

# 第3回DS概説演習 補足例題

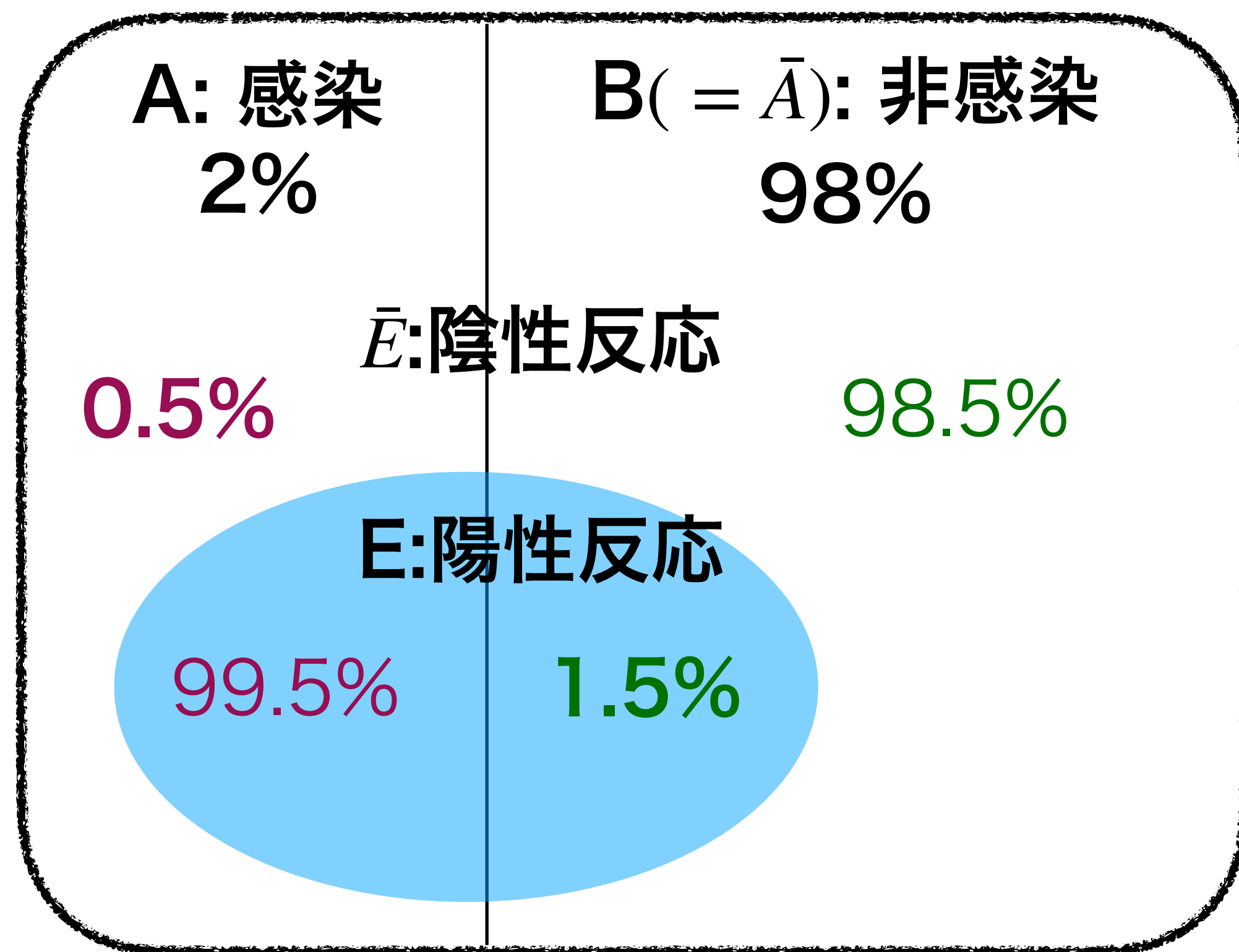
2021/10/18

小山居室：理学部A棟202, メール：[skoyama@create.niigata-u.ac.jp](mailto:skoyama@create.niigata-u.ac.jp)

# テキスト問題2-4 補足

疾病検査の信頼性：ある病気に感染しているかどうかを調べる検査法がある。この検査によると、被験者が非感染者なのに陽性になる確率が1.5%，感染者なのに陰性になる確率が0.5%ある。この病気は人口の2%が感染しているものとしよう。ある人が陽性と診断された。この人が本当に感染者である確率はどれだけか？

S：全人口100%



Aを感染者である事象、Bを非感染者である事象とすると、AとBは互いに排反で、かつ標本空間を尽くしている。検査で陽性反応の事象をEとする。

非感染者(=B)なのに陽性になる確率  $P(E|B) = 0.015$

感染者(=A)なのに陰性になる確率  $P(\bar{E}|A) = 0.005$

感染者で陽性になる確率  $1 - P(\bar{E}|A) = P(E|A) = 0.995$

ベイズの定理から、陽性反応(E)の上で感染者である(A)確率は、

$$\begin{aligned} P(A|E) &= \frac{P(A)P(E|A)}{P(A)P(E|A) + P(B)P(E|B)} \\ &= \frac{0.02 \times 0.995}{0.02 \times 0.995 + 0.98 \times 0.015} = 0.575 \end{aligned}$$

# ベイズの定理応用例題：迷惑メールの分類

迷惑メールに含まれている用語として、「無料」「当選」がある。これらは通常メールと迷惑メールの中に、それぞれ表の確率で含まれていることがわかっている。ある新着メールを調べたら、「無料」,「当選」という順番で各用語が見つかった。このとき、このメールを通常メール、迷惑メールのどちらに分類した方が良いか判定せよ。ただし、一般的には新着メールの中で、通常メールと迷惑メールの比率は6:4の割合とする。

| 用語 | 通常メール | 迷惑メール |
|----|-------|-------|
| 無料 | 0.2   | 0.3   |
| 当選 | 0.1   | 0.6   |

高校生のための統計学習教材 第1講より引用(p.33)  
<https://www.stat.go.jp/teacher/comp-learn-03.html>

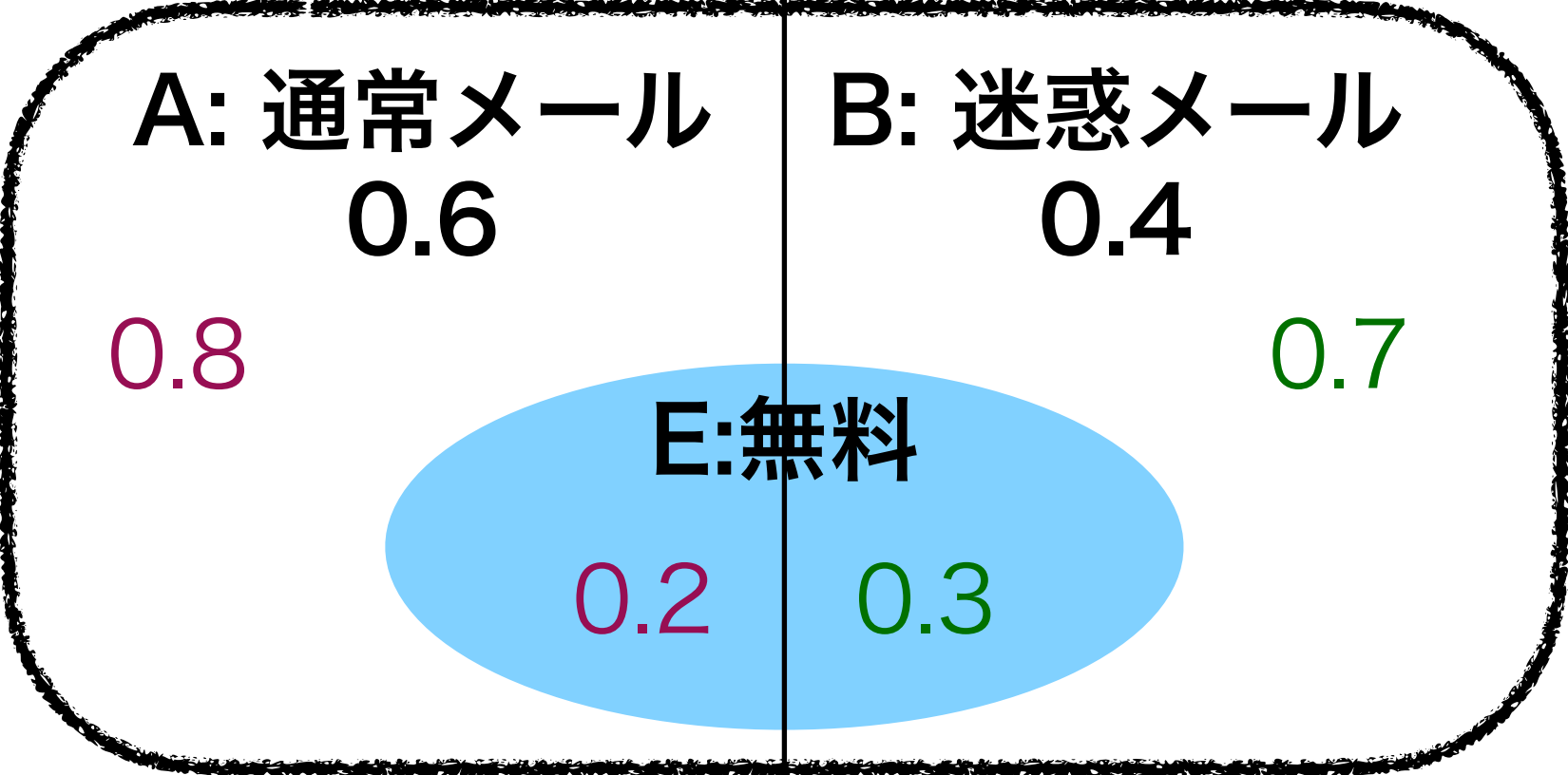


# ベイズの定理応用例題：迷惑メールの分類

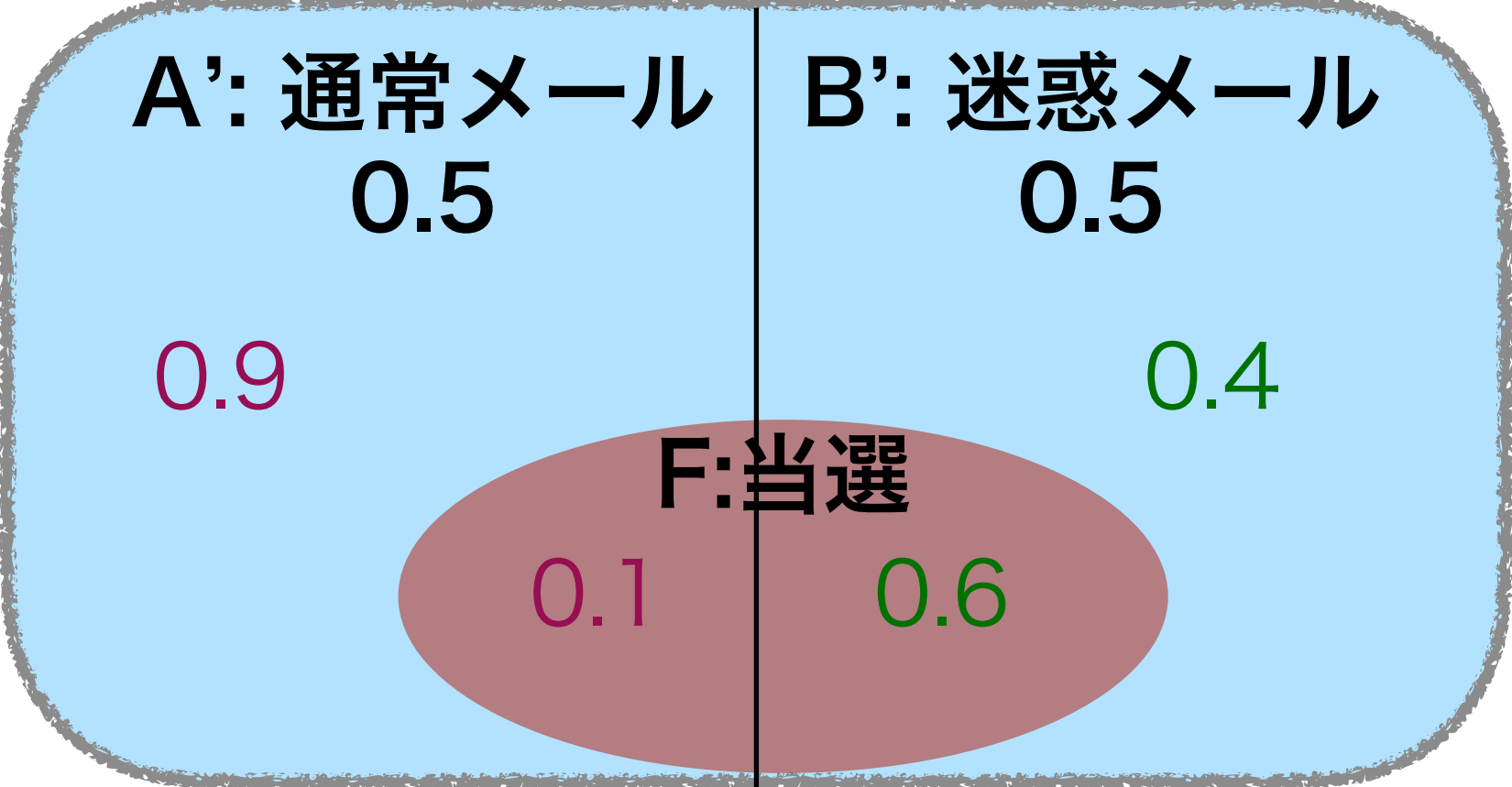
迷惑メールに含まれている用語として、「無料」「当選」がある。これらは通常メールと迷惑メールの中に、それぞれ表の確率で含まれていることがわかっている。ある新着メールを調べたら、「無料」，「当選」という順番で各用語が見つかった。このとき、このメールを通常メール、迷惑メールのどちらに分類した方が良いか判定せよ。ただし、一般的には新着メールの中で、通常メールと迷惑メールの比率は6:4の割合とする。

| 用語 | 通常メール | 迷惑メール |
|----|-------|-------|
| 無料 | 0.2   | 0.3   |
| 当選 | 0.1   | 0.6   |

S：新着メール



E：「無料」が含まれるメール



まず、「無料」が見つかる事象Eを考える。

「無料」が見つかったとき、それが通常メール(事象A)である確率

$$P(A|E) = \frac{P(A)P(E|A)}{P(A)P(E|A) + P(B)P(E|B)}$$
$$= \frac{0.6 \times 0.2}{0.6 \times 0.2 + 0.4 \times 0.3} = 0.5$$

「無料」が見つかったとき、  
迷惑メール(事象B)である確率

$$P(B|E) = 1 - P(A|E) = 0.5$$

次に、「無料」が見つかった中から、「当選」が見つかる事象Fを考える。

「当選」が見つかったとき、それが通常メール(事象A')である確率

$$P(A'|F) = \frac{P(A')P(F|A')}{P(A')P(F|A') + P(B')P(F|B')}$$
$$= \frac{0.5 \times 0.1}{0.5 \times 0.1 + 0.5 \times 0.6} = 0.14$$

ここでは、以下に注意：

$$P(A') = P(A|E) = 0.5$$

$$P(B') = P(B|E) = 0.5$$

迷惑メール(事象B')である確率

$$P(B'|F) = 1 - P(A'|F) = 0.86$$

→迷惑メールに分類した方が良い