ベイズ統計

古谷知之

講義概要

- 確率の諸問題
- 条件付き確率・同時確率・周辺確率
- ベイズの定理
- 逆確率
- 尤度
- 事前確率・事後確率

確率的な思考を身につける

- 世の中で起きていること（事象）をより（抽象的に）簡略化して考える
- 事象の発生数や場合の数を把握する
- 事象の因果関係を示す

確率の「定義」（ラプラスの定義）

事象Aの起こる確率$P(A)$
＝事象Aが起こる場合の数r /全ての場合の数N

$$P(A) = \frac{r}{N}$$

「全ての場合の数」なんて、定義できるのか？
発生回数が少なくても確率は安定するのか？
偶然的な現象にも適用できるのか？

確率

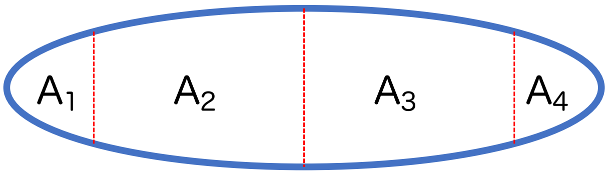
- イベント発生件数 / 生じうる全てのイベント件数

- $P(A)$ ：イベント（事象） Aの発生確率

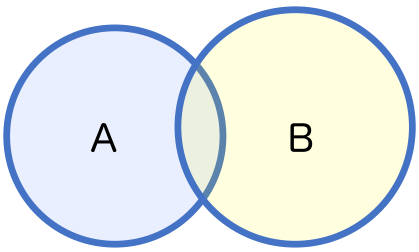
- 疾病による死亡確率
  - （疾病による死亡者数）/（全人口）
  - （疾病による死亡者数） /（疾病罹患者総数）

集合の重なり方

- 集合内の部分集合が互いに背反のとき

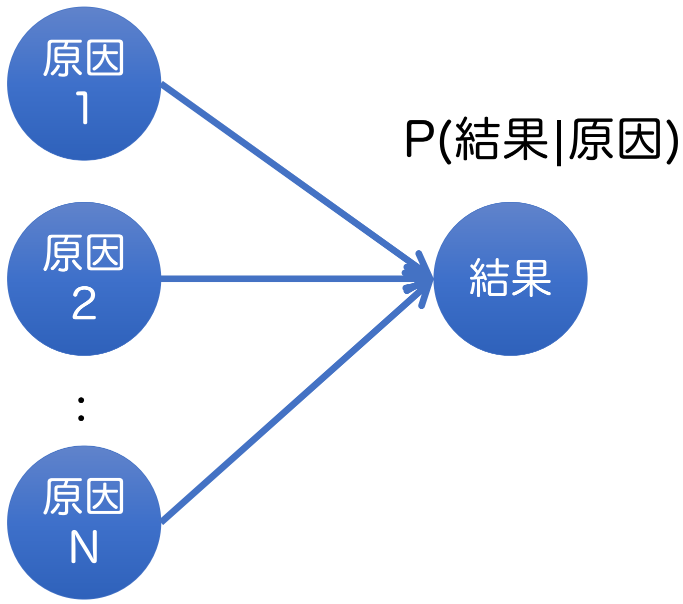


- 複数の集合が重なるとき



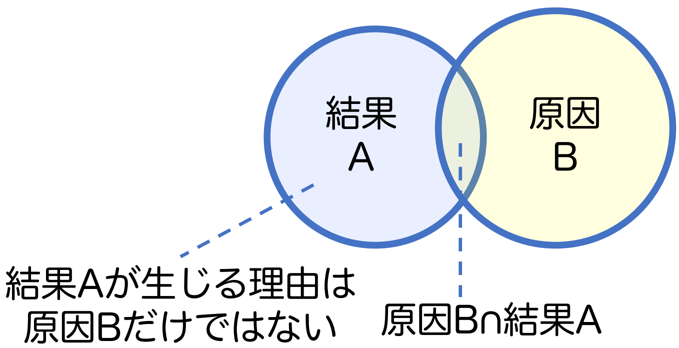
因果関係

- ものごとが発生するには必ず原因がある



因果関係とベン図

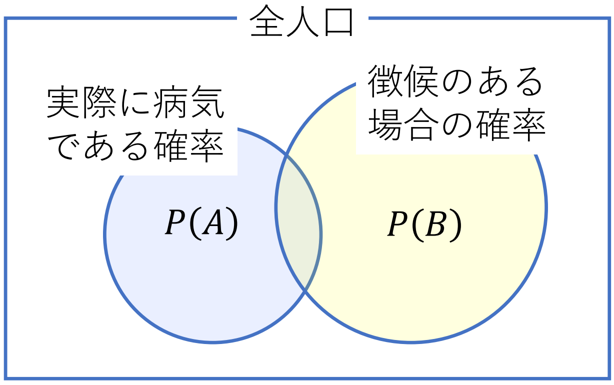
- 原因Bが理由となり結果Aが生じる場合
  =原因Bの割合に対する(原因B ∩結果A)が生じる割合



原因と結果の発生確率

- 病気の徴候がある場合に実際に病気となる確率

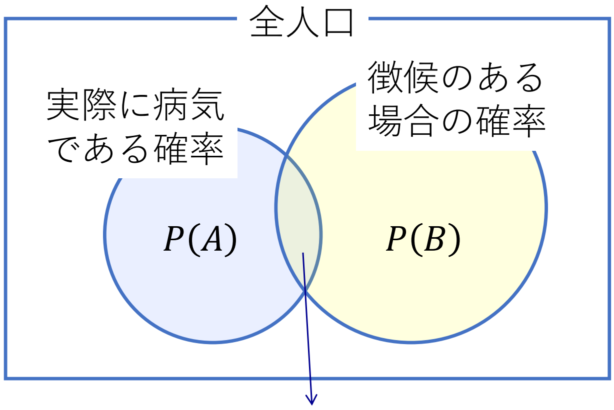
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(A \cap B)}{P(B)}$$



条件付き確率と同時確率

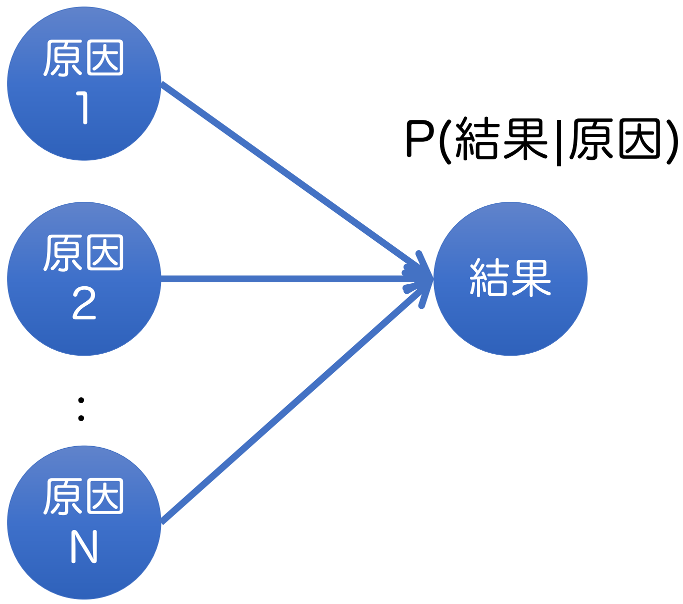
条件付き確率

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(A \cap B)}{P(B)}$$



同時確率 $P(A \cap B)$

因果関係



スクリーニングテストの例

- 兆候の有無：事象B
  - 兆候がある：B_+、兆候がない：B_-
- 健康状態：事象A
  - 健康である： $A_{健康}$、病気である： $A_{病気}$

健康状態	兆候（B）		
(A)	−	＋	合計
健康	800	100	900
病気	25	75	100
合計	825	175	1000

スクリーニングテストの例

- 無作為に一人選んだ人が、兆候が＋でかつ病気である確率（同時確率）　$P(A_{病気} \cap B_{+})$
- 無作為に選んだ人が病気である確率（周辺確率）$P(A_{病気})$
- 無作為に選んだ人が**兆候が+**のとき病気であるである確率（条件付き確率）　$P(A_{病気} | B_{+})$

健康状態	兆候（B）		
(A)	ー	＋	合計
健康	800	100	900
病気	25	75	100
合計	825	175	1000

S.セン（2005）、p.76

条件付き確率、同時確率、周辺確率

- 同時確率　$P(A_{病気} \cap B_{+}) = 75/1000$
- 周辺確率　$P(B_{+}) = 175/1000$
- 条件付き確率　$P(A_{病気} | B_{+}) = 75/175$
- 同時確率＝条件付き確率×周辺確率

健康状態	兆候（B）		
(A)	ー	＋	合計
健康	800	100	900
病気	25	75	100
合計	825	175	1000

S.セン（2005）、p.76

ベイズの定理の導き方

- より一般的に、二つの事象AとBがあるとする
- このとき、同時確率＝条件付き確率×周辺確率

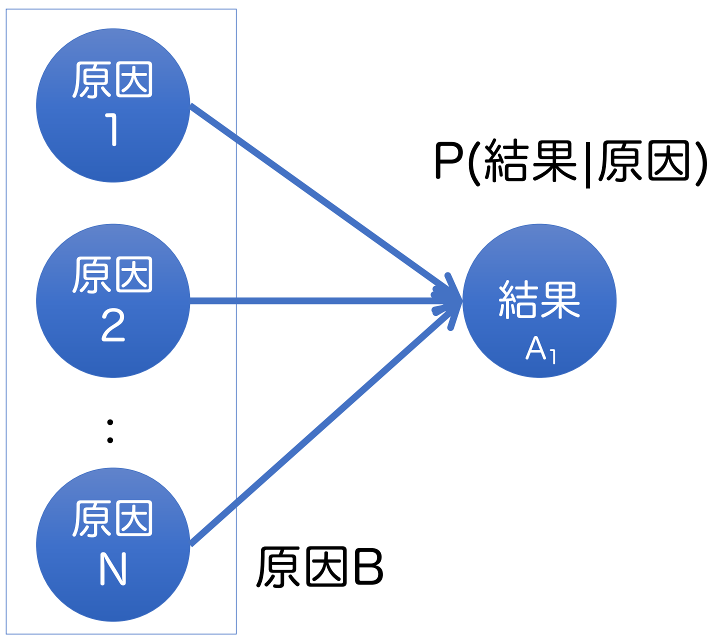
$$P(A \cap B) = P(A|B)P(B)$$

  が成立する。
- 同様に、 $P(B \cap A) = P(B|A)P(A)$が成立する

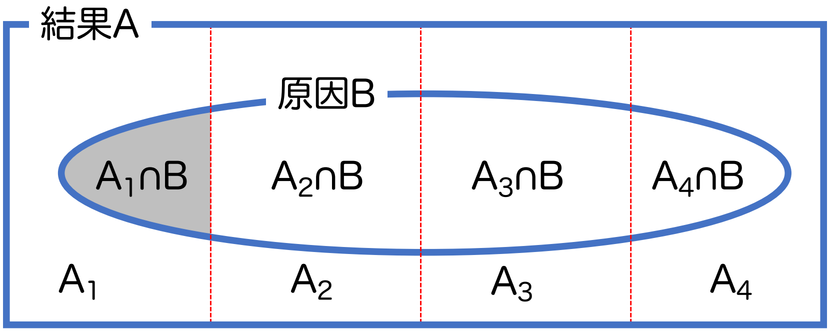
$$P(A \cap B) = P(B \cap A) \Leftrightarrow P(A|B)P(B) = P(B|A)P(A)$$
$$\Leftrightarrow P(A|B) = P(B|A)P(A)/P(B)$$

- これが「ベイズの定理」

因果関係と条件付き確率



原因と結果の関係



ベイズの定理

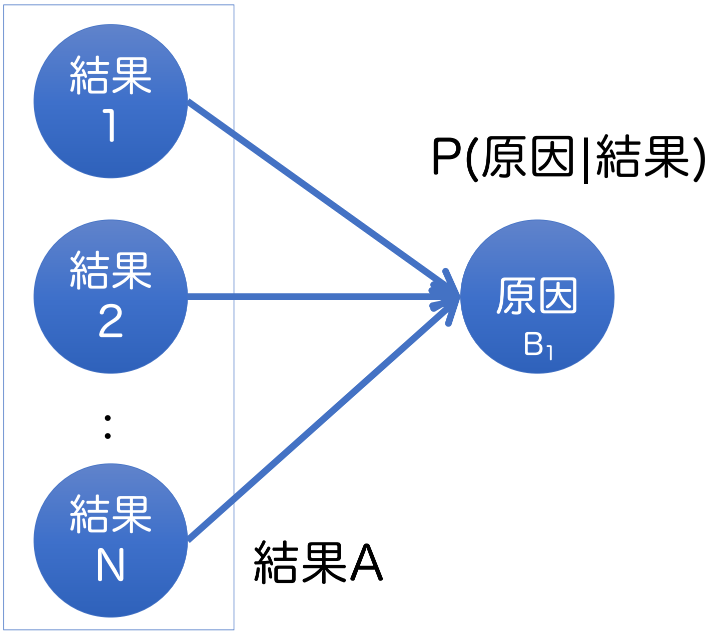
- 原因Bから結果A_1が生じうる確率

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{\sum_{i=1}^{n} P(B \cap A_i)}$$

$$= \frac{P(B|A_1)P(A_1)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

$$= \frac{P(B|A_1)P(A_1)}{P(B)}$$

逆確率による原因の推定



逆確率

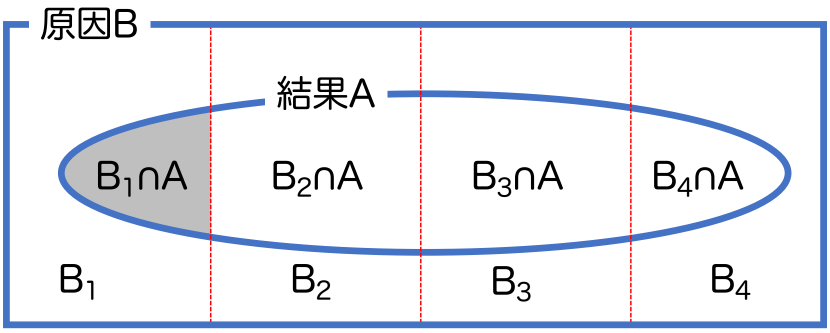
- 結果Aの原因BがB_1でありうる確率

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{\sum_{i=1}^{n} P(A \cap B_i)}$$

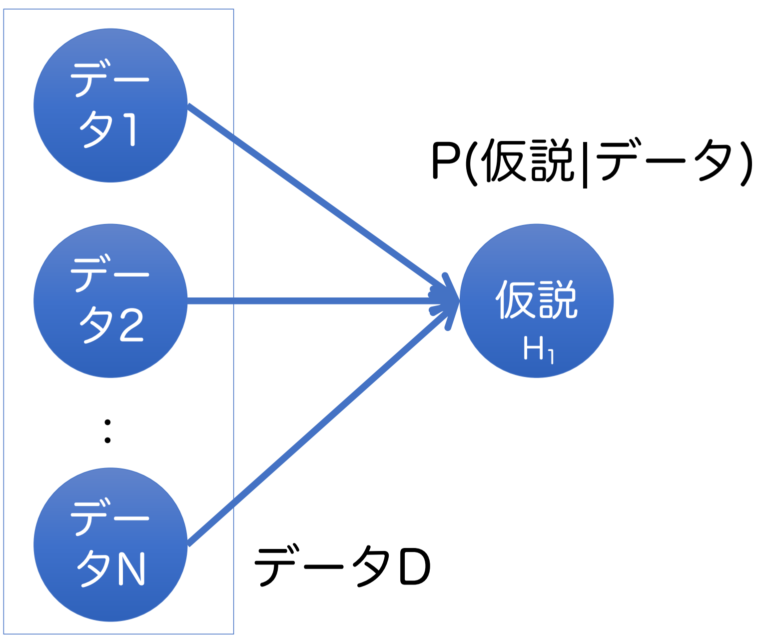
$$= \frac{P(A|B_1)P(B_1)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$$

$$= \frac{P(A|B_1)P(B_1)}{P(A)}$$

逆確率における原因と結果の関係



データから仮説（モデル）を推定する



データから仮説（モデル）を推定する

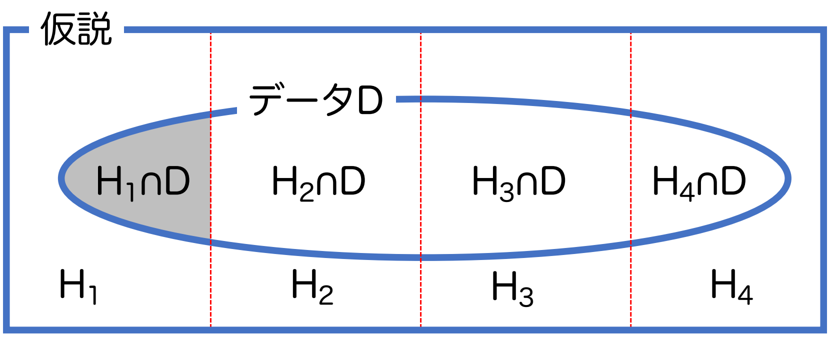
- データDから仮説H_1が成立する確率は以下のようになる

$$P(H_1|D) = \frac{P(D|H_1)P(H_1)}{\sum_{i=1}^{n} P(D \cap H_i)}$$

$$= \frac{P(D|H_1)P(H_1)}{\sum_{i=1}^{n} P(D|H_i)P(H_i)}$$

$$= \frac{P(D|H_1)P(H_1)}{P(D)}$$

データから仮説（モデル）を推定する



ベイズの定理とベイズ統計

- ベイズ統計でのモデル推定は、データが与えられた条件の下で仮定されたモデルの成立する確率P(仮説|データ)を求めていることに他ならない
- 確率的には与えられたデータのもとで様々なモデル（仮説）が成立する可能性がある

尤度（ゆうど）

- 原因から結果が生じる確率P(結果|原因)を尤度という
- 原因と結果を、データと仮説（モデル）と読み替えると、

ある仮説（モデル）を所与としてデータが得られる確率P(データ|仮説)といえる

- 仮説がデータに当てはまる当てはまりやすさ（尤もらしさ）を尤度という

事前確率・事後確率・尤度

- ベイズの定理にもとづいて、$P(H)$を事前確率、$P(D|H)$を尤度、$P(H|D)$を事後確率という

尤度　事前確率

事後確率

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

事前確率・事後確率・尤度

- データや仮説が複数（たくさん）あるとき、事前確率、事後確率、尤度はデータや仮説の値に対応した確率値をとる