パターン認識と機械学習

１．２　確率論

4月**18**日（金）
05**T**4051**R**
豊川　幸秀

役割

- パターン認識の基礎の中心
- 不確実性に関する定量化及び操作に対し一貫した枠組みを与える

確率の基本法則

- 確率の加法定理:

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

- 確率の乗法定理:

$$p(X, Y) = p(Y \mid X)\, p(X)$$

準備(1)

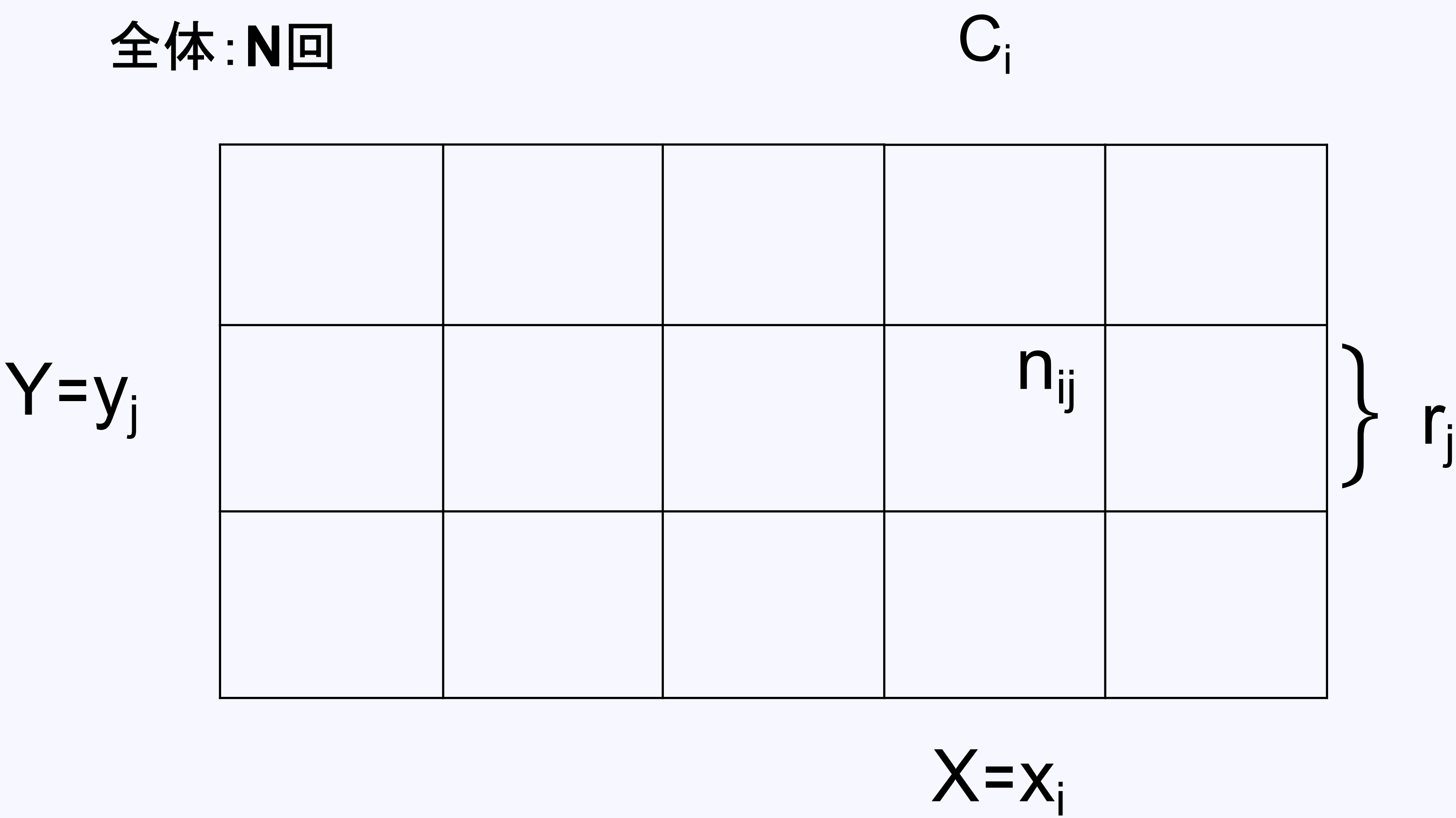
- X,Y:確率変数
- 試行はN回
- X,Yはそれぞれ任意の値x_i,y_jをとり、同時にこの2つになる時の確率を$p(X=x_i,Y=y_j)$、その試行の回数をn_{ij}とする。
- この時、この確率を$X=x_i,Y=y_j$の**同時確率**と呼ぶ。

準備（2）

- X＝x_iとなる確率をp(X=x_i)、その回数をc_i、Y=yjとなる確率をp(Y=y_j)、その回数をr_jとする。
- このときのp(X=x_i)、p(Y=y_j)は周辺確率と呼ばれる。
- p(Y=y_j|X=x_i)は、X=x_iという条件下におけるY=y_jが発生する確率で、条件付き確率と呼ばれる。

それぞれの値と回数の対応

全体：**N**回



確率の加法定理

- 図より、X＝x_i、Y=y_jの同時確率は

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

- また、Xがx_iを取る確率は

  $p(X=x_i) = c_i/N$

であり、同時に $c_i = \sum_j n_{ij}$ であるので、

$$p(X = x_i) = \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

確率の乗法定理

- $X=x_i$という条件下での$Y=y_j$の条件付き確率は、

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

よって、

$$p(X=x_i,Y=y_j) = n_{ij}/N$$
$$= n_{ij}/c_i \times c_i/N$$
$$=p(Y=y_j|X=x_i)p(X=x_i)$$

ベイズの定理

- $p(X=x_i, Y=y_j) \rightarrow p(X,Y)$、$p(Y=y_j|X=x_i) \rightarrow p(Y|X)$、$p(X=x_i) \rightarrow p(X)$
- 乗法定理及び$p(X,Y)=p(Y,X)$より、

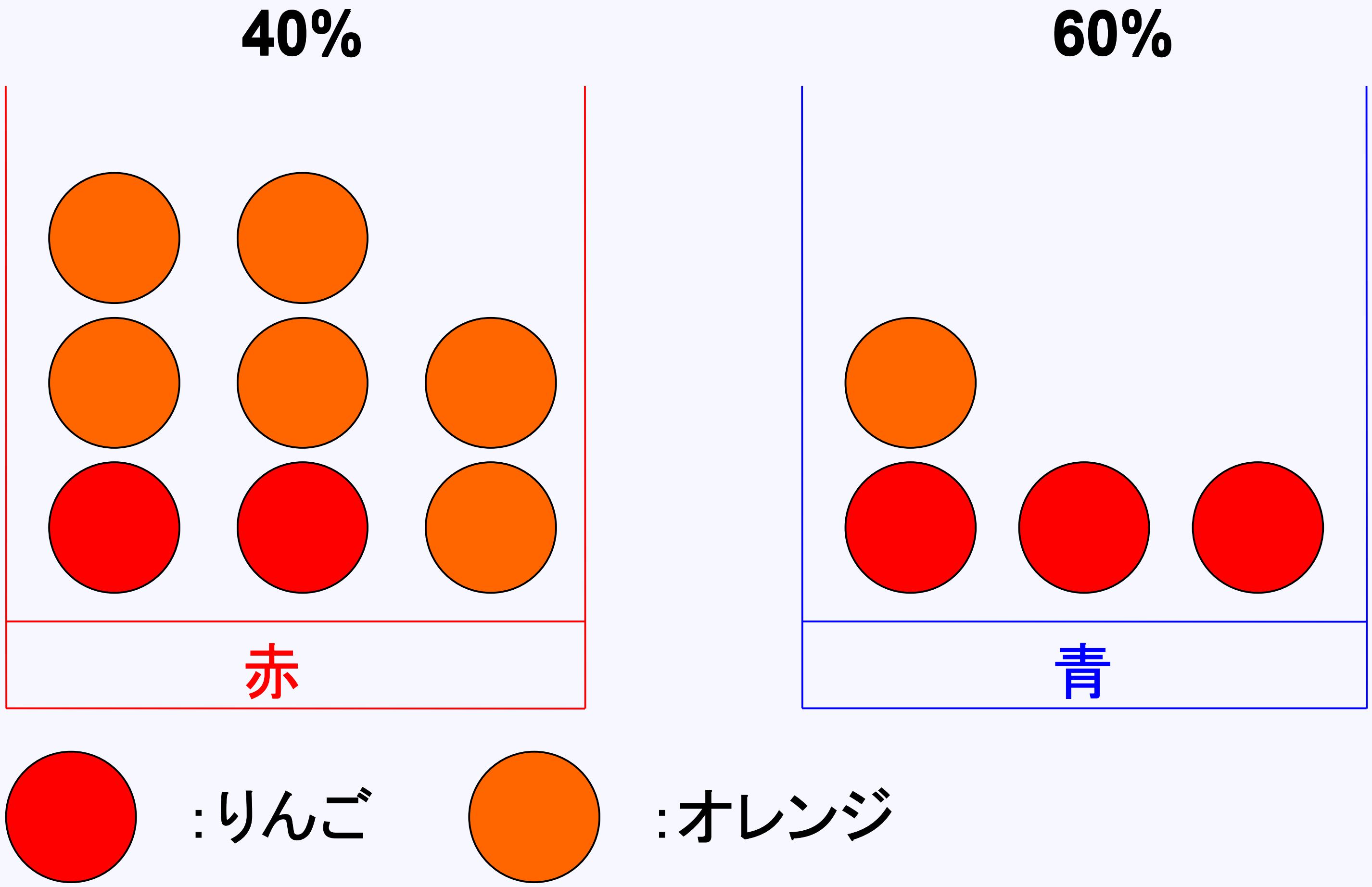
$$p(Y|X) = \frac{p(X|Y)\,p(Y)}{p(X)}$$

が導き出され、これを**ベイズの定理**と呼ぶ。

簡単な例による考察

- 赤と青の2つの箱があり、赤の箱にはりんご1個にオレンジ6個、青の箱にはりんご3個にオレンジ1個が入っている。
- 箱を1つランダムに選び、その後果物をランダムに1個取り出し記録するという作業を繰り返す。
- 赤の箱は40%、青の箱は60%の確率で選ばれる。

簡単な図



各確率変数

- どの箱を選ぶか、どの果物かという事象は確率変数となり、それぞれをBとFとする。
- Bはr(赤い箱)かb(青い箱)かとし、

  $p(B=r) = p(r) = 4/10$、$p(B=b) = p(b) = 6/10$

  となる。
- Fはa(りんご)かo(オレンジ)かとする。

果物の条件付き確率

- 赤い箱という条件下

  $p(F=a|B=r) = 1/4$:りんごが取られる確率

  $p(F=o|B=r) = 3/4$:オレンジが取られる確率

- 青い箱という条件下

  $p(F=a|B=b) = 3/4$:りんごが取られる確率

  $p(F=o|B=b) = 1/4$:オレンジが取られる確率

りんごの確率

- この条件付き確率と、確率の加法・乗法定理を利用することにより、りんごを選ぶ確率を計算することができる。

$$
\begin{aligned}
p(F=a) &= p(F=a|B=r)p(B=r) + \\
&\qquad p(F=a|B=b)p(B=b) \\
&= 1/4 \times 4/10 + 3/4 \times 6/10 \\
&= 11/20
\end{aligned}
$$

オレンジの確率

- 加法定理より、
  $p(F=o) = 1 - 11/20 = 9/20$

- このように、加法・乗法定理を利用することにより、多くの質問に答えることができる。

箱の条件付き確率

- 例:オレンジを選んだときに、どの箱から取り出されたか？

導出方法

- ベイズの定理より、

$$p(B=r\mid F=o)=\frac{p(F=o\mid B=r)\,p(B=r)}{p(F=o)}$$

= 3/4 × 4/10 × 20/9

= 2/3

となる。また、加法定理より、

p(B=b|F=o) = 1 - 2/3 = 1/3

事前･事後確率（1）

- 先の例におけるp(B)は、事前確率と呼ばれる。
  →どんな果物が取られるか分かる以前に与えられている、箱を選ぶ確率
- 一方、p(B|F)は事後確率と呼ばれる。
  →果物が取られた後に計算される、箱の確率

事前･事後確率（2）

- 果物が選ばれる前は赤い箱の確率4/10
- 選ばれた果物がオレンジの場合、赤い箱の確率2/3

↓

独立

- p(X,Y) = p(X)p(Y)となるとき、XとYは独立であるという。

乗法定理より、p(Y|X) = p(Y)

↓

Xがなんであろうと Yには影響しない