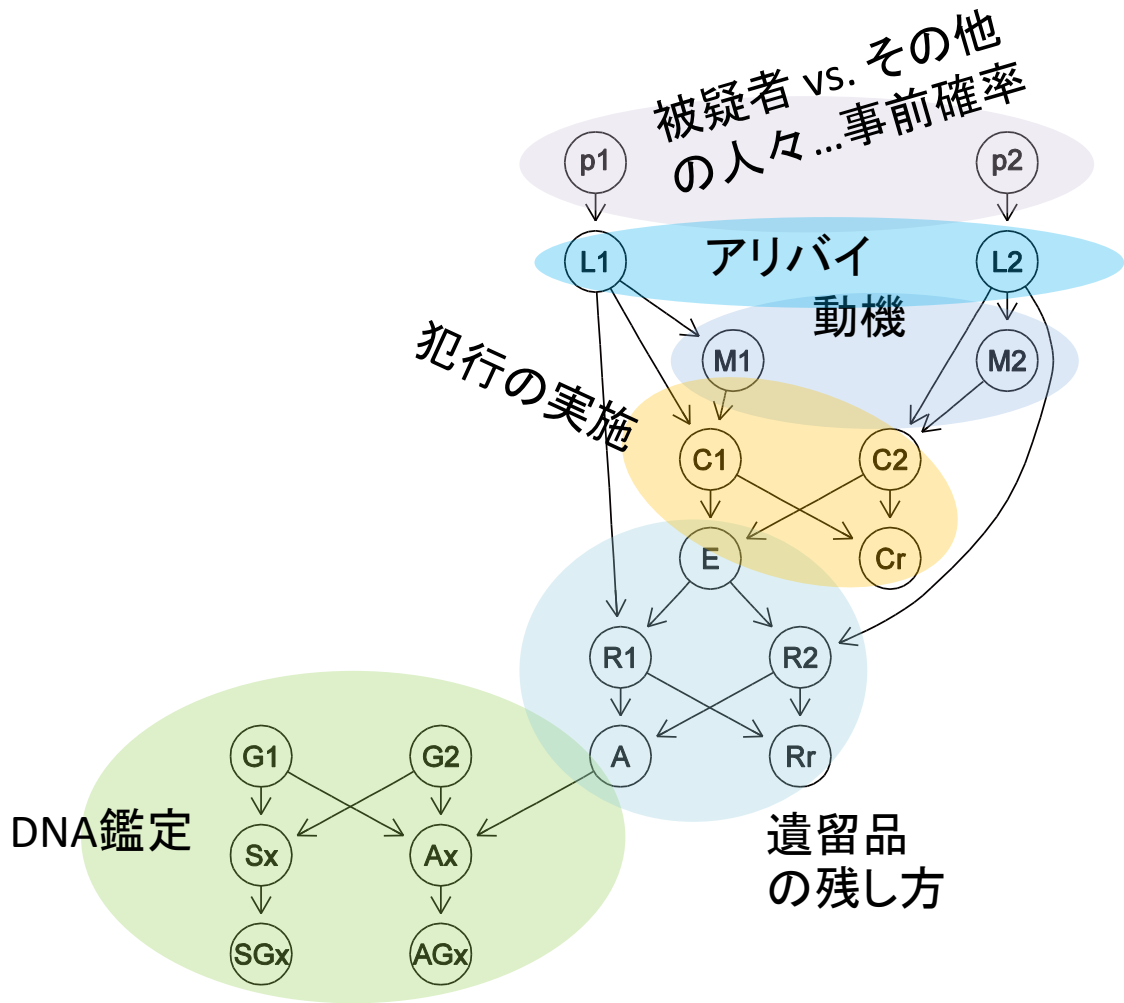
～改めて～
法廷のための
ベイジアン・ネットワーク

法数学勉強会

2013/09/28

京大(医)統計遺伝学分野

山田 亮

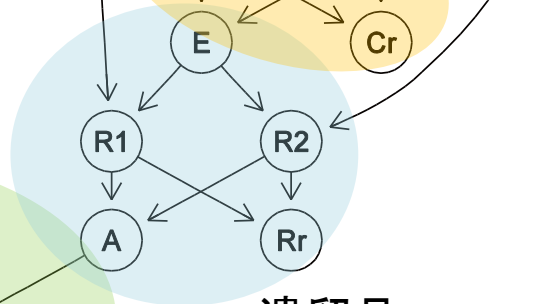


被疑者 vs. その他
の人々...事前確率

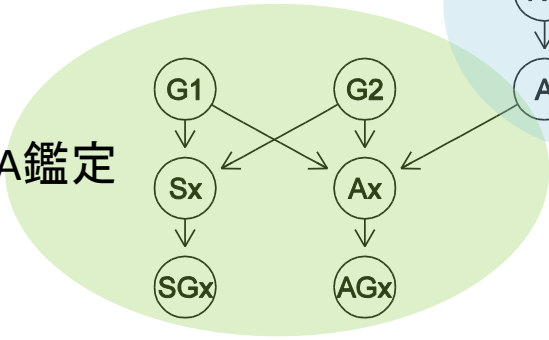
アリバイ

動機

犯行の実施



DNA鑑定

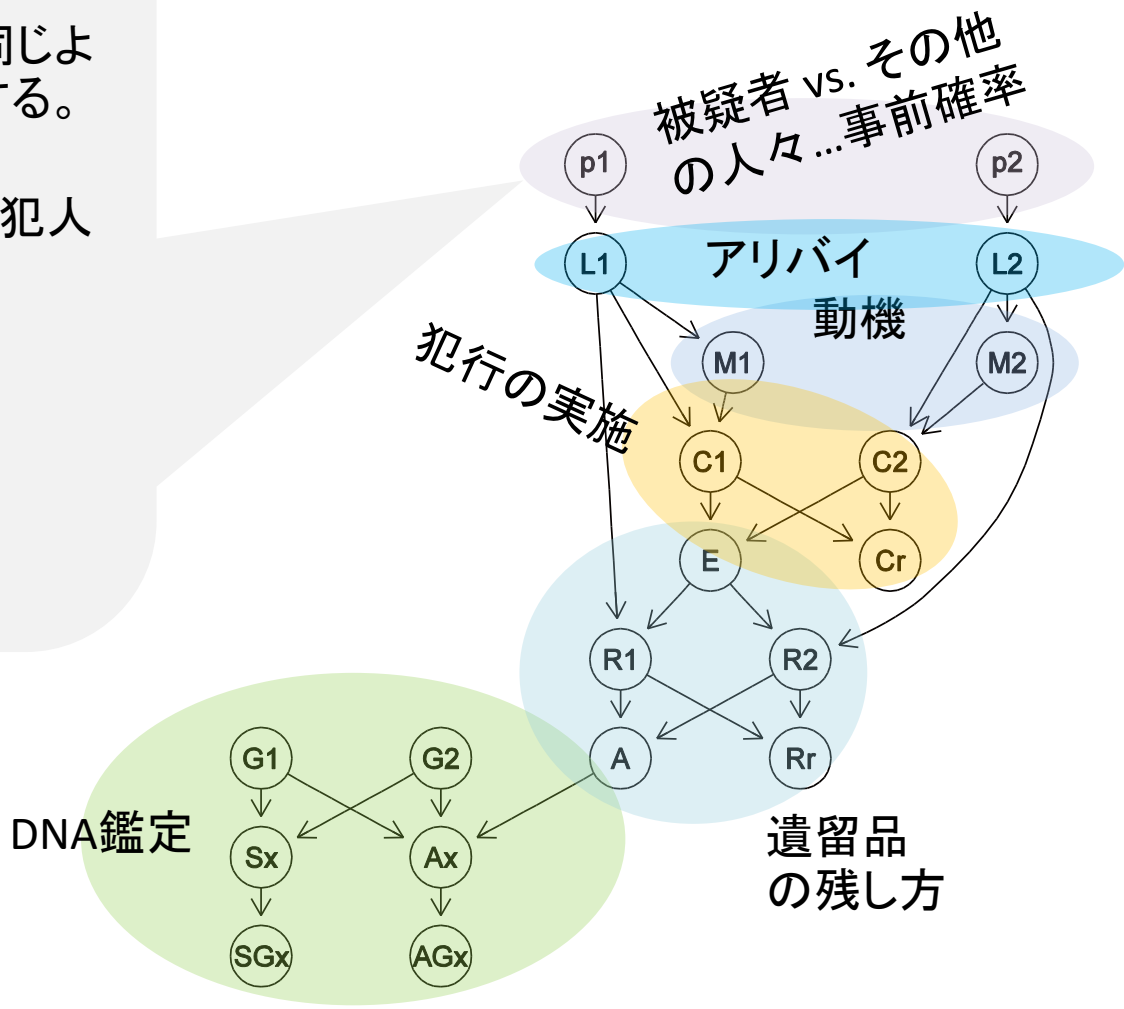


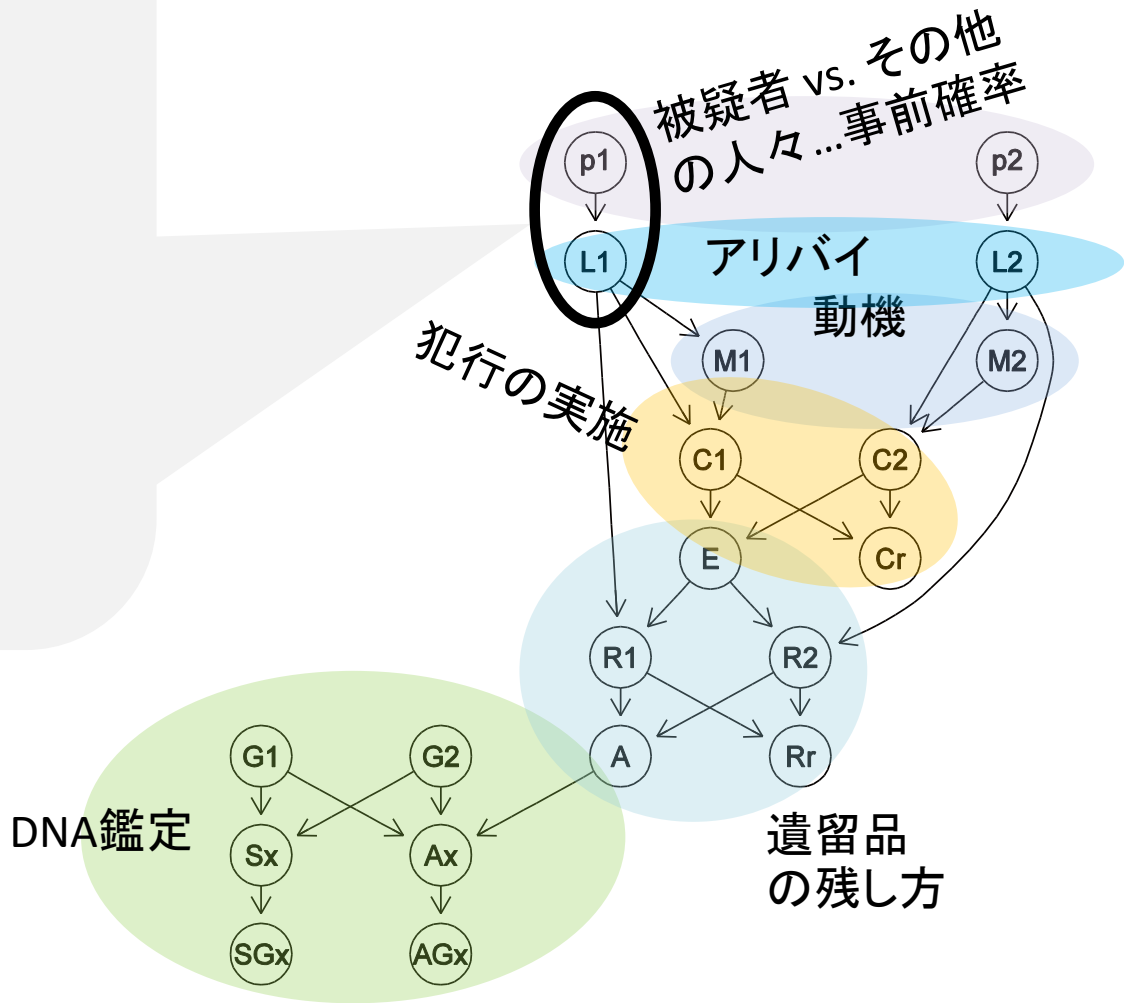
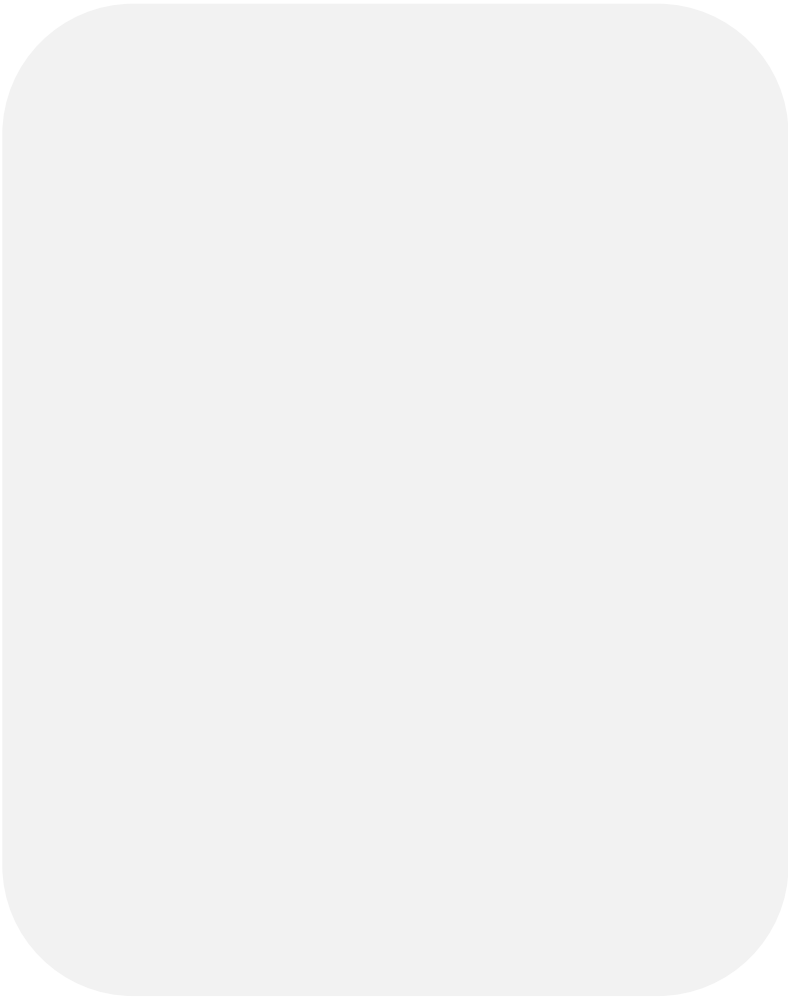
遺留品
の残し方

「被疑者」が1人
「そのほかに可能性のある同じよ
うな人」が100万人、いるとする。

「情報が皆無」なら、どちらを犯人
と考えるかは

1対100万





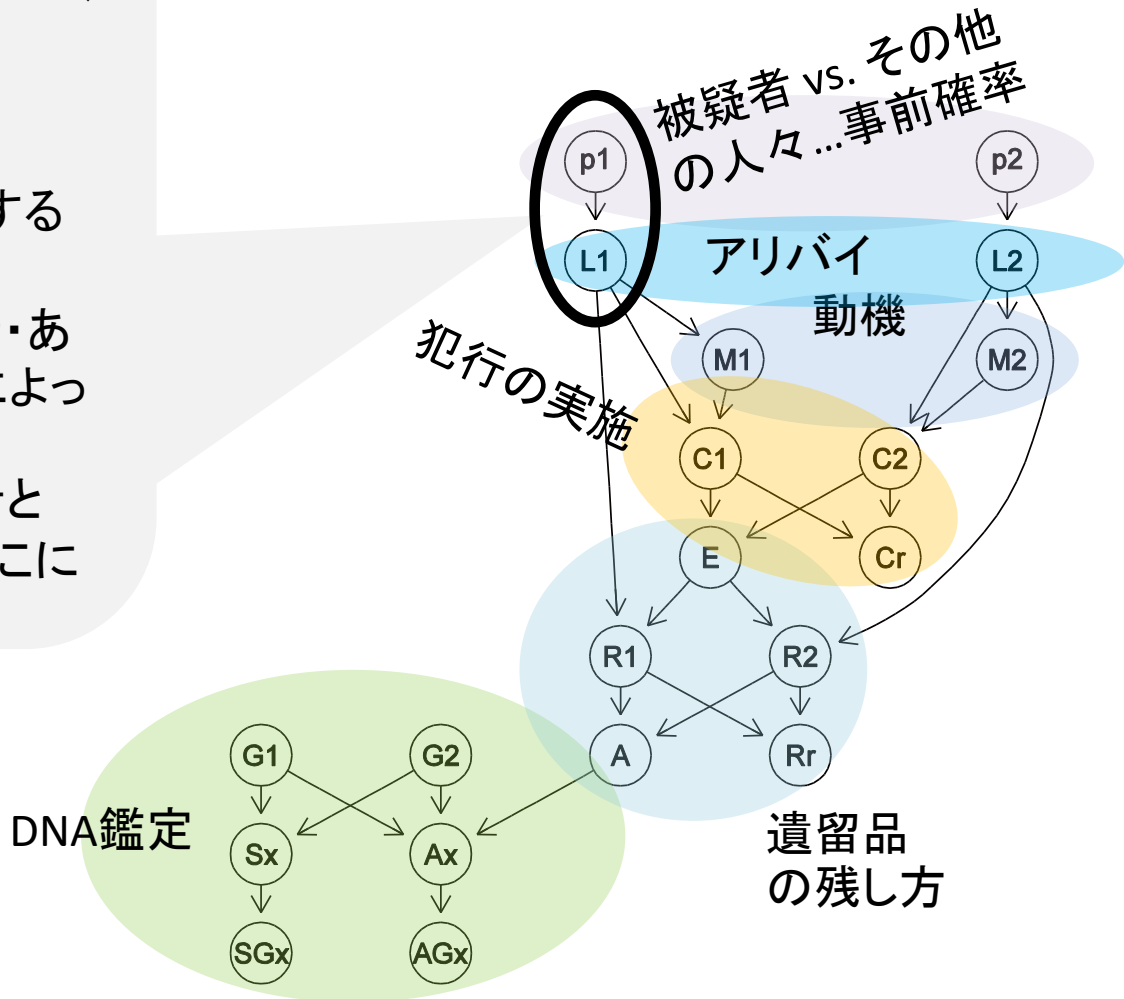
「犯行現場に犯行時刻に居られたわけがない」というアリバイがある

犯行現場にいた確率を0にしたなら、犯行は不可能になる

「アリバイ」に応じて「現場にいた確率」を小さくする

「その他大勢」も、「ある場所・ある時刻」に居た確率は状況によって上下する。

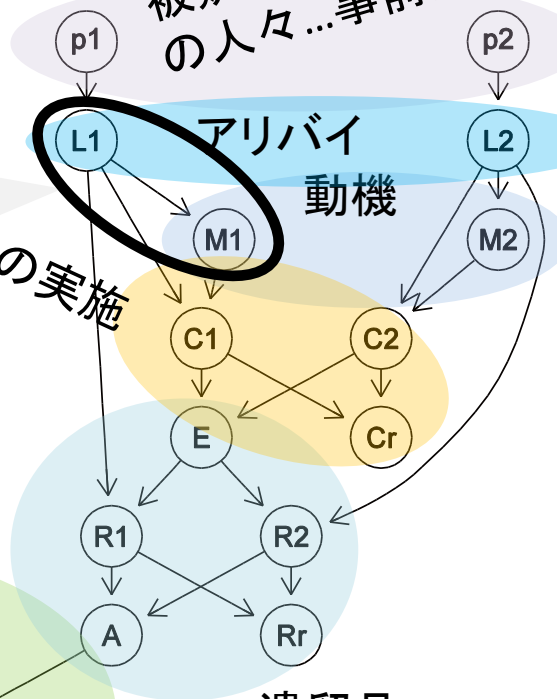
通勤ラッシュの駅での犯行と終電後の駅での犯行ではそこに居た人の人数が違う



被疑者 vs. その他
の人々...事前確率

アリバイ
動機

行の実施

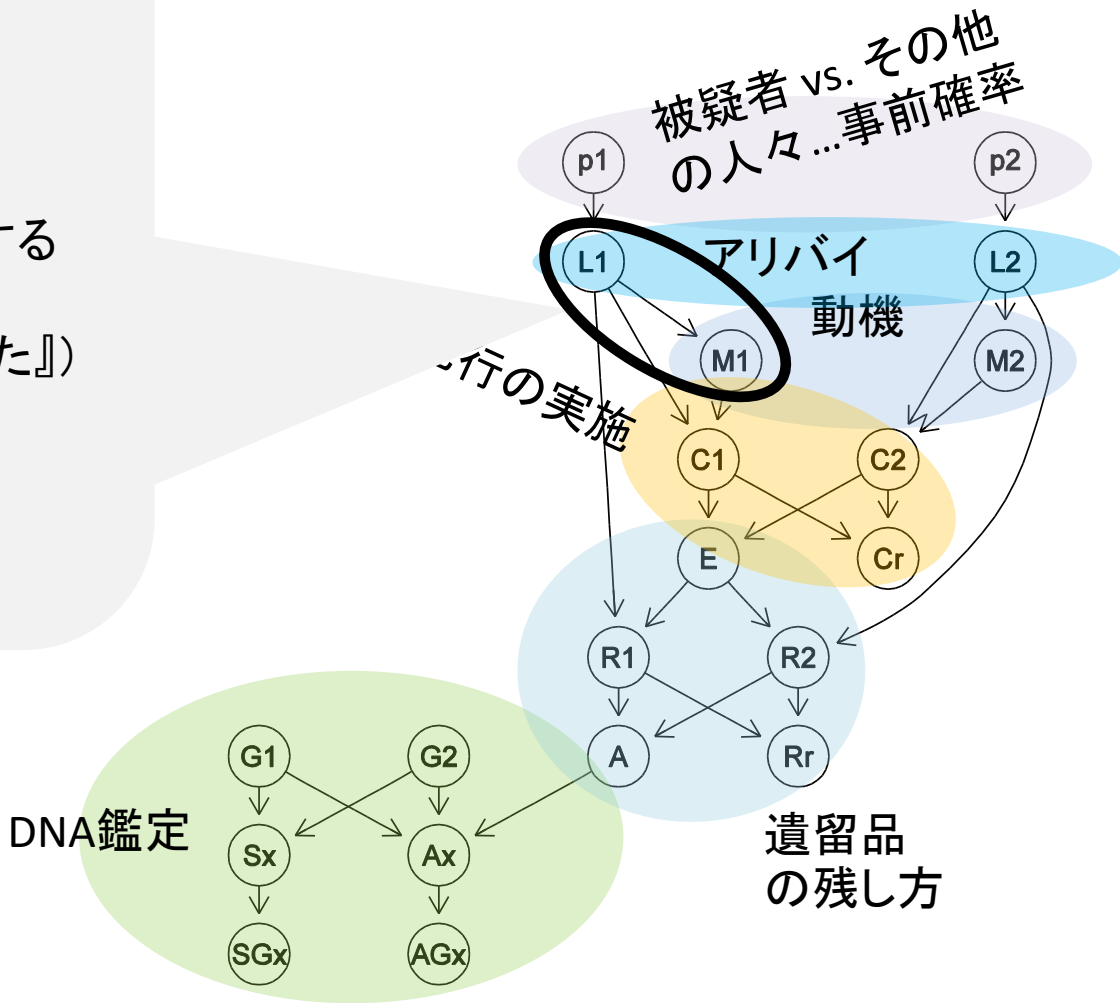


DNA鑑定

遺留品
の残し方

動機は
犯行現場に居なくても
ありえるし
(宿年の恨み！)

犯行現場でたまたま発生する
こともある
(足を踏まれて、『かっとなった』)



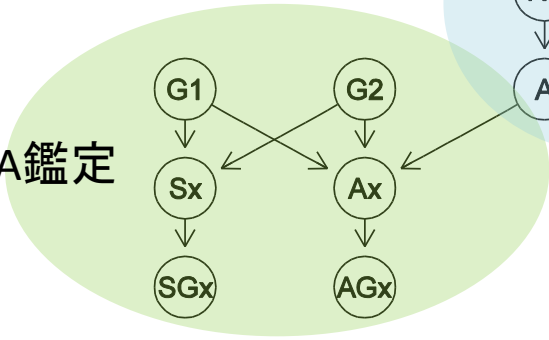
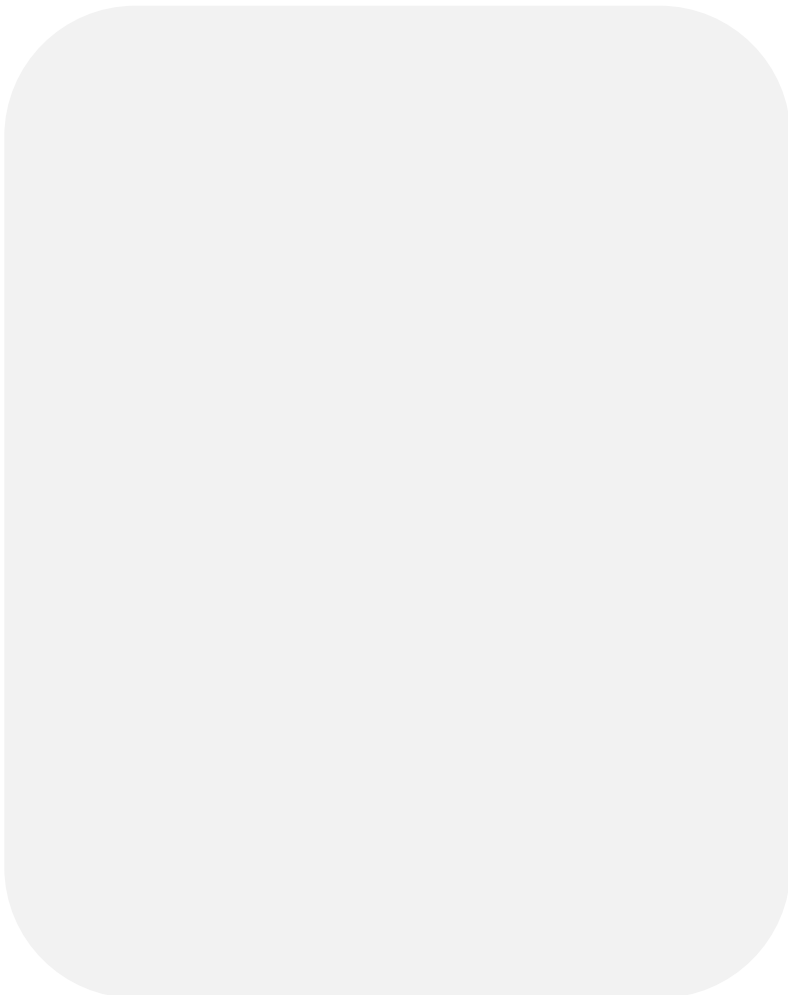
被疑者 vs. その他
の人々...事前確率

アリバイ
動機

犯行の実施

遺留品
の残し方

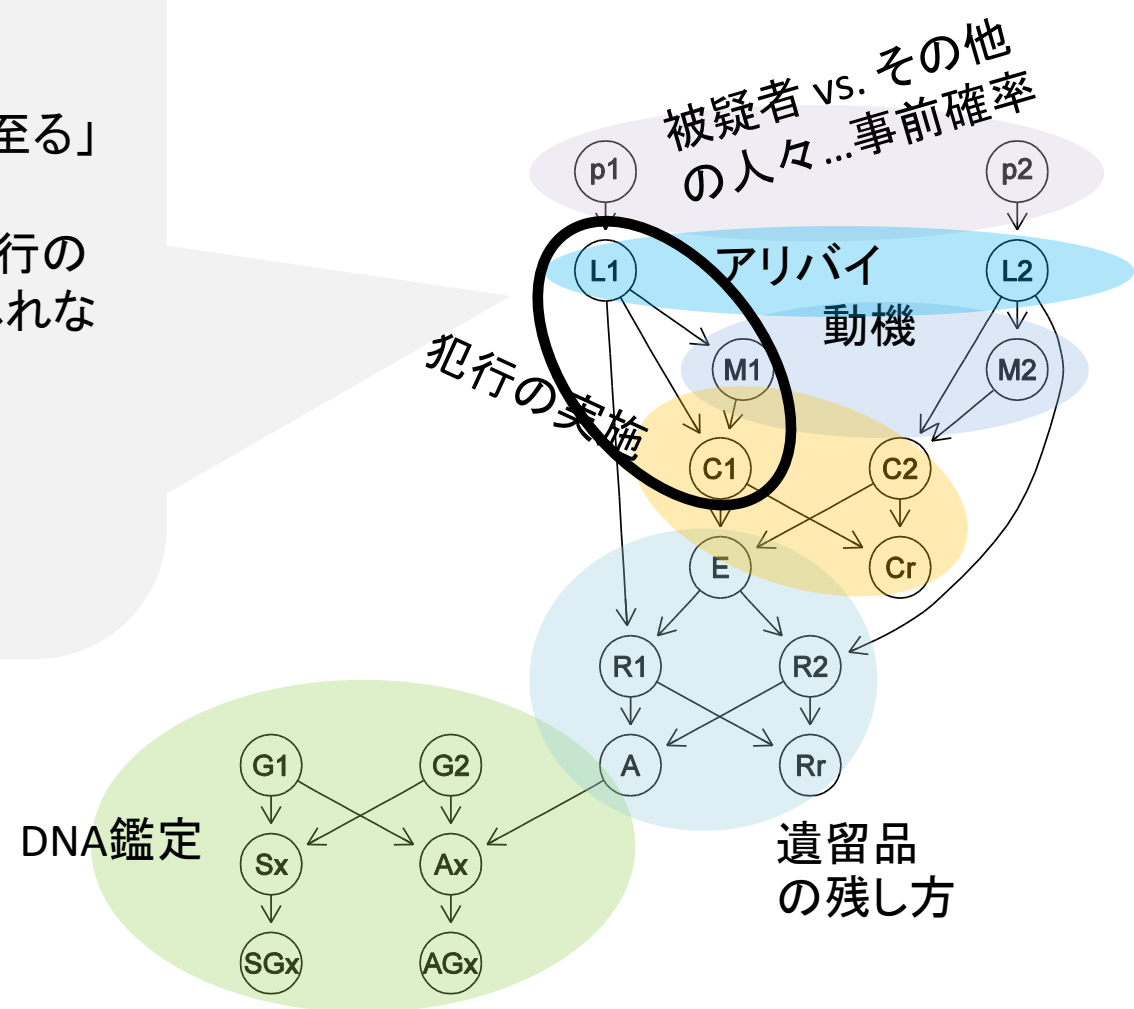
DNA鑑定

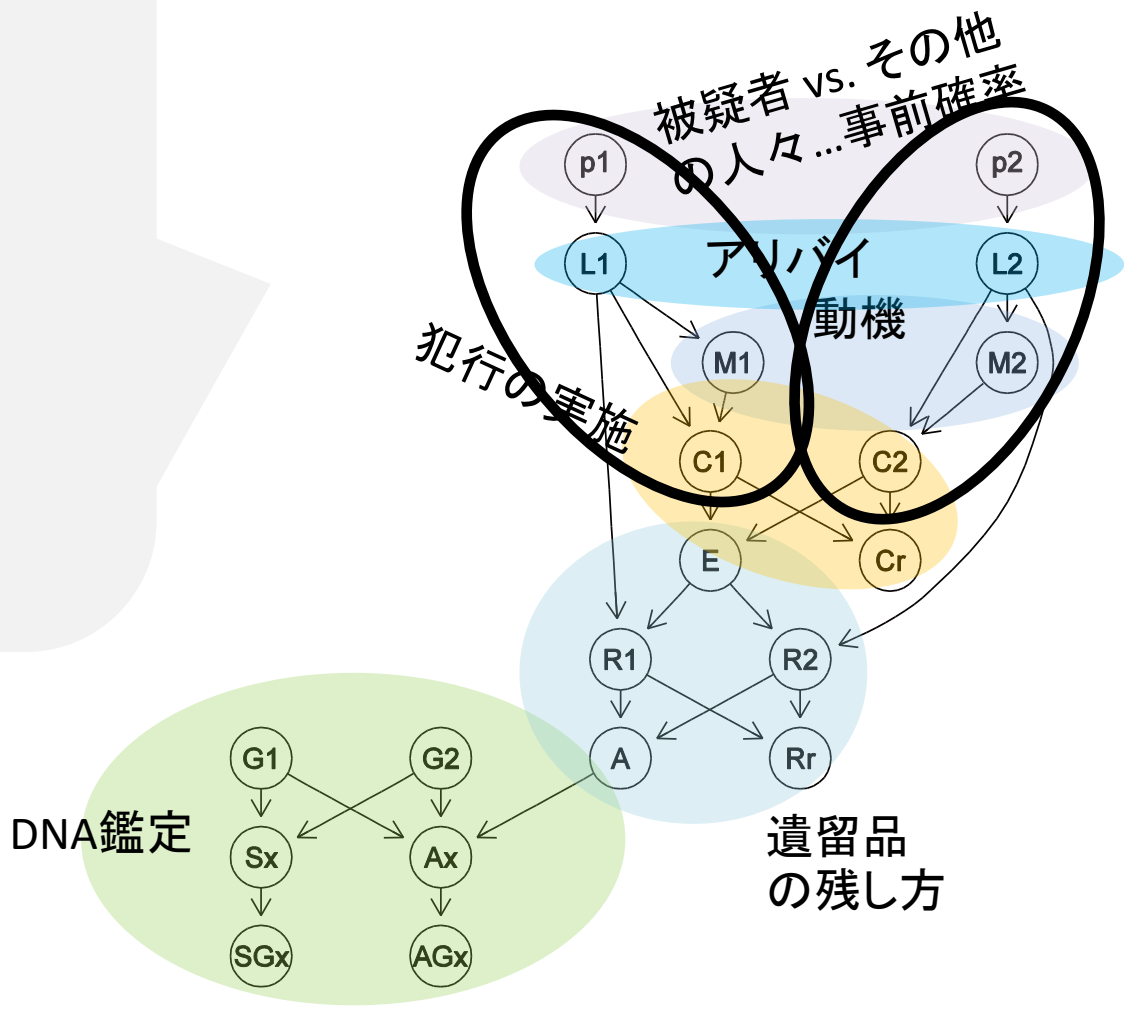


遺留品
の残し方

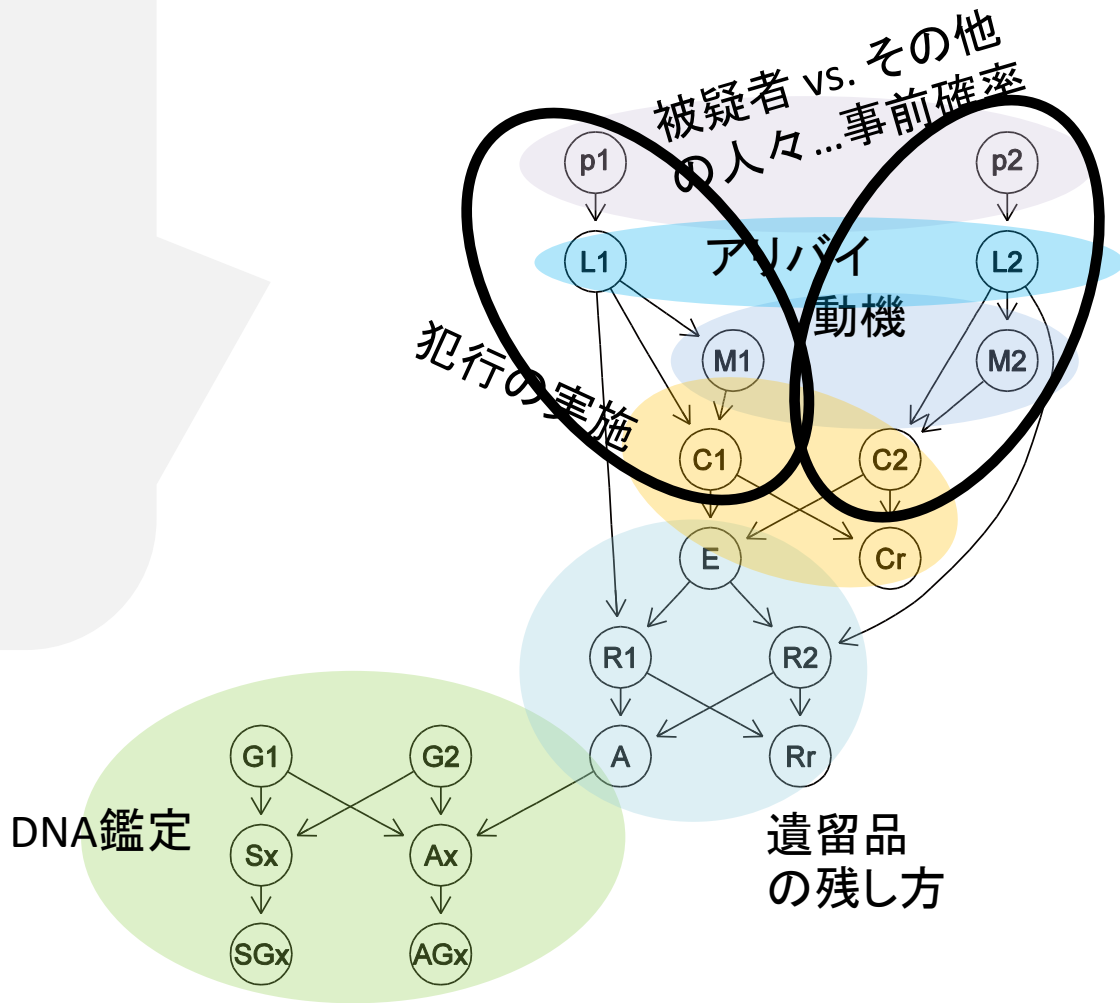
「現場に居て」
「動機がある」
と
「ある確率で犯行の実行に至る」

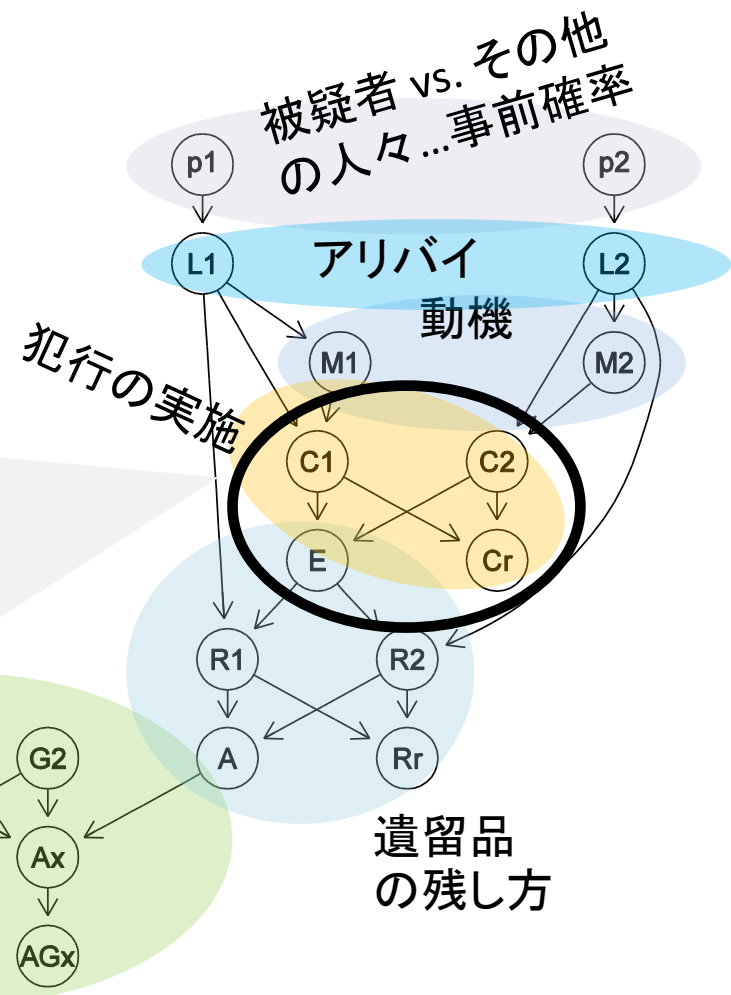
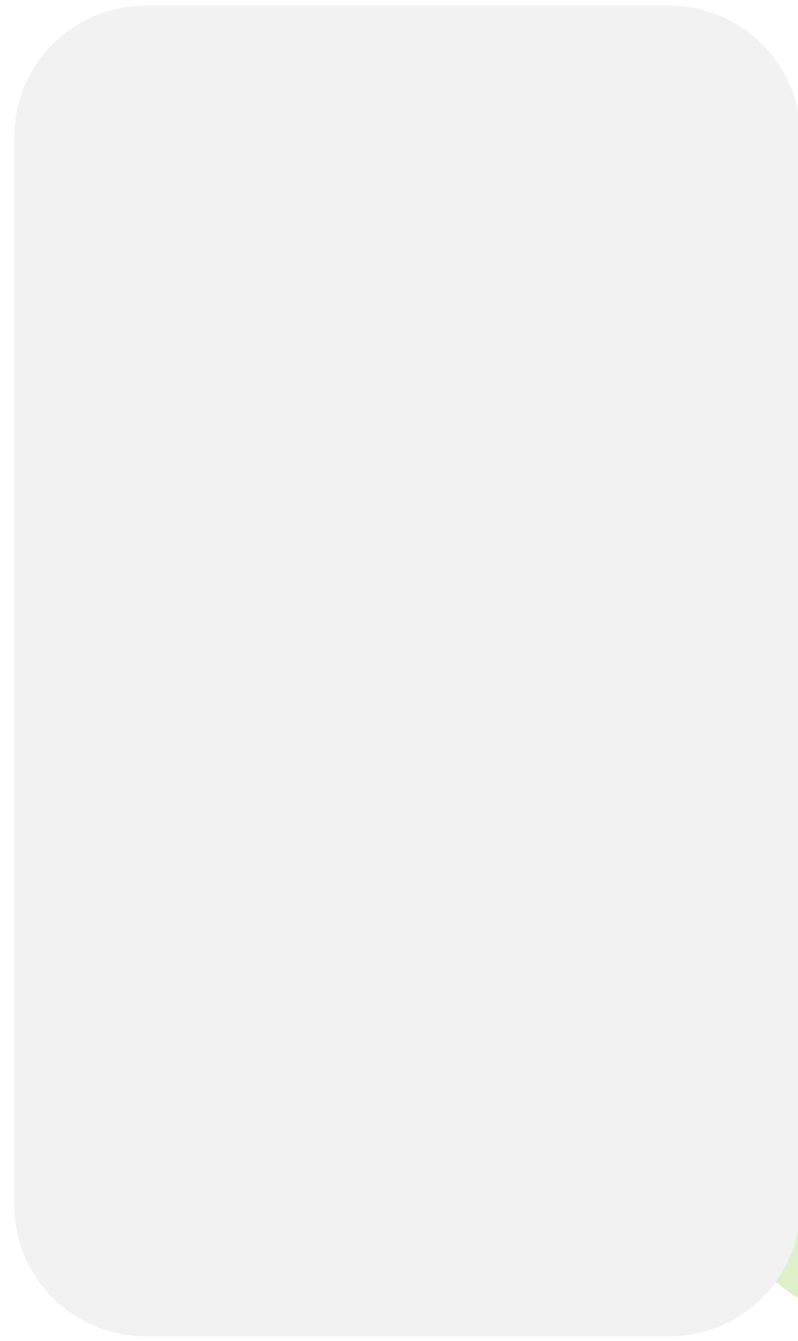
動機がなくても偶発的に犯行の
実行に至ることもあるかもしれ
ない





「被疑者」
「その他の人」
で対称的





少し、ベイジアン・ネットワーク構成上の技術的なことになるが...

「犯行を実行した」かどうかを、「被疑者」と「その他」の2系列で取り扱ってきた

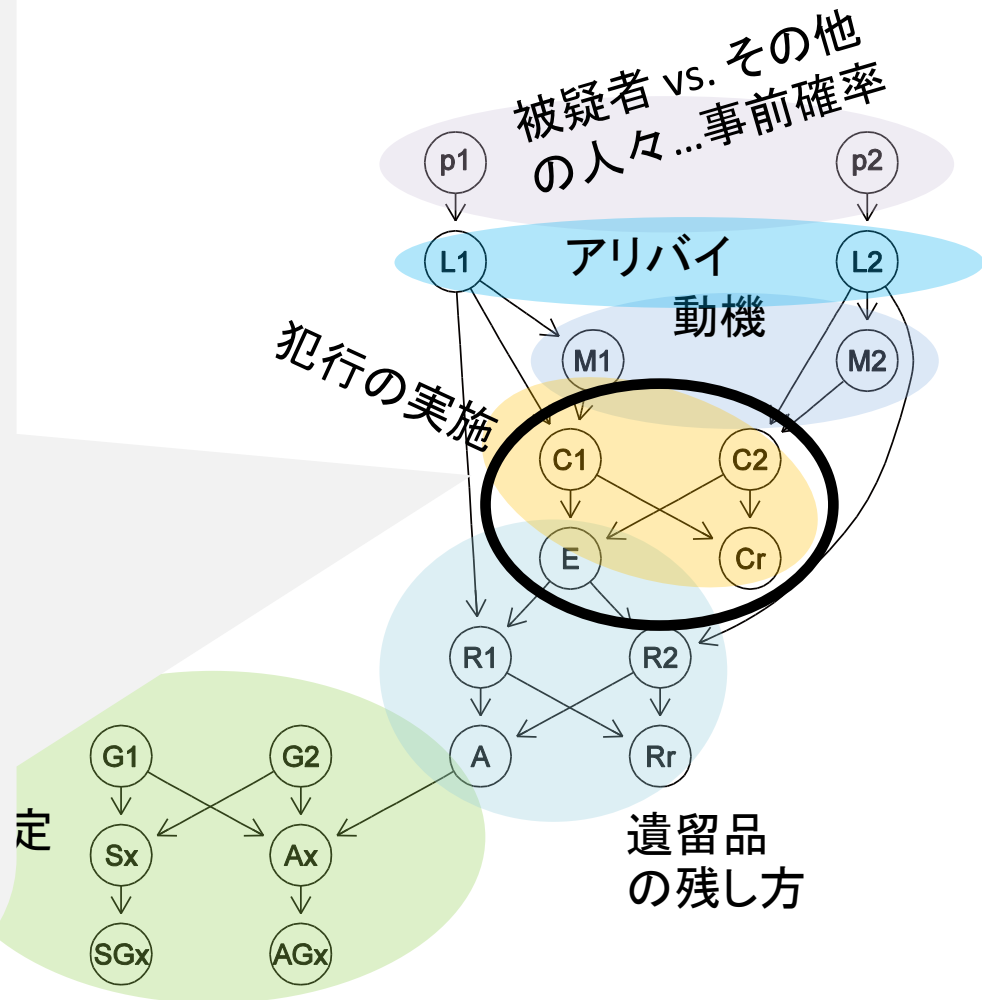
実際の犯行は1件なので

Cr:

C1,C2ともに、ある確率で「犯行の実行に至る」が、「実際に犯行が起きた」ということを変数に入れない

E:

C1,C2ともに、ある確率で「犯行の実行に至る」が、「実際に実行したのはどちらか片方」という制約を入れない



少し、ベイジアン・ネットワーク構成上の技術的なことになるが...

「犯行を実行した」かどうかを、「被疑者」と「その他」の2系列で取り扱ってきた

実際の犯行は1件なので

Cr:

C1,C2ともに、ある確率で「犯行の実行に至る」が、「実際に犯行が起きた」ということを変数に入れない

E:

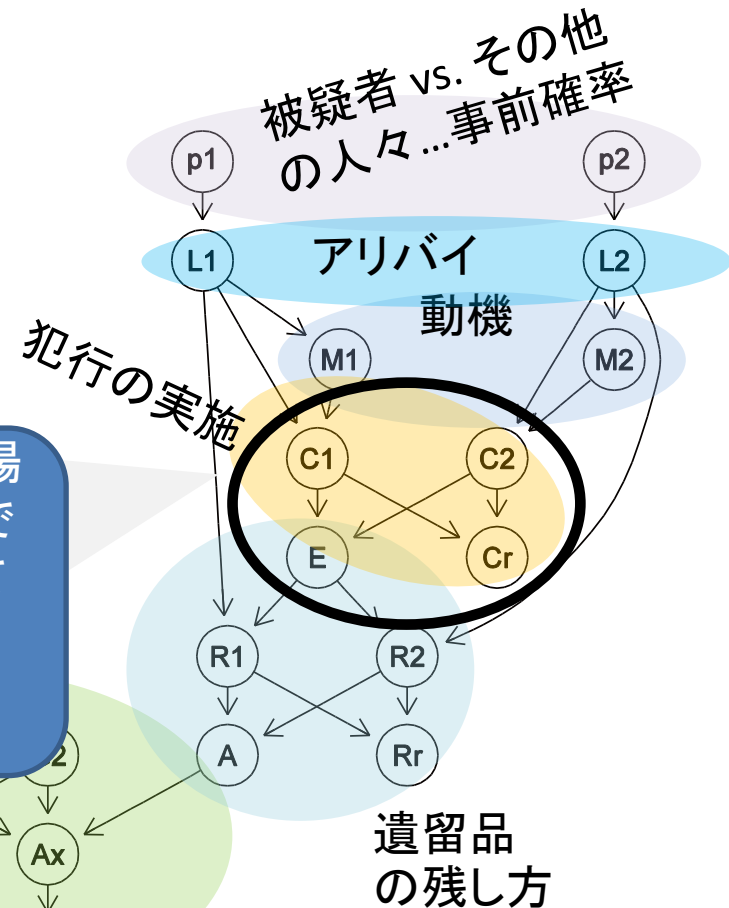
C1,C2ともに、ある確率で「犯行の実行に至る」が、「実際に実行したのはどちらか片方」という制約を入りたい

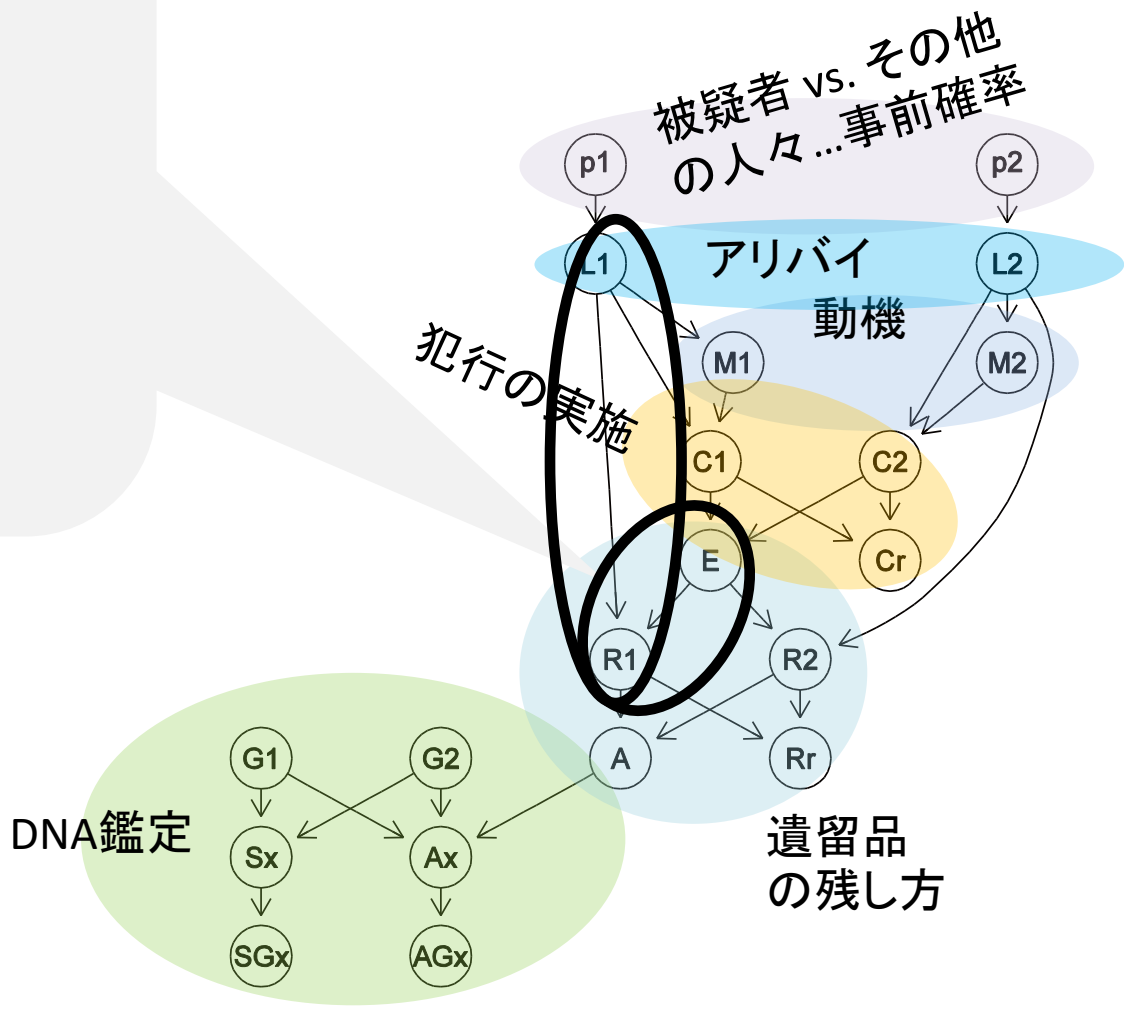
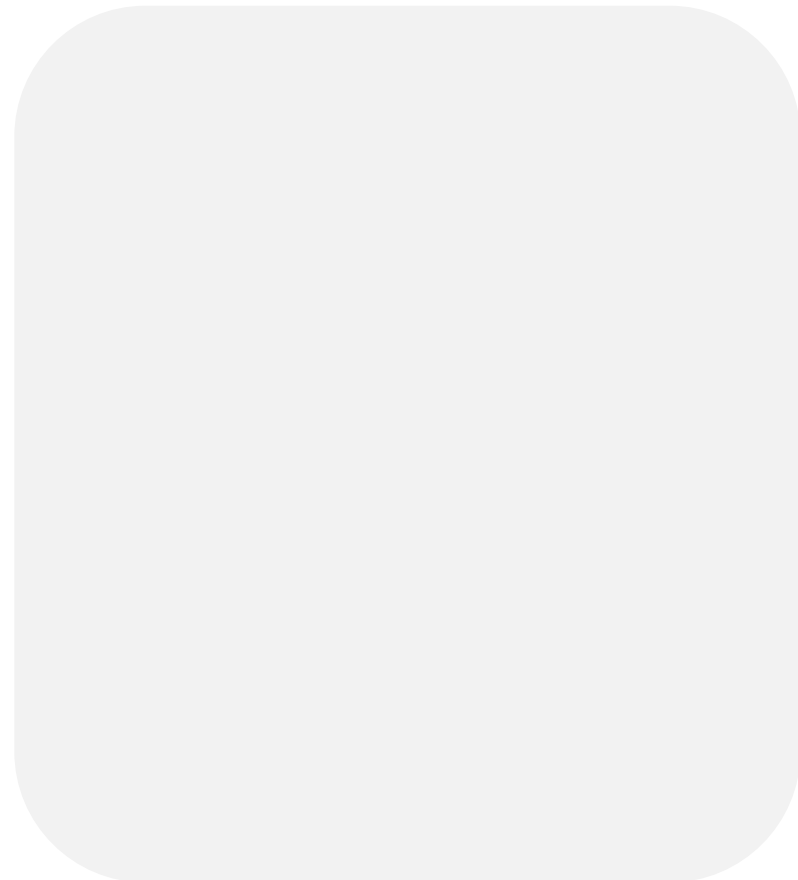
『強い動機で現場に行ったら、すでに犯行は起きていた！』
チャラリン

定

Sx
SGx

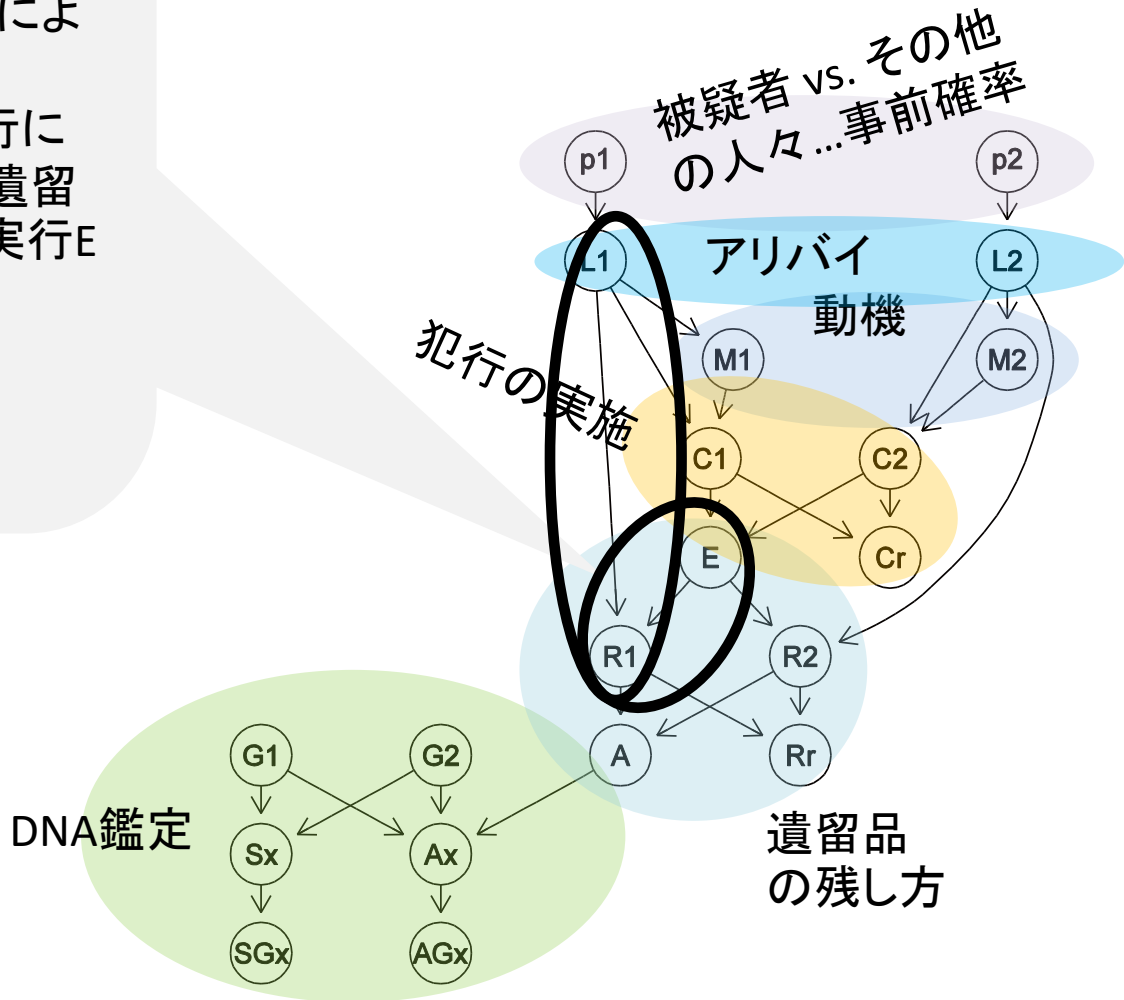
Ax
AGx

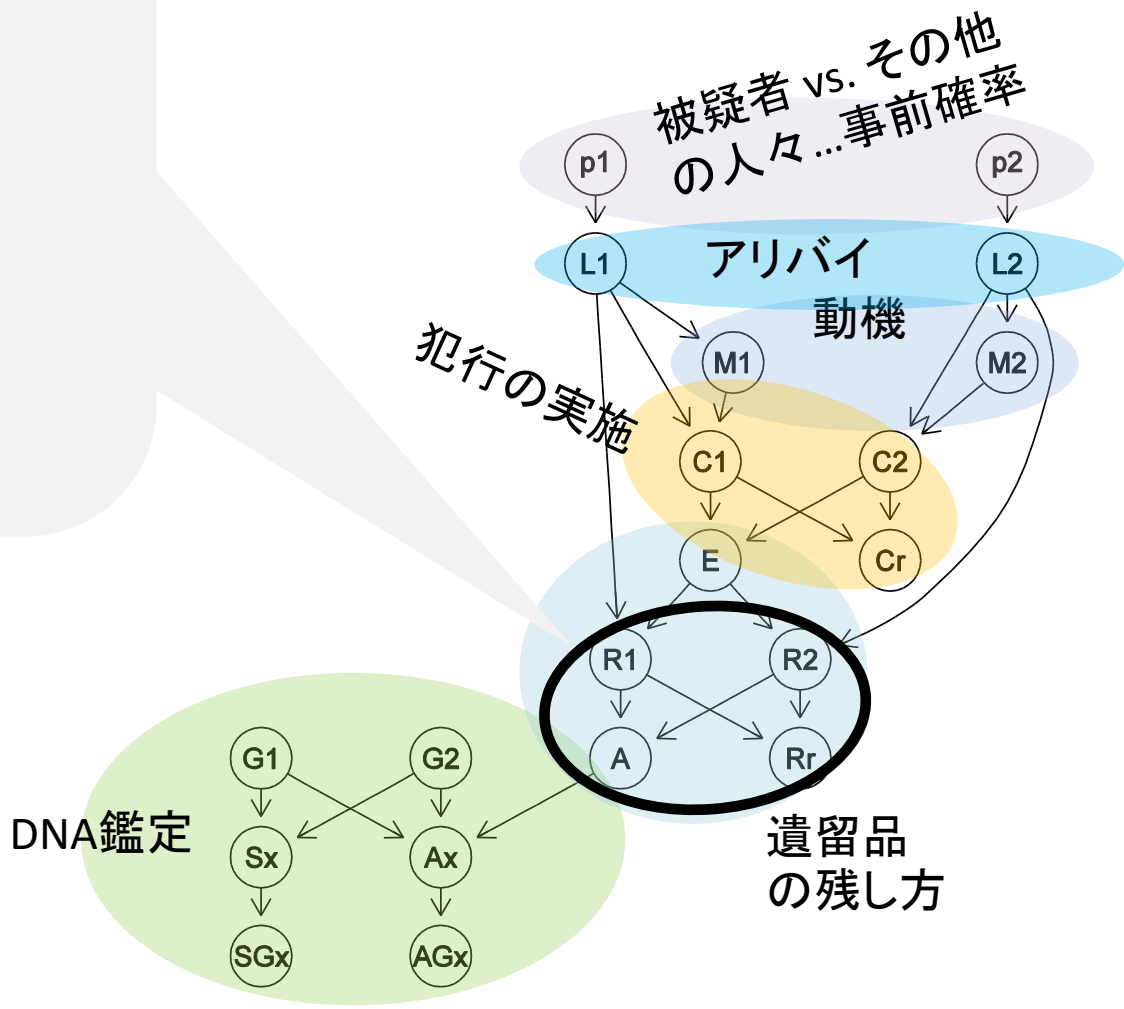




遺留品(にDNAがある)が現場にあるかどうかは、
「犯行現場に居た」かどうかによる

また、「いかにも犯行の実行に伴ってしか残らない」ような遺留品であれば、それは犯行の実行Eによって決まる

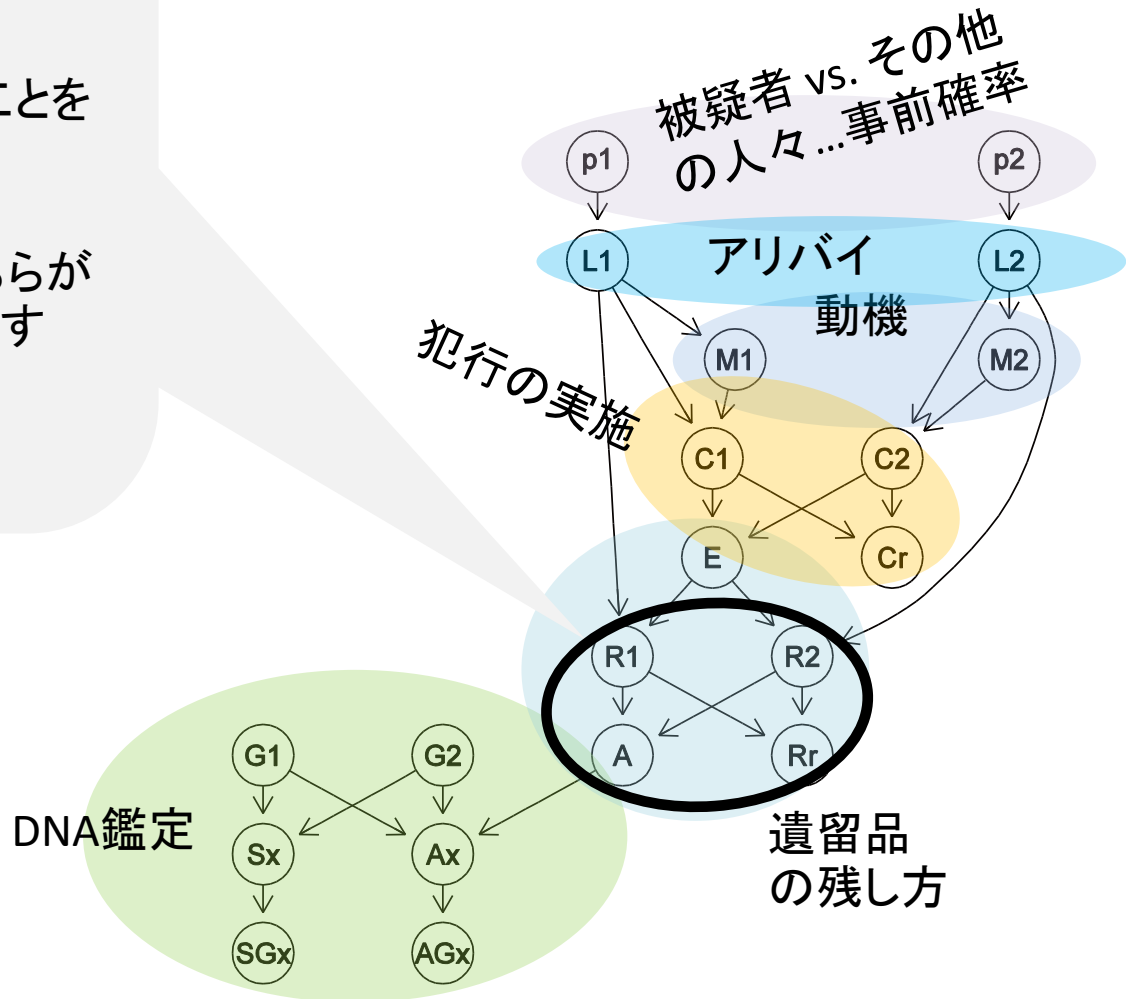




少し、ベイジアン・ネットワーク構成上の技術的なことになるが...

Rr:
確かに遺留品があるということを
確率変数にしたかった

A:
「被疑者」と「その他」と、どちらが
遺留品の『持ち主』かを表す



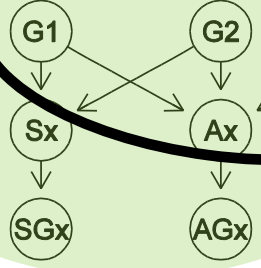
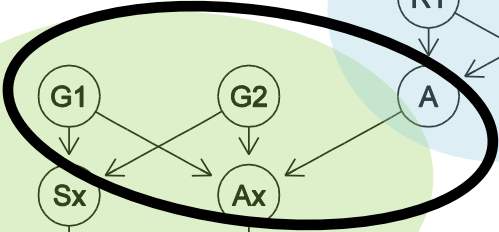
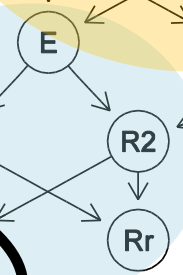
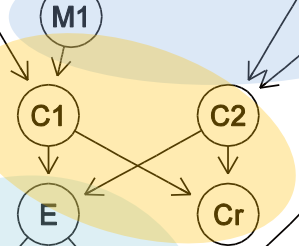
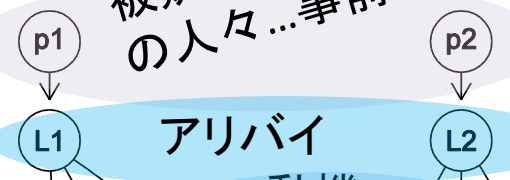
被疑者 vs. その他
の人々...事前確率

アリバイ
動機

犯行の実施

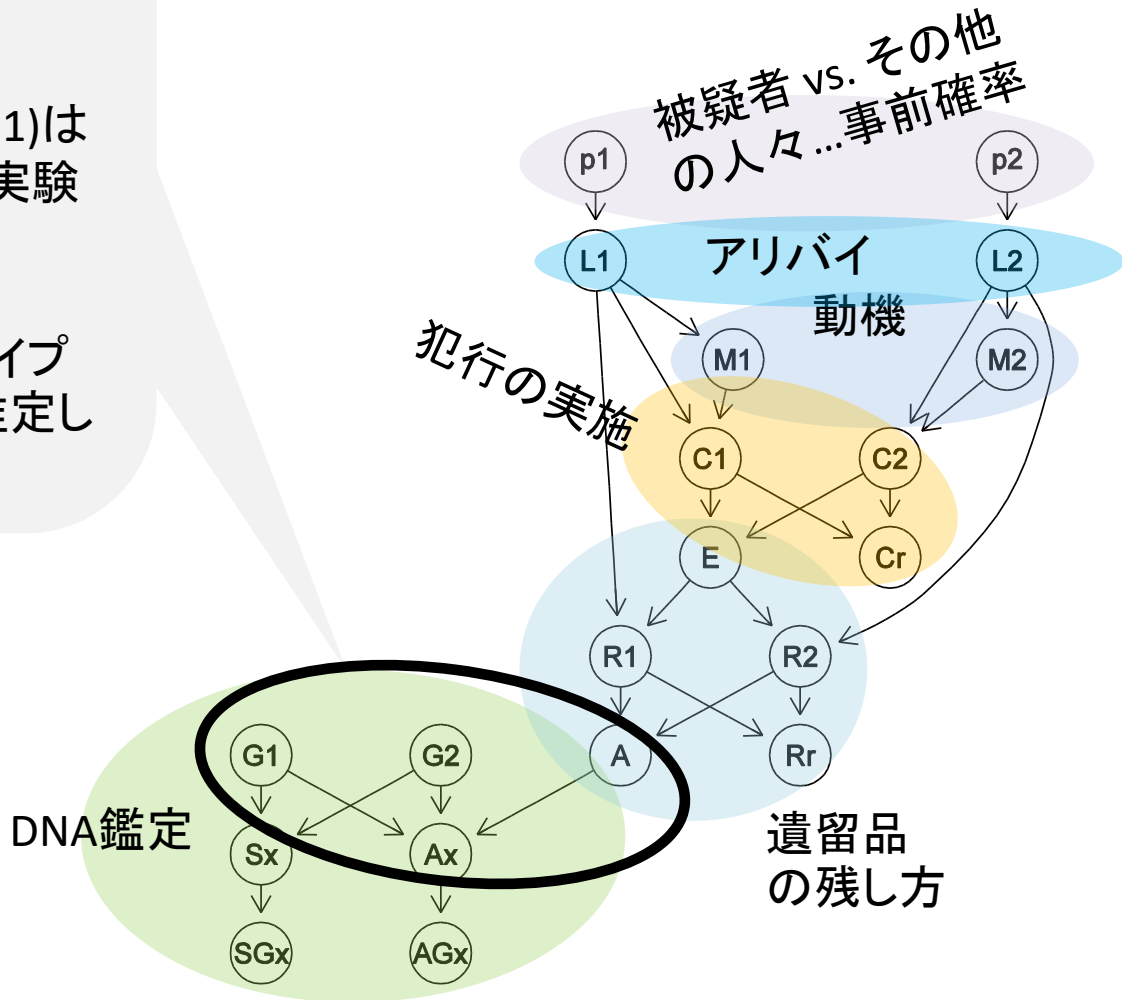
DNA鑑定

遺留品の
残し方



遺留DNAの『真の』ジェノタイプは
遺留品の持ち主(A)
持ち主候補(被疑者とその他の
人)のそれぞれの『真の』のジェノ
タイプ(G1,G2)
で決まる

被疑者の真のジェノタイプ(G1)は
被疑者からDNAを採取して実験
すれば
推定できるし
その他の人の真のジェノタイプ
(G2)は集団データなどから推定し
たものになっている



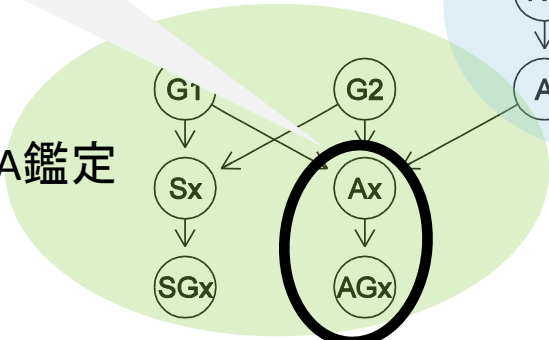
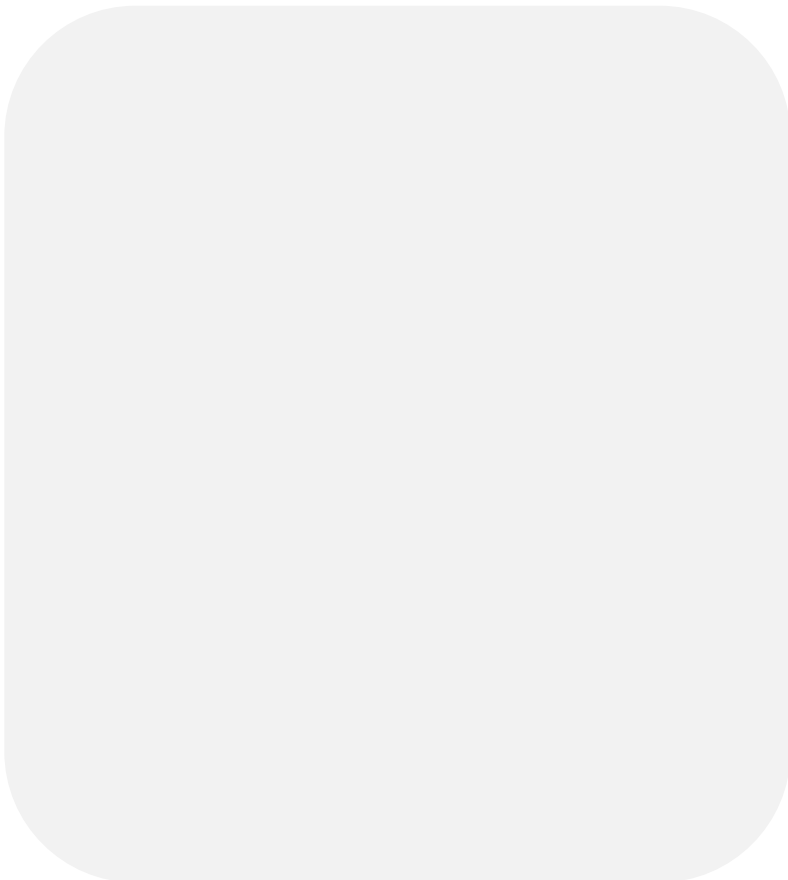
被疑者 vs. その他
の人々...事前確率

アリバイ
動機

犯行の実施

遺留品
の残し方

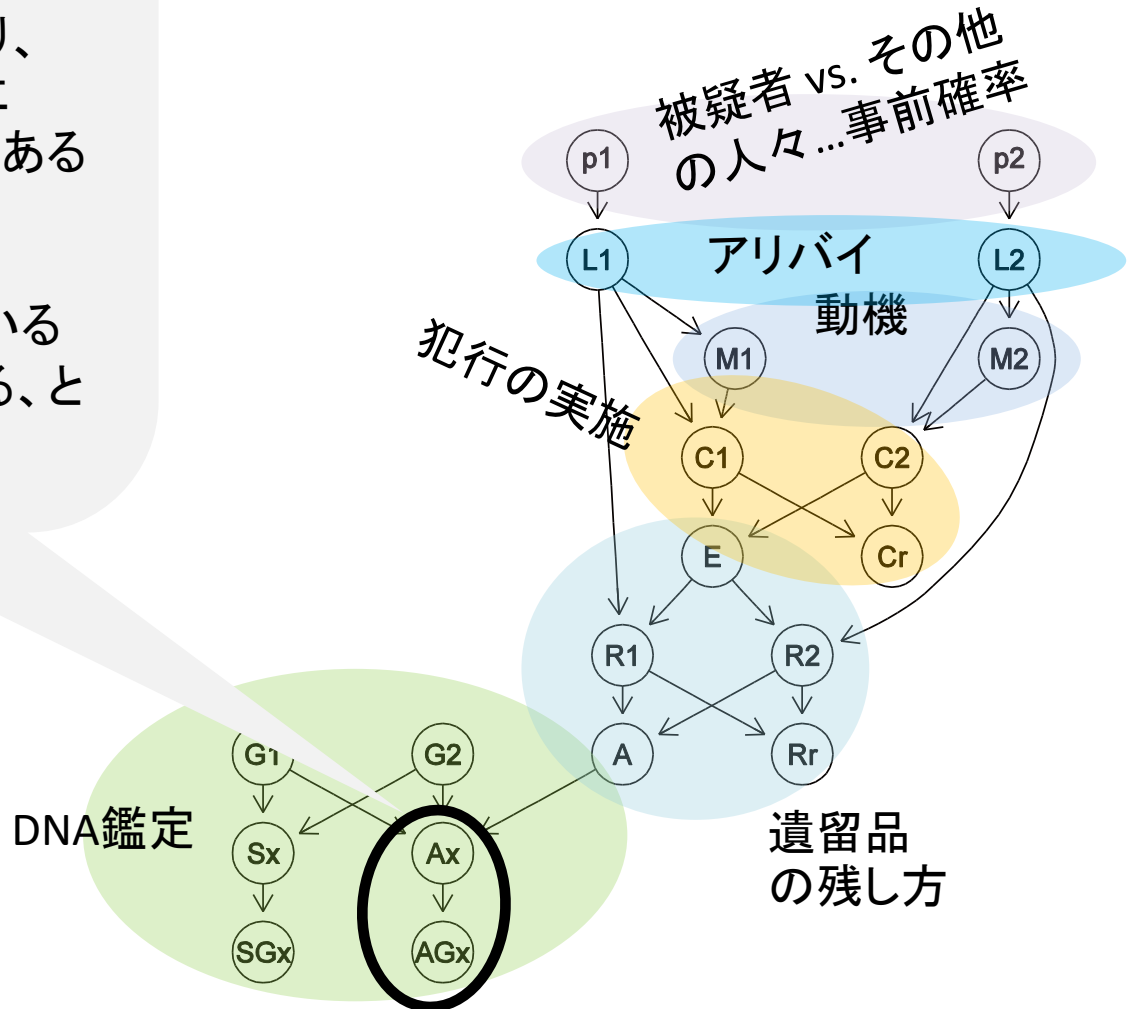
DNA鑑定



『真の』ジェノタイプというのは
実験結果から推定するもの

真のジェノタイプがAxのとき
実験結果のジェノタイプがAGxと
なることは確率的に決まり、
実験結果がAGxのときに
真のジェノタイプAxがなんである
かも確率的に決まる

実験エラーが組み込んでいる
もちろん、実験が完璧である、と
いう設定もできる



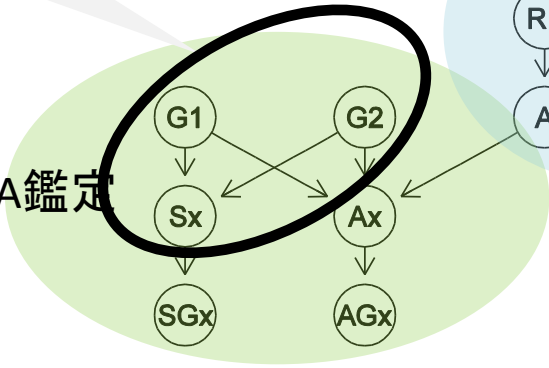
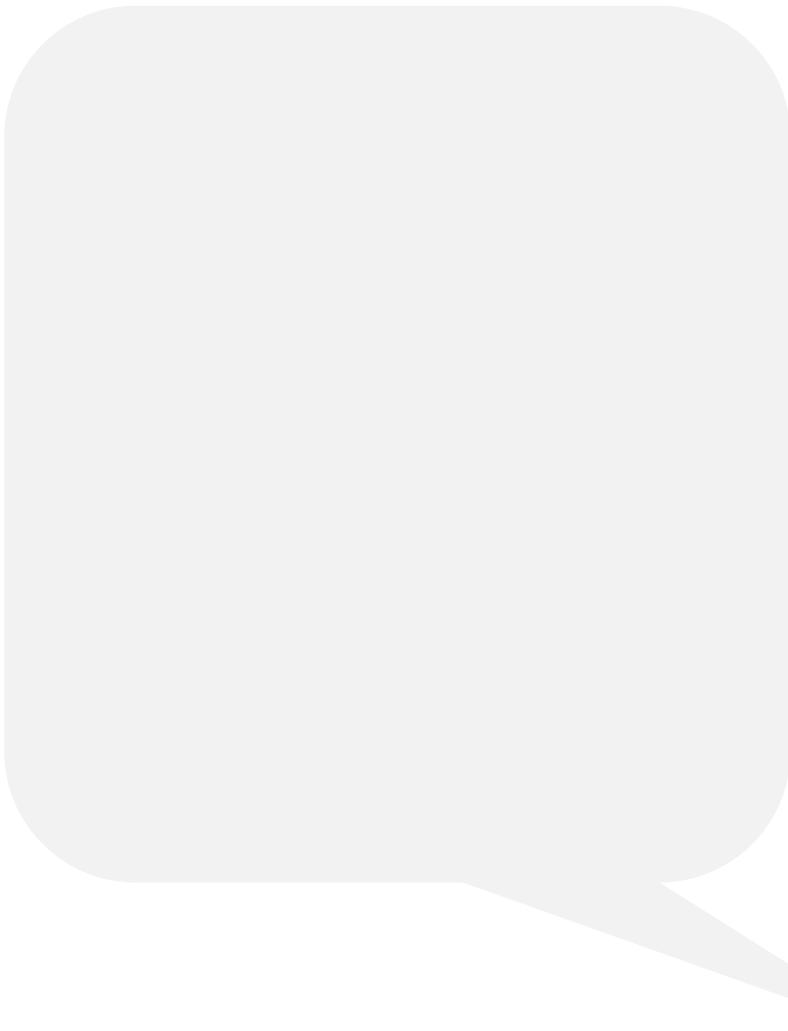
被疑者 vs. その他
の人々...事前確率

アリバイ
動機

犯行の実施

遺留品
の残し方

DNA鑑定



犯人からDNAを採取してタイピング
実験をする

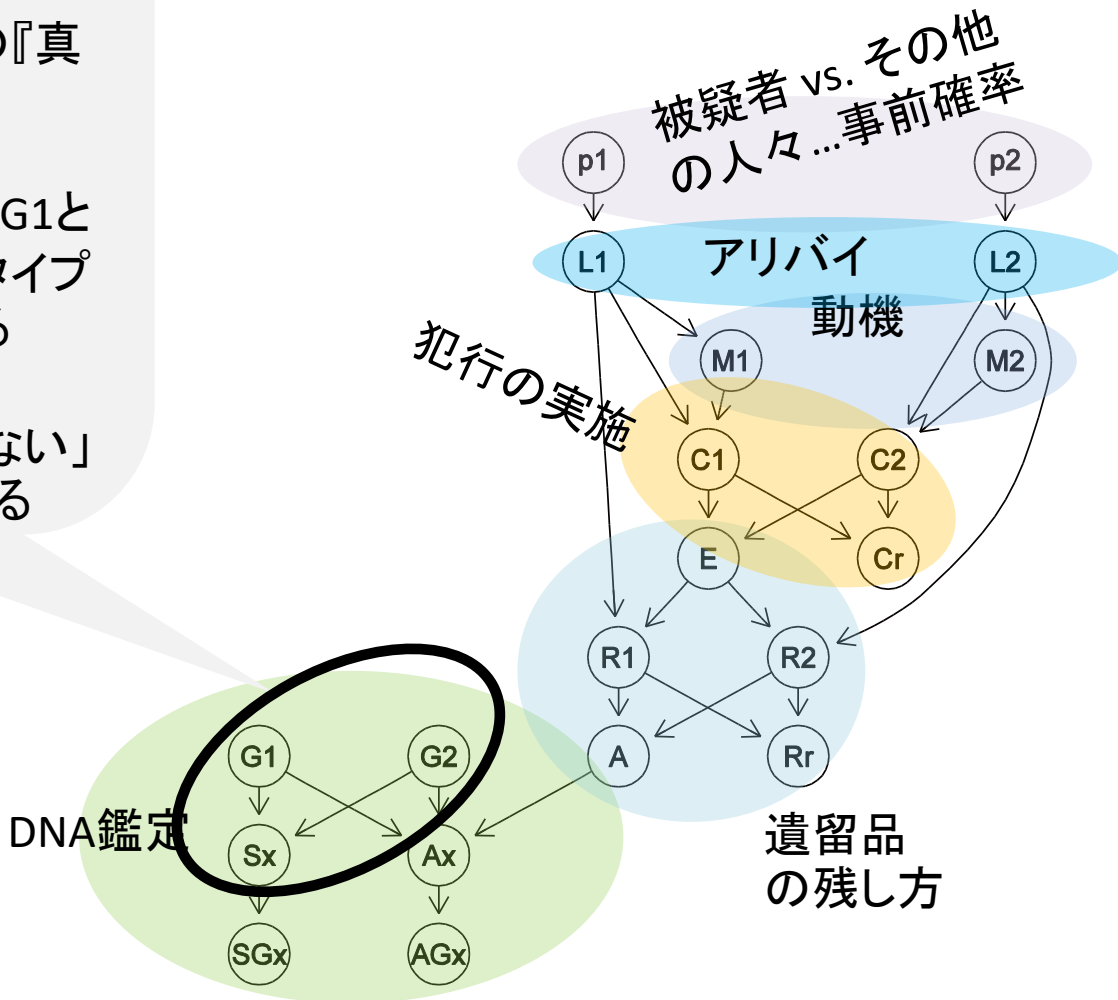
もしかしたらサンプルの取り違い
があるかもしれない

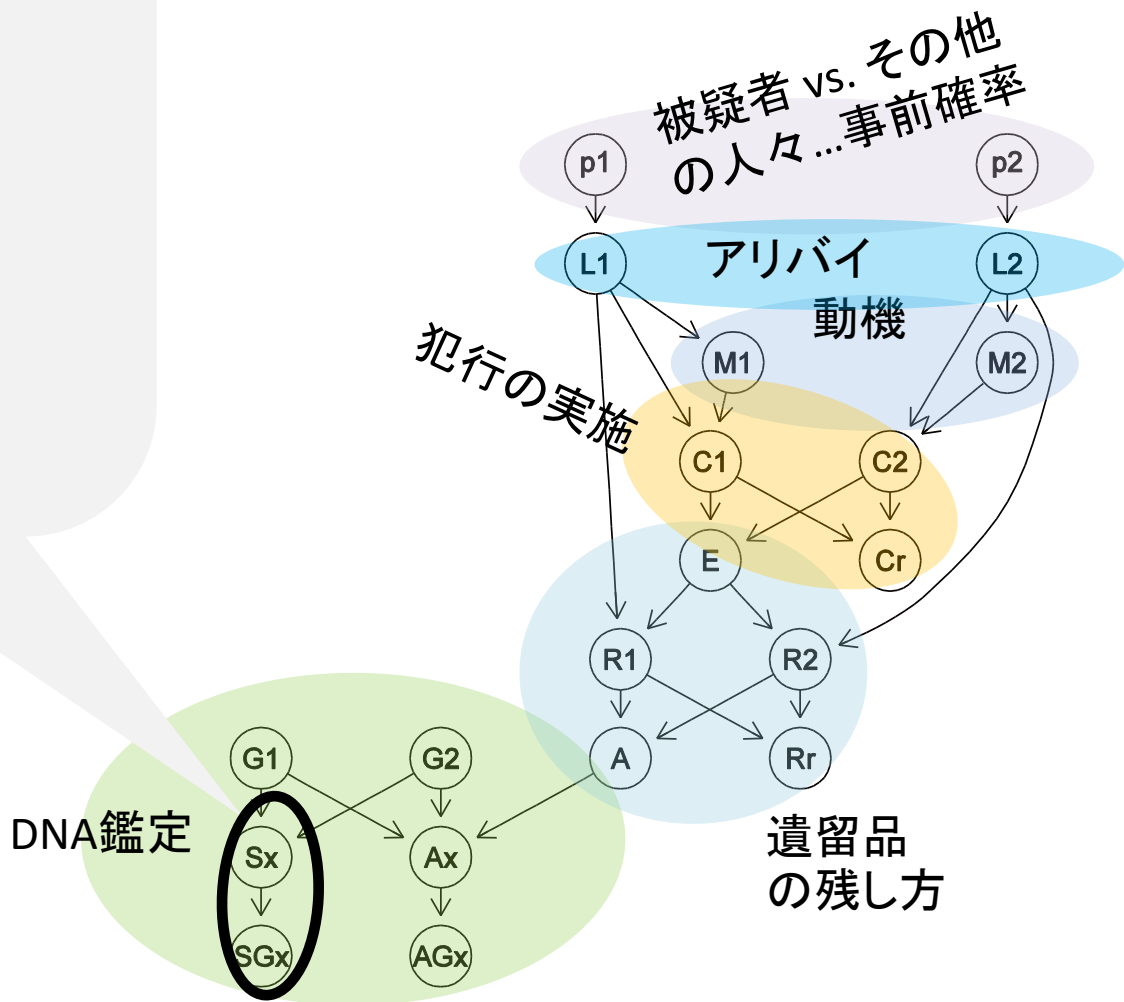
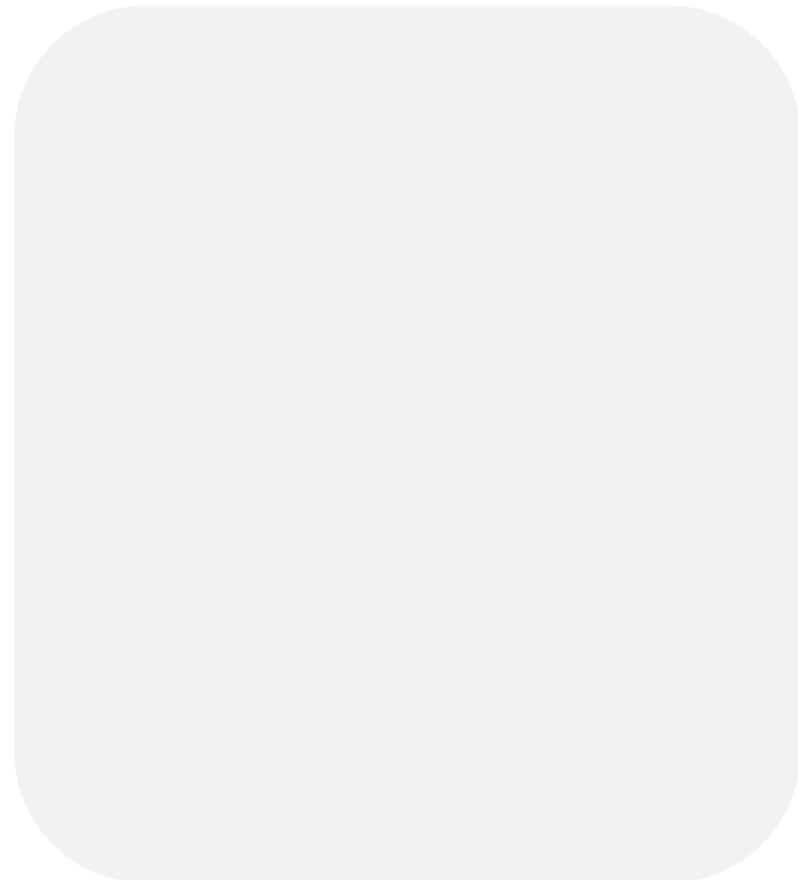
「実験に供されたサンプルの『真
の』ジェノタイプ: S_x

は

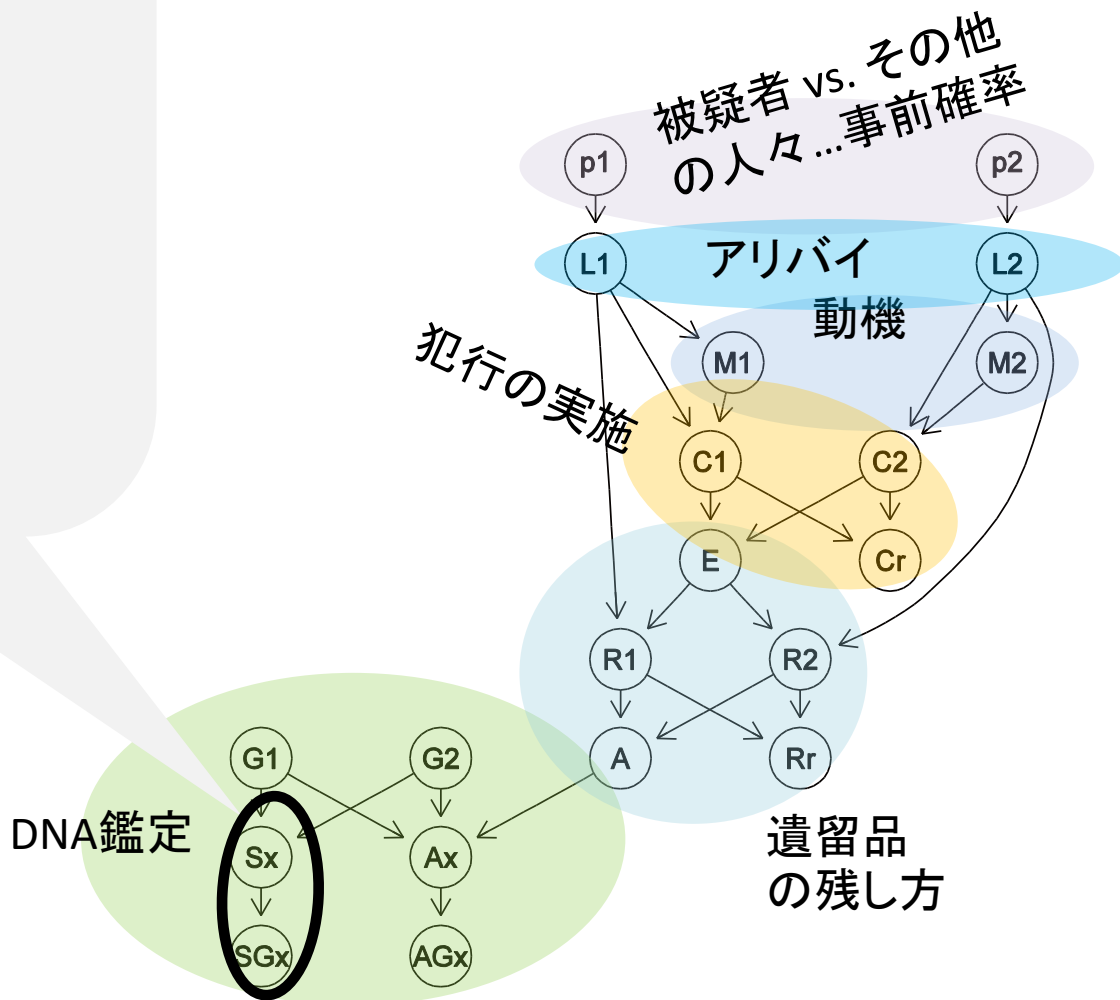
被疑者の『真の』ジェノタイプ G_1 と
その他の人の『真の』ジェノタイプ
 G_2 とから確率的に決まる

サンプル取り違いが「ありえない」
という設定ももちろんできる



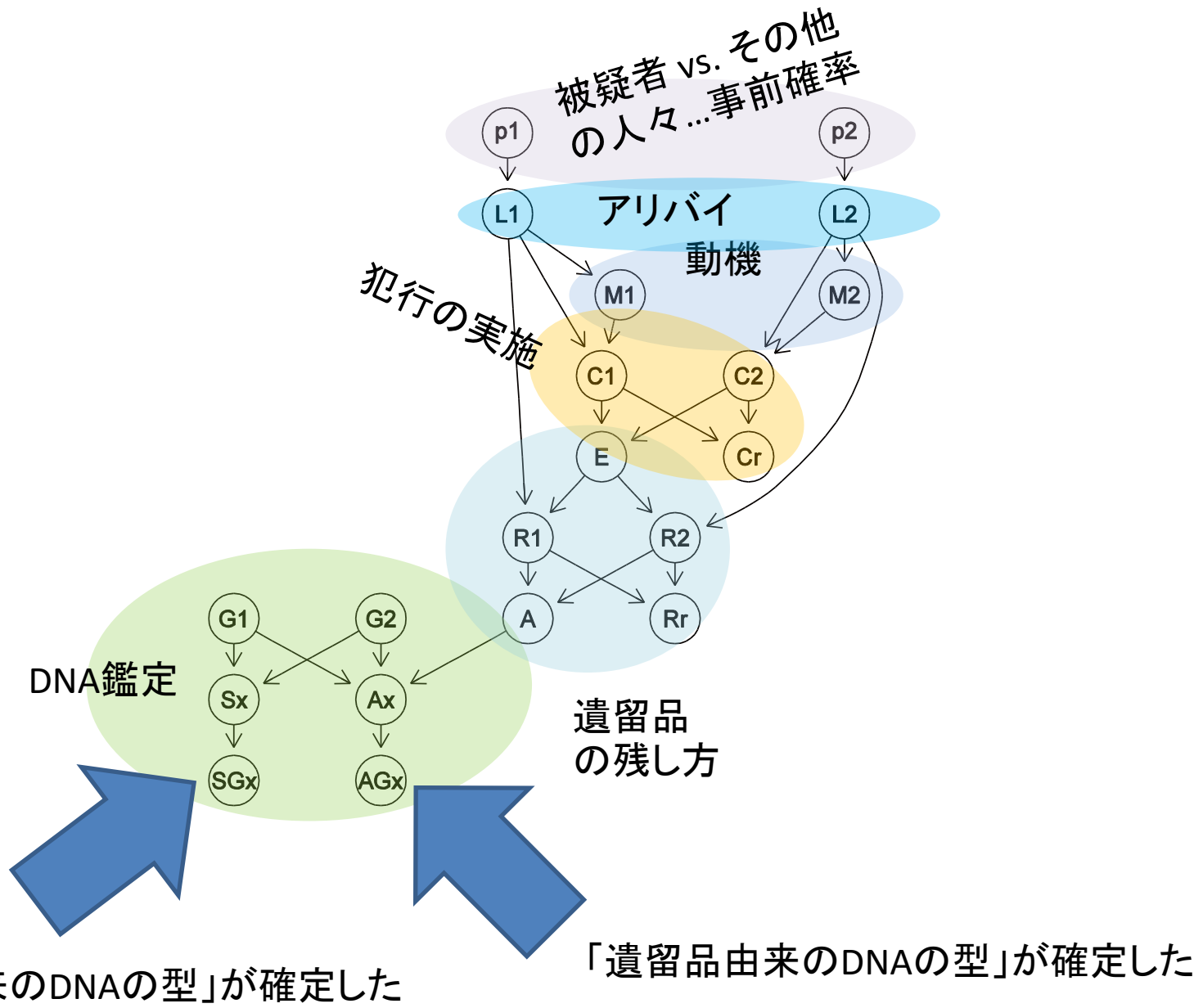


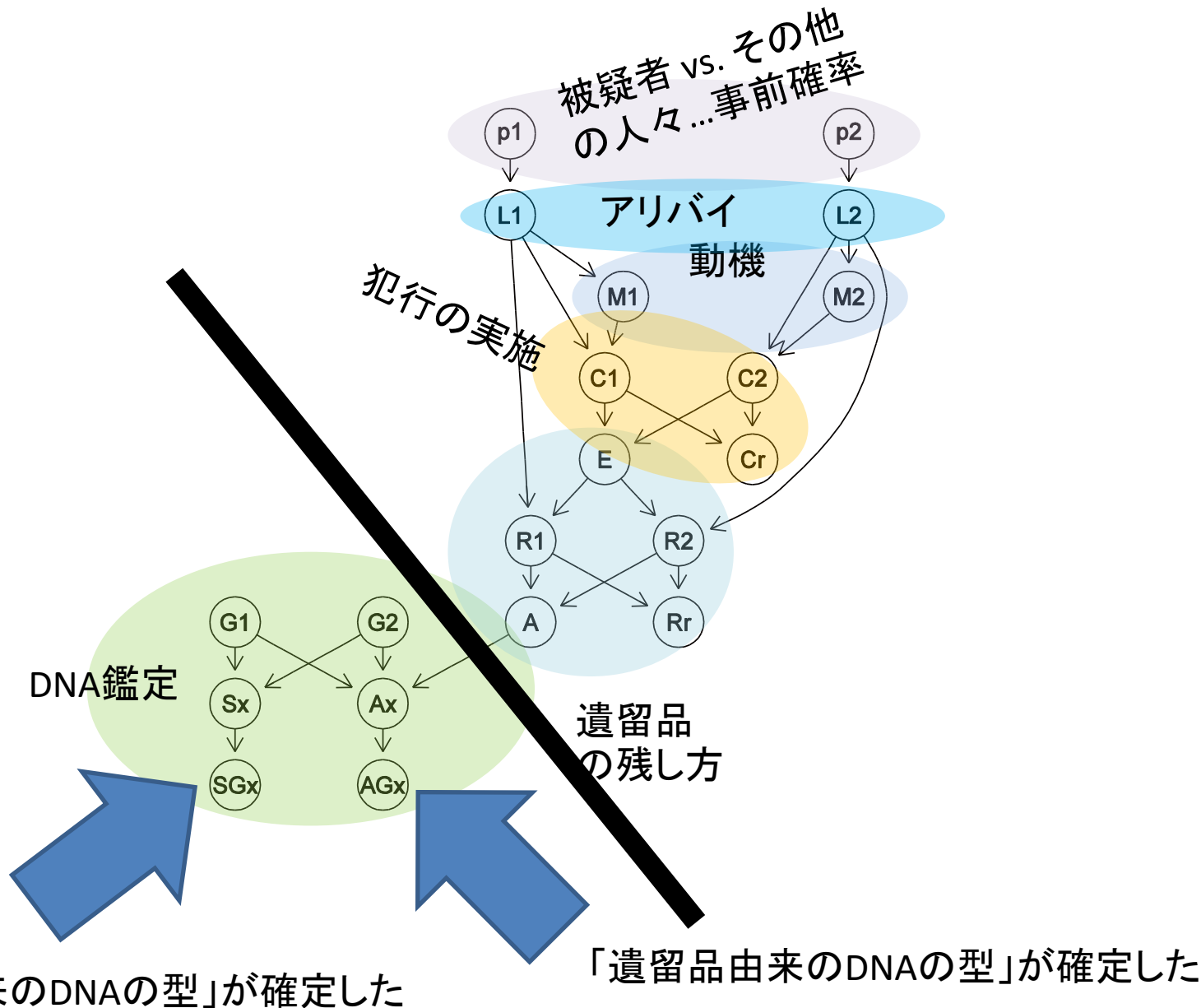
真のジェノタイプ
と
実験結果
との関係



終わった！

- 「親子関係」を逐一確認していっただけ
- ベイジアン・ネットワークは
 - 「子」が「親」になることでだんだんに出来上がって行った

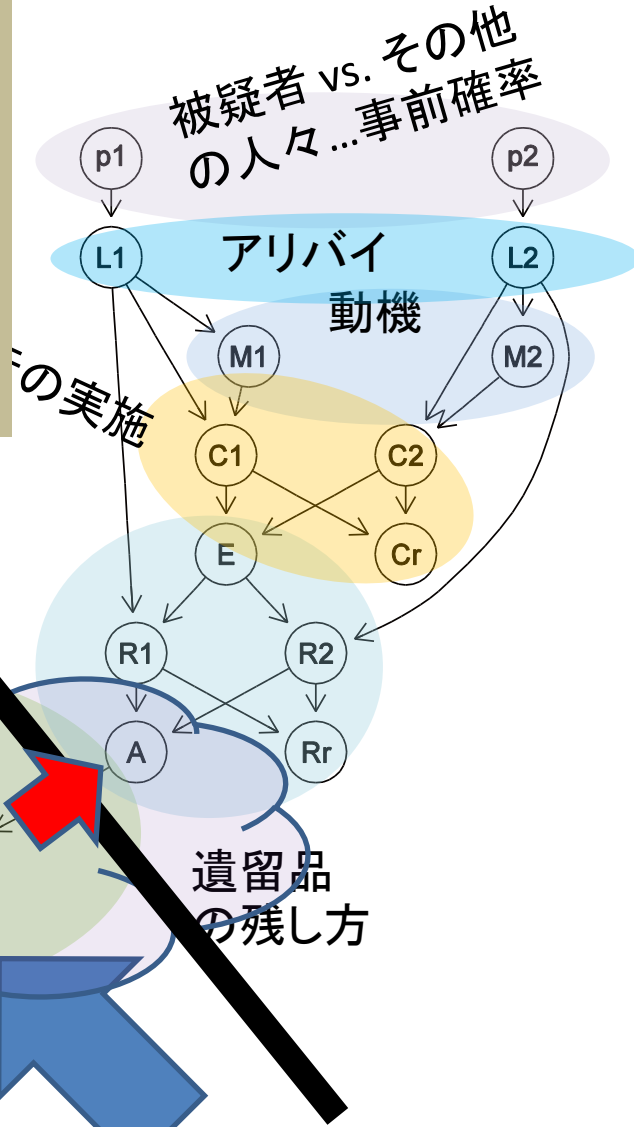




「被疑者由来のDNAの型」が確定した

「遺留品由来のDNAの型」が確定した

被疑者由来のDNA実験結果
と
遺留品由来のDNA実験結果
とを比較して
(それが一致していたときに)
『遺留品を残したのが被疑者である』
という
Aの事後確率が上がる



「被疑者由来のDNAの型」が確定した

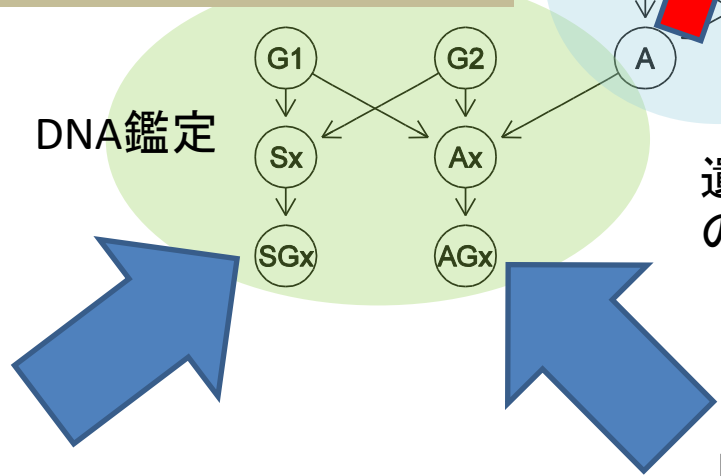
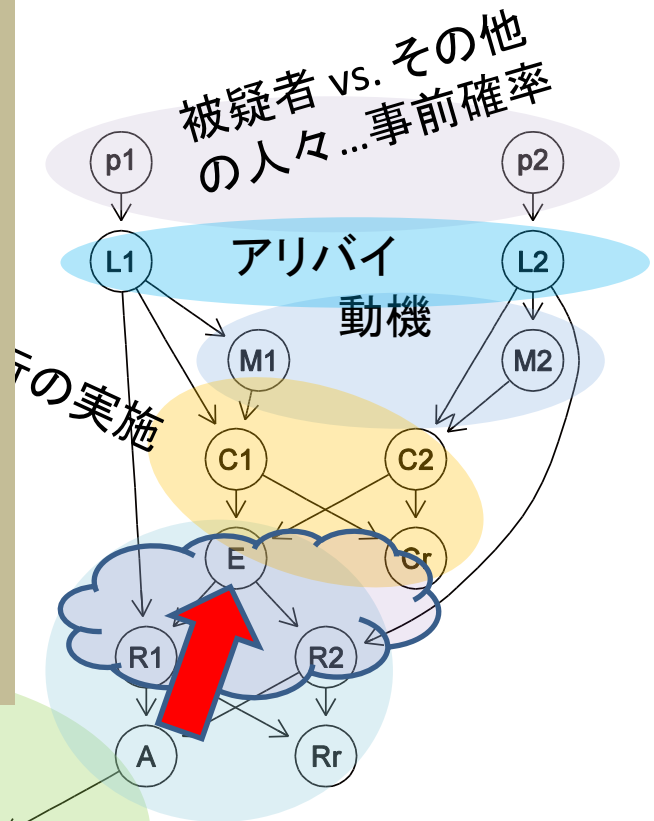
「遺留品由来のDNAの型」が確定した

DNA鑑定は遺留品の持ち主が誰か Aの事後確率に影響する

犯行の実行 E と遺留品の持ち主が誰か Aとの関係は

E -> R1, E->R2が強ければ強いがそうでなければ弱まる

遺留品が犯行によって残るものであれば強いが偶発的に残るものであれば相対的に弱い



遺留品の残し方

「被疑者由来のDNAの型」が確定した

「遺留品由来のDNAの型」が確定した

さて、実装

- フリーソフト『R』
- パッケージ “gRain”

- 今日の資料は

- http://www.genome.med.kyoto-u.ac.jp/wiki_tokyo/index.php/法数学勉強会2013年9月