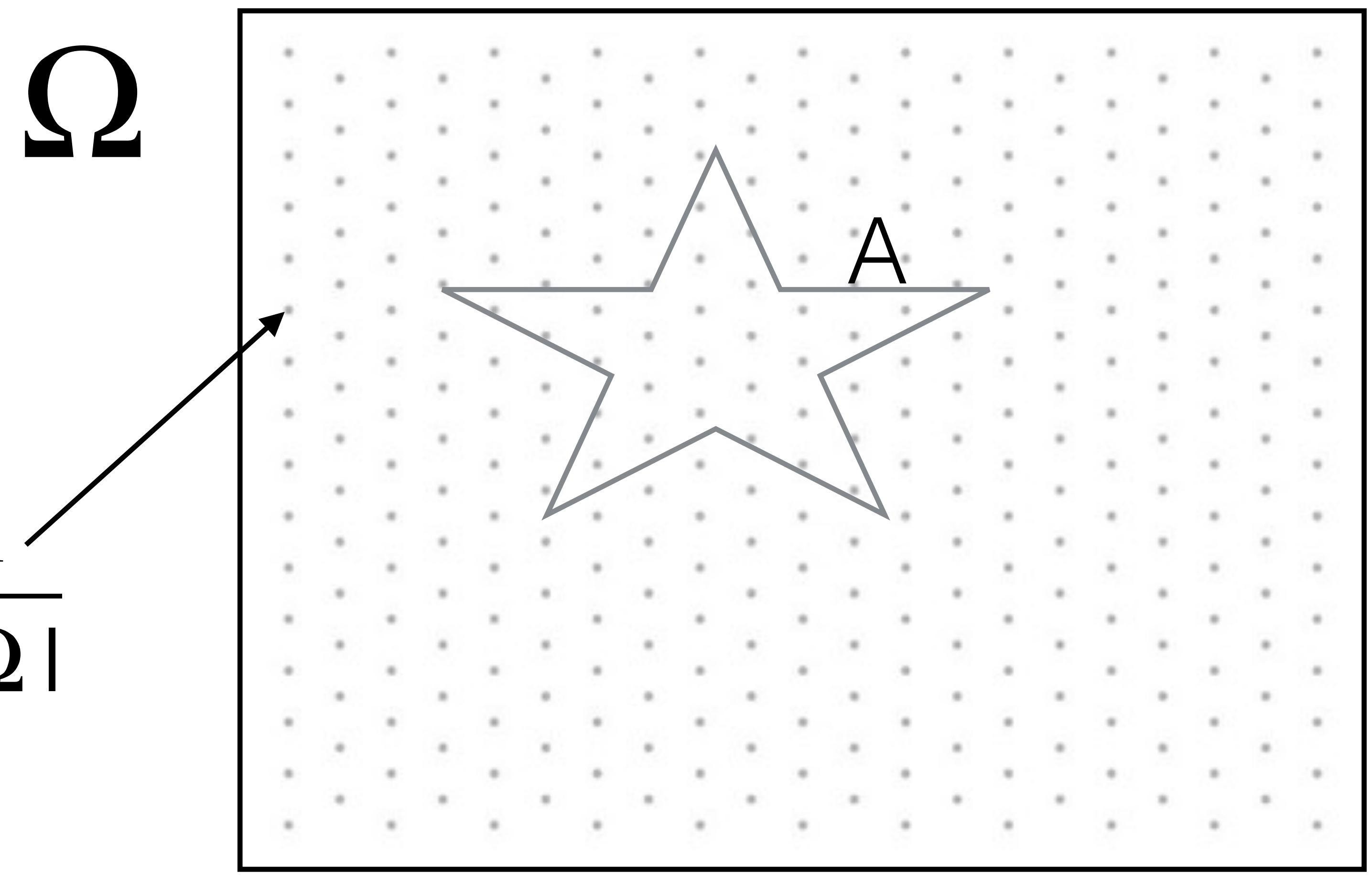
General finite distributions Conditional Probabilities and independence

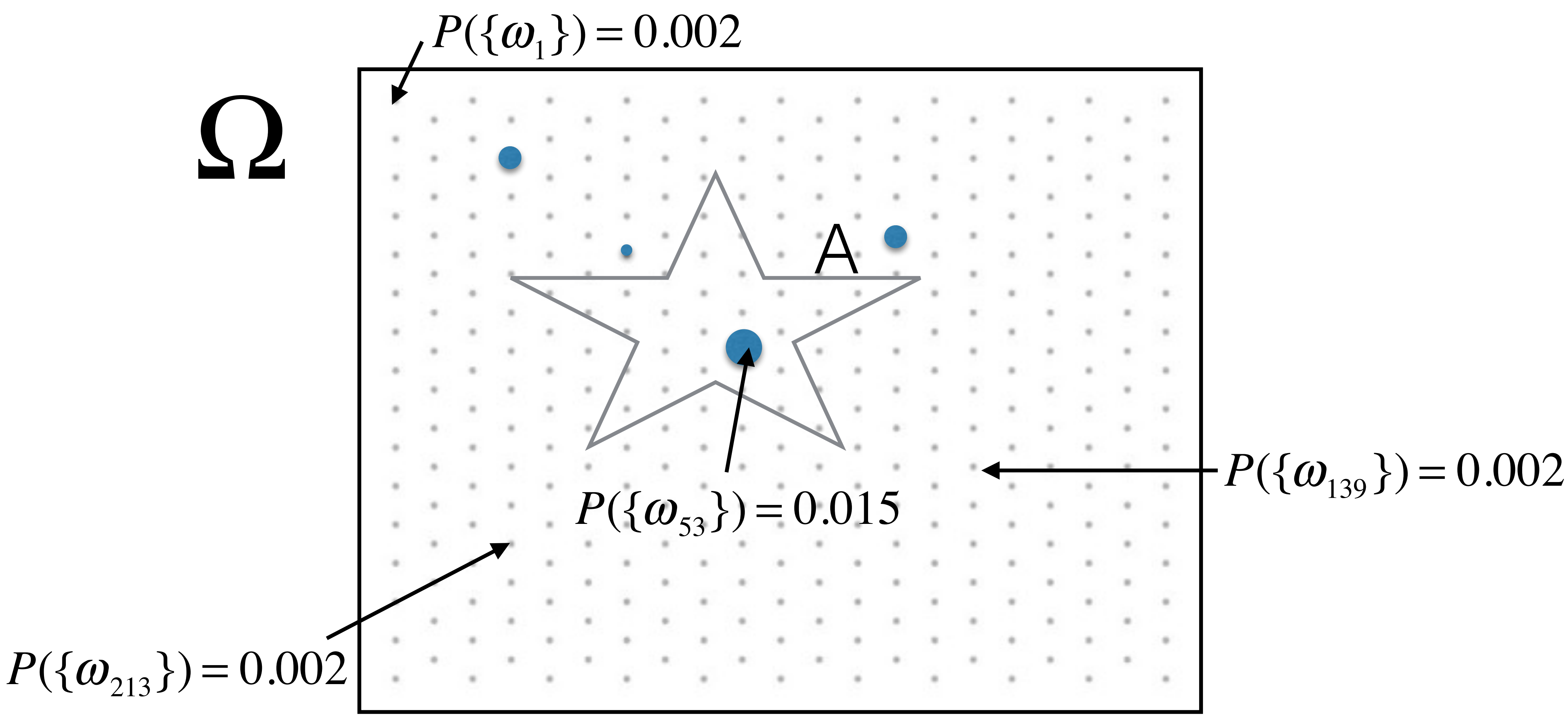
Uniform Dist. over finite outcome space spaces



$$P(\{\omega\}) = \frac{1}{|\Omega|}$$

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

Non-Uniform Dist. over finite outcome space spaces



Constraints: 1. $\forall \omega \in \Omega \quad 0 \le P(\{\omega\}) \le 1$ 2. $\sum_{\omega \in \Omega} P(\{\omega\}) = 1$

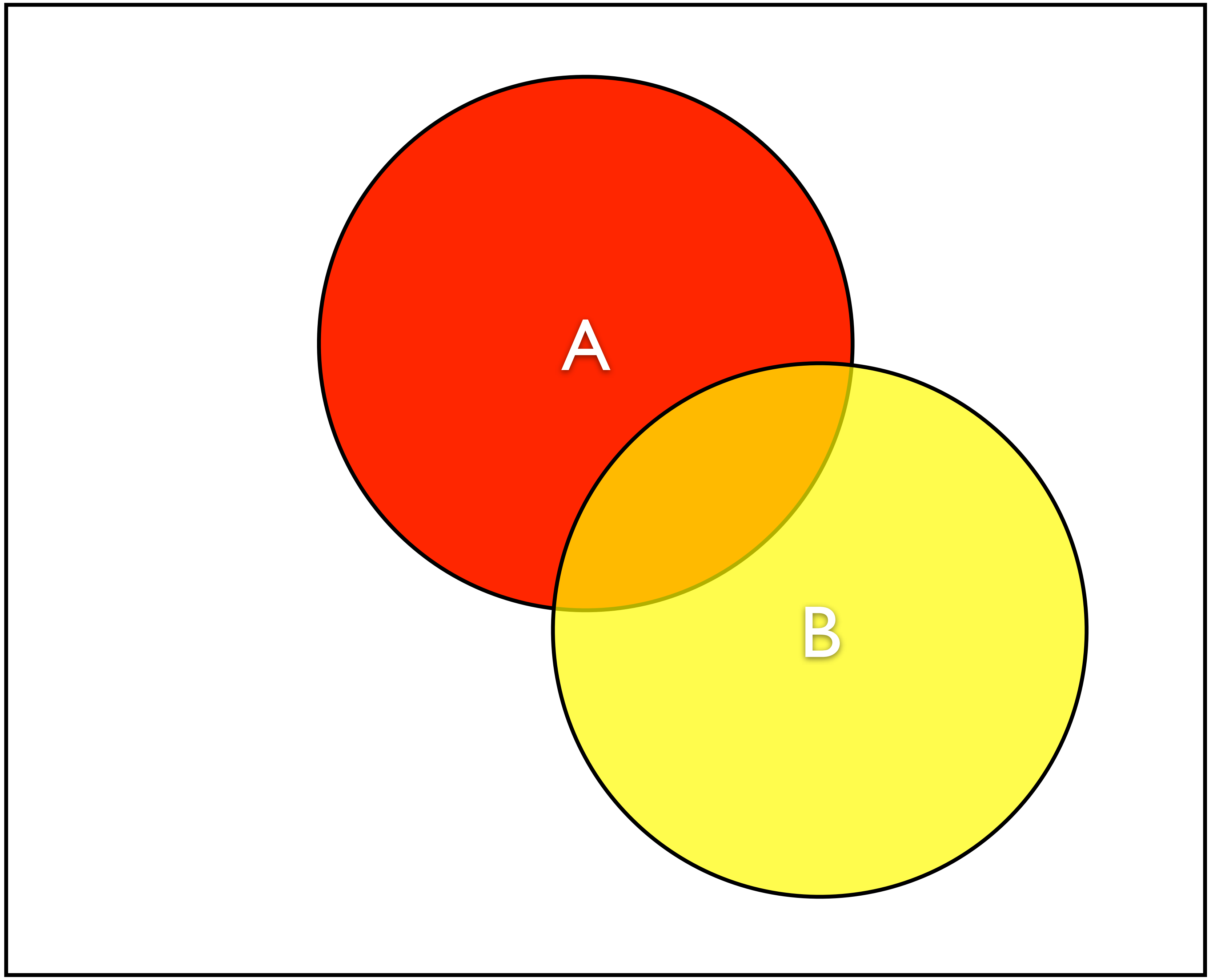
Probability Calculation: $\mathrm{P(A)} = \sum_{\omega \in A} P(\{\omega\})$

Example: Biased coin

- Outcome space= {Head,Tail}
- P({Head})=0.3
- P({Tail}) +P({Head})=1
- Therefor P({Tail})=1-0.3=0.7

Events

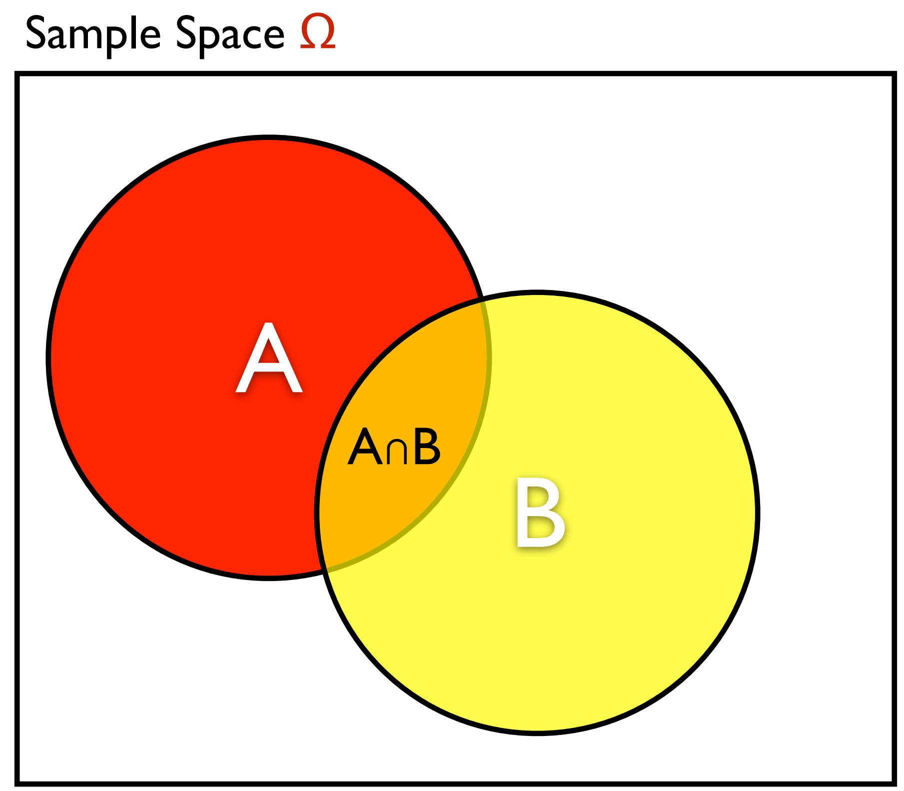
Sample Space Ω



Summation rules

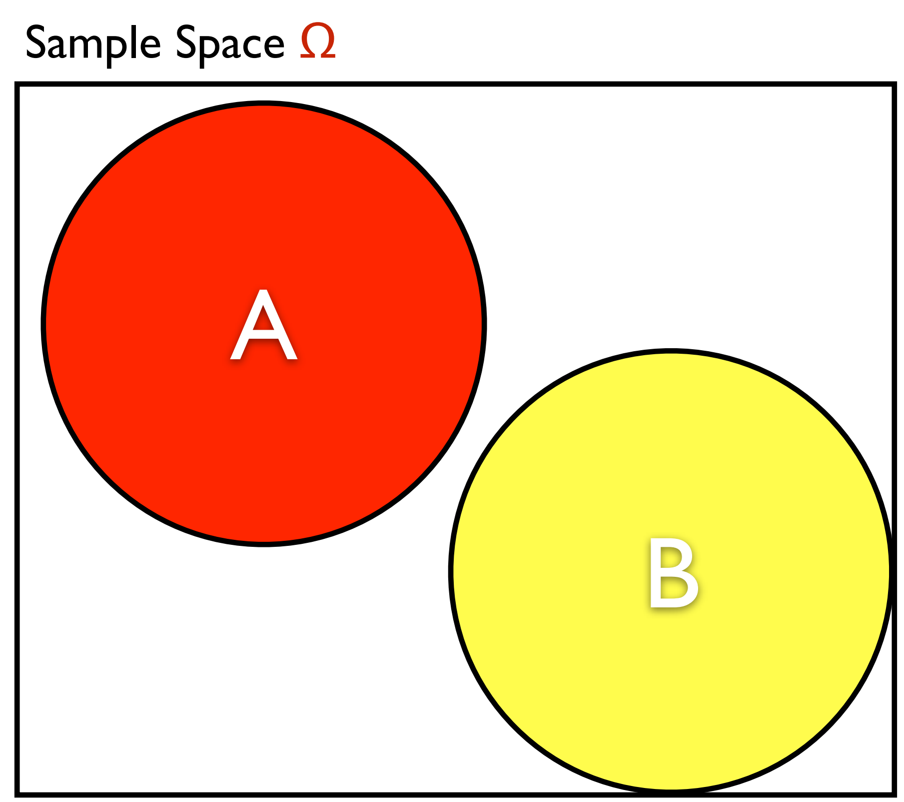
True always:

$P(A \cup B) \leq P(A) + P(B)$



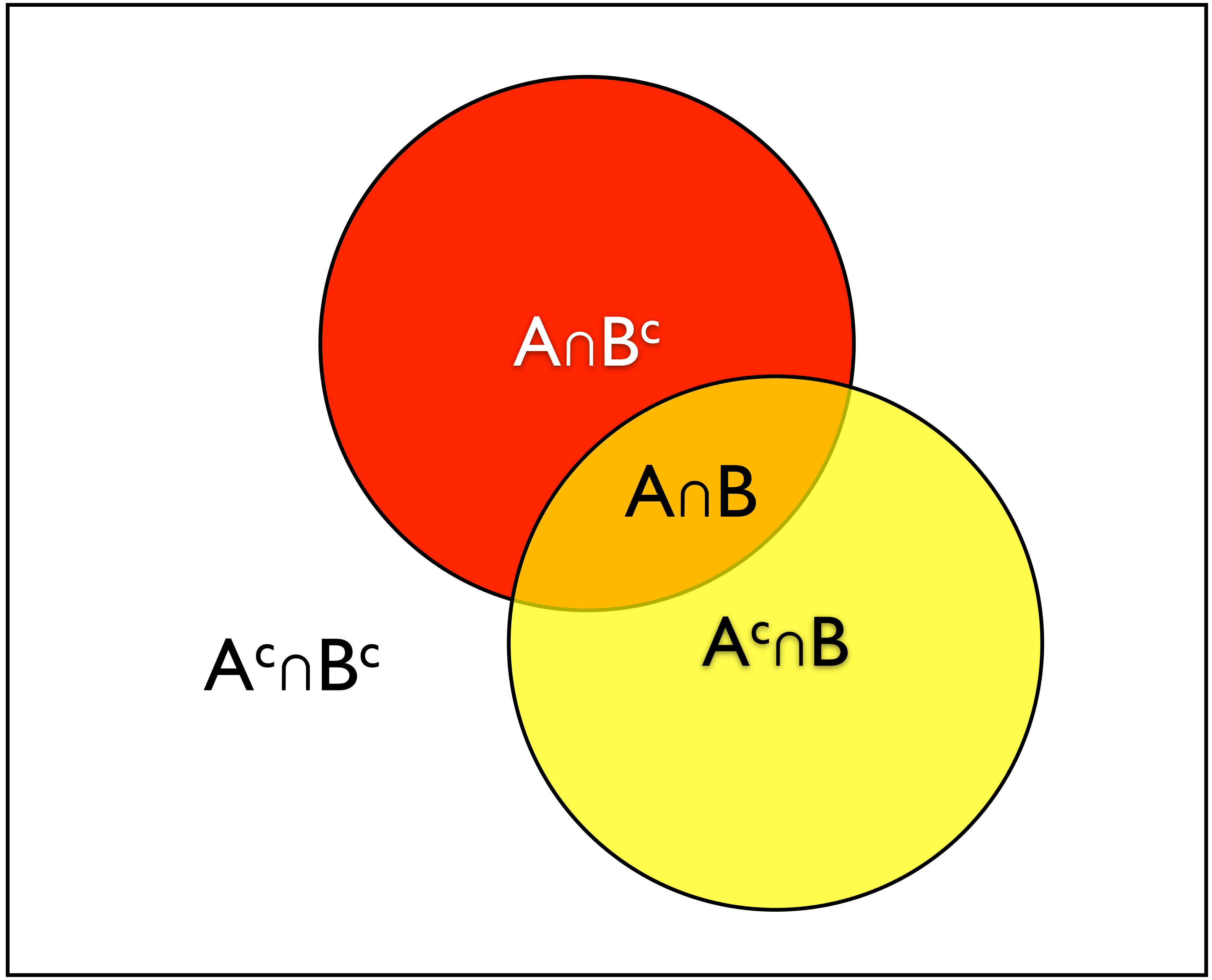
A and B are disjoint: $A \cap B = \varnothing$

implies: $P(A \cup B) = P(A) + P(B)$



Partitioning into disjoint events

Sample Space Ω

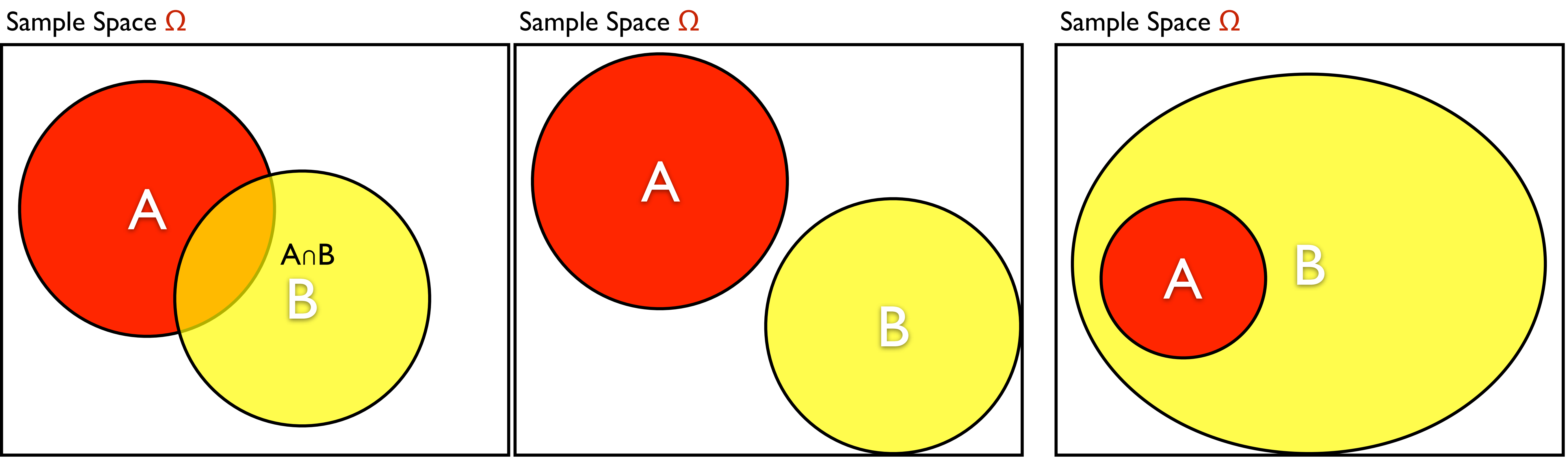


$$P(A \cup B) = P(A \cap B^c) + P(A^c \cap B) + P(A \cap B) = 1 - P(A^c \cap B^c)$$

What can we say about $P(A \cap B)$?

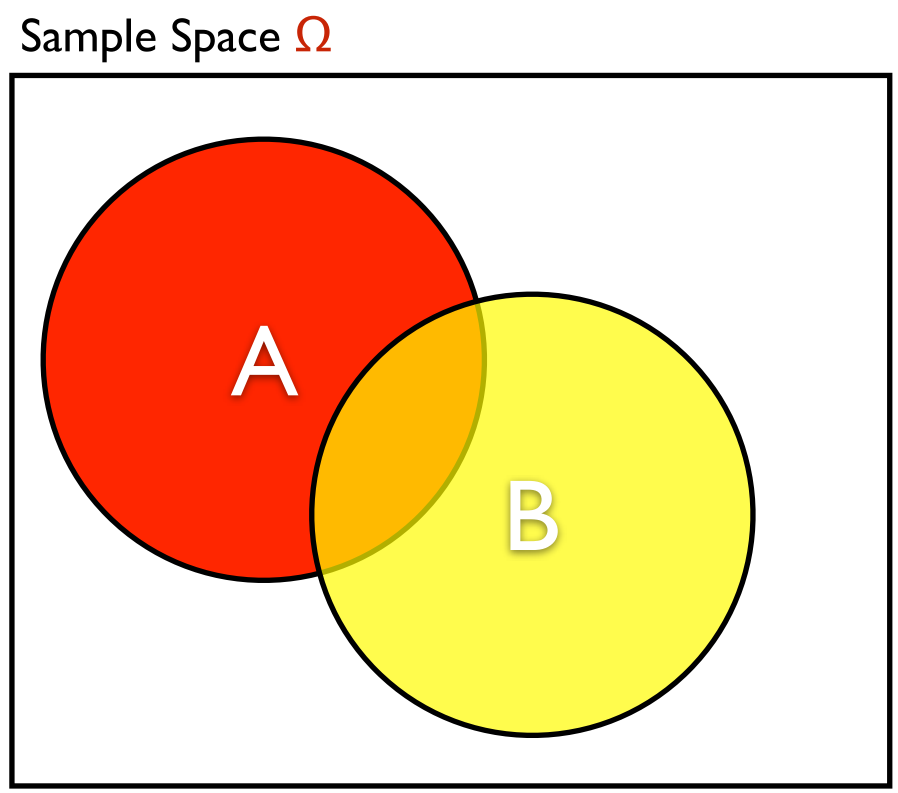
If A $\cap B = \varnothing$ then $P(A \cap B) = 0$ (other direction not quite - nonempty sets can have probability zero)

If A $\subset B$ then $P(A \cap B) = P(A)$ same as: $x \in A$ implies $x \in B$



Independence: Definition

$$P(A \cap B) = P(A) \times P(B) = P(A)P(B)$$



Example of Independence

Suppose we have a biased coin $P(H)=0.9, P(T)=0.1$

We flip the coin 10 times. What is the probability of

H,T,H,H,H,T,H,H,H,H ?

The coin flips define independent events, therefor

$P(H,T,H,H,H,T,H,H,H,H)=$

$=0.9\times 0.1\times 0.9\times 0.9\times 0.9\times 0.1\times 0.9\times 0.9\times 0.9\times 0.9$

$0.9^8\times 0.1^2$

Binomial distribution for biased coins

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i$$

Fair Coin

Suppose $a = b = \frac{1}{2}$ then we get:

$$1 = \left(\frac{1}{2}+\frac{1}{2}\right)^n = \sum_{i=0}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{n-i} = \left(\frac{1}{2}\right)^n \sum_{i=0}^{n} \binom{n}{i}$$

Biased Coin

$p =$ probability of heads

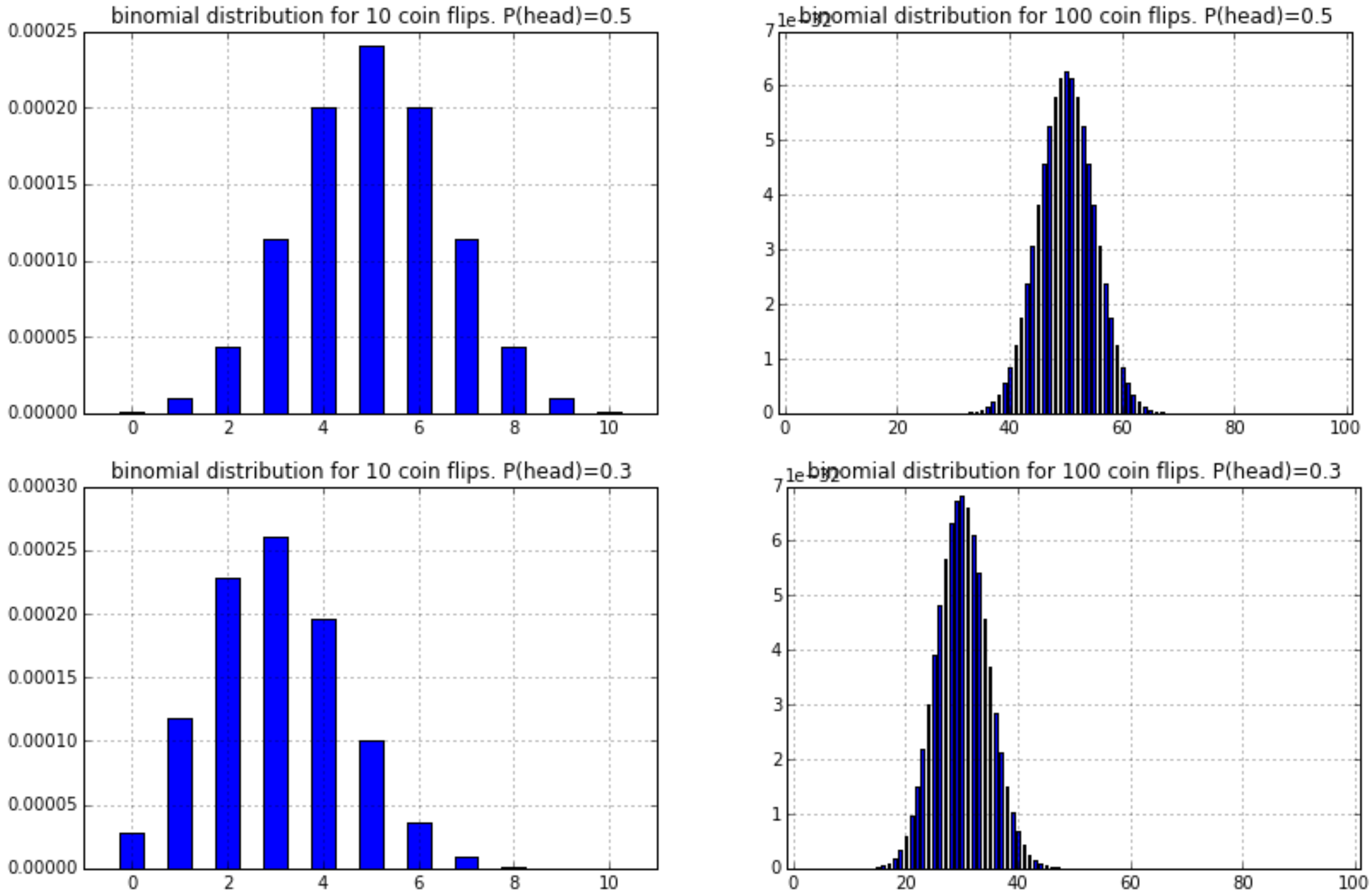
$q = 1 - p =$ probability of tails

$$1 = (p+q)^n = \sum_{m=0}^{n} \binom{n}{m} p^m q^{n-m}$$

Probability of a sequence of length n having m heads and n - m

General Binomial Distributions

n=10 n=100



p=0.5

p=0.3

Dependence: Example

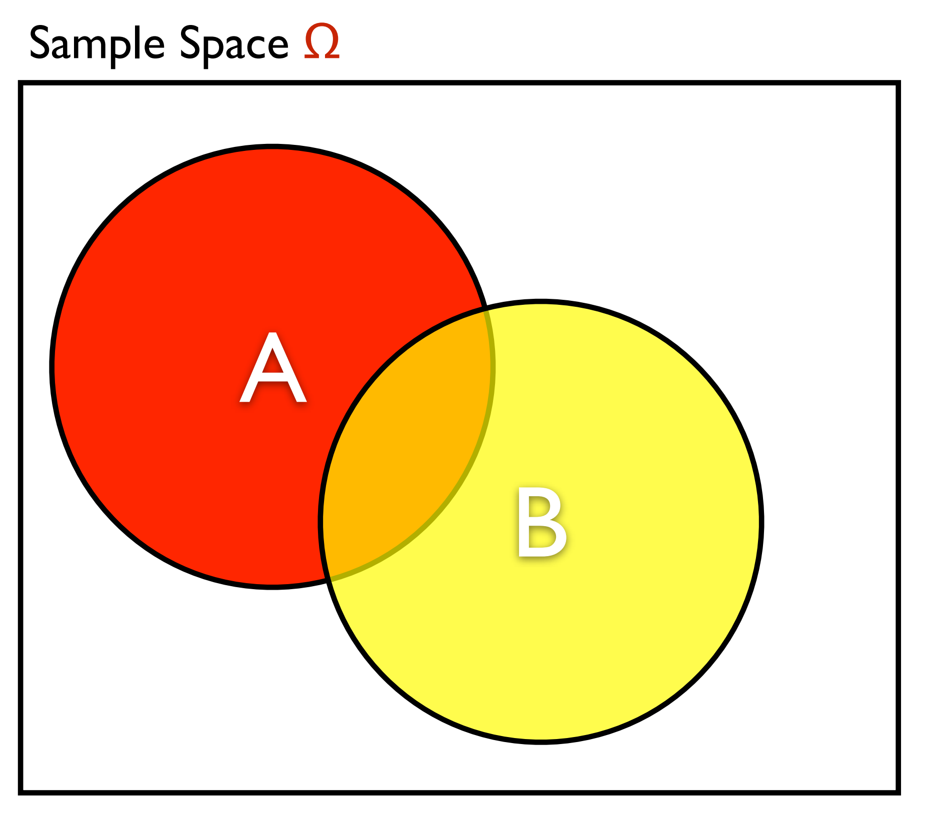
- Sample Space: all 22 year old men.
- Set A: men playing in the NBA
- Set B: men taller than 6' 5"
- $P(A \cap B) = P(\text{NBA players that are at least 6'5"}) \gg P(A)P(B)$

Quantifying dependence: Conditional probability

The probability of A given B is:

$$P(A \mid B) \doteq \frac{P(A \cap B)}{P(B)}$$

B replaces Ω



Dependence works both ways.

- If A depends on B then B depends on A
- Dependence/independence is **not** a directional relationship.
- Conditioning A on B rather than B on A is a decision of the observer - it is not inherent to the observed system.

Expressing independence using conditional prob.

Definition of independence: $P(A \cap B) = P(A)P(B)$

Implication 1: $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(A)P(B)}{P(B)} = P(A)$

Implication 2: $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{P(A)P(B)}{P(A)} = P(B)$

- Interpretation: If A and B are independent, knowing that A happened does not change the probability of B and vice versa

- Note: we are speaking of states of knowledge, NOT of causality.

Conditional Prob. in Poker

- Suppose your private cards (hole) are Ace ♡,2♤
- Suppose the community cards are 3♤,4♤,5♤,K♧,J♡
- What is the probability that a particular opponent has a Straight Flush? (All cards of the same suit and sequential)
- Pr(opponent's cards complete a straight flush
  | hole=(Ace ♡,2♤) and community=(3♤,4♤,5♤,K♧,J♡))
- There are three combination the 2 cards of the opponent that would give her a straight flush: (1♤,2♤),(2♤,6♤),(6♤,7♤). However the first two pairs are impossible because you hold the 2♤.
- The conditional probability of a straight flush is therefor:
  Pr(opponent's cards complete a straight flush
  | hole=(Ace ♡,2♤) and community=(3♤,4♤,5♤,K♧,J♡))=1/C(52-7,2)

independence in tabular form

	marginal on b	A	A^c
marginal on A		4/12 =1/3	8/12 =2/3
B	3/12 =1/4	1/12	2/12
B^c	9/12 =3/4	3/12	6/12

$$P(A \cap B) = 1/12$$

$$P(A) = 1/3$$

$$P(B) = 1/4$$

$$P(A \cap B) = P(A)P(B)$$

- A and B are independent
- Are A and B^c Independent?

Are A and B^c Independent?

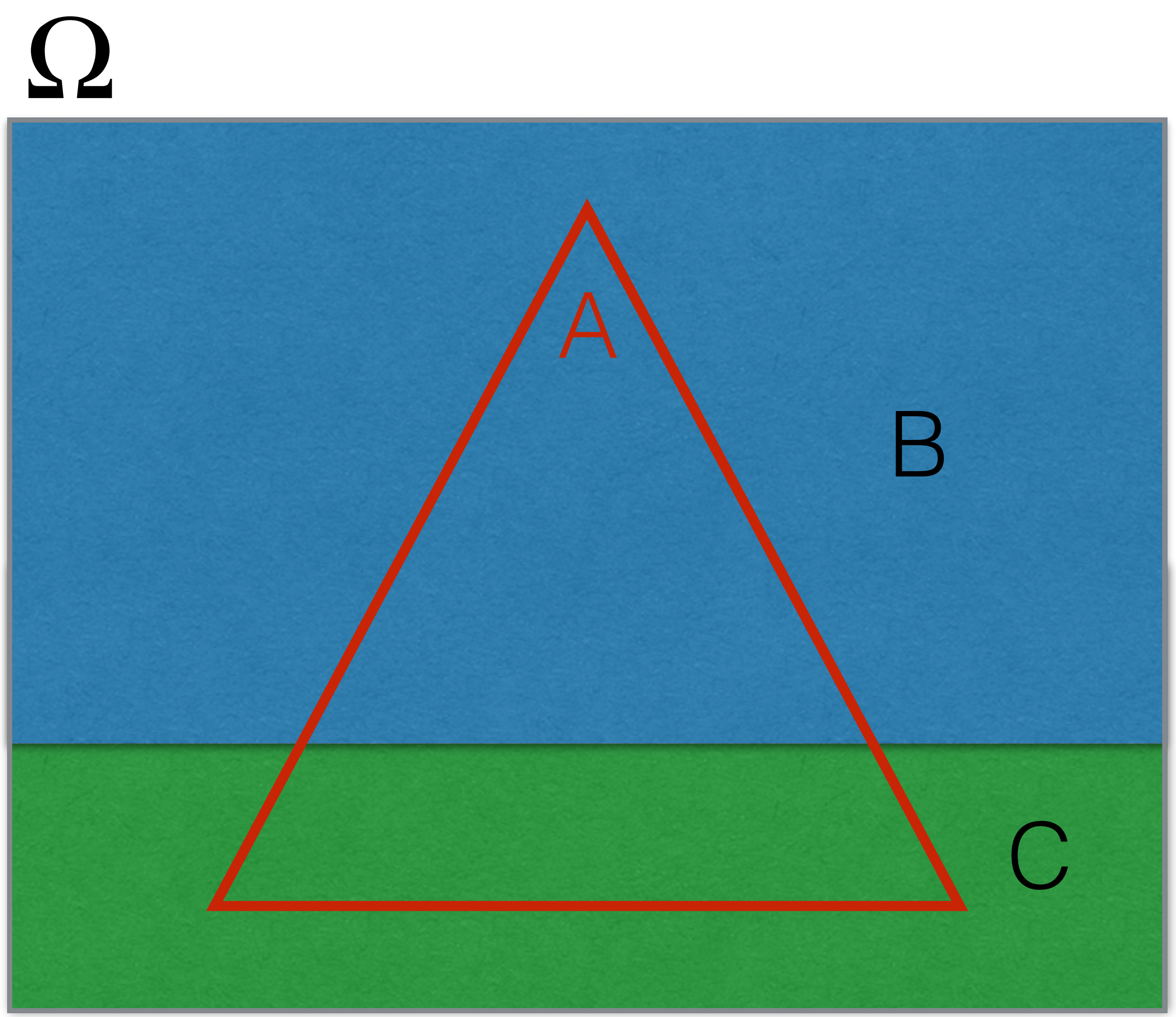
- Yes, because

$$P(A \cap B^c) = P(A) - P(A \cap B) =$$

$$= P(A) - P(A)P(B) = P(A)\big(1 - P(B)\big) = P(A)P(B^c)$$

- Same for: A^c and B , A^c and B^c

Conditional summation rule



Standard summation rule

$$P(B) + P(C) = 1$$

Conditional summation rule

$$P(B \mid A) + P(C \mid A) = 1$$

Because:

$$P(B, A) + P(C, A) = P(A)$$

Dividing all by $P(A)$ we get:

$$\frac{P(B, A)}{P(A)} + \frac{P(C, A)}{P(A)} = 1$$

pair-wise vs complete independence

- Suppose we have three coins.
- We flip two of the coins to get HH,HT,TH,TT
- We chose the side of the third coin so that the number of Heads is even. (i.e. H if HT or TH. T otherwise)
- We have therefor a uniform distribution over 4 possibilities: HHT,HTH,THH,TTT.
- Focusing on any pair of coins, the outcomes of the coins are independent (Check!)
- However, consider the combination HHH
    - $P(HHH)=0$, while $P(H)P(H)P(H)=1/8$
- Even though each pair of coins are independent, the set of 3 coins are not independent!

Bayesian Inference

Bayesian Inference

- Who crashed the car?
- Suppose the only possible drivers are Rob or Sarah.
- We know the following probabilities.
    - P(R rob drove)=10%, P(C crash | R rob drove)=50%
    - P(S Sarah drove)=90%,   P(C crash | S sarah drove)=1%

$$P(C) = P(C,R) + P(C,S) = P(C \mid R)P(R) + P(C \mid S)P(S)$$

$$P(R \mid C) = \frac{P(R,C)}{P(C)} = \frac{P(C \mid R)P(R)}{P(C)} = \frac{P(C \mid R)P(R)}{P(C \mid R)P(R) + P(C \mid S)P(S)}$$

$$= \frac{0.1 \times 0.5}{0.1 \times 0.5 + 0.9 \times 0.01} = 0.85 \qquad P(S \mid C) = 0.15$$

- Prior probabilities: Sarah drives the car 90% of the time, Rob 10% of the time.
- Posterior probabilities: given that there was an accident, the probability that the driver was Rob jumps to 85% because he is much more accident prone.

The following example is taken from *Probabilistic Reasoning in Intelligent Systems* by Judea Pearl:

> You wake up in the middle of the night to the shrill sound of your burglar alarm. What is the chance that a burglary attempt has taken place?

The relevant facts are:

- There is a 95% chance that an attempted burglary attempt will trigger the alarm. That is,

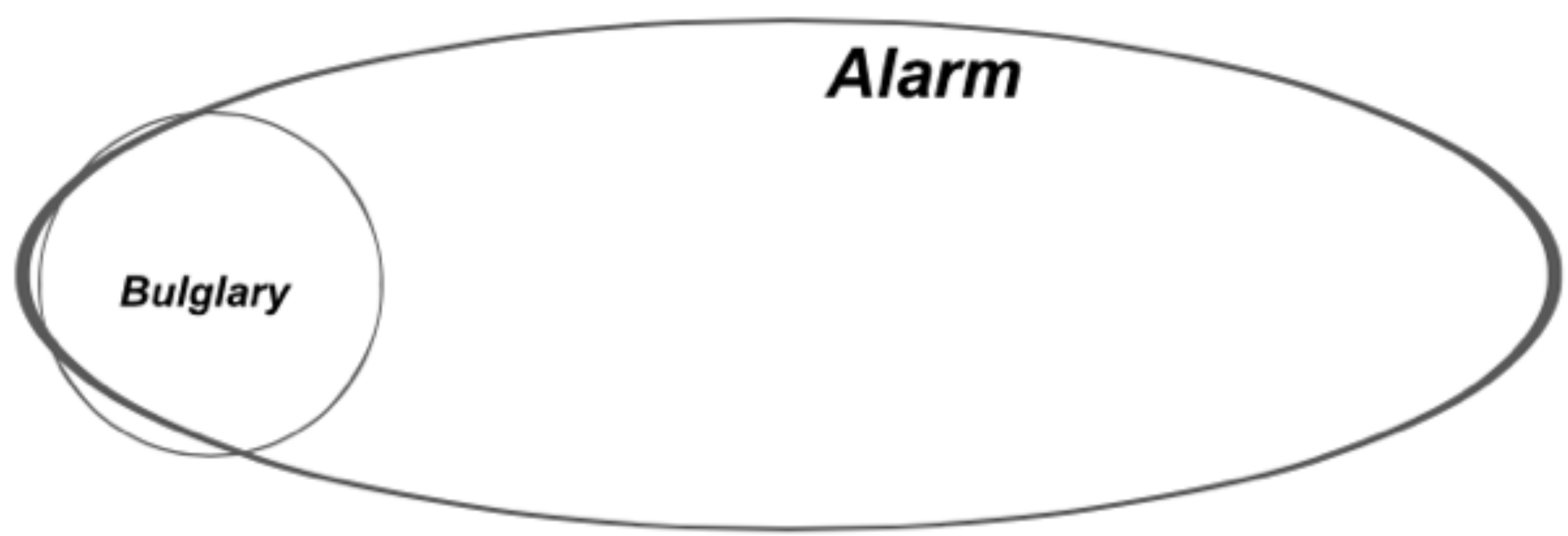
$$\Pr(\text{alarm}|\text{burglary}) = 0.95.$$

- There is a 1% chance of a false alarm.

$$\Pr(\text{alarm}|\text{no burglary}) = 0.01.$$

- Based on local crime statistics, there is a one-in-10,000 chance that a house will be burglarized on a given night.

$$\Pr(\text{burglary}) = 10^{-4}.$$



We are interested in the chance of a burglary given that the alarm has sounded. We can use the conditional probability formula for this:

$$\Pr(\text{burglary}|\text{alarm}) = \frac{\Pr(\text{burglary, alarm})}{\Pr(\text{alarm})} = \frac{\Pr(\text{alarm}|\text{burglary})\Pr(\text{burglary})}{\Pr(\text{alarm})}.$$

The one term we don't immediately know is Pr(alarm). By the summation rule,

$$\Pr(\text{alarm}) = \Pr(\text{alarm}|\text{burglary})\Pr(\text{burglary}) + \Pr(\text{alarm}|\text{no burglary})\Pr(\text{no burglary}).$$

Putting it all together,

$$\Pr(\text{burglary}|\text{alarm}) = \frac{0.95 \times 10^{-4}}{0.95 \times 10^{-4} + 0.01 \times (1 - 10^{-4})} = 0.00941,$$

about 0.94%. Thus our belief in a burglary has risen approximately a hundredfold from its default value of 10^{-4}, on account of the alarm.

It is frequently the case, as in this example, that we wish to update the chances of an event H based on new evidence E. In other words, we wish to know $\Pr(H|E)$. The derivation above implicitly uses the following formula, called **Bayes' rule**:

$$\Pr(H|E) = \frac{\Pr(E|H)\Pr(H)}{\Pr(E)}.$$

To calculate Pr(E) we use the summation rule

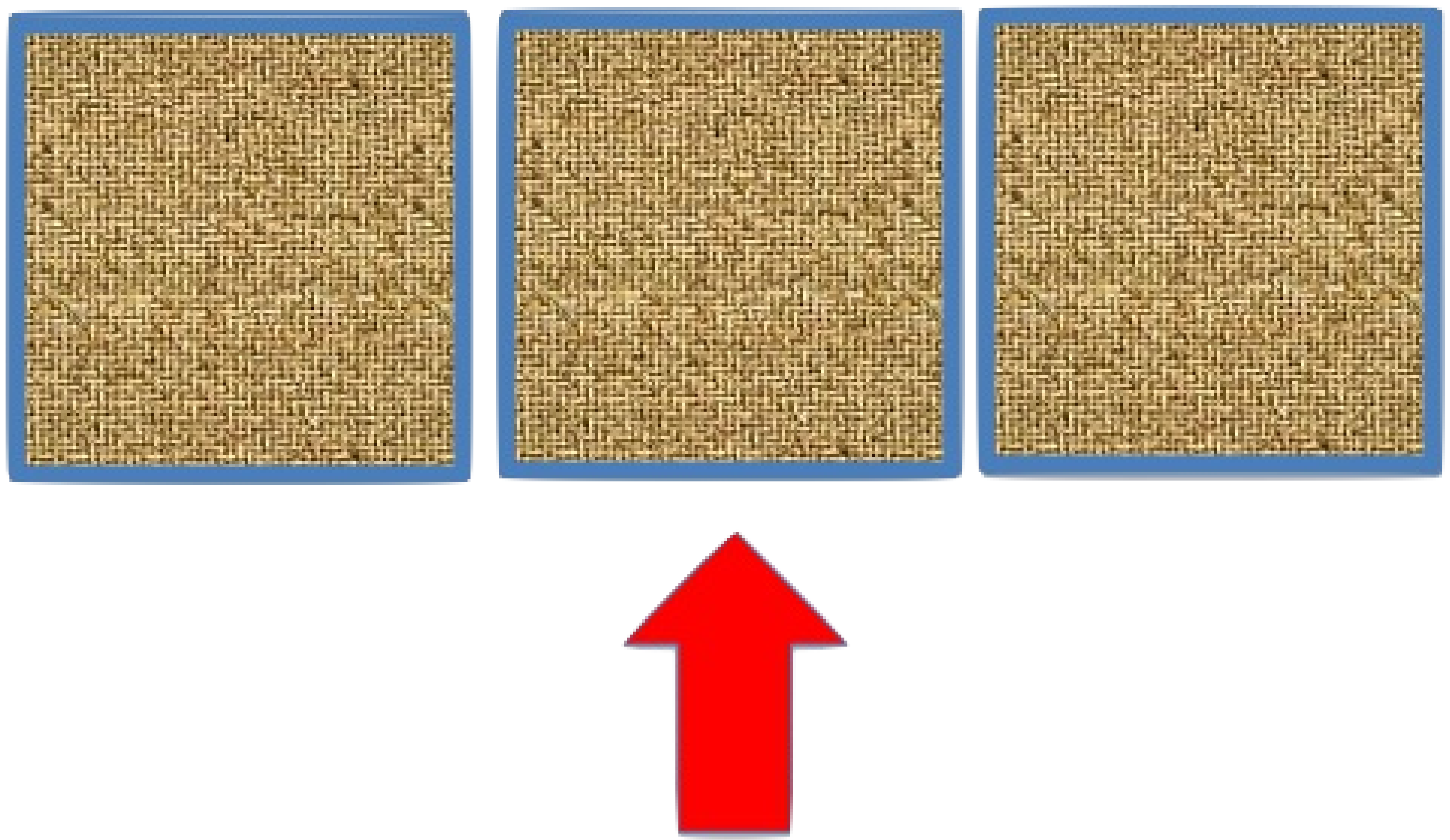
$$\Pr(H \mid E) = \frac{\Pr(E \mid H)\Pr(H)}{\Pr(E)} = \frac{\Pr(E \mid H)\Pr(H)}{\Pr(E \mid H)\Pr(H) + \Pr(E \mid H^c)\Pr(H^c)}$$

To calculate Pr(E) we use the summation rule

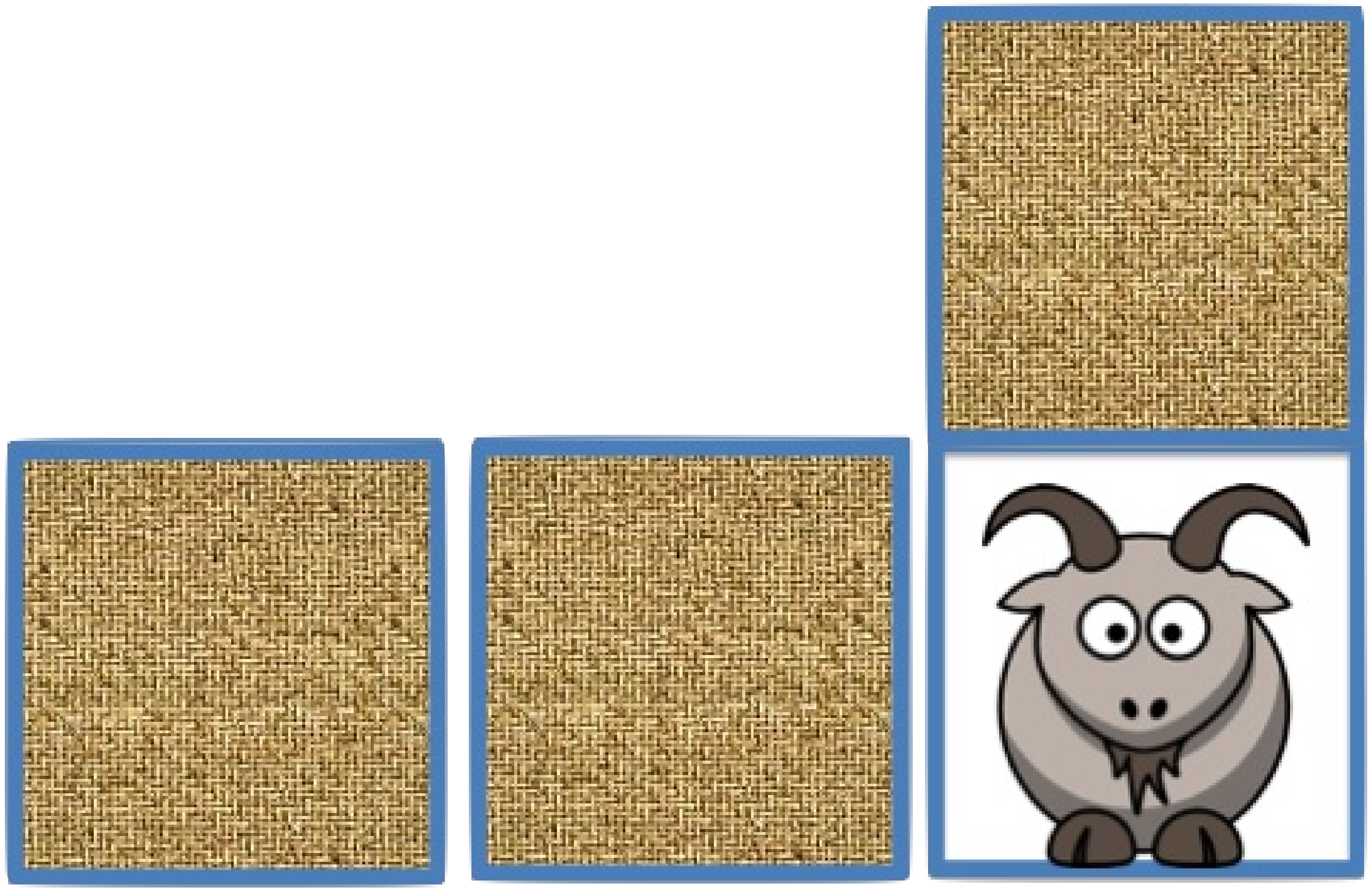
$$\Pr(H \mid E) = \frac{\Pr(E \mid H)\Pr(H)}{\Pr(E)} = \frac{\Pr(E \mid H)\Pr(H)}{\Pr(E \mid H)\Pr(H) + \Pr(E \mid H^c)\Pr(H^c)}$$

The Monty Hall Puzzle

- Monty Hall was a variety show on TV.
- In one of the games there are three doors, one hiding a treasure, two hiding goats.
- Your goal is to select the door with the treasure.



I am betting on this door



I am betting on this door

Monty opens this door

I am allowed to switch, should I?

Argument that it does not matter:

The chance that the treasure is behind each of the doors 50%.

As the probabilities are equal, it does not matter whether we switch or not.

Argument for choosing one of the two unopen doors at random.

Before I had to choose between 3 doors - my probability of sucsses was 1/3

Now I am choosing between between two doors, my probability of success is 1/2
So random is better than staying on the same door.

Argument for Switching.

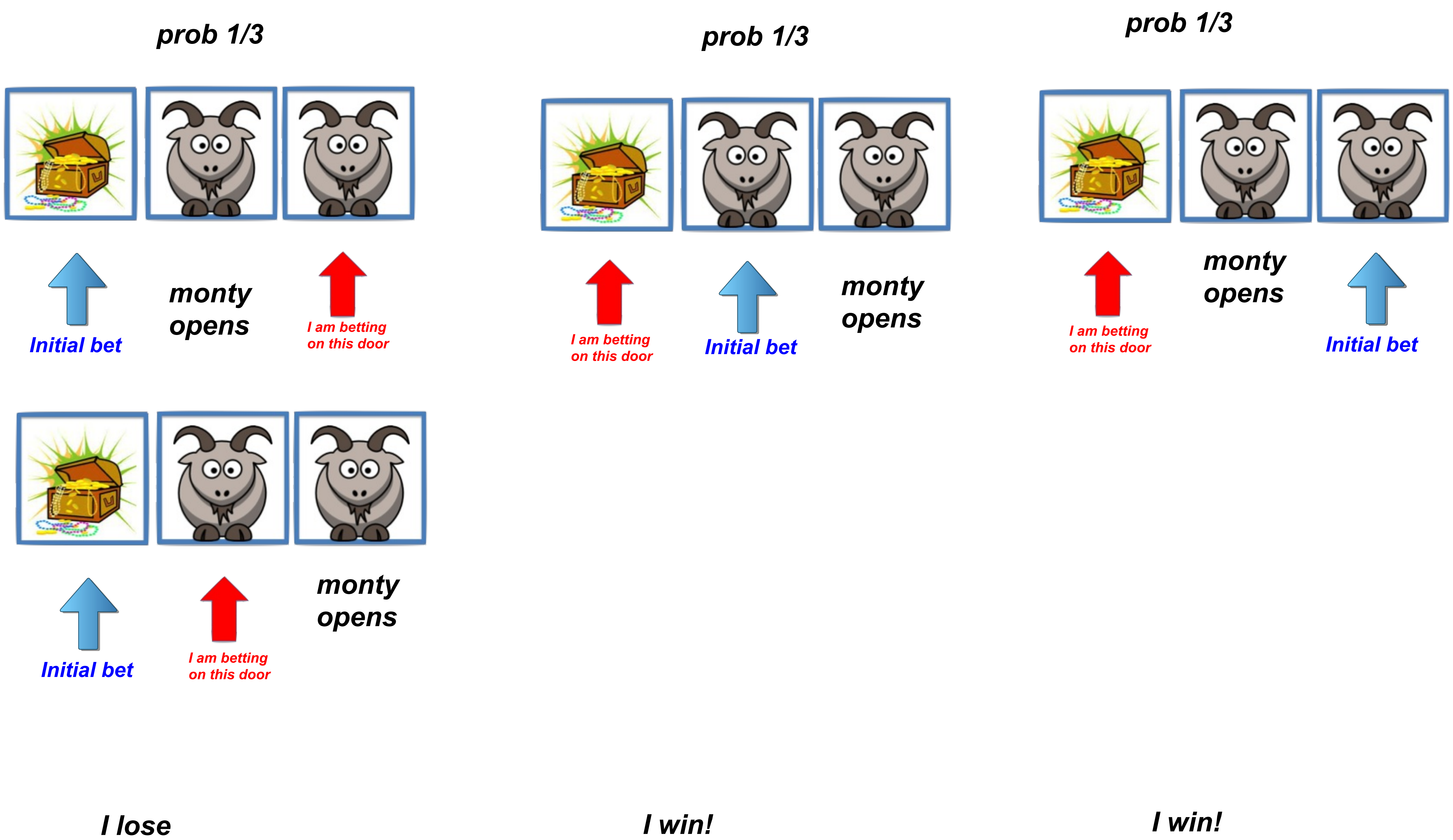
The probability that the treasure is behind the door I chose did not change.
Therefor the probability that switching will put me on the treasure must be 2/3:

$1/2*1/3+1/2*2/3 = 1/2$

Arguments against switching:

I know already that one of the other doors has a goat behind it. So getting the information does not tell me anything new.

Analysis for always switching



Hidden Assumption: ***monty always opens a door to reveal a goat.***

In fact, he might have his own goals:

If Monty wants us to lose: ***open door only when we choose the treasure door.***

If Monty wants us to win: ***open door only when we choose a goat door.***

For us the only SAFE thing to do is not to switch.

This is called the "Min-Max" strategy.

Min-Max is the strategy the guarantees us the best outcome in the worst case.

More on that - game theory.