Lecture - Day 2 (part 1)
Probability

MF9130E V24
2024.04.09

Chi Zhang
Oslo Center for Biostatistics and Epidemiology
chi.zhang@medisin.uio.no

Outline

Lecture 1: **introduction to probability, Bayes law**

Lecture 2: **Diagnostic testing, sensitivity, specificity**

Reading, demonstration, exercise

This afternoon we do NOT have data analysis practice, but you can use R as a calculator.

Probability

Probability is a fundamental tool for reasoning

… which can also be confusing

Aim for this lecture

- Develop an intuitive understanding via **visualization**: as little math as possible

- Introduce some concepts that you might encounter in the future

Probability

Probability expresses a potential for something to happen. Assessment of uncertainty,

Corresponds to ‘risk’ in medicine

Degree of belief that some event will occur: probability of rain tomorrow is 80%

Proportion of some outcomes happen in a large number of repeated events: proportion of girls among many new borns is roughly 50%

Law of large numbers

The average of some random outcomes from **a large number** of identical experiments converges to the true value

Example: coin toss. 1 as Heads, 0 as Tails. Assume it's fair, **probability of H is 0.5**
We care about the **probability of Heads**: p = #1 / #tosses (n)

Throw 1 time, (1). P = 1/1 = 1

Throw 3 times, (1,0,1). P = 2/3

Throw 10 times, (1,0,1,0,0,0,1,1,0,0). P = 4/10

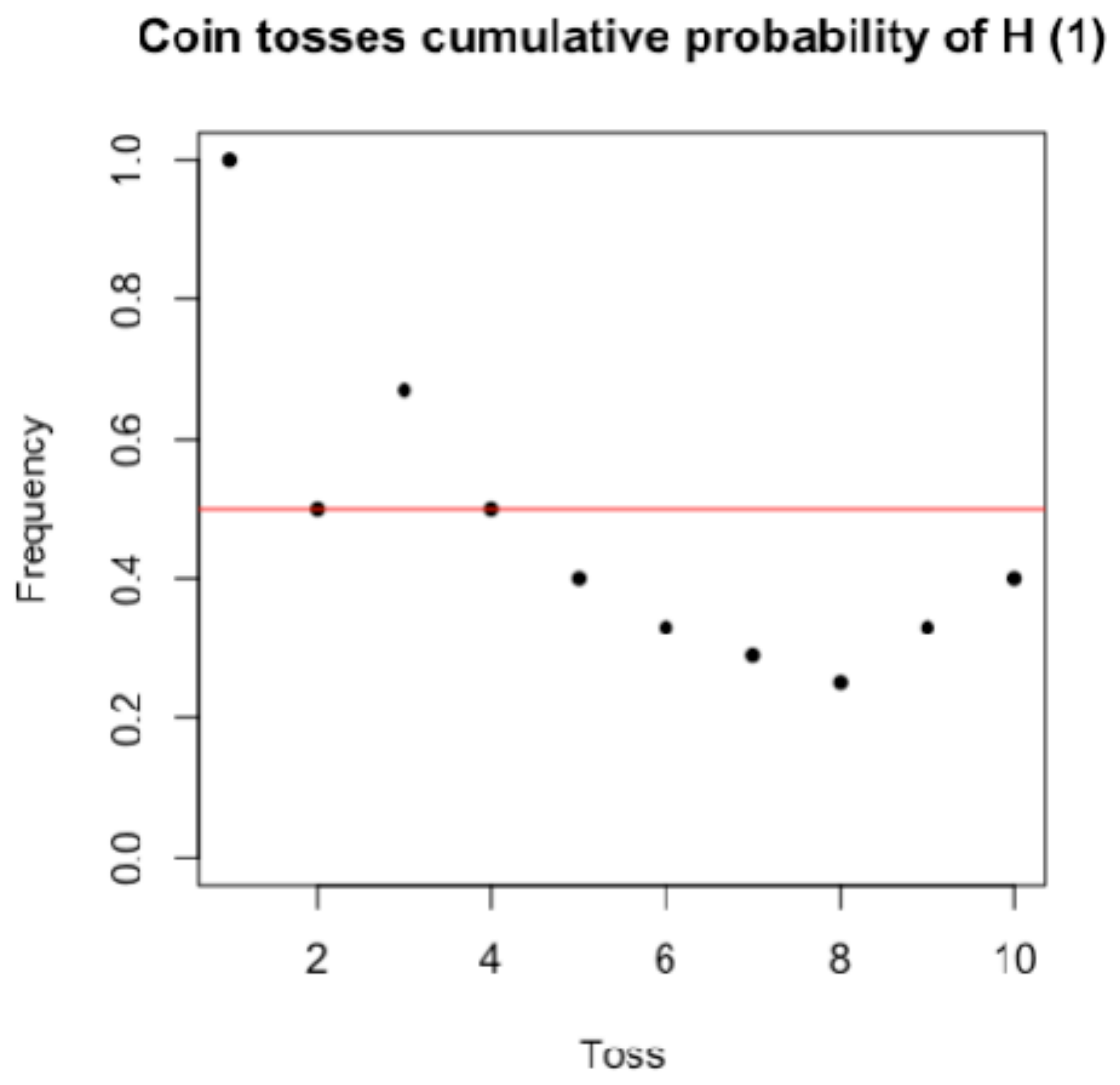
… Throw 100 times ?

It should be close to 0.5, if the coin is fair - equal probability of having 1 and 0.

Law of large numbers

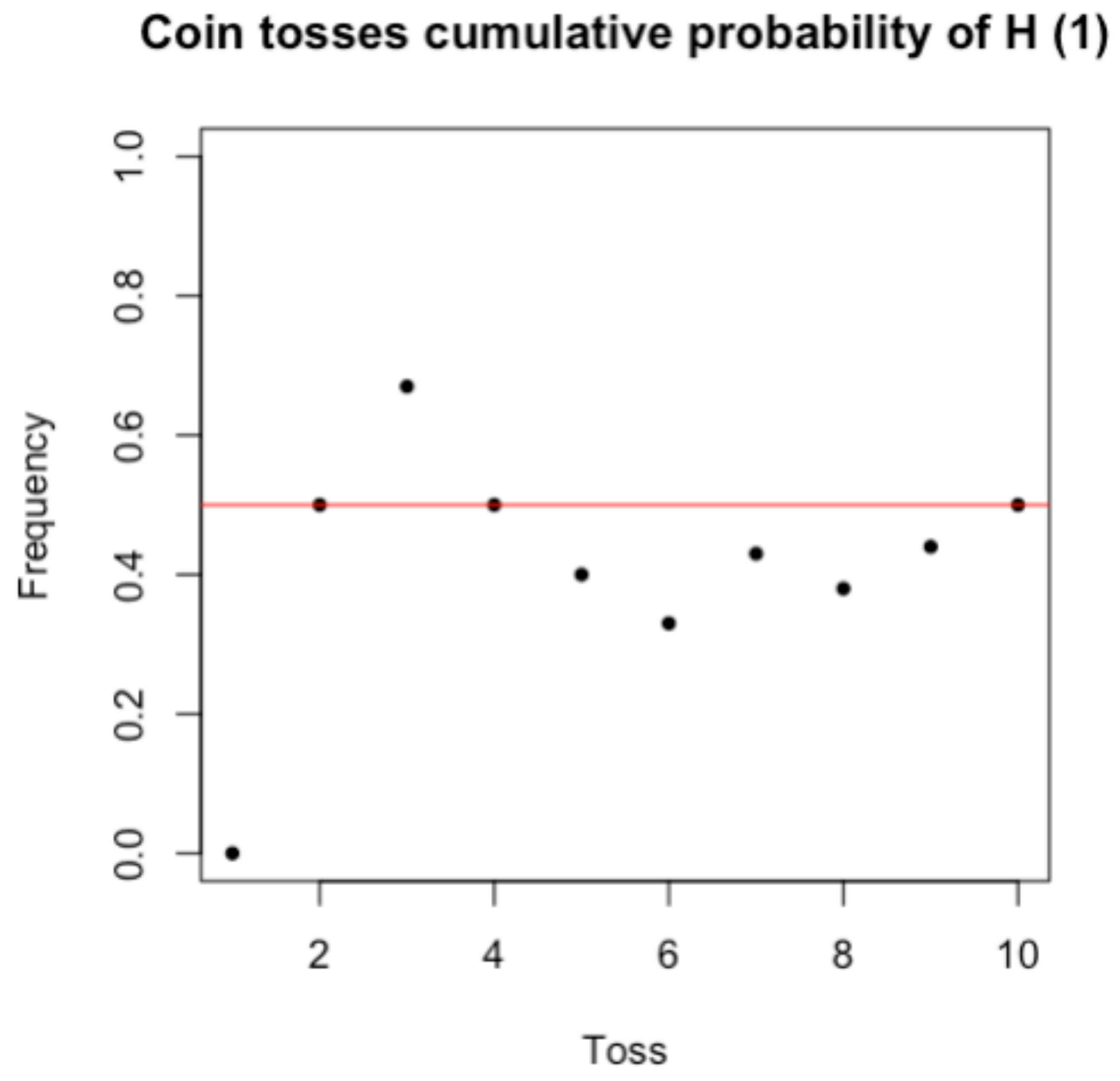
10 tosses

1 0 1 0 0 0 0 0 1 1

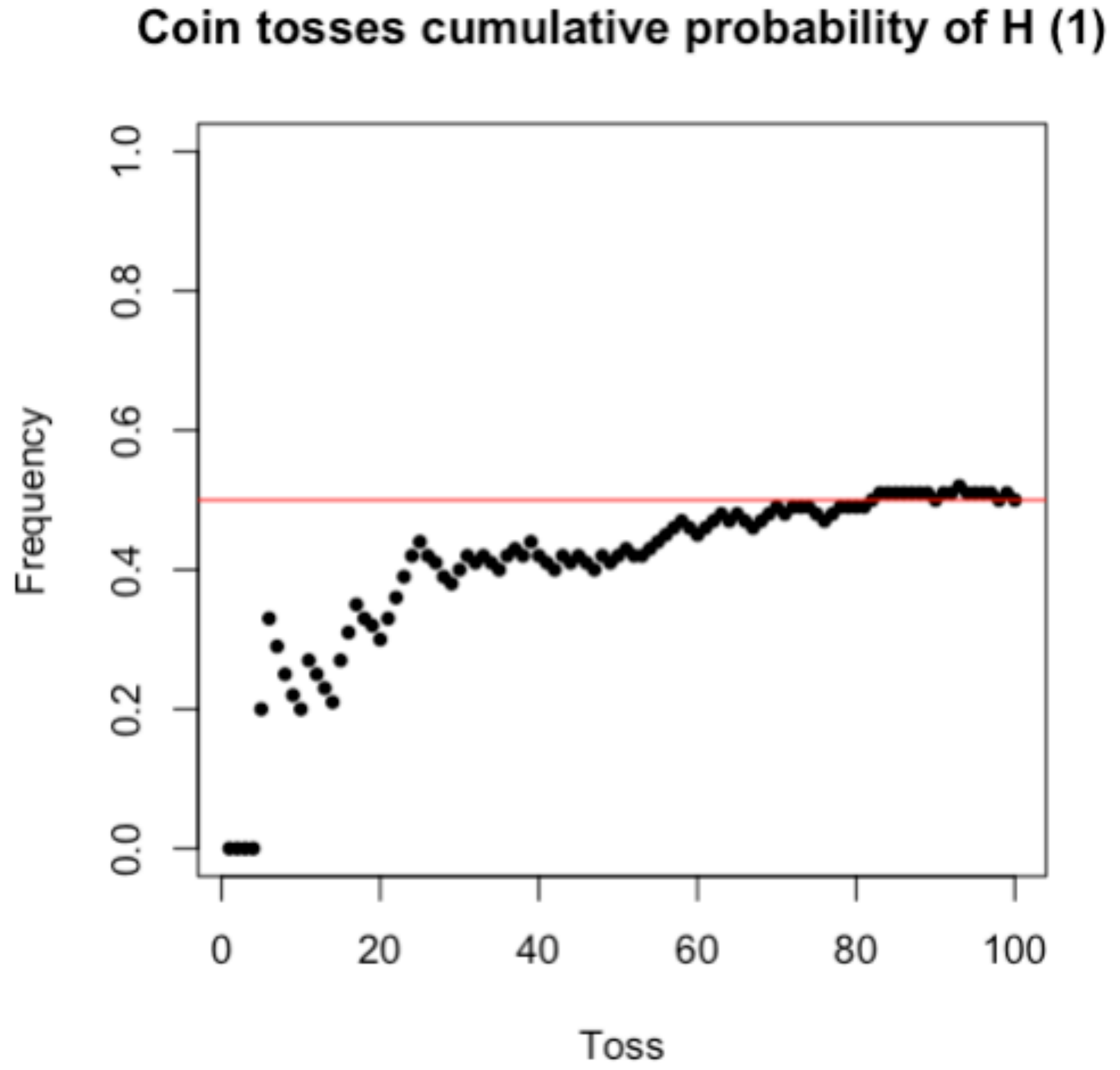


10 tosses

0 1 1 0 0 0 1 0 1 1



100 tosses



Sample space, events

Sample space means all possible outcomes; **event** is a collection of outcomes.

Throw one 6 sided dice: sample space (1,2,3,4,5,6)

Event: (1) - **Prob (1) = 1/6**
Event: (1,3,5) - **Prob (1,3,5)** = 3/6 = 1/2

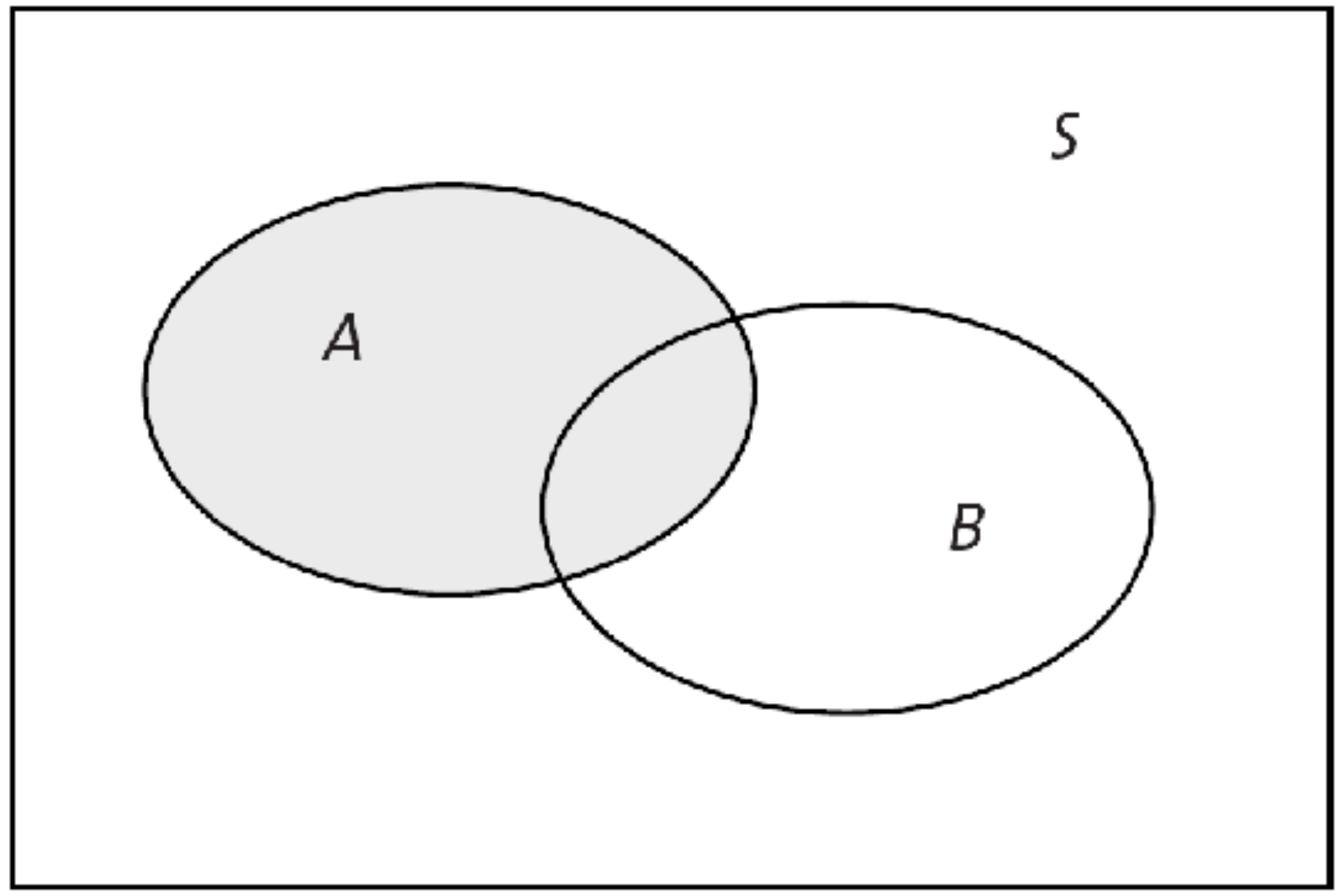
Event: not (1) - **Prob (not 1)** = 1 - 1/6 = 5/6

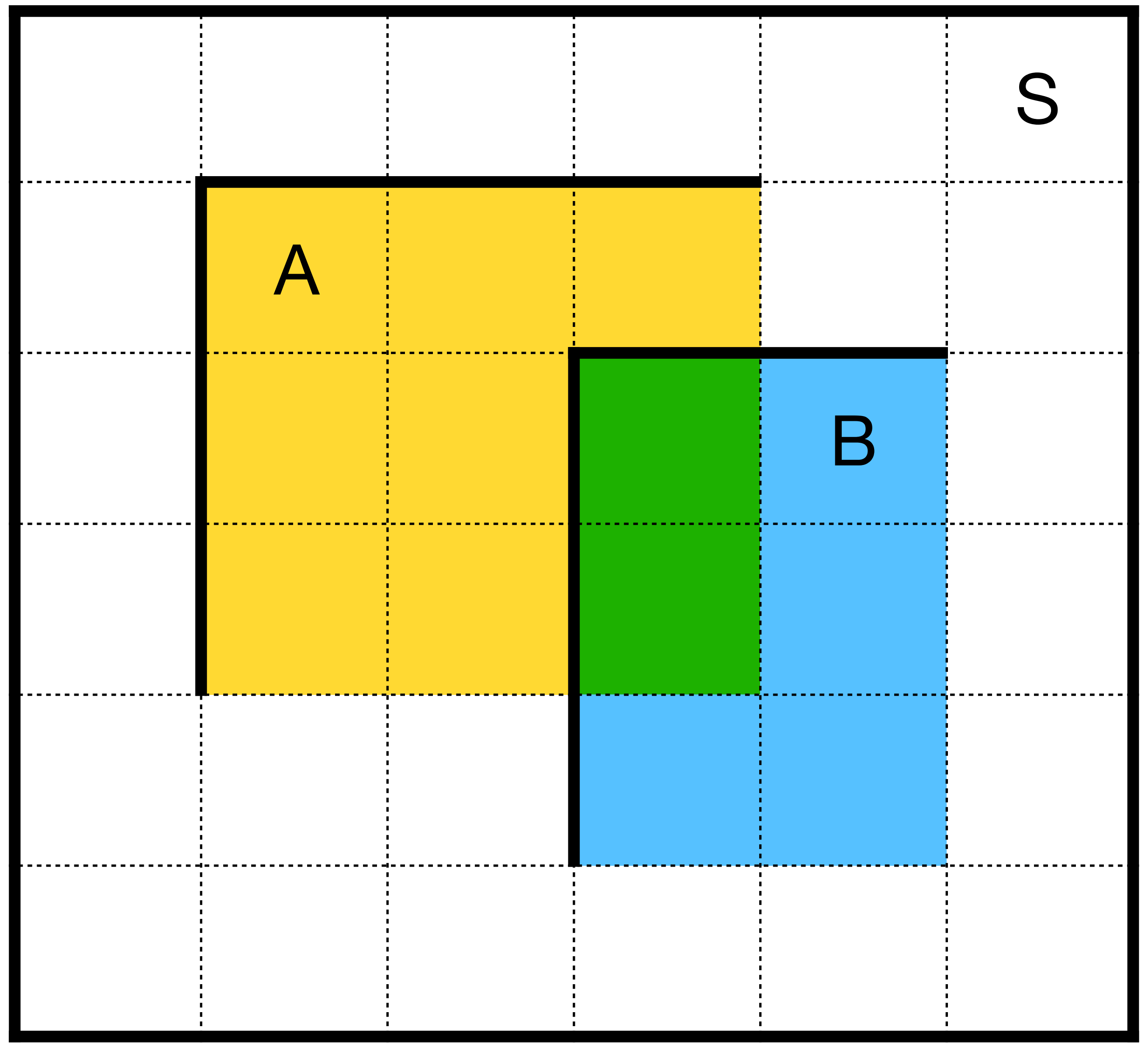
Throw two 6 sided dice, what is the probability of having **1 and 1** respectively?
1/6 * 1/6 = 1/36

There are 36 unique combinations. (1,2) is different from (2,1)

Probability of having **3 as a sum** is 1/36 + 1/36 = 2/36

Venn diagram



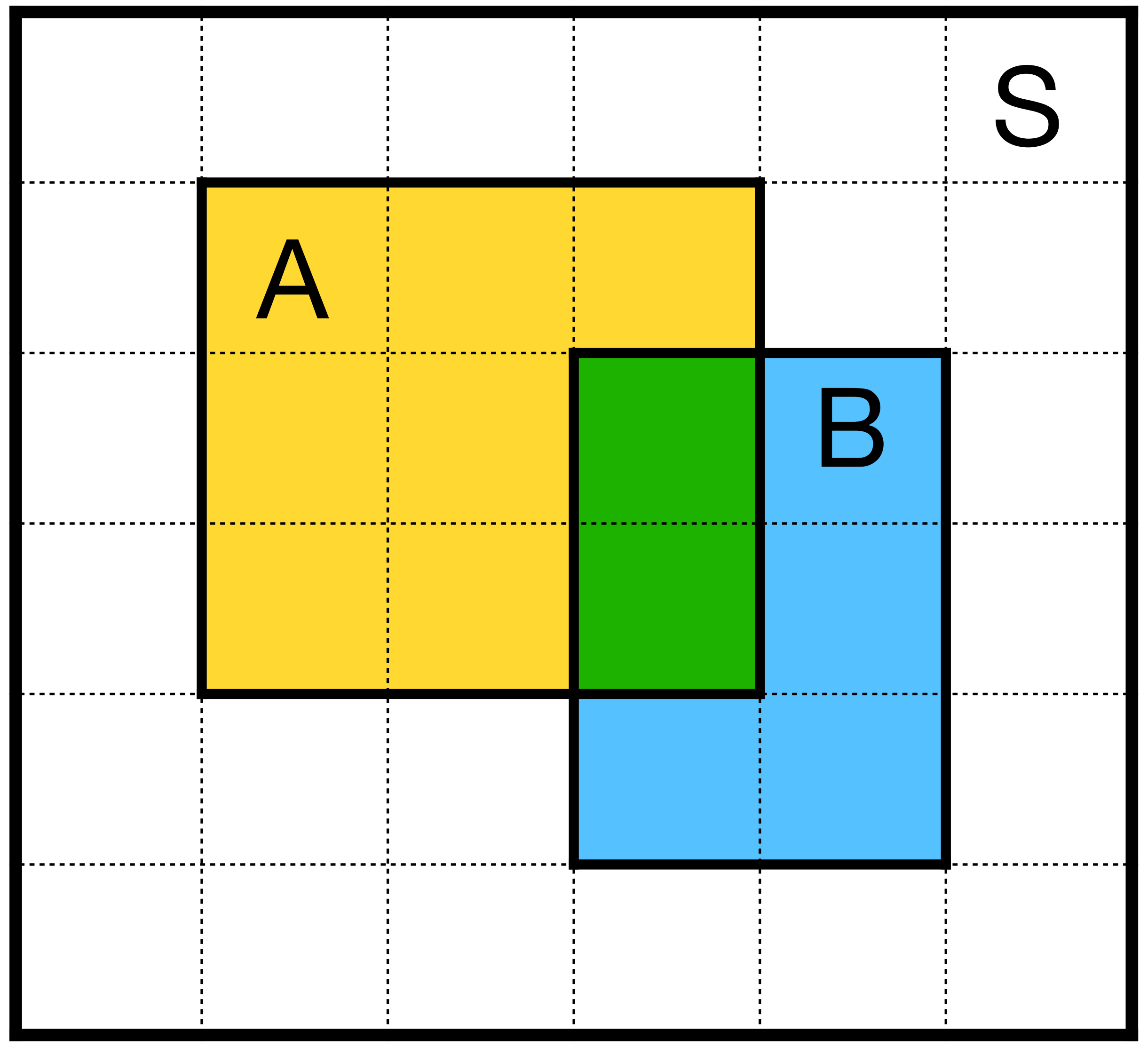


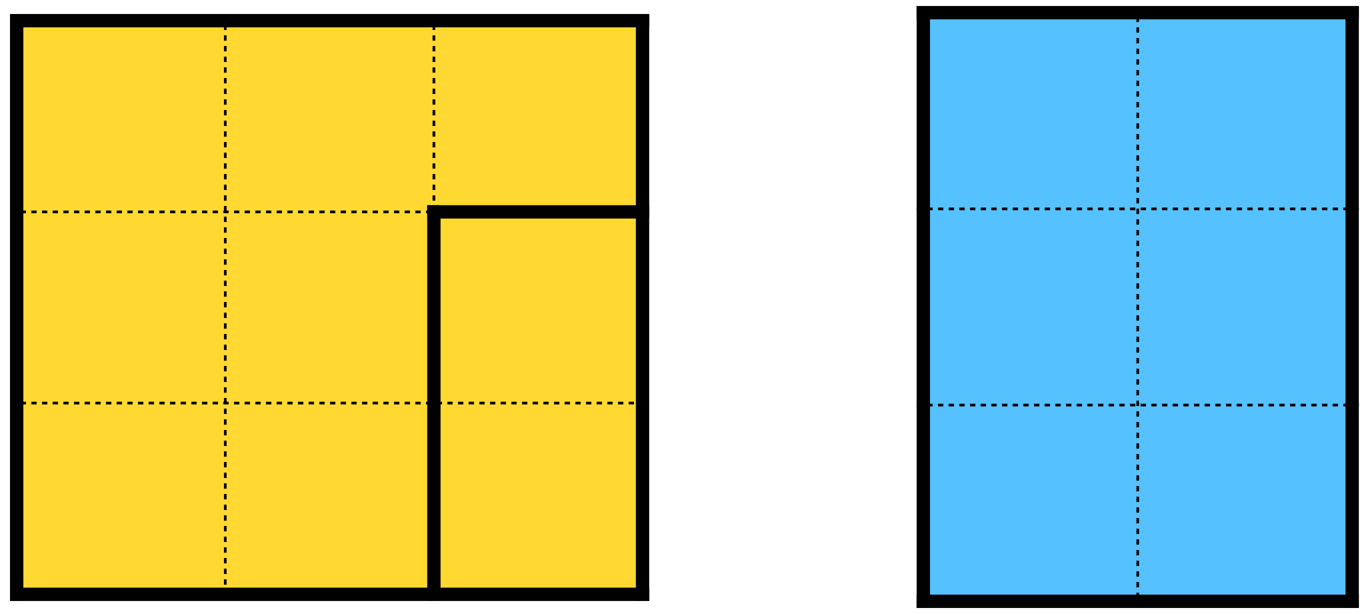
Venn diagram uses circles. S: sample space; A, B: event

We use rectangles in this course - easy to count

Marginal probability

Marginal probability: P(A), P(B)



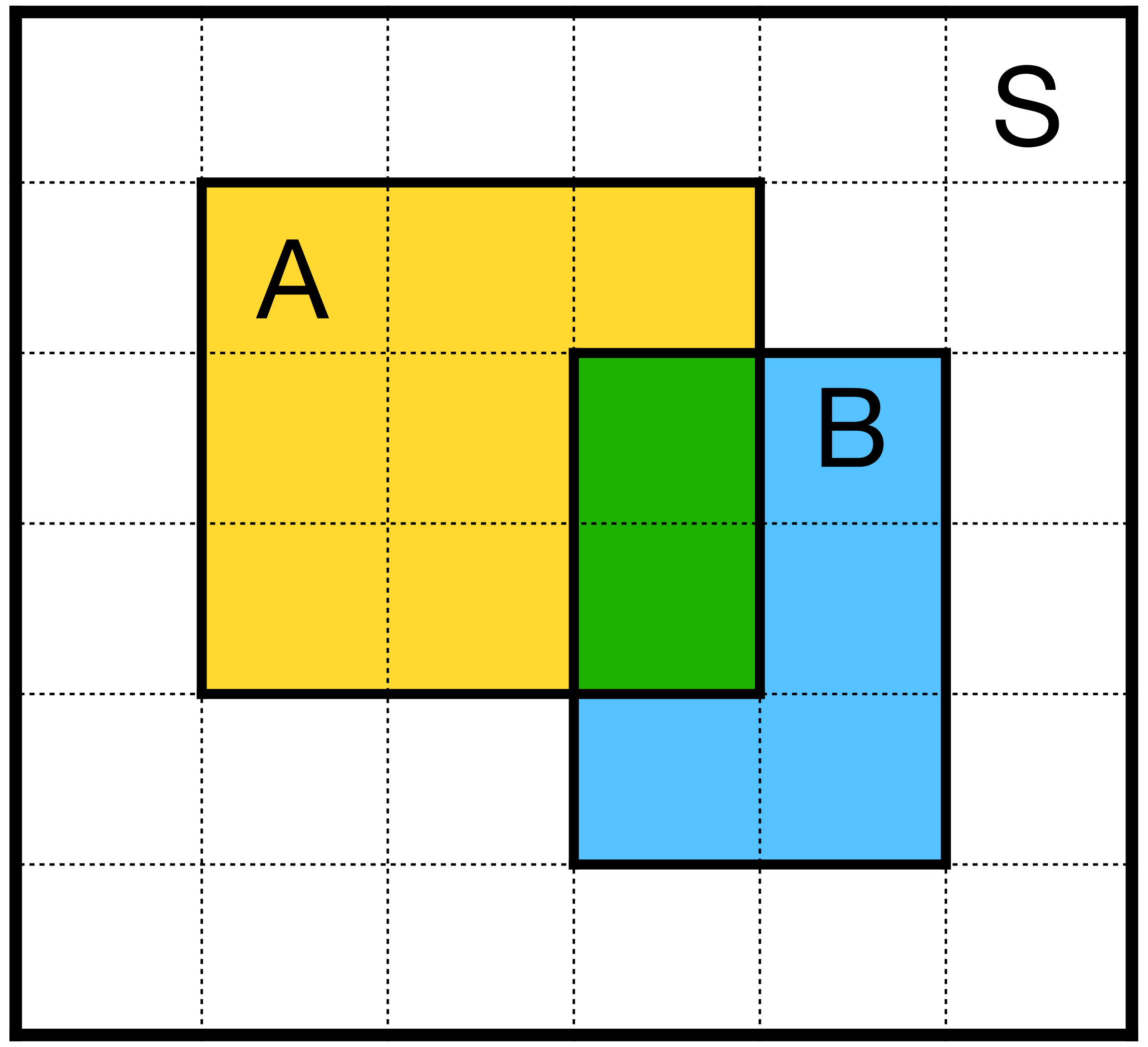


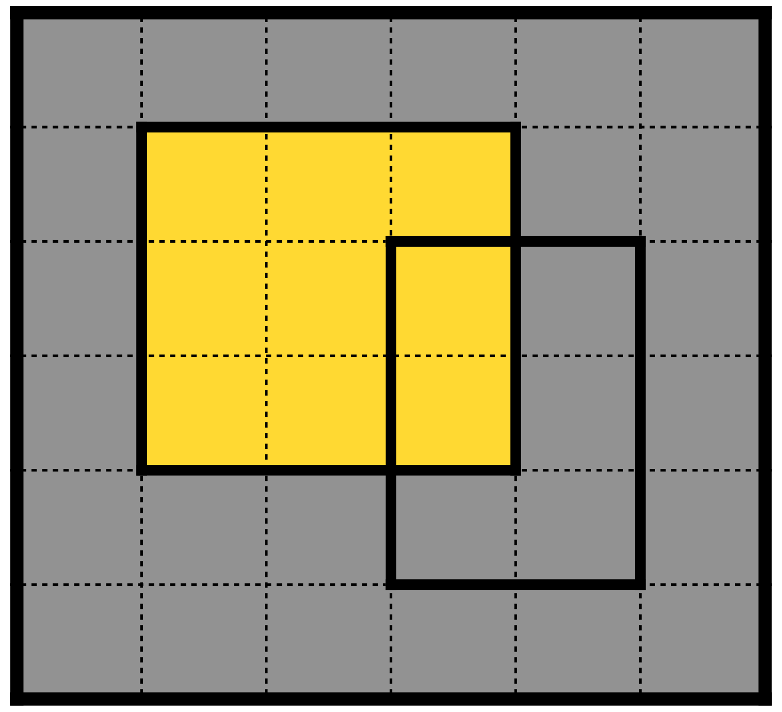
P(A) = (3x3)/(6x6) = 1/4

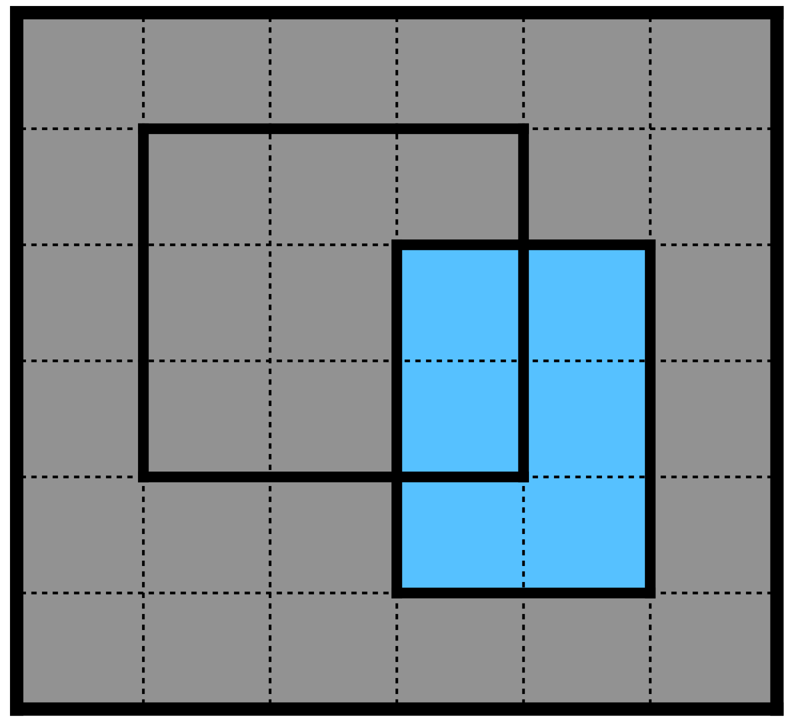
P(B) = (3x2)/(6x6) = 1/6

Marginal probability

Marginal probability: P(not A), P(not B)





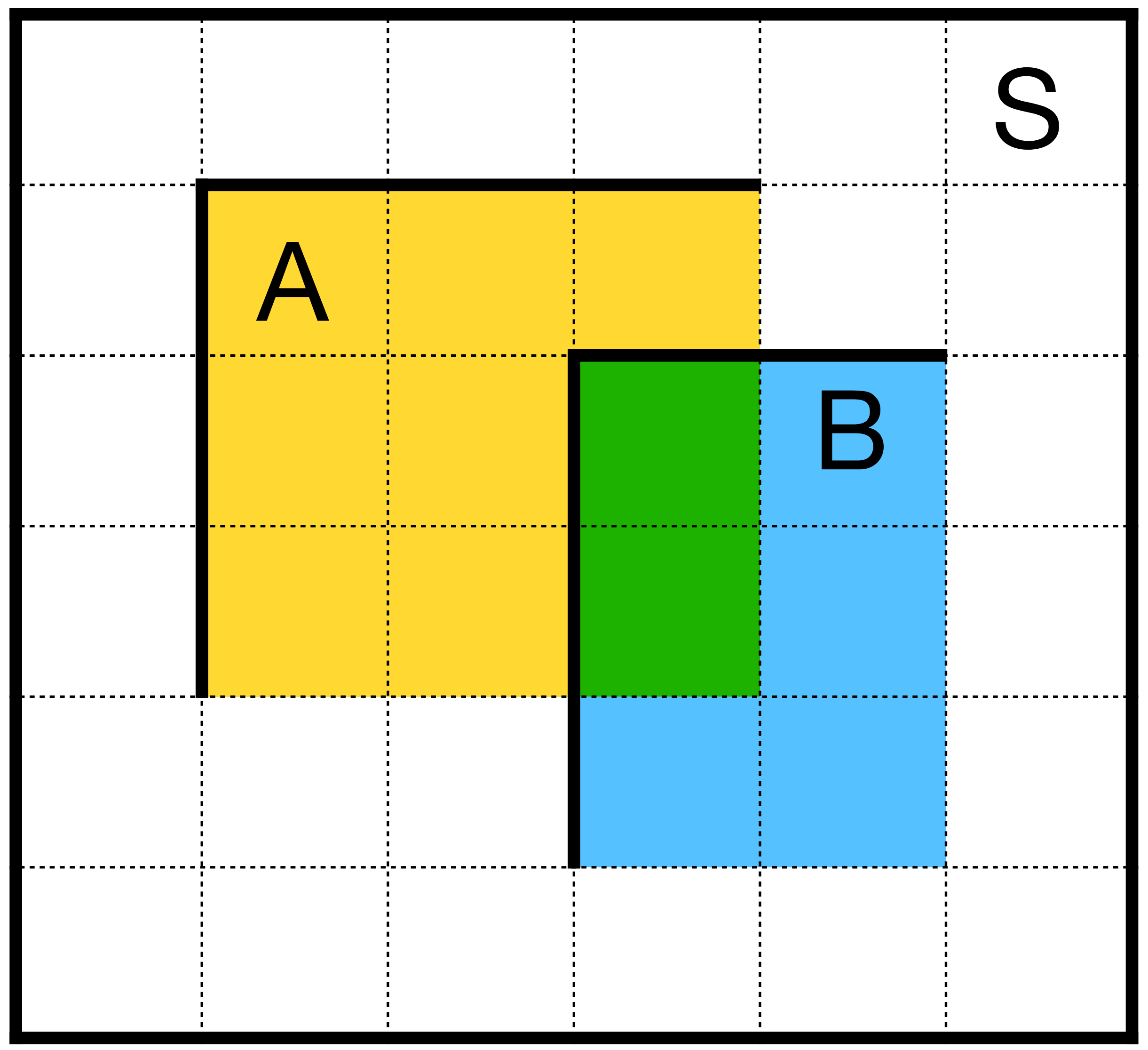


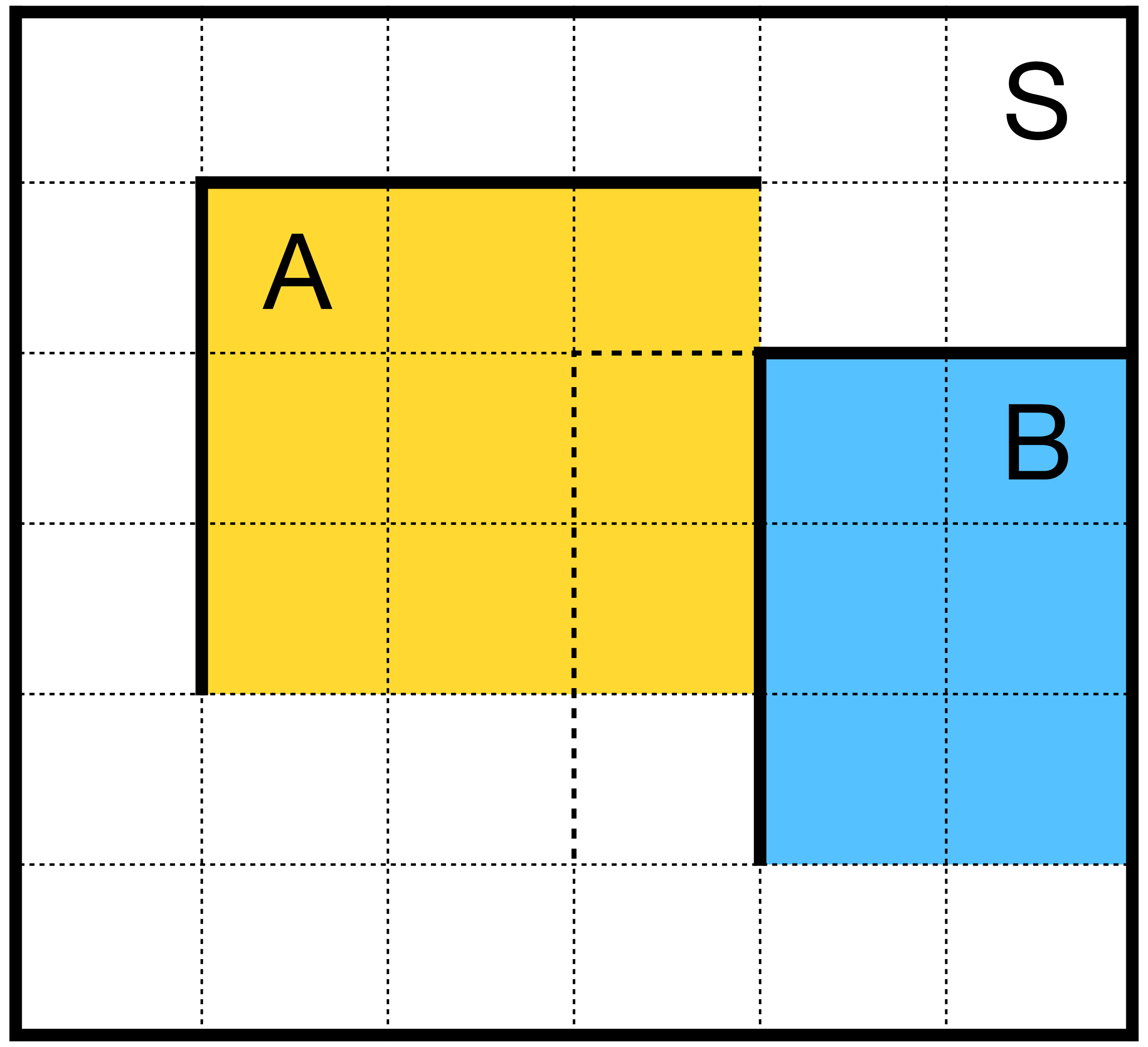
Complement rule

P(not A) = 1- P(A) = 3/4

P(not B) = 1- P(B) = 5/6

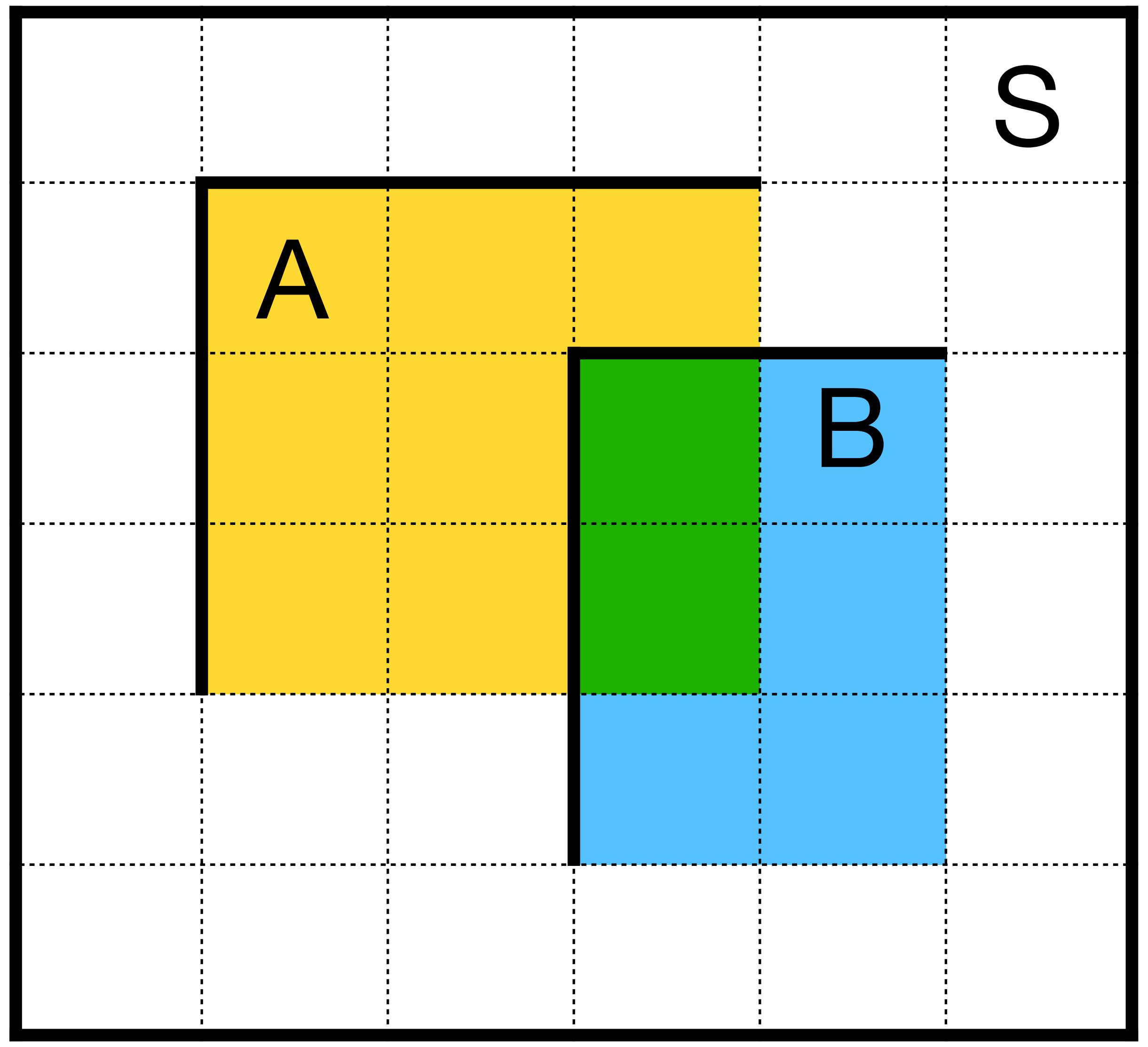
Joint probability, intersection





P(A and B) = 2/(6x6) = 1/18   P(A and B) = 0 (disjoint)

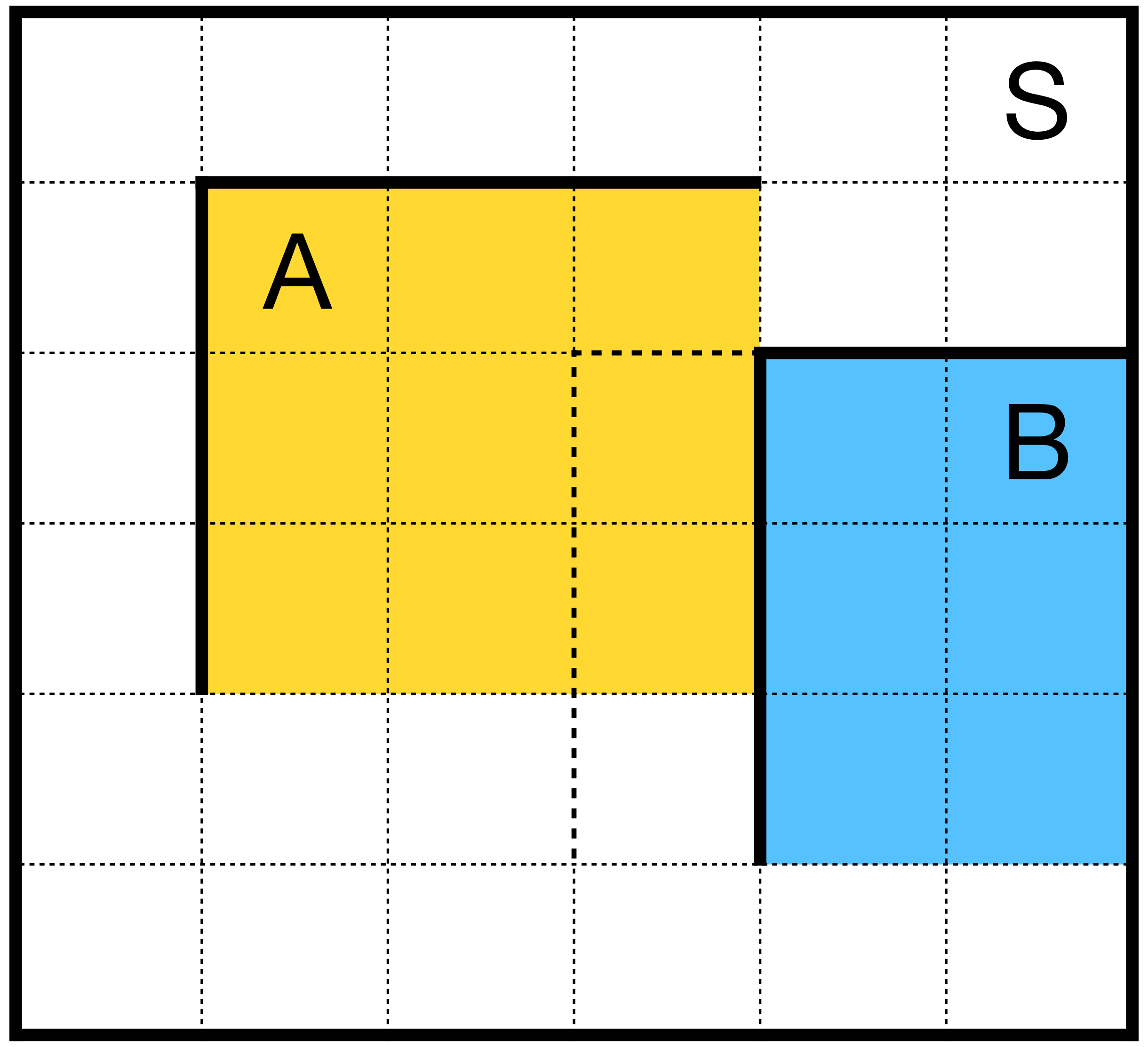
Union, additive rule



P(A or B) = P(A)+P(B)-P(A and B)

= 9/36 + 6/36 - 2/36

= 13/36



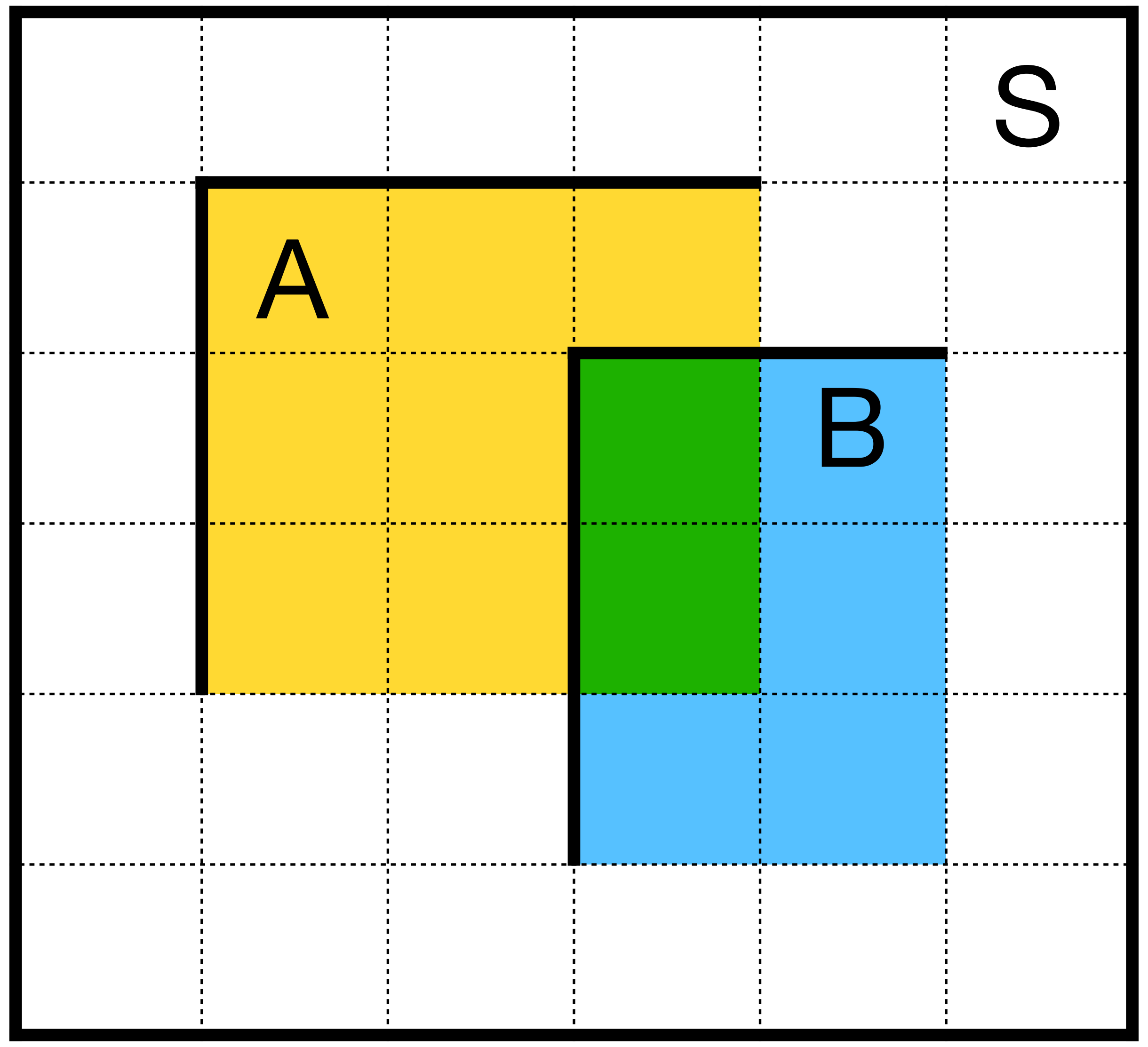
P(A or B) = P(A)+P(B)

= 9/36 + 6/36

= 15/36 = 5/12

Conditional probability

P(B|A) = ?



P(A and B) = P(B|A)P(A)

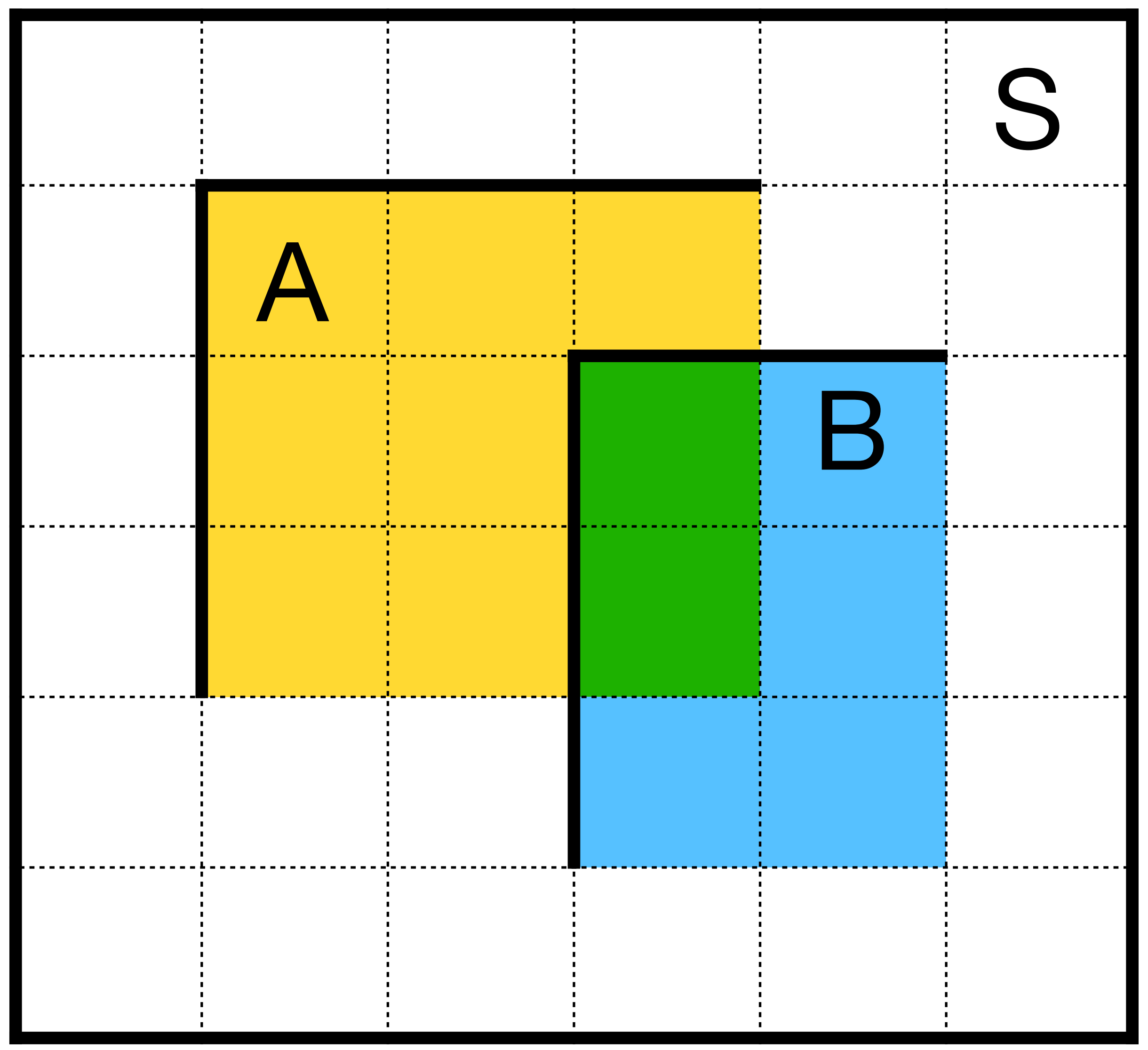
Probability of B given A =
how much proportion of B is in A

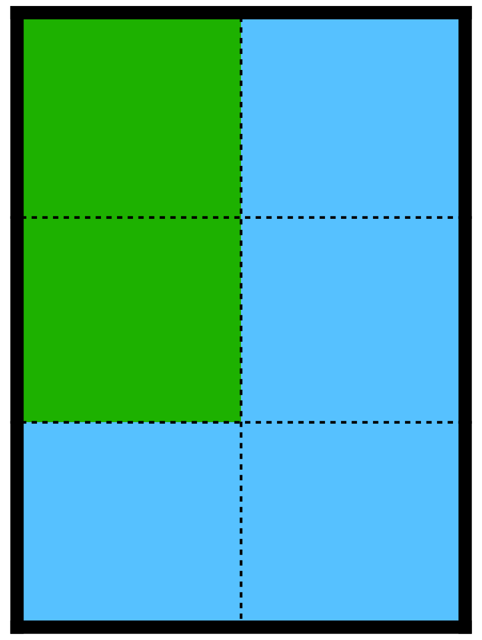
P(B|A) = P(A and B)/P(A)
= (1/18) / (1/4)
= 4/18 = 2/9

= 2/(3x3) = 2/9

Conditional probability

P(A|B) = ?



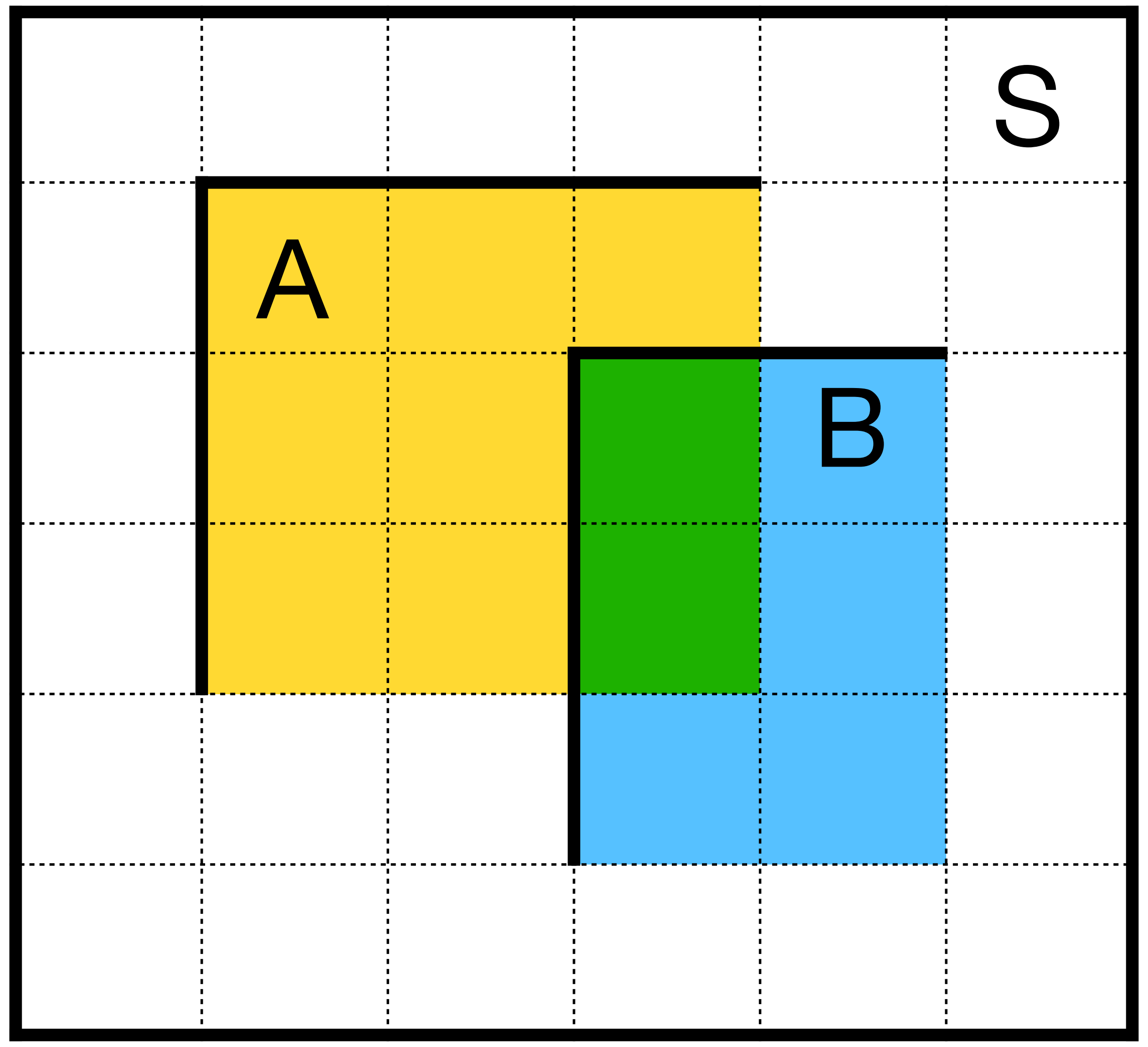


P(A and B) = P(A|B)P(B)

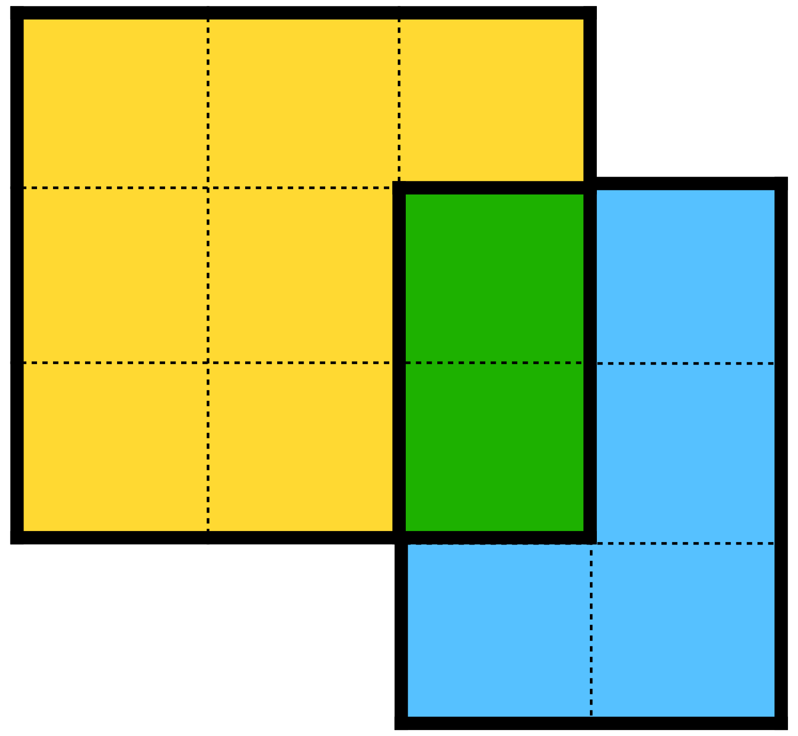
P(A|B) = P(A and B)/P(B)
= (1/18) / (1/6)
= 6/18 = 1/3

= 2/(2x3) = 1/3

Conditional probability



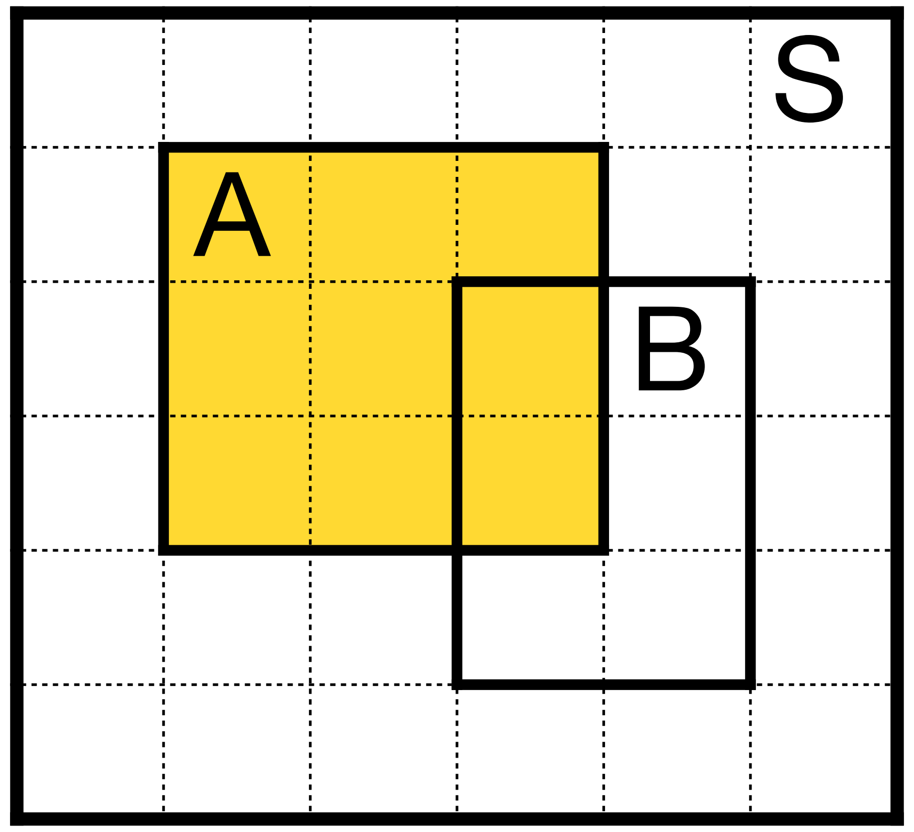
P(A and B) is the same



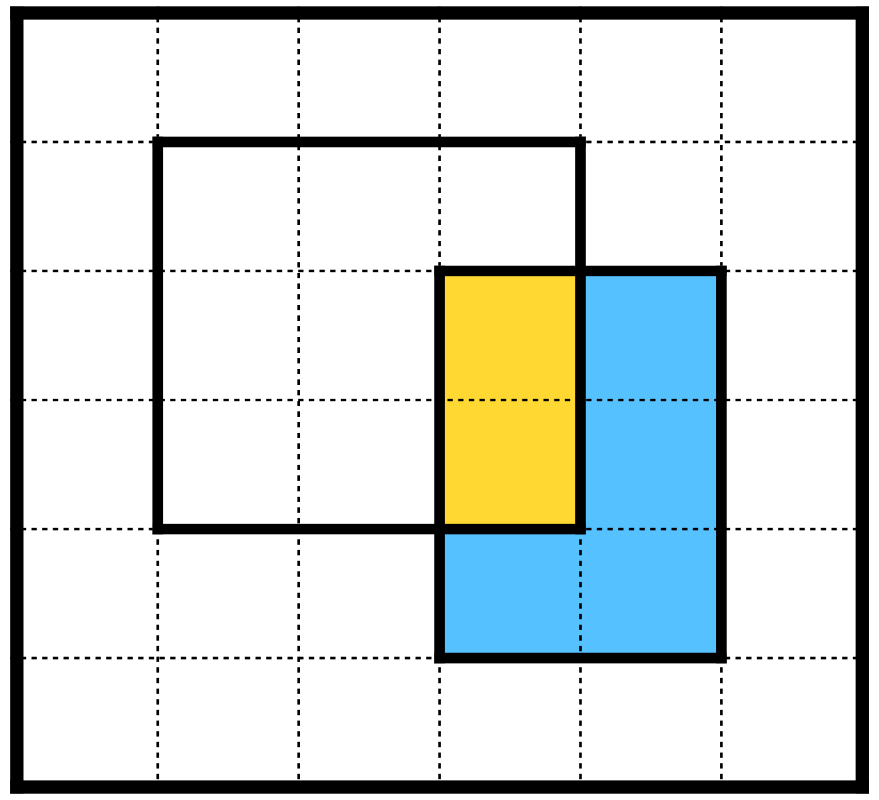
P(A and B) = P(A|B)P(B) = P(B|A)P(A)

Total probability

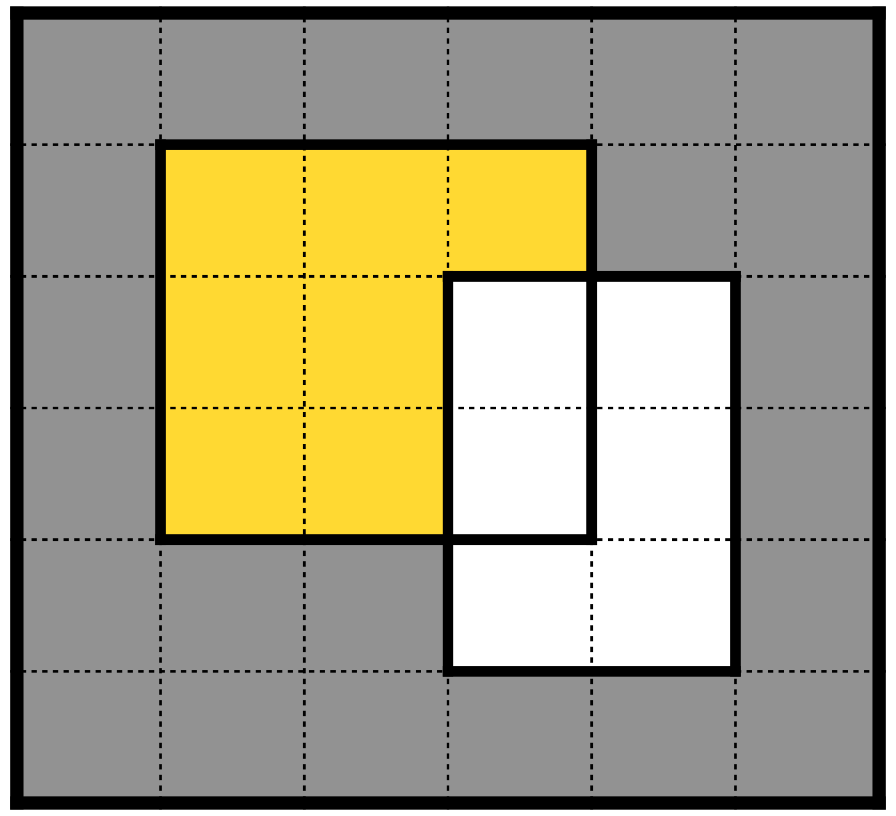
P(A) = ?



P(A and B)
=P(A|B)P(B)



P(A and not B)
=P(A|not B) P(not B)



$$P(A) = P(A|B)P(B) + P(A|\text{not } B)P(\text{not } B)$$

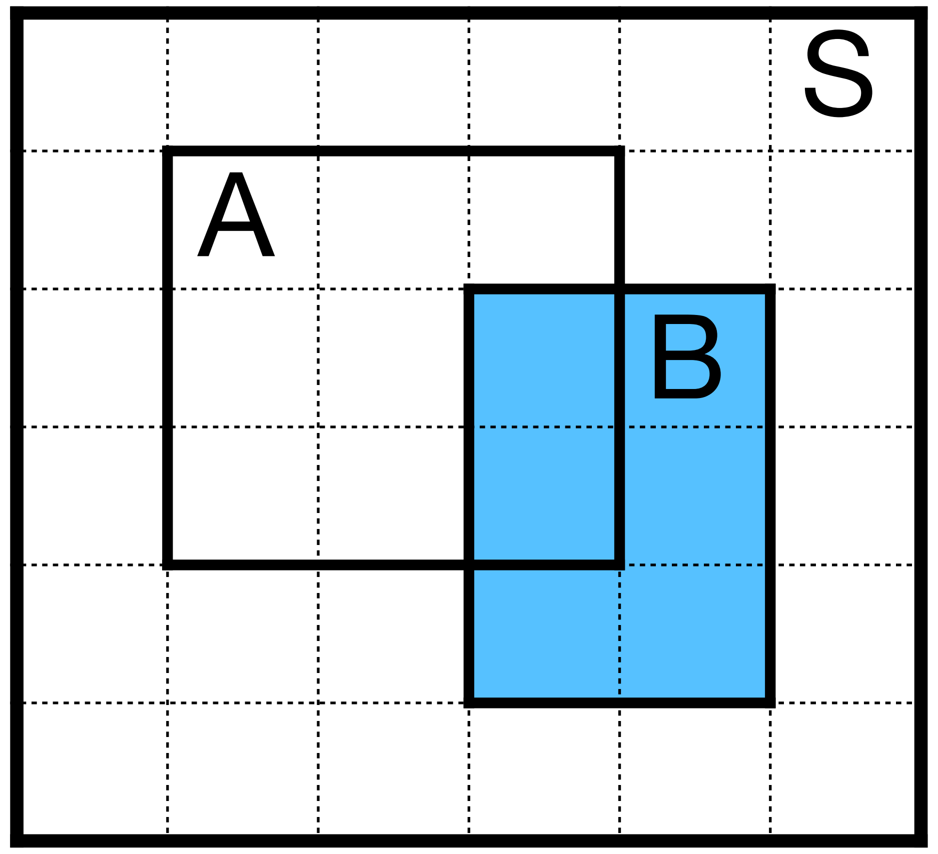
$= 2/36 + 7/36$
$= 9/36 = 1/4$

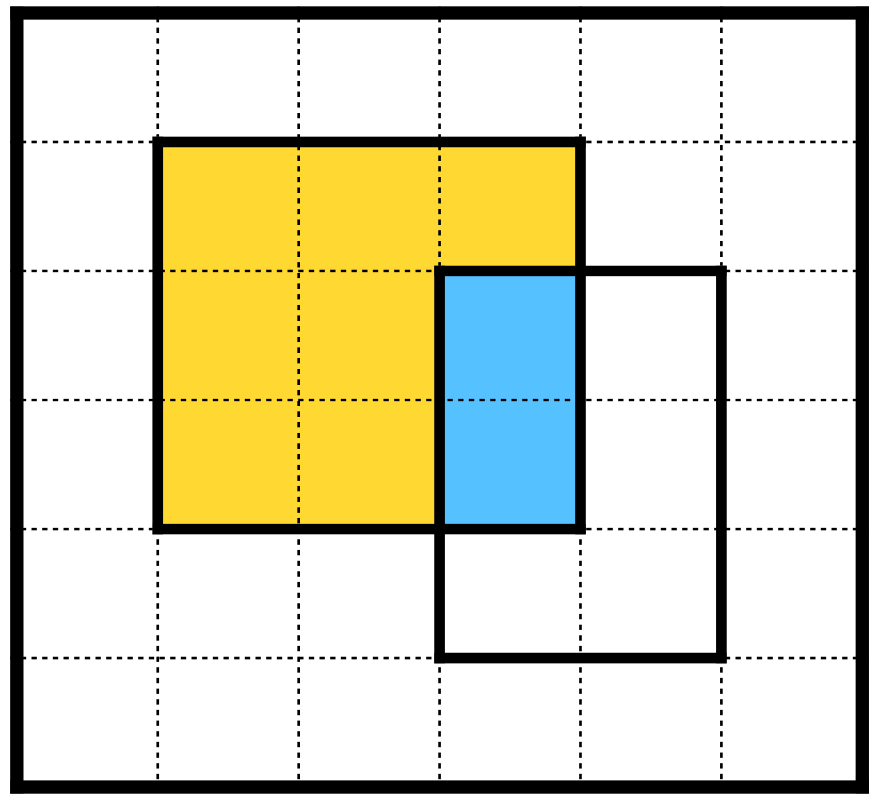
(Can double check P(A) calculated from previous slides)

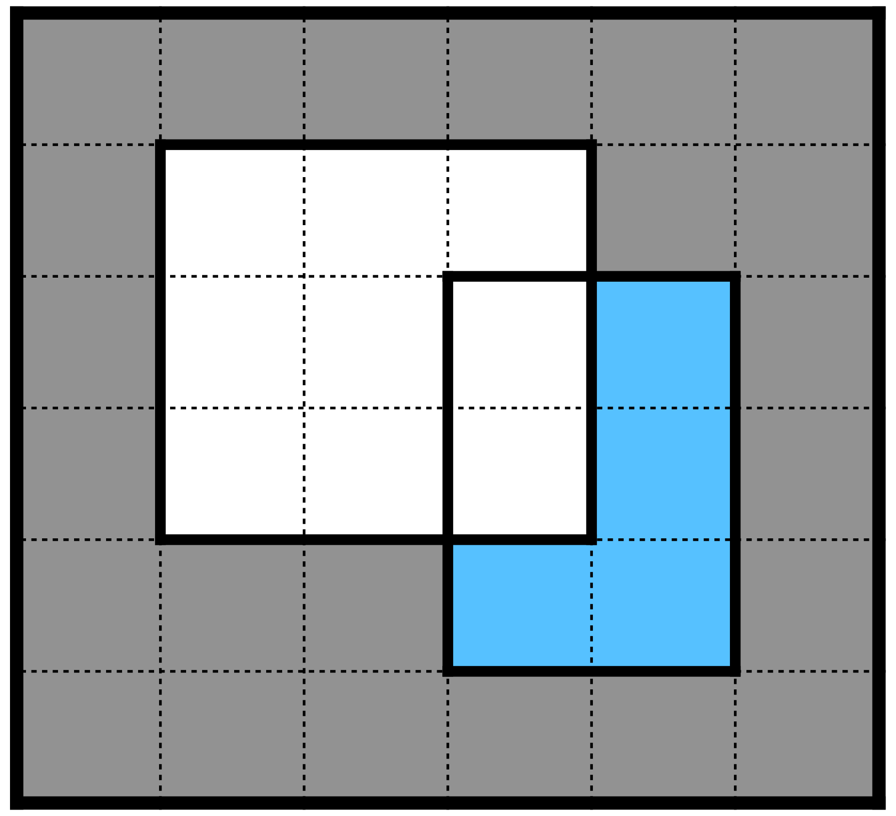
P(B) = 1/6
P(A|B) = 1/3
P(not B) = 5/6
P(A|not B) = 7/30

Total probability

P(B) = ?





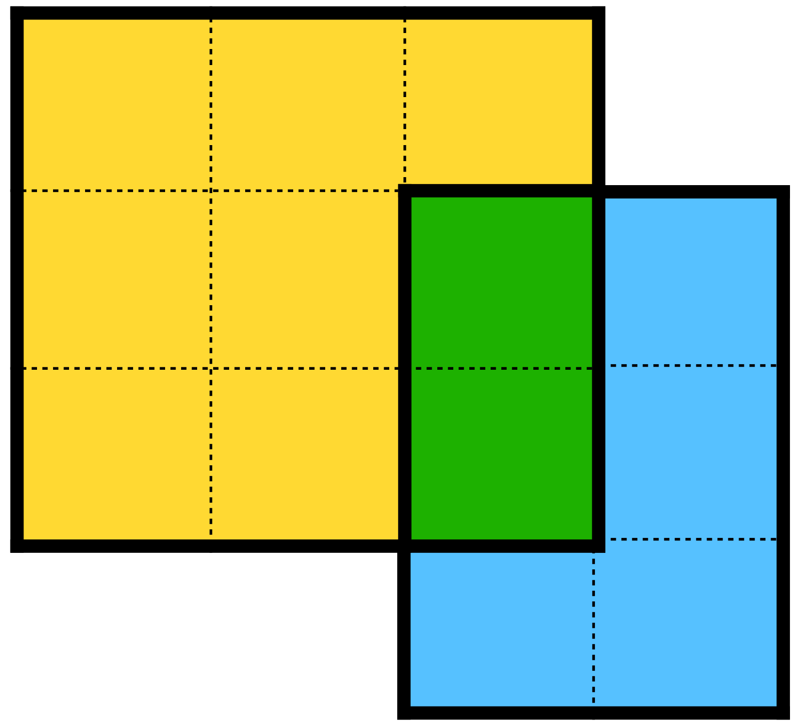


P(B) = P(B|A)P(A) + P(B|not A) P(not A)

= 2/36 + 4/36
= 6/36 = 1/6

Bayes Law: conditional probability reformulated

$$P(A \text{ and } B) = P(A|B)P(B) = P(B|A)P(A)$$



This is conditional probability

Equivalent to

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

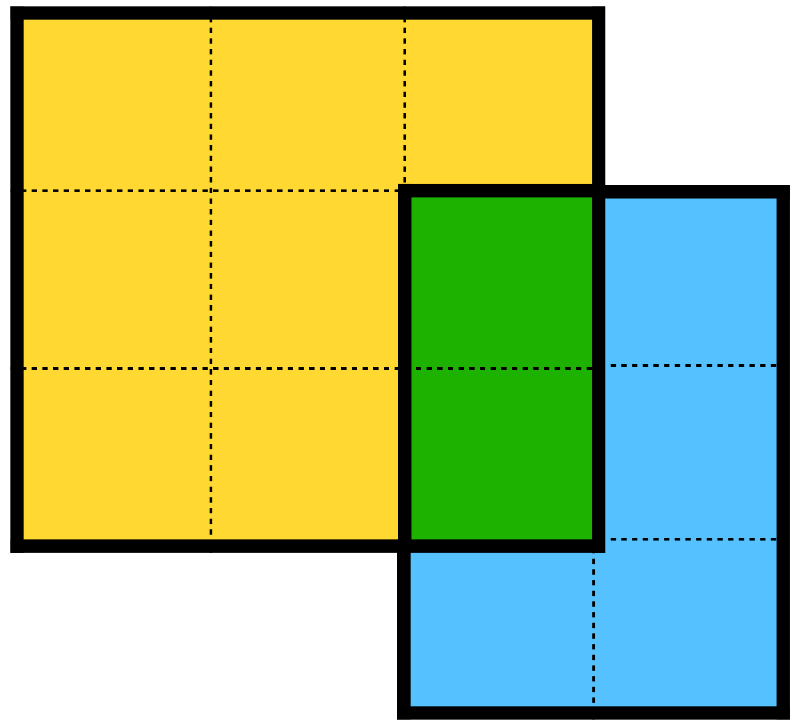
$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

$P(A|B)$, $P(B|A)$: conditional probability

$P(A)$, $P(B)$: marginal probability

Bayes Law: conditional probability reformulated

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



We get the marginal probability of B, P(B) via total probability

$$P(B) = P(B|A)P(A) + P(B|\text{not } A)\, P(\text{not } A)$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\text{not } A)\, P(\text{not } A)}$$

Bayes Law

Also called Bayes Theorem or Bayes Rule, formulated by English statistician Thomas Bayes (18th century)

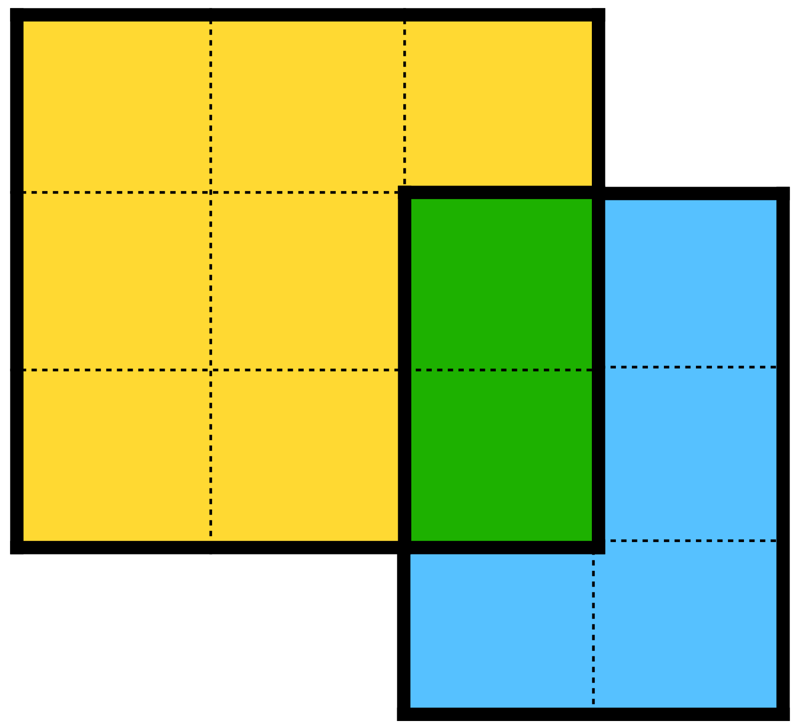
It is fundamental to Bayesian statistics and inference

Also useful to understand the relationship between concepts in diagnostic tests (next lecture):

E.g. how many people really have HIV after being test positive? How useful is my test for cancer screening? (it depends on how common the disease is)

Summary

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\text{not } A)\, P(\text{not } A)}$$