2 - Probability on Finite Sets

- Modeling physical phenomena
- Probability on finite sets
- The sample space and the pmf
- Events
- Unions, intersections and complements
- The axioms of probability
- Partitions
- Conditional probability
- Independence
- Dependent events

Modeling Physical Phenomena

- The deterministic way: write *differential equations*

  Differential equations are constructed from

  - physical principles
  - experiments

  Use the model to *predict the approximate behavior* of the system

- The probabilistic way: specify the probability of events

  Probabilities are derived from

  - physical principles, e.g., time symmetry for noise
  - experiments; i.e., observed frequency of events

  Again, we use the model to *predict the approximate behavior* of the system

  e.g., in approximately 95% of trials the hovercraft will deviate less than 3m from the given trajectory

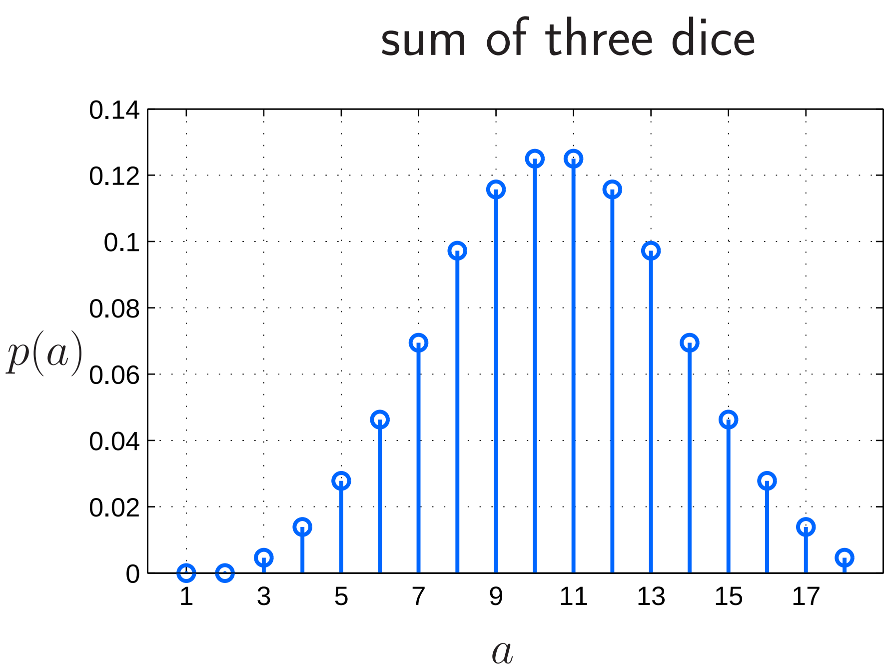
Probability on Finite Sets

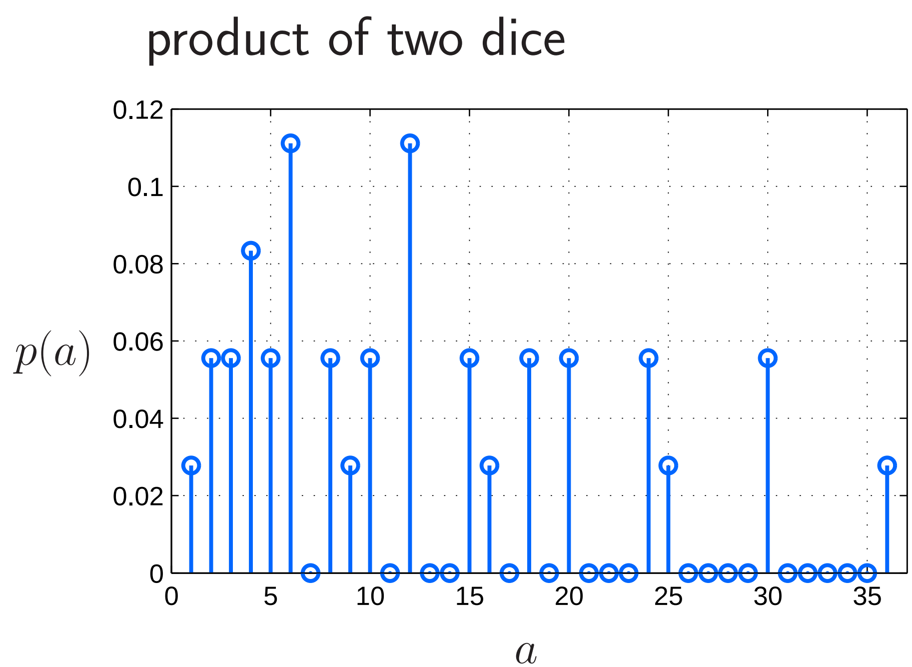
- The *sample space* is a finite set Ω; it's elements are called *outcomes*. Exactly one outcome occurs in every experiment.

- Function $p : \Omega \to [0,1]$ is called a *probability mass function (pmf)* if

$$p(a) \geq 0 \text{ for all } a \in \Omega \qquad \text{and} \qquad \sum_{a \in \Omega} p(a) = 1$$

Then $p(a)$ is the probability that outcome $a \in \Omega$ occurs





Events

An *event* is a subset of Ω

For example, if $\Omega = \{1, \dots, 2n\}$, the following are events

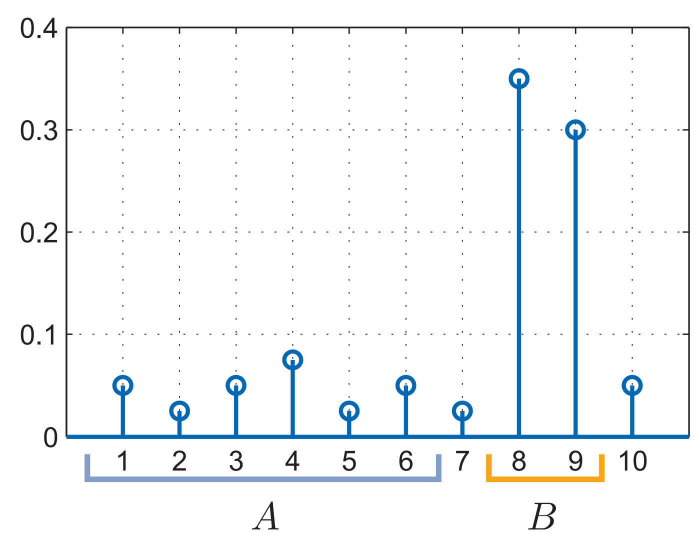
- $A = \{\, 2, 4, 6, \dots, 2n \,\}$, which we would call *the event that the outcome is even*
- $A = \{\, x \in \Omega \mid x \geq 32 \,\}$, which we would call *the event that the outcome is* ≥ 32

The *probability of an event* is

$$\mathbf{Prob}(A) = \sum_{b \in A} p(b)$$

$\mathbf{Prob} : 2^{\Omega} \to [0, 1]$ is called a *probability measure*

Example: $\mathbf{Prob}(A) = 0.275$, $\mathbf{Prob}(B) = 0.65$



Unions, Intersections and Complements

For any sets $A, B \subset \Omega$ we have

$$\boxed{\mathbf{Prob}(A \cup B) = \mathbf{Prob}(A) + \mathbf{Prob}(B) - \mathbf{Prob}(A \cap B)}$$

We interpret

$A \cup B$	is the event that A or B happens
$A \cap B$	is the event that A and B happens
$A^c = \{\, b \in \Omega \mid b \notin A \,\}$	is the event that A does not happen

Notation

Notice that $\mathbf{Prob}$ really depends on

- the sample space Ω
- and the probability mass function p

Sometimes we will write

$$\mathop{\mathbf{Prob}}_{\Omega,\, p}(A)$$

to specify which Ω and p are being used

Axioms of Probability

We have for all $A, B \subset \Omega$

(i) $\mathbf{Prob}(A) \geq 0$

(ii) $\mathbf{Prob}(\Omega) = 1$

(iii) if $A \cap B = \emptyset$ then $\mathbf{Prob}(A \cup B) = \mathbf{Prob}(A) + \mathbf{Prob}(B)$

- The above three conditions are called the *axioms of probability* for finite sets Ω
- If $\mathbf{Prob} : 2^\Omega \to \mathbb{R}$ satisfies the above, then we can construct a probability mass function via

$$p(b) = \mathbf{Prob}\big(\{b\}\big) \qquad \text{for all } b \in \Omega$$

and p will be positive and sum to one as required.

Partitions

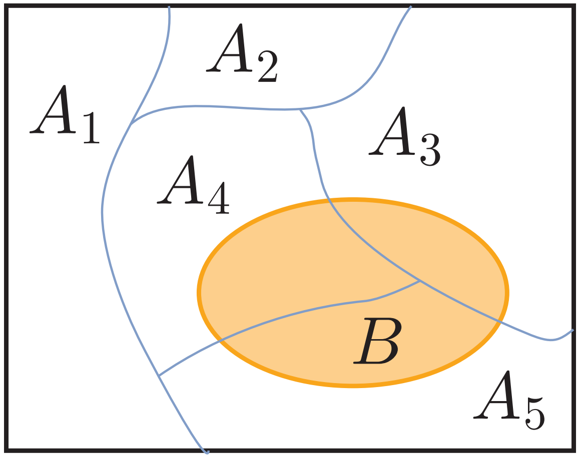
The set of events $A_1, A_2, \ldots, A_n$ is called a *partition* of Ω if

$A_i \cap A_j = \emptyset$	for all $i \neq j$	called *mutually exclusive*
$A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$		called *collectively exhaustive*

Then for any $B \subset \Omega$ we have

$$\mathbf{Prob}(B) = \sum_{i=1}^{n} \mathbf{Prob}(B \cap A_i)$$

called the *Law of Total Probability*

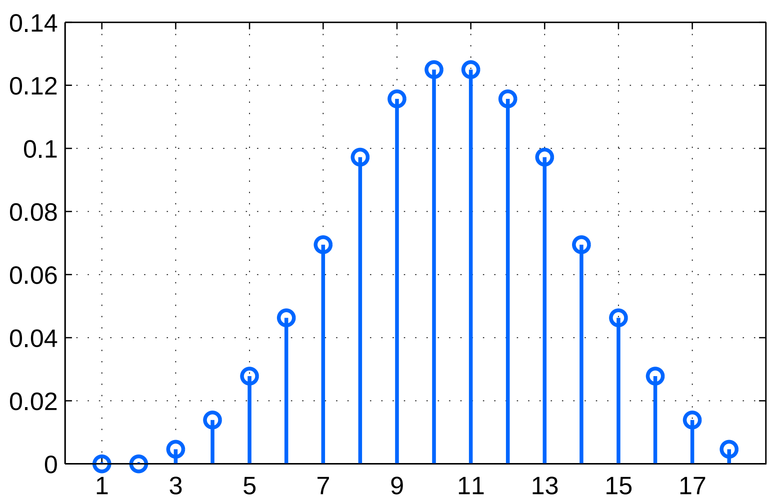


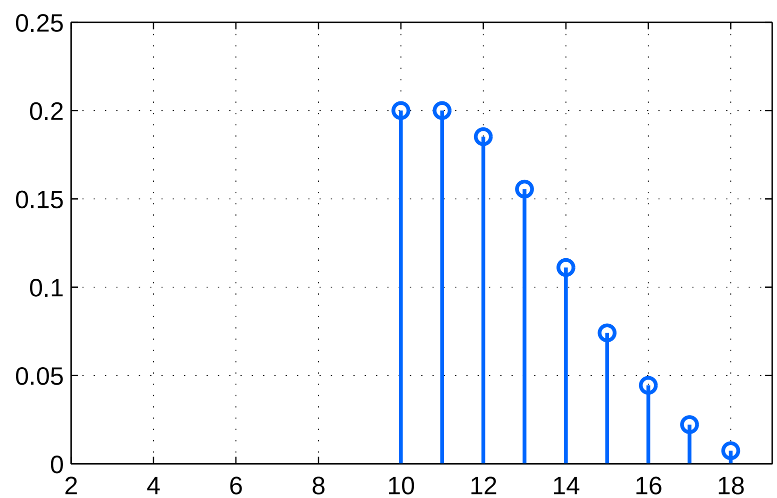
Conditional Probability

Suppose A and B are events, and $\mathbf{Prob}(B) \neq 0$. Define the *conditional probability of A given B* by

$$\boxed{\mathbf{Prob}(A \mid B) = \frac{\mathbf{Prob}(A \cap B)}{\mathbf{Prob}(B)}}$$

Example: suppose $B = \{\, x \in \Omega \mid x \geq 10 \}$





If we perform many repeated experiments, and throw away all $x \notin B$, then the observed frequency of outcomes $x \in B$ will increase.

Conditional Probability

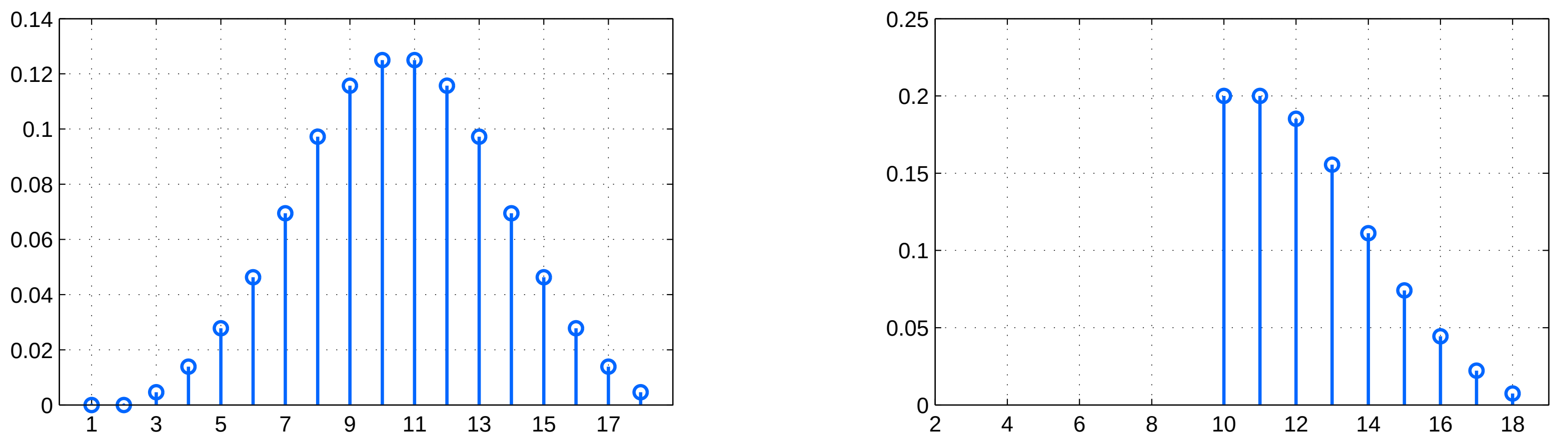
Conditioning defines a new probability mass function on Ω.

The *conditional pmf* is

$$p_2(a) = \begin{cases} \dfrac{p(a)}{\mathbf{Prob}(B)} & \text{if } a \in B \\ 0 & \text{otherwise} \end{cases}$$

Then we have, for any $A \subset \Omega$

$$\mathop{\mathbf{Prob}}_{\Omega,p}(A \mid B) = \mathop{\mathbf{Prob}}_{\Omega,p_2}(A)$$



Independence

Two events A and B are called *independent* if

$$\boxed{\mathbf{Prob}(A \cap B) = \mathbf{Prob}(A)\,\mathbf{Prob}(B)}$$

- If $\mathbf{Prob}(B) \neq 0$ this is equivalent to

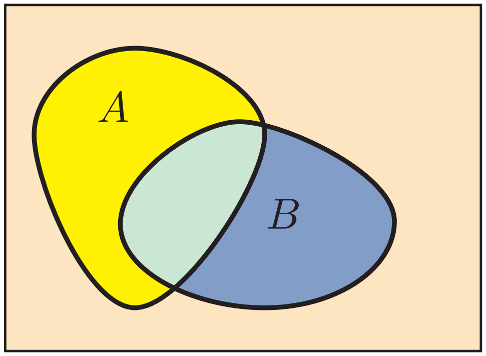
$$\mathbf{Prob}(A \mid B) = \mathbf{Prob}(A)$$

- if A and B are *dependent*, then knowing whether event A occurs also gives information regarding event B

Independence

Events A and B are independent if and only if $\mathbf{rank}(M) = 1$ where

$$M = \begin{bmatrix} \mathbf{Prob}(A \cap B) & \mathbf{Prob}(A \cap B^c) \\ \mathbf{Prob}(A^c \cap B) & \mathbf{Prob}(A^c \cap B^c) \end{bmatrix}$$



M is called the *joint probability matrix*.

- A and B are independent means the probabilities of A occurring do not change when we discard those outcomes when B occurs.

- The probabilities of A and A^c occurring are the row sums

$$\begin{bmatrix} \mathbf{Prob}(A) \\ \mathbf{Prob}(A^c) \end{bmatrix} = M\mathbf{1}$$

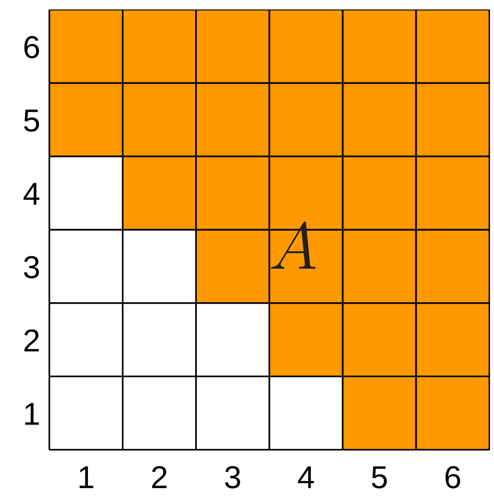
When $\mathbf{rank}(M) = 1$, each column is some multiple of $M\mathbf{1}$

$$M = \begin{bmatrix} \mathbf{Prob}(A) \\ \mathbf{Prob}(A^c) \end{bmatrix} \begin{bmatrix} \mathbf{Prob}(B) & \mathbf{Prob}(B^c) \end{bmatrix}$$

Example: two dice

Two dice. Pick sample space

$$\Omega = \{ (\omega_1, \omega_2) \mid \omega_i \in \{1, 2, \ldots, 6\} \}$$



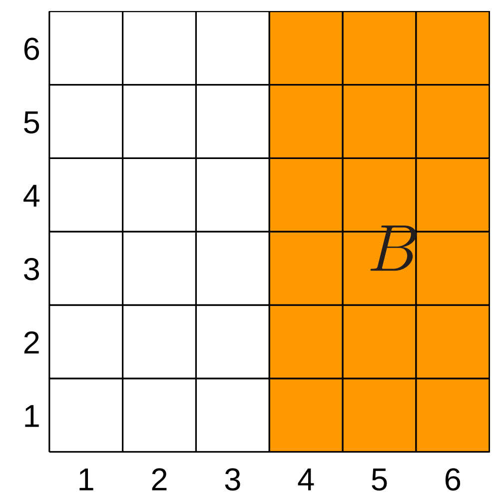
Two events are

- the sum is greater than 5

$$A = \{ \omega \in \Omega \mid \omega_1 + \omega_2 > 5 \}$$

- the first dice is greater than 3

$$B = \{ \omega \in \Omega \mid \omega_1 > 3 \}$$

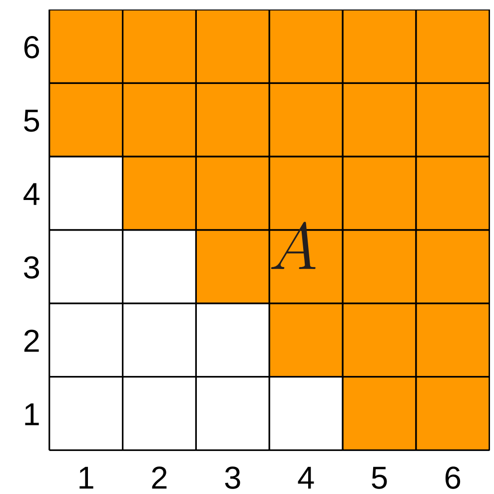


Example: two dice

By measuring B, we have information about A, because

$$\mathbf{Prob}(A) = \frac{26}{36}$$

$$\mathbf{Prob}(A \mid B) = \frac{17}{18}$$



- This is an example of *estimation*
- By measuring one random quantity, we have information about another
- More refined questions: what is the conditional distribution of the sum? What should we pick as an estimate?
- Later we will see problems of the form

$$y = Ax + w$$

w is random, we measure y, and would like to know x