Review of Background Material on Discrete Mathematics and Probability

Samson Abramsky

Set theory notations

We shall assume basic notations of elementary set theory:

set membership	$x \in S$
set inclusion	$S \subseteq T$
union	$S \cup T$, $\bigcup_{i \in I} S_i$
intersection	$S \cap T$, $\bigcap_{i \in I} S_i$
powerset	$\mathscr{P}(S)$
relative complement	S^c
set difference	$S \setminus T$

Ordered Pairs and Relations

We take as fundamental the notion of **ordered pair** (x, y). (This can be defined in set theory, but we shall not need this.)

Ordered Pairs and Relations

We take as fundamental the notion of **ordered pair** (x, y). (This can be defined in set theory, but we shall not need this.)

Ordered pairs have well-defined first and second components, namely x and y in the ordered pair (x, y). Two ordered pairs are equal if they have the same first and second components.

Ordered Pairs and Relations

We take as fundamental the notion of **ordered pair** (x, y). (This can be defined in set theory, but we shall not need this.)

Ordered pairs have well-defined first and second components, namely x and y in the ordered pair (x, y). Two ordered pairs are equal if they have the same first and second components.

The cartesian product of sets X, Y is defined as

$$X \times Y \ := \ \{(x, y) \mid x \in X \ \wedge \ y \in Y\}.$$

Ordered Pairs and Relations

We take as fundamental the notion of **ordered pair** (x, y). (This can be defined in set theory, but we shall not need this.)

Ordered pairs have well-defined first and second components, namely x and y in the ordered pair (x, y). Two ordered pairs are equal if they have the same first and second components.

The cartesian product of sets X, Y is defined as

$$X \times Y \;:=\; \{(x, y) \mid x \in X \;\wedge\; y \in Y\}.$$

A **relation** from a set X to a set Y is a subset $R \subseteq X \times Y$. We shall use the notation $R : X \multimap Y$ to indicate that R is a relation from X to Y.

Composition of Relations

Composition of Relations

Given relations $R : X \multimap Y$ and $S : Y \multimap Z$, we can compose them: $R; S : X \multimap Z$ is defined by

$$R; S \;:=\; \{(x, z) \mid \exists y.\, (x, y) \in R \;\wedge\; (y, z) \in S\}.$$

Composition of Relations

Given relations $R : X \multimap Y$ and $S : Y \multimap Z$, we can compose them:
$R; S : X \multimap Z$ is defined by

$$R; S \;:=\; \{(x, z) \mid \exists y.\, (x, y) \in R \;\wedge\; (y, z) \in S\}.$$

Proposition

Composition of relations is associative, and has the diagonal relations as identities. Given $R : X \multimap Y$, $S : Y \multimap Z$, $T : Z \multimap W$:

$$(R; S); T \;=\; R; (S; T)$$

and

$$\mathrm{id}_X; R \;=\; R \;=\; R; \mathrm{id}_Y$$

where $\mathrm{id}_X := \{(x, x) \mid x \in X\}$.

Converse

Converse

Given a relation $R : X \multimap Y$, we define the **converse relation** $R^{\cup} : Y \multimap X$ by $R^{\cup} := \{(y, x) \mid (x, y) \in R\}$.

Converse

Given a relation $R : X \multimap Y$, we define the **converse relation** $R^{\cup} : Y \multimap X$ by $R^{\cup} := \{(y, x) \mid (x, y) \in R\}$.

Proposition

We have $R^{\cup\cup} = R$, *and* $(R; S)^{\cup} = S^{\cup}; R^{\cup}$.

Functions

Functions

A **function** f from X to Y is a relation from X to Y with the following additional properties:

Single-valuedness: $(x, y) \in f$ and $(x, z) \in f$ implies that $y = z$.

Totality: for all $x \in X$, for some $y \in Y$, $(x, y) \in f$.

Functions

A **function** f from X to Y is a relation from X to Y with the following additional properties:

Single-valuedness:	$(x, y) \in f$ and $(x, z) \in f$ implies that $y = z$.
Totality:	for all $x \in X$, for some $y \in Y$, $(x, y) \in f$.

We write $f : X \to Y$ to indicate that f is a function from X to Y, and $f(x) = y$ or $f : x \mapsto y$ to indicate that y is the unique element of Y such that $(x, y) \in f$.

Functions

A **function** f from X to Y is a relation from X to Y with the following additional properties:

Single-valuedness:	$(x, y) \in f$ and $(x, z) \in f$ implies that $y = z$.
Totality:	for all $x \in X$, for some $y \in Y$, $(x, y) \in f$.

We write $f : X \to Y$ to indicate that f is a function from X to Y, and $f(x) = y$ or $f : x \mapsto y$ to indicate that y is the unique element of Y such that $(x, y) \in f$.

We refer to X as the **domain** of f, and Y as the **codomain**.

Functions

A **function** f from X to Y is a relation from X to Y with the following additional properties:

Single-valuedness:	$(x, y) \in f$ and $(x, z) \in f$ implies that $y = z$.
Totality:	for all $x \in X$, for some $y \in Y$, $(x, y) \in f$.

We write $f : X \to Y$ to indicate that f is a function from X to Y, and $f(x) = y$ or $f : x \mapsto y$ to indicate that y is the unique element of Y such that $(x, y) \in f$.

We refer to X as the **domain** of f, and Y as the **codomain**.

We write $Y^X := \{f \mid f : X \to Y\}$ for the set of all functions from X to Y.

Functions

A **function** f from X to Y is a relation from X to Y with the following additional properties:

Single-valuedness:	$(x, y) \in f$ and $(x, z) \in f$ implies that $y = z$.
Totality:	for all $x \in X$, for some $y \in Y$, $(x, y) \in f$.

We write $f : X \to Y$ to indicate that f is a function from X to Y, and $f(x) = y$ or $f : x \mapsto y$ to indicate that y is the unique element of Y such that $(x, y) \in f$.

We refer to X as the **domain** of f, and Y as the **codomain**.

We write $Y^X := \{f \mid f : X \to Y\}$ for the set of all functions from X to Y.

Given functions $f : X \to Y$ and $g : Y \to Z$, we can compose them:

$$g \circ f : X \to Z, \qquad g \circ f(x) := g(f(x)).$$

This agrees with the composition of g and f as relations.

Injective and Surjective

Injective and Surjective

A function $f : X \to Y$ is **injective** if for all $x, x' \in X$, $f(x) = f(x')$ implies $x = x'$.

Injective and Surjective

A function $f : X \to Y$ is **injective** if for all $x, x' \in X$, $f(x) = f(x')$ implies $x = x'$.

It is **surjective** if for all $y \in Y$, for some $x \in X$, $f(x) = y$.

Injective and Surjective

A function $f : X \to Y$ is **injective** if for all $x, x' \in X$, $f(x) = f(x')$ implies $x = x'$.

It is **surjective** if for all $y \in Y$, for some $x \in X$, $f(x) = y$.

It is **bijective** if there is a function $g : Y \to X$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. We write $X \cong Y$ if there is a bijection $f : X \to Y$.

Injective and Surjective

A function $f : X \to Y$ is **injective** if for all $x, x' \in X$, $f(x) = f(x')$ implies $x = x'$.

It is **surjective** if for all $y \in Y$, for some $x \in X$, $f(x) = y$.

It is **bijective** if there is a function $g : Y \to X$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. We write $X \cong Y$ if there is a bijection $f : X \to Y$.

Exercises

1. Show that a function is bijective if and only if it is injective and surjective.
2. Show that $\mathscr{P}(X) \cong \mathbf{2}^X$, where $\mathbf{2} := \{0, 1\}$.
3. Show that a relation $R : X \multimap Y$ is single-valued if and only if $R^\cup ; R \subseteq \mathrm{Id}_Y$.
4. Show that R is total if and only if $\mathrm{Id}_X \subseteq R ; R^\cup$.
5. Show that a function $f : X \to Y$ is injective if and only if $f ; f^\cup \subseteq \mathrm{Id}_X$.
6. Show that a function $f : X \to Y$ is surjective if and only if $\mathrm{Id}_Y \subseteq f^\cup ; f$.

Disjoint families, restriction

Disjoint families, restriction

Given a function $f : X \to Y$, there is a function $f^{-1} : \mathscr{P}(Y) \to \mathscr{P}(X)$, the **inverse image** of f, defined by:

$$f^{-1}(B) := \{x \in X \mid f(x) \in B\}.$$

Disjoint families, restriction

Given a function $f : X \to Y$, there is a function $f^{-1} : \mathscr{P}(Y) \to \mathscr{P}(X)$, the **inverse image** of f, defined by:

$$f^{-1}(B) := \{x \in X \mid f(x) \in B\}.$$

In case $B = \{y\}$ is a singleton, we just write $f^{-1}(y)$ rather than $f^{-1}(\{y\})$.

Disjoint families, restriction

Given a function $f : X \to Y$, there is a function $f^{-1} : \mathscr{P}(Y) \to \mathscr{P}(X)$, the **inverse image** of f, defined by:

$$f^{-1}(B) := \{x \in X \mid f(x) \in B\}.$$

In case $B = \{y\}$ is a singleton, we just write $f^{-1}(y)$ rather than $f^{-1}(\{y\})$.

A family of sets $\{X_i\}_{i \in I}$ is **pairwise disjoint** if $X_i \cap X_j = \varnothing$ whenever $i \neq j$. The union of such a pairwise disjoint family is called a **disjoint union**: given any $x \in \bigcup_{i \in I} X_i$, there is a unique i such that $x \in X_i$.

Disjoint families, restriction

Given a function $f : X \to Y$, there is a function $f^{-1} : \mathscr{P}(Y) \to \mathscr{P}(X)$, the **inverse image** of f, defined by:

$$f^{-1}(B) \;:=\; \{x \in X \mid f(x) \in B\}.$$

In case $B = \{y\}$ is a singleton, we just write $f^{-1}(y)$ rather than $f^{-1}(\{y\})$.

A family of sets $\{X_i\}_{i \in I}$ is **pairwise disjoint** if $X_i \cap X_j = \varnothing$ whenever $i \neq j$. The union of such a pairwise disjoint family is called a **disjoint union**: given any $x \in \bigcup_{i \in I} X_i$, there is a unique i such that $x \in X_i$.

Given a function $f : X \to Y$ and $U \subseteq X$, the **restriction** of f to U, written $f|_U : U \to Y$, is defined by

$$f|_U(x) \;:=\; f(x), \quad x \in U.$$

Disjoint families, restriction

Given a function $f : X \to Y$, there is a function $f^{-1} : \mathscr{P}(Y) \to \mathscr{P}(X)$, the **inverse image** of f, defined by:

$$f^{-1}(B) \;:=\; \{x \in X \mid f(x) \in B\}.$$

In case $B = \{y\}$ is a singleton, we just write $f^{-1}(y)$ rather than $f^{-1}(\{y\})$.

A family of sets $\{X_i\}_{i \in I}$ is **pairwise disjoint** if $X_i \cap X_j = \varnothing$ whenever $i \neq j$. The union of such a pairwise disjoint family is called a **disjoint union**: given any $x \in \bigcup_{i \in I} X_i$, there is a unique i such that $x \in X_i$.

Given a function $f : X \to Y$ and $U \subseteq X$, the **restriction** of f to U, written $f|_U : U \to Y$, is defined by

$$f|_U(x) \;:=\; f(x), \quad x \in U.$$

Exercise Suppose we have functions $f : U \to Y$ and $g : V \to Y$. Give a necessary and sufficient condition for there to exist a function $h : U \cup V \to Y$ such that $h|_U = f$ and $h|_V = g$.

Probability distributions and measures on finite sets

Probability distributions and measures on finite sets

Let X be a finite set. A **probability distribution** on X is a function $d : X \to [0, 1]$ such that

$$\sum_{x \in X} d(x) \;=\; 1.$$

Probability distributions and measures on finite sets

Let X be a finite set. A **probability distribution** on X is a function $d : X \to [0, 1]$ such that

$$\sum_{x \in X} d(x) \;=\; 1.$$

A probability distribution extends to a function $p : \mathscr{P}(X) \to [0, 1]$ on subsets of X:

$$p(A) \;:=\; \sum_{x \in A} d(x).$$

Probability distributions and measures on finite sets

Let X be a finite set. A **probability distribution** on X is a function $d : X \to [0, 1]$ such that

$$\sum_{x \in X} d(x) \;=\; 1.$$

A probability distribution extends to a function $p : \mathscr{P}(X) \to [0, 1]$ on subsets of X:

$$p(A) \;:=\; \sum_{x \in A} d(x).$$

This function satisfies the following properties:

$$(P1) \quad p(X) \;=\; 1$$

$$(P2) \quad A \cap B = \varnothing \;\Rightarrow\; p(A \cup B) \;=\; p(A) + p(B)$$

Probability distributions and measures on finite sets

Let X be a finite set. A **probability distribution** on X is a function $d : X \to [0, 1]$ such that

$$\sum_{x \in X} d(x) \;=\; 1.$$

A probability distribution extends to a function $p : \mathscr{P}(X) \to [0, 1]$ on subsets of X:

$$p(A) \;:=\; \sum_{x \in A} d(x).$$

This function satisfies the following properties:

$$(P1) \quad p(X) \;=\; 1$$

$$(P2) \quad A \cap B = \varnothing \;\Rightarrow\; p(A \cup B) \;=\; p(A) + p(B)$$

We call a function $p : \mathscr{P}(X) \to [0, 1]$ satisfying (P1) and (P2) a **probability measure**.

Equivalence of distributions and measures

Equivalence of distributions and measures

By induction, $(P2)$ extends to any finite disjoint union: if $A_1, \ldots, A_n$ are pairwise disjoint, then

$$p(A_1 \cup \cdots \cup A_n) \;=\; \sum_{i=1}^{n} p(A_i).$$

Equivalence of distributions and measures

By induction, $(P2)$ extends to any finite disjoint union: if $A_1, \ldots, A_n$ are pairwise disjoint, then

$$p(A_1 \cup \cdots \cup A_n) \;=\; \sum_{i=1}^{n} p(A_i).$$

Proposition

Each probability distribution $d : X \to [0,1]$ determines a probability measure $p : \mathscr{P}(X) \to [0,1]$ defined by

$$p(A) \;:=\; \sum_{x \in A} d(x).$$

Conversely, each probability measure $p : \mathscr{P}(X) \to [0,1]$ determines a probability distribution $d : X \to [0,1]$, defined by

$$d(x) \;=\; p(\{x\}).$$

These maps between probability distributions and measures are mutually inverse.

Additional properties of probability measures

Additional properties of probability measures

Proposition

A probability measure satisfies the following additional properties:

$$(P3) \quad p(A^c) = 1 - p(A)$$

$$(P4) \quad p(\varnothing) = 0$$

$$(P5) \quad p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

$$(P6) \quad A \subseteq B \Rightarrow p(A) \leq p(B)$$

$$(P7) \quad p(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} p(A_i)$$

Additional properties of probability measures

Proposition

A probability measure satisfies the following additional properties:

$$
\begin{aligned}
&(P3) \quad p(A^c) \;=\; 1 - p(A) \\
&(P4) \quad p(\varnothing) \;=\; 0 \\
&(P5) \quad p(A \cup B) \;=\; p(A) + p(B) - p(A \cap B) \\
&(P6) \quad A \subseteq B \;\Rightarrow\; p(A) \leq p(B) \\
&(P7) \quad p\left(\textstyle\bigcup_{i=1}^{n} A_i\right) \;\leq\; \sum_{i=1}^{n} p(A_i)
\end{aligned}
$$

Proposition

Let $B_1, \ldots, B_n$ be a pairwise disjoint family of sets, with $\bigcup_{i=1}^{n} B_i = X$. Then, for any set A:

$$p(A) \;=\; \sum_{i=1}^{n} p(A \cap B_i).$$

Conditional Probability

Conditional Probability

Suppose we have sets A and B with $p(B) > 0$. Then we define the **conditional probability of A given B**:

$$p(A|B) := \frac{p(A \cap B)}{p(B)}.$$

Conditional Probability

Suppose we have sets A and B with $p(B) > 0$. Then we define the **conditional probability of A given B**:

$$p(A|B) := \frac{p(A \cap B)}{p(B)}.$$

Proposition

If $p(A) > 0$ and $p(B) > 0$, then:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}.$$

Conditional Probability

Suppose we have sets A and B with $p(B) > 0$. Then we define the **conditional probability of A given B**:

$$p(A|B) := \frac{p(A \cap B)}{p(B)}.$$

Proposition

If $p(A) > 0$ and $p(B) > 0$, then:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}.$$

Proposition (Law of Total Probability)

Let $B_1, \ldots, B_n$ be a pairwise disjoint family of sets, with $\bigcup_{i=1}^{n} B_i = X$. Then, for any set A:

$$p(A) = \sum_{i=1}^{n} p(A|B_i)p(B_i).$$

Bayes' Theorem

Bayes' Theorem

We now state Bayes' Theorem in its most commonly encountered form.

Proposition (Bayes' Theorem)

Let $B_1, \ldots, B_n$ be a pairwise disjoint family of sets, with $\bigcup_{i=1}^{n} B_i = X$. Then, for any set A:

$$p(B_k|A) = \frac{p(A|B_k)p(B_k)}{\sum_{i=1}^{n} p(A|B_i)p(B_i)}.$$

Example of use of Bayes' Theorem

An HIV test gives a positive result with probability 98% when the patient is indeed affected by HIV, while it gives a negative result with 99% probability when the patient is not affected by HIV. If a patient is drawn at random from a population in which 0.1% of individuals are affected by HIV and he is found positive, what is the probability that he is indeed affected by HIV?

Example of use of Bayes' Theorem

An HIV test gives a positive result with probability 98% when the patient is indeed affected by HIV, while it gives a negative result with 99% probability when the patient is not affected by HIV. If a patient is drawn at random from a population in which 0.1% of individuals are affected by HIV and he is found positive, what is the probability that he is indeed affected by HIV?

The given information can be summarised as follows:

$$
\begin{aligned}
p(\text{positive}|\text{HIV}) &= 0.98 \\
p(\text{positive}|\text{NO HIV}) &= 1 - 0.99 = 0.01 \\
p(\text{HIV}) &= 0.001 \\
p(\text{NO HIV}) &= 0.999
\end{aligned}
$$

Example ctd

The probability of being found positive can be derived using the law of total probability:

$$\begin{aligned} p(\text{positive}) &= p(\text{positive}|\text{HIV})p(\text{HIV}) + p(\text{positive}|\text{NO HIV})p(\text{NO HIV}) \\ &= 0.98 \cdot 0.001 + 0.01 \cdot 0.999 \\ &= 0.00098 + 0.00999 \\ &= 0.01097 \end{aligned}$$

Example ctd

The probability of being found positive can be derived using the law of total probability:

$$\begin{aligned} p(\text{positive}) &= p(\text{positive}|\text{HIV})p(\text{HIV}) + p(\text{positive}|\text{NO HIV})p(\text{NO HIV}) \\ &= 0.98 \cdot 0.001 + 0.01 \cdot 0.999 \\ &= 0.00098 + 0.00999 \\ &= 0.01097 \end{aligned}$$

So, by Bayes' Theorem:

$$\begin{aligned} p(\text{HIV}|\text{positive}) &= \frac{p(\text{positive}|\text{HIV})p(\text{HIV})}{p(\text{positive}} \\ &= \frac{0.98 \cdot 0.001}{0.01097} \\ &= \frac{0.00098}{0.01097} \simeq 0.08933. \end{aligned}$$

Example ctd

The probability of being found positive can be derived using the law of total probability:

$$\begin{aligned} p(\text{positive}) &= p(\text{positive}|\text{HIV})p(\text{HIV}) + p(\text{positive}|\text{NO HIV})p(\text{NO HIV}) \\ &= 0.98 \cdot 0.001 + 0.01 \cdot 0.999 \\ &= 0.00098 + 0.00999 \\ &= 0.01097 \end{aligned}$$

So, by Bayes' Theorem:

$$\begin{aligned} p(\text{HIV}|\text{positive}) &= \frac{p(\text{positive}|\text{HIV})p(\text{HIV})}{p(\text{positive}} \\ &= \frac{0.98 \cdot 0.001}{0.01097} \\ &= \frac{0.00098}{0.01097} \simeq 0.08933. \end{aligned}$$

So despite the conditional accuracy of the test, the unconditional probability of being affected by HIV when found positive is less than 10%.

Independence

Independence

Proposition

Let A and B be events with $p(A) > 0$ and $p(B) > 0$. The following conditions are equivalent:

$$(1) \quad p(A|B) = p(A)$$

$$(2) \quad p(B|A) = p(B)$$

$$(3) \quad p(A \cap B) = p(A)p(B)$$

Independence

Proposition

Let A and B be events with $p(A) > 0$ and $p(B) > 0$. The following conditions are equivalent:

$$(1) \quad p(A|B) = p(A)$$

$$(2) \quad p(B|A) = p(B)$$

$$(3) \quad p(A \cap B) = p(A)p(B)$$

If any of these equivalent conditions are satisfied, we say that A and B are **independent**. We write $A \perp\!\!\!\perp B$.

Conditional Independence

Conditional Independence

Proposition

Let A, B and C be events with $p(A \cap C) > 0$ and $p(B \cap C) > 0$. The following conditions are equivalent:

$$(1) \quad p(A|B,C) = p(A|C)$$

$$(2) \quad p(B|A,C) = p(B|C)$$

$$(3) \quad p(A,B|C) = p(A|C)p(B|C)$$

Conditional Independence

Proposition

Let A, B and C be events with $p(A \cap C) > 0$ and $p(B \cap C) > 0$. The following conditions are equivalent:

$$(1) \quad p(A|B,C) = p(A|C)$$

$$(2) \quad p(B|A,C) = p(B|C)$$

$$(3) \quad p(A,B|C) = p(A|C)p(B|C)$$

In this proposition, we use the notation A, B instead of $A \cap B$. This is standard.

Conditional Independence

Proposition

Let A, B and C be events with $p(A \cap C) > 0$ and $p(B \cap C) > 0$. The following conditions are equivalent:

$$(1) \quad p(A|B,C) = p(A|C)$$

$$(2) \quad p(B|A,C) = p(B|C)$$

$$(3) \quad p(A,B|C) = p(A|C)p(B|C)$$

In this proposition, we use the notation A, B instead of $A \cap B$. This is standard.

If any of these equivalent conditions are satisfied, we say that A and B are **conditionally independent given** C. We write $A \perp\!\!\!\perp B \mid C$.

Products and Marginals

Products and Marginals

We consider a probability distribution $d : X \times Y \to [0, 1]$ on the cartesian product of finite sets X, Y. This induces a probability distribution $d_X : X \to [0, 1]$, defined by:

$$d_X(x) := \sum_{y \in Y} d(x, y).$$

Products and Marginals

We consider a probability distribution $d : X \times Y \to [0, 1]$ on the cartesian product of finite sets X, Y. This induces a probability distribution $d_X : X \to [0, 1]$, defined by:

$$d_X(x) := \sum_{y \in Y} d(x, y).$$

This is the **marginal** of d on X. Similarly, there is a marginal d_Y on Y.

Random Variables

Random Variables

Given a probability distribution d on X, a **random variable** on X is a function $R : X \to V$.

Random Variables

Given a probability distribution d on X, a **random variable** on X is a function $R : X \to V$.

We write $p(R = v)$for the probability that the random variable R takes the value $v \in V$. This is defined by:

$$p(R = v) \ := \ p(\{x \in X \mid R(x) = v\}).$$

Random Variables

Given a probability distribution d on X, a **random variable** on X is a function $R : X \to V$.

We write $p(R = v)$for the probability that the random variable R takes the value $v \in V$. This is defined by:

$$p(R = v) \ := \ p(\{x \in X \mid R(x) = v\}).$$

If we have two random variables $R : X \to V$, $S : X \to W$ defined on X, we have a **joint distribution**:

$$p(R = v, S = w) \ := \ p(\{x \in X \mid R(x) = v \ \wedge \ S(x) = w\}).$$

Probability distributions as a datatype

Probability distributions as a datatype

Given a finite set X, we write $\mathrm{Prob}(X)$ for the set of all probability distributions on X.

Probability distributions as a datatype

Given a finite set X, we write $\mathrm{Prob}(X)$ for the set of all probability distributions on X.

Given $f : X \to Y$, we define a function $\mathrm{Prob}(f) : \mathrm{Prob}(X) \to \mathrm{Prob}(Y)$, which maps probability distributions on X to probability distributions on Y, by "pushing them forwards along f". This is defined by $\mathrm{Prob}(f)(d) = d'$, where $d' \in \mathrm{Prob}(Y)$ is defined by:

$$d'(y) \;=\; \sum_{f(x)=y} d(x) \;=\; p(f^{-1}(y)),$$

where p is the probability measure $p : \mathscr{P}(X) \to [0, 1]$ corresponding to the distribution d.

Probability distributions as a datatype

Given a finite set X, we write $\mathrm{Prob}(X)$ for the set of all probability distributions on X.

Given $f : X \to Y$, we define a function $\mathrm{Prob}(f) : \mathrm{Prob}(X) \to \mathrm{Prob}(Y)$, which maps probability distributions on X to probability distributions on Y, by "pushing them forwards along f". This is defined by $\mathrm{Prob}(f)(d) = d'$, where $d' \in \mathrm{Prob}(Y)$ is defined by:

$$d'(y) \;=\; \sum_{f(x)=y} d(x) \;=\; p(f^{-1}(y)),$$

where p is the probability measure $p : \mathscr{P}(X) \to [0,1]$ corresponding to the distribution d.

Proposition

If $f : X \to Y$ *and* $g : Y \to Z$, *then* $\mathrm{Prob}(g \circ f) = \mathrm{Prob}(g) \circ \mathrm{Prob}(f)$. *Also,* $\mathrm{Prob}(\mathrm{id}_X) = \mathrm{id}_{\mathrm{Prob}(X)}$.

Complex numbers

Real numbers $\mathbb{R}$ don't suffice to solve all algebraic (polynomial) equations:

$$x^2 + 1 = 0.$$

Need i such that $i^2 = -1$.

Complex numbers

Real numbers $\mathbb{R}$ don't suffice to solve all algebraic (polynomial) equations:

$$x^2 + 1 = 0.$$

Need i such that $i^2 = -1$.

Complex numbers $\mathbb{C}$: numbers of the form $a + ib$, $a, b \in \mathbb{R}$.

Complex numbers

Real numbers $\mathbb{R}$ don't suffice to solve all algebraic (polynomial) equations:

$$x^2 + 1 = 0.$$

Need i such that $i^2 = -1$.

Complex numbers $\mathbb{C}$: numbers of the form $a + ib$, $a, b \in \mathbb{R}$.

Addition:

$$(a + ib) + (c + id) = (a + c, i(c + d)).$$

Complex numbers

Real numbers $\mathbb{R}$ don't suffice to solve all algebraic (polynomial) equations:

$$x^2 + 1 = 0.$$

Need i such that $i^2 = -1$.

Complex numbers $\mathbb{C}$: numbers of the form $a + ib$, $a, b \in \mathbb{R}$.

Addition:

$$(a + ib) + (c + id) = (a + c, i(c + d)).$$

Multiplication:

$$\begin{aligned}(a + ib)(c + id) &= ac + iad + iad + i^2 bd \\ &= (ac - bd, i(ad + bc))\end{aligned}$$

Complex numbers

Real numbers $\mathbb{R}$ don't suffice to solve all algebraic (polynomial) equations:

$$x^2 + 1 = 0.$$

Need i such that $i^2 = -1$.

Complex numbers $\mathbb{C}$: numbers of the form $a + ib$, $a, b \in \mathbb{R}$.

Addition:

$$(a + ib) + (c + id) = (a + c, i(c + d)).$$

Multiplication:

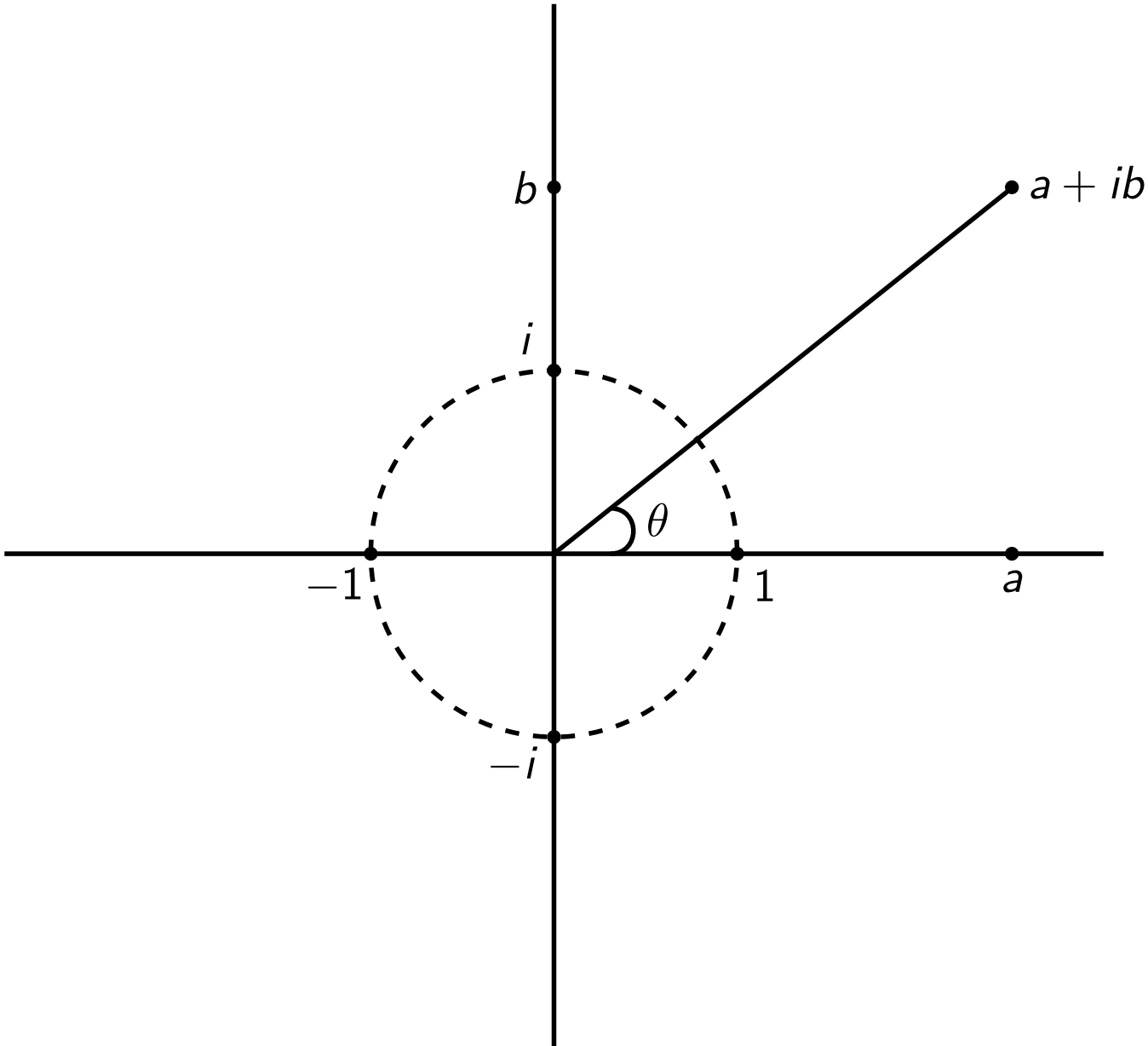
$$\begin{aligned}(a + ib)(c + id) &= ac + iad + iad + i^2 bd \\ &= (ac - bd, i(ad + bc))\end{aligned}$$

Remarkably, **all** algebraic equations over $\mathbb{C}$ have solutions in $\mathbb{C}$. (The "Fundamental Theorem of Algebra", or in modern terminology, $\mathbb{C}$ is algebraically closed.)

The Complex Plane

The Complex Plane

Complex numbers can be viewed geometrically, as 2-dimensional vectors.



Argument and modulus

Argument and modulus

We can represent complex numbers $z = a + ib$ as (r, θ), where $r = \sqrt{a^2 + b^2}$, the **magnitude** of z, *i.e.* its distance from the origin.

Argument and modulus

We can represent complex numbers $z = a + ib$ as (r, θ), where $r = \sqrt{a^2 + b^2}$, the **magnitude** of z, *i.e.* its distance from the origin.

We have the operation of **complex conjugation**:

$$z \mapsto \bar{z}, \qquad a + ib \mapsto a - ib.$$

Note that

$$\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}, \qquad \overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}.$$

Argument and modulus

We can represent complex numbers $z = a + ib$ as (r, θ), where $r = \sqrt{a^2 + b^2}$, the **magnitude** of z, *i.e.* its distance from the origin.

We have the operation of **complex conjugation**:

$$z \mapsto \bar{z}, \qquad a + ib \mapsto a - ib.$$

Note that

$$\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}, \qquad \overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}.$$

Also,

$$z\bar{z} \;=\; (a + ib)(a - ib) \;=\; a^2 + b^2.$$

So $z\bar{z} = |z|^2$.

Complex phases

Complex phases

Note that the reals of modulus 1, *i.e.* $|r| = 1$, are just two discrete points, $\{+1, -1\}$.

Complex phases

Note that the reals of modulus 1, *i.e.* $|r| = 1$, are just two discrete points, $\{+1, -1\}$.

By contrast, the complex numbers of modulus one form a **circle**.

Complex phases

Note that the reals of modulus 1, *i.e.* $|r| = 1$, are just two discrete points, $\{+1, -1\}$.

By contrast, the complex numbers of modulus one form a **circle**.

These are all numbers of the form $e^{i\theta}$, $\theta \in [0, 2\pi]$.

Complex phases

Note that the reals of modulus 1, *i.e.* $|r| = 1$, are just two discrete points, $\{+1, -1\}$.

By contrast, the complex numbers of modulus one form a **circle**.

These are all numbers of the form $e^{i\theta}$, $\theta \in [0, 2\pi]$.

Euler's formula: $e^{i\theta} = \cos\theta + i \sin\theta$**.**

Complex phases

Note that the reals of modulus 1, *i.e.* $|r| = 1$, are just two discrete points, $\{+1, -1\}$.

By contrast, the complex numbers of modulus one form a **circle**.

These are all numbers of the form $e^{i\theta}$, $\theta \in [0, 2\pi]$.

Euler's formula: $e^{i\theta} = \cos\theta + i \sin\theta$.

Pythagoras' theorem: $\cos^2\theta + \sin^2\theta = 1$.

Complex phases

Note that the reals of modulus 1, *i.e.* $|r| = 1$, are just two discrete points, $\{+1, -1\}$.

By contrast, the complex numbers of modulus one form a **circle**.

These are all numbers of the form $e^{i\theta}$, $\theta \in [0, 2\pi]$.

Euler's formula: $e^{i\theta} = \cos\theta + i\sin\theta$.

Pythagoras' theorem: $\cos^2\theta + \sin^2\theta = 1$.

So

$$\begin{aligned} |e^{i\theta}|^2 &= e^{i\theta}\overline{e^{i\theta}} \\ &= (\cos\theta + i\sin\theta)(\cos\theta - i\sin\theta) \\ &= \cos^2\theta + \sin^2\theta \;=\; 1. \end{aligned}$$