

2.4.2 ベイズの定理

条件つき確率を用いると、たとえば文字 j の後に文字 a が来る確率 $p(Y = a | X = j)$ は、式(2.24)のように計算することができます。それでは逆に、 j の前に来ることができる文字には、どんなものがあるでしょうか^{*14}。

確率で書けば、これは、前に来る文字を x として

$$(2.27) \quad p(X = x | Y = j) \quad (x \text{ は前に来る任意の 1 文字})$$

を求めたい、ということです。条件つき確率の定義式(2.19)から、上の確率は

$$(2.28) \quad p(X = x | Y = j) = \frac{p(X = x, Y = j)}{p(Y = j)}$$

と表すことができます。ここで分子に式(2.21)の確率の連鎖則を用いれば、

$$(2.29) \quad p(X = x | Y = j) = \frac{p(Y = j | X = x)p(X = x)}{p(Y = j)}$$

と分解することができます。すべての文字 x について $p(Y = j | X = x)$ は式(2.24)のように計算することができ、文字のユニグラム確率 $p(X = x)$ 、 $p(Y = j)$ は2.1節で計算したように文字の頻度から求められますから、式(2.29)は簡単に計算することができます。

たとえば、式(2.24)より $p(Y = j | X = a) = 0.0014$ ですが、式(2.1)より $p(a) = 0.0663$ 、 $p(j) = 0.0011$ なので、^{*15}

$$(2.30) \quad p(X = a | Y = j) = \frac{p(Y = j | X = a)p(X = a)}{p(Y = j)} = \frac{0.0014 \times 0.0663}{0.0011} = 0.0844$$

と計算されます。同様にほかの文字についても計算してみると、

```
r = {}
for y in chars:      # c[x][y] = p(y|x)
```

^{*14} 筆者は大学時代にエスペラント研究会に所属していましたが、世界共通語を目指して作られた人工語であるエスペラント語では j は複数を表す接尾辞で、*libroj* (本たち) や *ĉiuj* (すべて) のように頻繁に使われています。しかし英語では、 j が単語の末尾に用いられることは稀です。それでは、 j はどんな場合に使われているのでしょうか？

^{*15} Y の方は範囲に $\$$ も含みますので、厳密にはこれと少しだけ異なります。

```

r[y] = {}          # r[y][x] = p(x|y)
for x in chars:
    r[y][x] = c[x][y] * p0[x] / p0[y]
r['j']
⇒ {'a': 0.0821917808219178,
    'l': 0.0,
    'i': 0.0,
    'c': 0.0,
    'e': 0.0,
    's': 0.0,
    ' ': 0.7808219178082191,
    :
    'h': 0.0136986301369863,
    'p': 0.0,
    'b': 0.04794520547945205,
    :
    'z': 0.0}

```

となり, `alice.txt` では空白, すなわち単語の先頭を除くと, “aj” および “bj” の確率が高いことがわかります. (ただし, これらは “majesty” と “subject” によるものですので, 一般にはもう少し大きなテキストで調べる必要があります.)

式(2.30)をよく見ると, これは条件つき確率 $p(X=a|Y=j)$ を, 逆方向の条件つき確率 $p(Y=j|X=a)$ を使って計算することができる, ということを意味しています. こうして**条件つき確率を引っくり返す**ことができる式(2.30), あるいはより一般に

$$(2.31) \quad p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)} \quad (\text{ベイズの定理})$$

を, **ベイズの定理** (Bayes' Theorem) といいます. ^{*16} ベイズの定理を使えば, 条件つき確率 $p(X|Y)$ を逆の条件つき確率 $p(Y|X)$ で表すことができます. たとえば X が「原因」, Y が「結果」のとき, 結果 Y がわかったときの原因 X の確率 $p(X|Y)$ は, 原因から結果が得られる確率 $p(Y|X)$ を使って計算できる, と

^{*16} この公式の特別な場合はイギリスの牧師 Thomas Bayes (c.1701–1761) によって発見され, 死後に友人によって発表されたため, ベイズの定理と呼ばれています. ただし, 実際の発展や一般化は, その後にフランスの数学者ラプラス (1749–1827) によって行われたものです[25]. とはいえ, ベイズの定理を積極的に利用するベイジアンにとって, ロンドン市内にあるベイズの墓石は今でも聖地となっています.