OKAYAMA UNIVERSITY

ベイズの定理

岡山大学 異分野基礎科学研究所

大槻純也

RIIS

ベイズ統計の目的

- 観測事実からどれだけの結論が出せるか
  - いま出せる結論を確率として出す
  - わからないことはわからない
- 通常の統計学との違い（例えば、実験によりあるパラメータθを決めたいとする）
  - 通常の統計学は
    - 統計集団を特徴づける（平均値、標準偏差など）
    - 応用例：保険の価格設定
    - θの真値があって実験により誤差が生じた（誤差が確率的）
  - ベイズ統計は
    - 観測事実からθの確率分布を出す
    - 確率が確信度を表す（主観的）
    - 応用例：スパムメールフィルタ

例：

ある病気の罹患者に対して95%の確率で陽性となる検査があったとする。この検査で陽性と診断されたときに、実際に罹患している確率はどのくらいか？（この問題は後程扱います）

これを応用すると…

- 実験データをフィットする複数のモデルのどれが妥当か（モデル選択）
- 実験データからモデルハミルトニアンを決める（モデル選択）
- 物質合成の最適条件を効率よく決める（ベイズ最適化）
- どこから優先的に測定するか（実験計画）

まずはベイズの定理を理解する

準備：条件付き確率

確率変数 (random variable)

$$X = x_1, x_2, \cdots, x_i, \cdots, x_M$$

$$Y = y_1, y_2, \cdots, y_j, \cdots, y_L$$

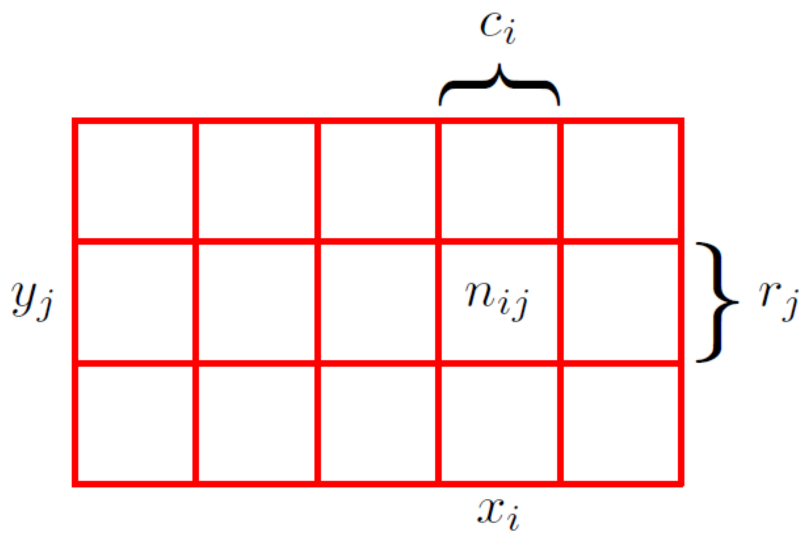
事象

例：サイコロ

偶数、奇数

素数、素数じゃない

XとYを決める試行 (trial) をN回行い、$X = x_i$, $Y = y_j$がn_{ij}回得られた



PRML Fig.1.10

同時確率 (joint probability)

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

周辺確率 (marginal probability)

$$p(X = x_i) = \frac{c_i}{N}, \quad p(Y = y_j) = \frac{r_j}{N}$$

Yは何でもよい　　Xは何でもよい

条件付き確率 (conditional probability)

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

$X = x_i$ が得られた下で、 $Y = y_j$が得られる確率

「probability of $Y = y_j$ given $X = x_i$」と読む

準備：条件付き確率に関する重要な関係式

OKAYAMA UNIVERSITY

総和則 (sum rule)

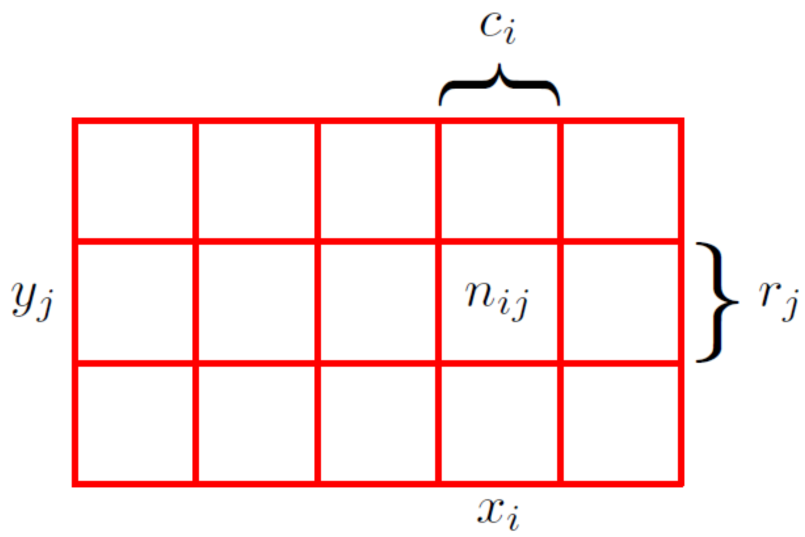
$$p(X) = \sum_{Y} p(X, Y)$$

周辺化

積則 (product rule)

$$p(X, Y) = p(Y|X)p(X)$$

XとYを決める試行 (trial) をN回行い、$X = x_i$, $Y = y_j$がn_{ij}回得られた



PRML Fig.1.10

ベイズの定理

Product rule $p(X,Y) = p(Y|X)p(X) = p(X|Y)p(Y)$ より

$$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)}$$

ベイズの定理

p(Y)は規格化定数なので省略すると

$$p(X|Y) \propto p(Y|X)p(X)$$

（ 事後確率 尤度関数 事前確率 ）

ただし、これらの呼び名は
確率変数X, Yに意味を持たせて
初めて意味を成す
（例題を解きながら説明していきます）

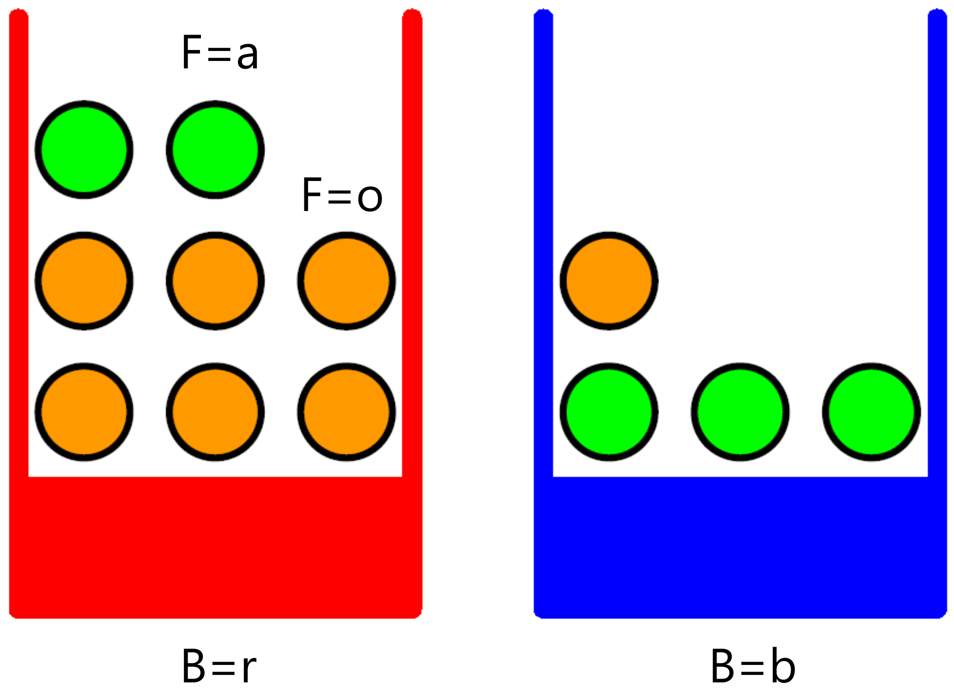
例：箱に入った玉

状況設定

　箱が２つあり、赤い箱にはオレンジが6つ、りんごが2つ入っている。青い箱にはオレンジが1つ、りんごが3つ入ってるとする。赤い箱と青い箱をそれぞれ40%, 60%の確率で選び、その箱から果物をひとつ選ぶ。

Q1.
オレンジが選ばれる確率は？

Q2.
オレンジが選ばれたときに、
青い箱を選んでいた確率は？

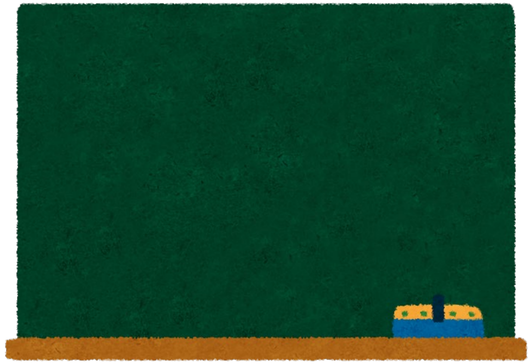


PRML Fig.1.9

解答

Q1.
オレンジが選ばれる確率は？

Q2.
オレンジが選ばれたときに、
青い箱を選んでいた確率は？



「板書で説明」マーク
要所でたびたび登場します

ベイズの定理の見方

選んだ箱がB=bである確率

観測によって
確率が変わった

Fを観測する前

$$p(B=b)=0.6$$

事前確率
prior probability

Fを観測

F=oを観測した後

$$p(B=b|F=o)=0.333$$

事後確率
posterior probability

ベイズの定理

$$\underline{p(B|F)} \propto p(F|B)\underline{p(B)}$$

B: 知りたいこと
F: 観測結果

ベイズ統計は答えを確率として出す。
確率は答えの確信度を表す

ここまでのまとめ：ベイズの定理

Bayesian theorem

$$\underline{p(X|Y)} = \frac{p(Y|X)\underline{p(X)}}{p(Y)}$$

規格化定数

sum rule + product rule

$$p(Y) = \sum_X p(Y|X)p(X)$$

事後確率
posterior probability

事前確率
prior probability

X: 知りたいこと（青い箱から選んだか）

Y: 観測結果（選ばれたのはオレンジorりんご）

例題

OKAYAMA UNIVERSITY

ベイズ統計では直感に反する結果が得られる場合があることを見る。

日本人の罹患率が0.1%の病気（例えばガン）を考えます。ある検査で陽性が出る確率が、罹患者に対しては95%、罹患していない人に対しては2%（誤診）とします。この検査で陽性と診断されたとき、実際に罹患している確率はどのくらいでしょうか？

ヒント：95%ではない

「完全独習・ベイズ統計学入門」小島寛之著　ダイヤモンド社より

計算手順

1. 確率変数を定義

確率変数X
- $X = a$: 罹患している
- $X = \bar{a}$: 罹患していない

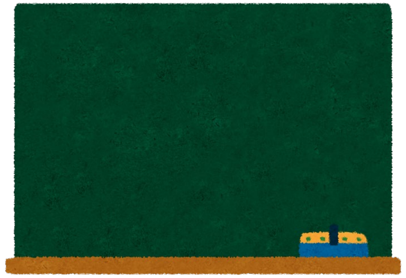
確率変数Y
- $Y = b$: 検査で陽性
- $Y = \bar{b}$: 検査で陰性

2. 既知の情報をまとめる

$$p(Y|X) \qquad p(X)$$

3. ベイズの定理を使って逆確率を計算

$$p(X|Y)$$



罹患している確率4.5%をどう見るか

OKAYAMA UNIVERSITY

- 「95%だと思ってたら実は4.5%！安心！」

$$p(Y = b | X = a) \qquad p(X = a | Y = b)$$

見るべきものが違っただけ
ベイズ統計を知らない人はこう考える

- ベイズ的な見方

検査前 $p(X = a) = 0.1\%$ 事前確率 (prior probability)

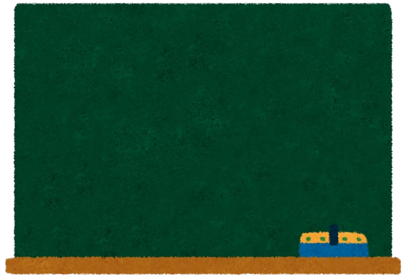
↓ 検査 $Y = a$

検査後 $p(X = a | Y = a) = 4.5\%$ 事後確率 (posterior probability)

検査(観測)で確率が更新された

なぜ直感とは異なる結果が得られたか

ベイズの定理を図で理解する



まとめ：ベイズの定理

ベイズの定理

$$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)}$$

X : 知りたいもの
→ モデルパラメータ $\boldsymbol{w}$ と表す

Y : そのために測定するもの
→ 観測データ $\mathcal{D}$ と表す

XとYに意味を持たせる

$$p(\boldsymbol{w}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{w})p(\boldsymbol{w})}{p(\mathcal{D})}$$

知りたいもの　観測事実

事前確率
prior probability
$p(\boldsymbol{w})$

事後確率
posterior probability
$p(\boldsymbol{w}|\mathcal{D})$

測定により確率が更新される

$p(\mathcal{D}|\boldsymbol{w})$ の意味については次回