横浜国立大学集中講義「現代物理学」
第1回　2021年11月24日

OKAYAMA UNIVERSITY

ベイズ統計への導入：条件付き確率

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

OKAYAMA UNIVERSITY

- 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？（この問題は次回扱います）

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
- どこから優先的に測定するか（実験計画）

今回から3回でベイズの定理を理解する

準備：条件付き確率

OKAYAMA UNIVERSITY

確率変数 (random variable)

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

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

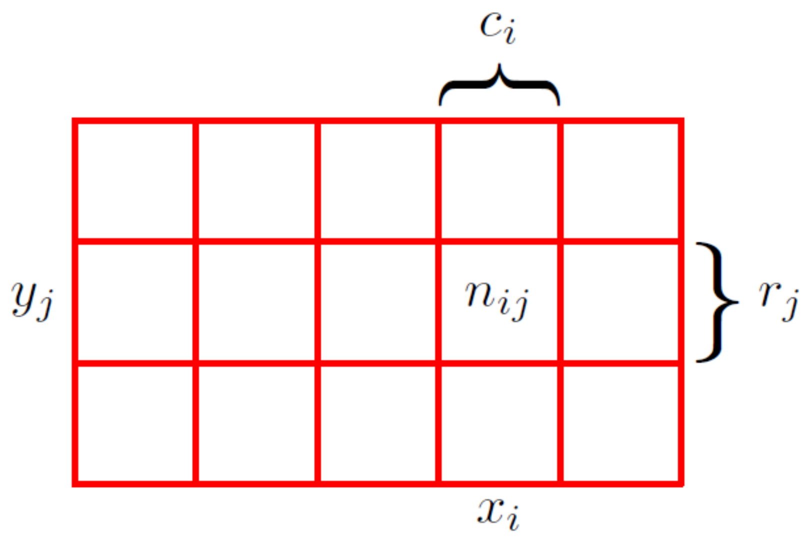
事象

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：条件付確率

重要な関係式

総和則 (sum rule)

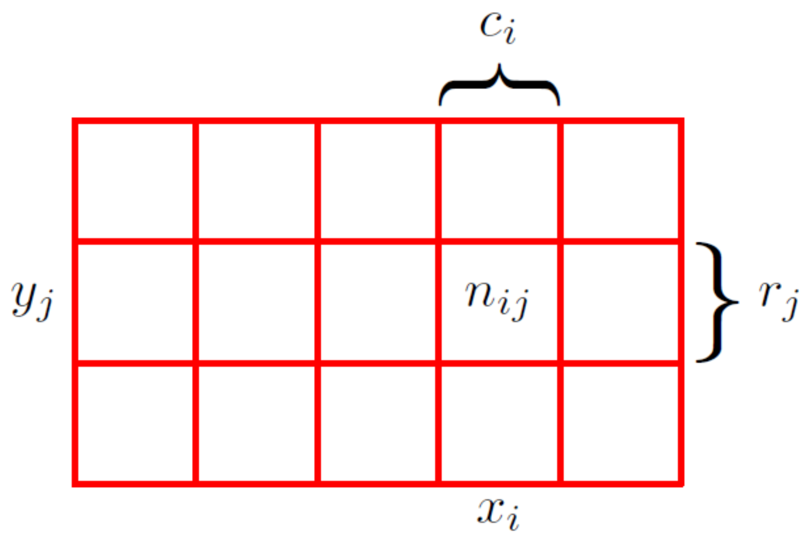
$$p(X) = \sum_{Y} p(X, Y)$$

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）

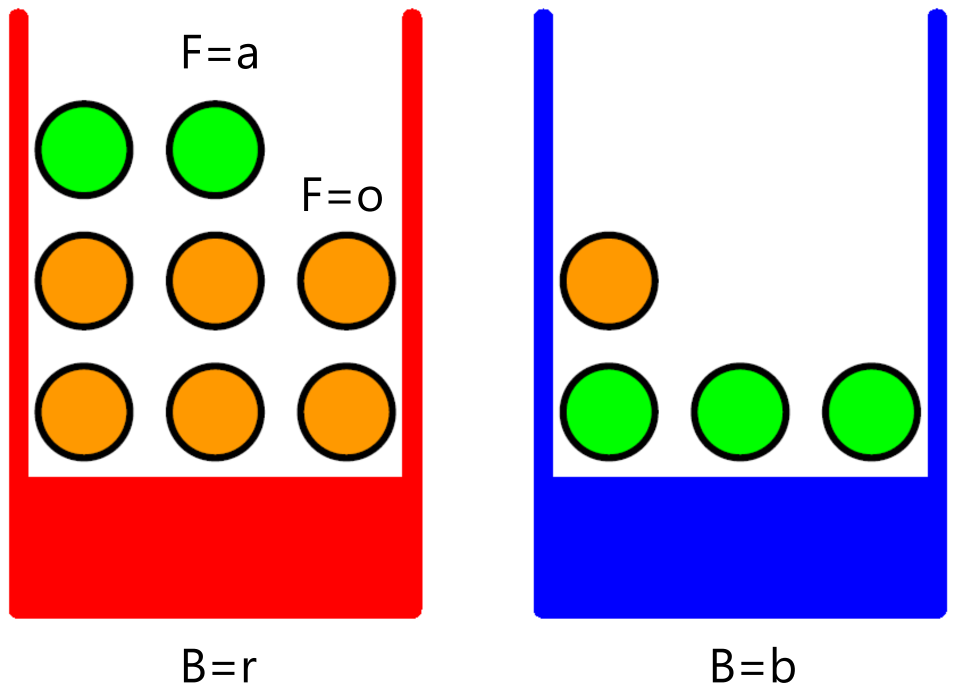
例：箱に入った玉

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？

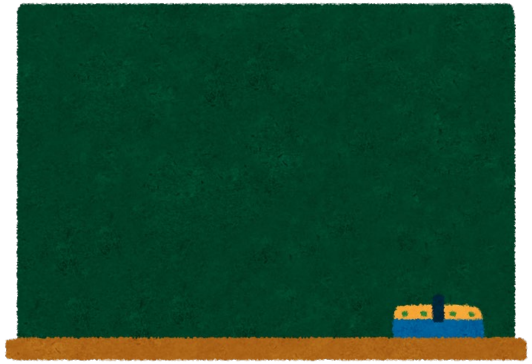


PRML Fig.1.9

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

Fを観測

F=oを観測した後

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

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

事前確率
prior probability

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

次回の予告

ベイズ推定では直感と異なる結果が得られることがある

例題：
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？（95%ではない）