Model Counting for Logical Theories

Tuesday

Dmitry Chistikov Rayna Dimitrova

Department of Computer Science
University of Oxford, UK

Max Planck Institute for Software Systems (MPI-SWS)
Kaiserslautern and Saarbrücken, Germany

ESSLLI 2016

$$
\begin{tikzcd}
\begin{array}{c}\mathrm{SAT}\\ \text{satisfiability}\end{array} \arrow[r] \arrow[d] & \begin{array}{c}\mathrm{SMT}\\ \text{satisfiability}\\ \text{modulo theories}\end{array} \arrow[d] \\
\begin{array}{c}\#\mathrm{SAT}\\ \text{model counting}\end{array} \arrow[r] & \begin{array}{c}\#\mathrm{SMT}\\ \text{model counting}\\ \text{modulo theories}\end{array}
\end{tikzcd}
$$

Agenda

Tuesday	computational complexity, probability theory
Wednesday	randomized algorithms, Monte Carlo methods
Thursday	hashing-based approach to model counting
Friday	from discrete to continuous model counting

Outline

1. Complexity theory: $\mathbf{P}$, $\mathbf{NP}$, and $\#\mathbf{P}$

2. Probability theory
   - Probability theory: Events
   - Probability theory: Random variables

Decision problems and algorithms

Decision problem:
$L \subseteq \{0,1\}^*$ (encodings of yes-instances)

Algorithm for L:
says "yes" on every $x \in L$, "no" on every $x \in \{0,1\}^* \setminus L$

Time complexity

- of algorithm $\mathcal{A}$ on input x
- of algorithm $\mathcal{A}$ on inputs of length n (worst-case)
- of decision problem L

Complexity class $\mathbf{P}$ and efficient algorithms

Cobham–Edmonds thesis:

Efficiently computable in a reasonable computational model

$=$

Computable in polynomial time on a Turing machine

$$\mathbf{P} = \bigcup_{d \geq 1} \bigcup_{c \geq 1} \mathrm{DTIME}(c \cdot n^d)$$

Problems with efficiently verifiable solutions: $\mathbf{NP}$

- Definition via certificates
- Definition via nondeterministic machines

Reductions and $\mathbf{NP}$-complete problems

- Polynomial-time reduction
- $\mathbf{NP}$-hard and $\mathbf{NP}$-complete problems

From decision to counting problems

Real-valued problem: $f\colon \{0,1\}^* \to \mathbb{R}$
Counting problem: $f\colon \{0,1\}^* \to \{0,1,2,\ldots\}$

$\#\mathbf{P}$: consists of problems that count the number of certificates to instances of $\mathbf{NP}$-problems

Complexity classes: brief summary

$\mathbf{P}$: polynomial time (efficiently solvable)

$\mathbf{NP}$: nondeterministic polynomial time (with efficiently verifiable solutions)

$\#\mathbf{P}$: counting polynomial time

Outline

1. Complexity theory: $\mathbf{P}$, $\mathbf{NP}$, and $\#\mathbf{P}$

2. Probability theory
   - Probability theory: Events
   - Probability theory: Random variables

Recap: Measured theories

Measured theories and model count

A logical theory $\mathcal{T}$ is measured if every $[\![\varphi]\!]$ is measurable.

The model count of a formula φ is $\mathsf{mc}(\varphi) = \mu([\![\varphi]\!])$.

σ-algebras

σ-algebra $(D, \mathcal{F})$: domain D, set of subsets $\mathcal{F} \subseteq 2^D$ such that

$$\begin{array}{rcll} & & \varnothing \in \mathcal{F} & \text{(the empty set is an element)} \\ A \in \mathcal{F} & \implies & D \setminus A \in \mathcal{F} & \text{(closure under complementation)} \\ A_i \in \mathcal{F} & \implies & \bigcup_i A_i \in \mathcal{F} & \text{(closure under countable union)} \end{array}$$

Examples

- finite set D, $\mathcal{F} = 2^D$
- $D = \mathbb{R}$, $\mathcal{F}$ obtained from the set of all open intervals by adding all complements and countable unions iteratively until the closure properties are met (Borel hierarchy)

Measure: How big is a set?

Measure μ for $(D, \mathcal{F})$: maps each $A \in \mathcal{F}$ to a real number $\mu(A) \geq 0$

More formally

$$\begin{array}{lcl} A \in \mathcal{F} & \implies & \mu(A) \geq \mu(\varnothing) = 0 \\ A_i \in \mathcal{F} \text{ disjoint} & \implies & \mu(\bigcup_i A_i) = \sum_i \mu(A_i) \end{array}$$

Measure space $(D, \mathcal{F}, \mu)$: σ-algebra $(D, \mathcal{F})$, measure $\mu : \mathcal{F} \to \mathbb{R}$

Probability theory:
Events

Probability spaces

Definition

A triple $(\Omega, \mathcal{F}, \mathsf{P})$ is a **probability space**
if $\mathcal{F}$ is a σ-algebra of subsets of Ω
and P is a measure on $(\Omega, \mathcal{F})$ that satisfies $\mathsf{P}(\Omega) = 1$.

Probability spaces

Definition

A triple $(\Omega, \mathcal{F}, \mathsf{P})$ is a **probability space**
if $\mathcal{F}$ is a σ-algebra of subsets of Ω
and $\mathsf{P}\colon \mathcal{F} \to \mathbb{R}$ satisfies the following properties:

- $\mathsf{P}(A) \geq 0$ for all $A \in \mathcal{F}$.
- $\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i)$ if $A_i \cap A_j = \varnothing$ for $i \neq j$.
- $\mathsf{P}(\Omega) = 1$.

Probability spaces

Definition

A triple $(\Omega, \mathcal{F}, \mathsf{P})$ is a **probability space**
if $\mathcal{F}$ is a σ-algebra of subsets of Ω
and $\mathsf{P}\colon \mathcal{F} \to \mathbb{R}$ satisfies the following properties:

- $\mathsf{P}(A) \geq 0$ for all $A \in \mathcal{F}$.
- $\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i)$ if $A_i \cap A_j = \varnothing$ for $i \neq j$.
- $\mathsf{P}(\Omega) = 1$.

Discrete probability space: Ω is finite.
For such spaces it's usually convenient to pick $\mathcal{F} = 2^{\Omega}$.

Law of Sum

If events A and B are disjoint ($A \cap B = \varnothing$),
then $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B)$.

If $A_1, \ldots, A_n$ are pairwise disjoint events
($A_i \cap A_j = \varnothing$ for all $i \neq j$),
then $\mathsf{P}(A_1 \cup \ldots \cup A_n) = \sum_{i=1}^{n} \mathsf{P}(A_i)$.

Probability of Union

Is it true that $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B)$?

Probability of Union

Is it true that $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B)$?

Answer:
This is only true if $\mathsf{P}(A \cap B) = 0$.

Example

A fair die gives each $k \in \{1, 2, 3, 4, 5, 6\}$ with probability $1/6$.
Consider $A = \{1, 2, 3\}$ and $B = \{2, 4, 6\}$.
$\mathsf{P}[\,\text{outcome} \leq 3\,] = \mathsf{P}[A] = 1/2$.
$\mathsf{P}[\,\text{outcome even}\,] = \mathsf{P}[B] = 1/2$.
$\mathsf{P}[\,\text{outcome} \leq 3 \text{ or even}\,] = \mathsf{P}[A \cup B] = \mathsf{P}[\{1, 2, 3, 4, 6\}] = 5/6 \neq 1/2 + 1/2$.

Probability of Union

Is it true that $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B)$?

Answer:
This is only true if $\mathsf{P}(A \cap B) = 0$.

In general,

$$\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B).$$

Why?

$$\begin{aligned}
\mathsf{P}(A \cup B) &= \mathsf{P}(A) + \mathsf{P}(\overline{A} \cap B) \\
\mathsf{P}(B) &= \mathsf{P}(\overline{A} \cap B) + \mathsf{P}(A \cap B) \\
\mathsf{P}(A \cup B) - \mathsf{P}(B) &= \mathsf{P}(A) - \mathsf{P}(A \cap B)
\end{aligned}$$

The Union Bound

We know that

$$\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B).$$

Observe that, as a corollary,

$$\mathsf{P}(A \cup B) \leq \mathsf{P}(A) + \mathsf{P}(B).$$

Generalize this:

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathsf{P}(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} \mathsf{P}(A_{i_1} A_{i_2}) + \\ + \ldots + (-1)^{n+1} \mathsf{P}(A_1 \ldots A_n) \qquad \text{(difficult)}$$

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathsf{P}(A_i) \qquad \text{(easy; union bound)}$$

Law of Complement

$$\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A).$$

What about Product?

Does the equality $\mathsf{P}(A \cap B) = \mathsf{P}(A) \cdot \mathsf{P}(B)$ hold?

What about Product?

Does the equality $\mathsf{P}(A \cap B) = \mathsf{P}(A) \cdot \mathsf{P}(B)$ hold?

The answer is NO (in the general case).
(Although it does hold in an important special case, to be discussed later.)

Conditional probability

Conditional probability of A **given** B
(probability that event A occurs given that event B occurs)
is defined as

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

if $\mathsf{P}(B) > 0$ (and undefined otherwise).

Example

A fair die gives each $k \in \{1, 2, 3, 4, 5, 6\}$ with probability $1/6$.
$\mathsf{P}[\,\text{outcome} \leq 5 \mid \text{outcome even}\,] = ?$

Conditional probability

Conditional probability of A **given** B
(probability that event A occurs given that event B occurs)
is defined as

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

if $\mathsf{P}(B) > 0$ (and undefined otherwise).

Example

A fair die gives each $k \in \{1, 2, 3, 4, 5, 6\}$ with probability $1/6$.
Consider $A = \{1, 2, 3, 4, 5\}$ and $B = \{2, 4, 6\}$.
$\mathsf{P}[\,\text{outcome} \leq 5\,] = \mathsf{P}[A] = 5/6$.
$\mathsf{P}[\,\text{outcome even}\,] = \mathsf{P}[B] = 1/2$.
$\mathsf{P}[\,\text{outcome} \leq 5 \mid \text{outcome even}\,] = \mathsf{P}[A \mid B] =$

$$= \frac{\mathsf{P}[\,\text{outcome} \leq 5 \text{ and even}\,]}{\mathsf{P}[\,\text{outcome even}\,]} = \frac{\mathsf{P}[\,\{2, 4\}\,]}{\mathsf{P}[\,\{2, 4, 6\}\,]} = \frac{2/6}{3/6} = 2/3.$$

Conditional probability

Conditional probability of A **given** B
(probability that event A occurs given that event B occurs)
is defined as

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

if $\mathsf{P}(B) > 0$ (and undefined otherwise).

Example

A fair die gives each $k \in \{1, 2, 3, 4, 5, 6\}$ with probability $1/6$.
$\mathsf{P}[\,\text{outcome} \leq 5 \mid \text{outcome even}\,] = 2/3$
$\mathsf{P}[\,\text{outcome even} \mid \text{outcome} \leq 3\,] = ?$

Conditional probability

Conditional probability of A **given** B
(probability that event A occurs given that event B occurs)
is defined as

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

if $\mathsf{P}(B) > 0$ (and undefined otherwise).

Example

A fair die gives each $k \in \{1,2,3,4,5,6\}$ with probability $1/6$.
Consider $A = \{1,2,3,4,5\}$ and $B = \{2,4,6\}$.
$\mathsf{P}[\,\text{outcome} \leq 5\,] = \mathsf{P}[A] = 5/6.$
$\mathsf{P}[\,\text{outcome even}\,] = \mathsf{P}[B] = 1/2.$
$\mathsf{P}[\,\text{outcome even} \mid \text{outcome} \leq 5\,] = \mathsf{P}[B \mid A] =$

$$= \frac{\mathsf{P}[\,\text{outcome even and} \leq 5\,]}{\mathsf{P}[\,\text{outcome} \leq 5\,]} = \frac{\mathsf{P}[\,\{2,4\}\,]}{\mathsf{P}[\,\{1,2,3,4,5\}\,]} = \frac{2/6}{5/6} = 2/5.$$

Conditional probability

Conditional probability of A **given** B
(probability that event A occurs given that event B occurs)
is defined as

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

if $\mathsf{P}(B) > 0$ (and undefined otherwise).

Example

A fair die gives each $k \in \{1, 2, 3, 4, 5, 6\}$ with probability $1/6$.
$\mathsf{P}[\,\text{outcome} \le 5 \mid \text{outcome even}\,] = 2/3$
$\mathsf{P}[\,\text{outcome even} \mid \text{outcome} \le 3\,] = 2/5$

Conditional probability is probability!

If $\mathsf{P}(B) > 0$, then the function $\mathsf{Q}\colon 2^{\Omega} \to [0,1]$ defined by

$$\mathsf{Q}(A) = \mathsf{P}(A \mid B)$$

is a probability measure.

- What does this mean?

Conditional probability is probability!

If $\mathsf{P}(B) > 0$, then the function $\mathsf{Q}\colon 2^{\Omega} \to [0,1]$ defined by

$$\mathsf{Q}(A) = \mathsf{P}(A \mid B)$$

is a probability measure.

- What does this mean?
- Why? (homework)

Conditional probability is probability!

If $\mathsf{P}(B) > 0$, then the function $\mathsf{Q}\colon 2^\Omega \to [0,1]$ defined by

$$\mathsf{Q}(A) = \mathsf{P}(A \mid B)$$

is a probability measure.

- What does this mean?
- Why? (homework)
- So what?

Conditional probability is probability!

If $\mathsf{P}(B) > 0$, then the function $\mathsf{Q}\colon 2^\Omega \to [0,1]$ defined by

$$\mathsf{Q}(A) = \mathsf{P}(A \mid B)$$

is a probability measure.

- What does this mean?
- Why? (homework)
- So what?

$$\mathsf{P}(A \cup C \mid B) = \mathsf{P}(A \mid B) + \mathsf{P}(C \mid B) - \mathsf{P}(A \cap C \mid B)$$

$$\mathsf{P}\Big(\bigcup_{i=1}^{n} A_i \mid B\Big) \leq \sum_{i=1}^{n} \mathsf{P}(A_i \mid B)$$

$$\cdots$$

Law of Total Probability

Let $B_1, \ldots, B_m$ be a **partition**:
$\mathsf{P}(B_i \cap B_j) = 0$ for $i \neq j$ and $\mathsf{P}(B_1 \cup \ldots \cup B_m) = 1$.

Suppose $\mathsf{P}(B_i) > 0$ for all i.

Then for any event A

$$\mathsf{P}(A) = \sum_{i=1}^{m} \mathsf{P}(A \mid B_i) \cdot \mathsf{P}(B_i).$$

Independent events: Example

Example

A fair die gives each $k \in \{1, 2, 3, 4, 5, 6\}$ with probability $1/6$.
$\mathrm{P}[\,\text{outcome} \leq 4 \mid \text{outcome even}\,] = ?$

Independent events: Example

Example

A fair die gives each $k \in \{1, 2, 3, 4, 5, 6\}$ with probability $1/6$.

$\mathsf{P}[\,\text{outcome} \le 4 \mid \text{outcome even}\,] = ?$

Consider $A = \{1, 2, 3, 4\}$ and $B = \{2, 4, 6\}$.

$\mathsf{P}[\,\text{outcome} \le 4\,] = \mathsf{P}[A] = 2/3$.

$\mathsf{P}[\,\text{outcome even}\,] = \mathsf{P}[B] = 1/2$.

$\mathsf{P}[\,\text{outcome} \le 4 \mid \text{outcome even}\,] = \mathsf{P}[A \mid B] =$

$$= \frac{\mathsf{P}[\,\text{outcome} \le 4 \text{ and even}\,]}{\mathsf{P}[\,\text{outcome even}\,]} = \frac{\mathsf{P}[\,\{2, 4\}\,]}{\mathsf{P}[\,\{2, 4, 6\}\,]} = \frac{2/6}{3/6} = 2/3.$$

Independent events: Example

Example

A fair die gives each $k \in \{1, 2, 3, 4, 5, 6\}$ with probability $1/6$.

$\mathsf{P}[\,\text{outcome} \le 4 \mid \text{outcome even}\,] = ?$

Consider $A = \{1, 2, 3, 4\}$ and $B = \{2, 4, 6\}$.

$\mathsf{P}[\,\text{outcome} \le 4\,] = \mathsf{P}[A] = 2/3$.

$\mathsf{P}[\,\text{outcome even}\,] = \mathsf{P}[B] = 1/2$.

$\mathsf{P}[\,\text{outcome} \le 4 \mid \text{outcome even}\,] = \mathsf{P}[A \mid B] =$

$$= \frac{\mathsf{P}[\,\text{outcome} \le 4 \text{ and even}\,]}{\mathsf{P}[\,\text{outcome even}\,]} = \frac{\mathsf{P}[\,\{2, 4\}\,]}{\mathsf{P}[\,\{2, 4, 6\}\,]} = \frac{2/6}{3/6} = 2/3.$$

In this example $\mathsf{P}[A \mid B] = \mathsf{P}[A]$.

Independent events

When $\mathsf{P}(A \mid B) = \mathsf{P}(A)$?

This equality asserts that $\dfrac{\mathsf{P}(AB)}{\mathsf{P}(B)} = \mathsf{P}(A)$.

Assuming $\mathsf{P}(B) > 0$, rewrite this as $\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B)$.

Independent events

When $\mathsf{P}(A \mid B) = \mathsf{P}(A)$?

This equality asserts that $\dfrac{\mathsf{P}(AB)}{\mathsf{P}(B)} = \mathsf{P}(A)$.

Assuming $\mathsf{P}(B) > 0$, rewrite this as $\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B)$.

Definition

Events A and B are called **independent** if $\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B)$.

Independent events

When $\mathsf{P}(A \mid B) = \mathsf{P}(A)$?

This equality asserts that $\dfrac{\mathsf{P}(AB)}{\mathsf{P}(B)} = \mathsf{P}(A)$.

Assuming $\mathsf{P}(B) > 0$, rewrite this as $\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B)$.

Definition

Events A and B are called **independent** if $\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B)$.

This definition usually helps to define P.

Independent events: A standard example

Definition

Events A and B are called **independent** if $\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B)$.

Example

A fair coin is tossed twice so that the second toss does not depend on the outcome of the first.

$\mathsf{P}[\,\text{tails, tails}\,] = ?$

Independent events: A standard example

Definition
Events A and B are called **independent** if $\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B)$.
Example
A fair coin is tossed twice so that the second toss does not depend on the outcome of the first.

$\mathsf{P}[\,\text{tails, tails}\,] = ?$

Model the outcome of each toss as 0 (heads) or 1 (tails).
Four possible scenarios: $\Omega = \{00, 01, 10, 11\}$. How to define P?

Independent events: A standard example

Definition

Events A and B are called **independent** if $\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B)$.

Example

A fair coin is tossed twice so that the second toss does not depend on the outcome of the first.

$\mathsf{P}[\,\text{tails, tails}\,] = ?$

Model the outcome of each toss as 0 (heads) or 1 (tails).
Four possible scenarios: $\Omega = \{00, 01, 10, 11\}$. How to define P?

Consider events $A = \{10, 11\}$ and $B = \{01, 11\}$:
"first coin lands tails" and "second coin lands tails" respectively.

We want these events to have probability $1/2$ each and to be independent.

Independent events: A standard example

Definition

Events A and B are called **independent** if $\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B)$.

Example

A fair coin is tossed twice so that the second toss does not depend on the outcome of the first.

$\mathsf{P}[\,\text{tails, tails}\,] = ?$

Model the outcome of each toss as 0 (heads) or 1 (tails).
Four possible scenarios: $\Omega = \{00, 01, 10, 11\}$. How to define P?

Consider events $A = \{10, 11\}$ and $B = \{01, 11\}$:
"first coin lands tails" and "second coin lands tails" respectively.

We want these events to have probability $1/2$ each and to be independent.

Then $\mathsf{P}[\,\text{tails, tails}\,] = \mathsf{P}[\{11\}] = \mathsf{P}[A \cap B] = \mathsf{P}[A]\,\mathsf{P}[B] = 1/4$.

Three or more independent events

Events $A_1, \ldots, A_n$ are called **independent**
if for any subset $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$
$$\mathsf{P}(A_{i_1} \ldots A_{i_k}) = \mathsf{P}(A_1) \ldots \mathsf{P}(A_k).$$

Example

From the set of strings $\{000, 001, 002, \ldots, 999\}$
a string $X_1X_2X_3$ is picked uniformly at random.
Are the events $X_1 = 5$, $X_2 = 5$, and $X_3 = 5$ independent?

Three or more independent events

Events $A_1, \ldots, A_n$ are called **independent**
if for any subset $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$
$\mathsf{P}(A_{i_1} \ldots A_{i_k}) = \mathsf{P}(A_1) \ldots \mathsf{P}(A_k)$.

Example

From the set of strings $\{000, 001, 002, \ldots, 999\}$
a string $X_1X_2X_3$ is picked uniformly at random.
Are the events $X_1 = 5$, $X_2 = 5$, and $X_3 = 5$ independent?
Yes: $\mathsf{P}[X_i = 5] = 1/10$, $\mathsf{P}[X_i = 5, X_j = 5] = 1/100$ if $i \neq j$, and
$\mathsf{P}[X_1 = X_2 = X_3 = 5] = 1/1000$.

Three or more independent events

Events $A_1, \dots, A_n$ are called **independent**
if for any subset $\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$
$$\mathsf{P}(A_{i_1} \dots A_{i_k}) = \mathsf{P}(A_1) \dots \mathsf{P}(A_k).$$

Example

From the set of strings $\{000, 001, 002, \dots, 999\}$
a string $X_1X_2X_3$ is picked uniformly at random.
Are the events $X_1 = 5$, $X_2 = 5$, and $X_3 = 5$ independent?
Yes: $\mathsf{P}[X_i = 5] = 1/10$, $\mathsf{P}[X_i = 5, X_j = 5] = 1/100$ if $i \neq j$, and
$\mathsf{P}[X_1 = X_2 = X_3 = 5] = 1/1000$.
What if the string 999 is excluded from the set? (homework)

Independence and pairwise independence

Example

Consider a pyramid (a tetrahedron) with facets colored

red, blue, green, red-blue-green.

Suppose the pyramid lands on each facet with probability $1/4$.
Consider events R, B, G asserting that the facet the pyramid lands on has color red, blue, green on it, respectively.

Are these events independent?

Independence and pairwise independence

Example

Consider a pyramid (a tetrahedron) with facets colored

red, blue, green, red-blue-green.

Suppose the pyramid lands on each facet with probability $1/4$. Consider events R, B, G asserting that the facet the pyramid lands on has color red, blue, green on it, respectively.

Are these events independent?

$\mathsf{P}(R) = \mathsf{P}(B) = \mathsf{P}(G) = 1/2$.

Independence and pairwise independence

Example

Consider a pyramid (a tetrahedron) with facets colored

red, blue, green, red-blue-green.

Suppose the pyramid lands on each facet with probability $1/4$.
Consider events R, B, G asserting that the facet the pyramid lands on has color red, blue, green on it, respectively.

Are these events independent?

$\mathsf{P}(R) = \mathsf{P}(B) = \mathsf{P}(G) = 1/2.$

$\mathsf{P}(RB) = \mathsf{P}(RG) = \mathsf{P}(BG) = 1/4 = (1/2)^2.$

Independence and pairwise independence

Example

Consider a pyramid (a tetrahedron) with facets colored

red, blue, green, red-blue-green.

Suppose the pyramid lands on each facet with probability $1/4$.
Consider events R, B, G asserting that the facet the pyramid lands on has color red, blue, green on it, respectively.

Are these events independent?

$\mathsf{P}(R) = \mathsf{P}(B) = \mathsf{P}(G) = 1/2$.

$\mathsf{P}(RB) = \mathsf{P}(RG) = \mathsf{P}(BG) = 1/4 = (1/2)^2$.

$\mathsf{P}(RBG) = 1/4 \neq (1/2)^3$.

These events are NOT independent, but only pairwise independent.

Probability theory:
Random variables, distributions

From events to random variables

Given a probability space $(\Omega, 2^\Omega, \mathsf{P})$, we can talk about events $A \in 2^\Omega$ and their probability $\mathsf{P}(A)$.
However, it is often more convenient to talk about functions of the form $X\colon \Omega \to \mathbb{R}$, which are called **random variables**.

Example:
Bernoulli trial:
$\Omega = \{\text{heads}, \text{tails}\}$, $\mathsf{P}(\{\text{tails}\}) = p \in [0,1]$, $\mathsf{P}(\{\text{heads}\}) = 1 - p$
Define

$$X(\omega) = \begin{cases} 1 & \text{if } \omega = \text{tails}, \\ 0 & \text{if } \omega = \text{heads}. \end{cases}$$

We say that the random variable X has **Bernoulli distribution** with parameter p.

From events to random variables

Given a probability space $(\Omega, 2^\Omega, \mathsf{P})$, we can talk about events $A \in 2^\Omega$ and their probability $\mathsf{P}(A)$.
However, it is often more convenient to talk about functions of the form $X\colon \Omega \to \mathbb{R}$, which are called **random variables**.

Example:
Let $A \in 2^\Omega$ be an event.
The **indicator function** of A is defined as

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

$\mathbf{1}_A(\omega)$ is a random variable that has Bernoulli distribution with parameter $\mathsf{P}(A)$. (Why?)

From events to random variables

Given a probability space $(\Omega, 2^\Omega, \mathsf{P})$, we can talk about events $A \in 2^\Omega$ and their probability $\mathsf{P}(A)$.
However, it is often more convenient to talk about functions of the form $X : \Omega \to \mathbb{R}$, which are called **random variables**.

Example:
Let $A \in 2^\Omega$ be an event.
The **indicator function** of A is defined as

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

Sometimes the indicator function of A is denoted by "$[A]$", e.g., "[the coin gives heads]".

Binomial distribution

Suppose a coin is tossed n times, the outcome of all tosses are independent, and each gives tails with probability p.

Define

$$X_i = \begin{cases} 1 & \text{if the } i\text{th toss gives tails,} \\ 0 & \text{otherwise.} \end{cases}$$

(X_i has **Bernoulli distribution** with parameter p.)

Define

$$X = X_1 + \ldots + X_n.$$

We say that the random variable X has **binomial distribution** with parameters n, p.

Distributions of random variables

X has **Bernoulli distribution** with parameter p:

x	0	1
$\mathsf{P}(X = x)$	$1 - p$	p

X has **binomial distribution** with parameters $3, p$:

x	0	1	2	3
$\mathsf{P}(X = x)$	$(1 - p)^3$	$3p\,(1 - p)^2$	$3p^2\,(1 - p)$	p^3

X has **uniform distribution** on the set $\{0, 1, 2, 3\}$:

x	0	1	2	3
$\mathsf{P}(X = x)$	$1/4$	$1/4$	$1/4$	$1/4$

Expectation

$$\mathsf{E}\,X = \sum_k k \cdot \mathsf{P}(X = k)$$

Examples

X has **uniform distribution** on the set $\{1, 2, 3, \ldots, n\}$:

x	1	2	3	$\ldots$	n
$\mathsf{P}(X = x)$	$1/n$	$1/n$	$1/n$	$\ldots$	$1/n$

$\mathsf{E}\,X = 1/n \cdot 1 + 1/n \cdot 2 + 1/n \cdot 3 + \ldots + 1/n \cdot n = (n+1)/2$.

Expectation

$$\mathsf{E}\, X = \sum_k k \cdot \mathsf{P}(X = k)$$

Examples

X has Bernoulli distribution with parameter p:

x	0	1
$\mathsf{P}(X = x)$	$1 - p$	p

$\mathsf{E}\, X = 0 \cdot (1 - p) + 1 \cdot p = p$.
$\mathsf{E}\, X = 1/2$ if and only if the coin is unbiased.
So $\mathsf{E}\, X$ is the **expected** number of tails in one flip.

Expectation

$$\mathsf{E}\,X = \sum_k k \cdot \mathsf{P}(X = k)$$

Examples

X has **binomial distribution** with parameters n, p:

x	0	1	2	$\ldots$	n
$\mathsf{P}(X = x)$	$(1-p)^n$	$np\,(1-p)^{n-1}$	$\frac{n(n-1)}{2}\,p^2\,(1-p)^{n-2}$	$\ldots$	p^n

$\mathsf{E}\,X = ?$

Linearity of expectation

$$\mathsf{E}\,X = \sum_k k \cdot \mathsf{P}(X = k)$$

$$\begin{aligned} \mathsf{E}(X_1 + \ldots + X_n) &= \mathsf{E}\,X_1 + \ldots + \mathsf{E}\,X_n \\ \mathsf{E}(c_1 X_1 + \ldots + c_n X_n) &= c_1\,\mathsf{E}\,X_1 + \ldots + c_n\,\mathsf{E}\,X_n \end{aligned}$$

(if c_i are fixed, i.e., non-random)

Linearity of expectation

$$\mathsf{E}\,X = \sum_k k \cdot \mathsf{P}(X = k)$$

$$\mathsf{E}(X_1 + \ldots + X_n) = \mathsf{E}\,X_1 + \ldots + \mathsf{E}\,X_n$$
$$\mathsf{E}(c_1 X_1 + \ldots + c_n X_n) = c_1\,\mathsf{E}\,X_1 + \ldots + c_n\,\mathsf{E}\,X_n$$

(if c_i are fixed, i.e., non-random)

Example

If X has **binomial distribution** with parameters n, p:

x	0	1	2	$\ldots$	n
$\mathsf{P}(X = x)$	$(1-p)^n$	$np\,(1-p)^{n-1}$	$\frac{n(n-1)}{2}\,p^2\,(1-p)^{n-2}$	$\ldots$	p^n

…then X has the same distribution as $Y_1 + \ldots + Y_n$
where each Y_i has Bernoulli distribution with parameter p.

Linearity of expectation

$$\mathsf{E}\, X = \sum_k k \cdot \mathsf{P}(X = k)$$

$$\mathsf{E}(X_1 + \ldots + X_n) = \mathsf{E}\, X_1 + \ldots + \mathsf{E}\, X_n$$
$$\mathsf{E}(c_1 X_1 + \ldots + c_n X_n) = c_1\, \mathsf{E}\, X_1 + \ldots + c_n\, \mathsf{E}\, X_n$$

(if c_i are fixed, i.e., non-random)

Example

If X has **binomial distribution** with parameters n, p:

x	0	1	2	$\ldots$	n
$\mathsf{P}(X = x)$	$(1-p)^n$	$np\,(1-p)^{n-1}$	$\frac{n(n-1)}{2}\, p^2\, (1-p)^{n-2}$	$\ldots$	p^n

…then X has the same distribution as $Y_1 + \ldots + Y_n$
where each Y_i has Bernoulli distribution with parameter p.

But $\mathsf{E}(Y_1 + \ldots + Y_n) = \mathsf{E}\, Y_1 + \ldots + \mathsf{E}\, Y_n = n \cdot p$, so $\mathsf{E}\, X = n \cdot p$.

Properties of expectation: Summary

$$
\begin{aligned}
\mathsf{E}(\mathbf{1}_A) &= \mathsf{P}(A) && \text{for any event } A \\
\mathsf{E}(X+Y) &= \mathsf{E}\,X + \mathsf{E}\,Y \\
\mathsf{E}(cX) &= c \cdot \mathsf{E}\,X && \text{for any constant } c \\
\mathsf{E}\,c &= c && \text{for any constant } c \\
\mathsf{E}\,X &\geq \mathsf{E}\,Y \quad \text{if } X \geq Y \\
\mathsf{E}\,f(X) &= \sum_x f(x) \cdot \mathsf{P}(X=x) \\
\text{If } \mathsf{E}\,X &= 0 \text{ and } X \geq 0, \text{ then } \mathsf{P}(X=0) = 1.
\end{aligned}
$$

Variance

$$\mathsf{Var}\, X = \mathsf{E}(X - \mathsf{E}\, X)^2 \geq 0$$

Variance

$$\mathsf{Var}\, X = \mathsf{E}(X - \mathsf{E}\, X)^2 \geq 0$$

$$\begin{aligned}\mathsf{Var}\, X &= \mathsf{E}(X^2 - 2X \cdot \mathsf{E}\, X + (\mathsf{E}\, X)^2) \\ &= \mathsf{E}\, X^2 - 2\ \mathsf{E}\, X \cdot \mathsf{E}\, X + (\mathsf{E}\, X)^2 \\ &= \mathsf{E}\, X^2 - (\mathsf{E}\, X)^2\end{aligned}$$

Variance

$$\operatorname{Var} X = \mathsf{E}(X - \mathsf{E}\, X)^2 \geq 0$$

$$\begin{aligned}
\operatorname{Var} X &= \mathsf{E}(X^2 - 2X \cdot \mathsf{E}\, X + (\mathsf{E}\, X)^2) \\
&= \mathsf{E}\, X^2 - 2\, \mathsf{E}\, X \cdot \mathsf{E}\, X + (\mathsf{E}\, X)^2 \\
&= \mathsf{E}\, X^2 - (\mathsf{E}\, X)^2
\end{aligned}$$

$\operatorname{Var} X = 0$ iff there exists a constant c such that $\mathsf{P}(X = c) = 1$.

For all c we have $\operatorname{Var}(cX) = c^2 \cdot \operatorname{Var} X$ and $\operatorname{Var}(X + c) = \operatorname{Var} X$.

Variance of the sum: Example

Let X and Y be Bernoulli random variables associated to independent Bernoulli trials with parameter $1/2$.
Define $Z = 1 - X$.

$$\mathsf{E}(X+Y) = \mathsf{E}\,X + \mathsf{E}\,Y = 1$$
$$\mathsf{E}(X+Z) = \mathsf{E}\,X + \mathsf{E}\,Z = 1$$

Variance of the sum: Example

Let X and Y be Bernoulli random variables associated to independent Bernoulli trials with parameter $1/2$.
Define $Z = 1 - X$.

$$\mathsf{E}(X + Y) = \mathsf{E}\,X + \mathsf{E}\,Y = 1$$
$$\mathsf{E}(X + Z) = \mathsf{E}\,X + \mathsf{E}\,Z = 1$$

In fact, $X + Z \equiv 1$.

Variance of the sum: Example

Let X and Y be Bernoulli random variables associated to independent Bernoulli trials with parameter $1/2$.
Define $Z = 1 - X$.

$$\mathsf{E}(X+Y) = \mathsf{E}\,X + \mathsf{E}\,Y = 1$$
$$\mathsf{E}(X+Z) = \mathsf{E}\,X + \mathsf{E}\,Z = 1$$

In fact, $X + Z \equiv 1$.

k	0	1	2
$\mathsf{P}(X+Y=k)$	$1/4$	$1/2$	$1/4$
$\mathsf{P}(X+Z=k)$	0	1	0

Variance of the sum: Example

Let X and Y be Bernoulli random variables associated to independent Bernoulli trials with parameter $1/2$.
Define $Z = 1 - X$.

$$\mathsf{E}(X+Y) = \mathsf{E}\,X + \mathsf{E}\,Y = 1$$
$$\mathsf{E}(X+Z) = \mathsf{E}\,X + \mathsf{E}\,Z = 1$$

In fact, $X + Z \equiv 1$.

k	0	1	2
$\mathsf{P}(X+Y=k)$	$1/4$	$1/2$	$1/4$
$\mathsf{P}(X+Z=k)$	0	1	0

$$\mathsf{Var}(X+Y) = \mathsf{E}(X+Y-1)^2 = 1/2$$
$$\mathsf{Var}(X+Z) = \mathsf{E}(X+Z-1)^2 = 0$$

Variance of the sum

$$
\begin{aligned}
\mathsf{Var}(X+Y) &= \mathsf{E}\left((X+Y)-\mathsf{E}(X+Y)\right)^2 \\
&= \mathsf{E}\left((X-\mathsf{E}\,X)+(Y-\mathsf{E}\,Y)\right)^2 \\
&= \mathsf{E}\left((X-\mathsf{E}\,X)^2+2\,(X-\mathsf{E}\,X)(Y-\mathsf{E}\,Y)+(Y-\mathsf{E}\,Y)^2\right) \\
&= \mathsf{E}(X-\mathsf{E}\,X)^2+\mathsf{E}(Y-\mathsf{E}\,Y)^2+2\,\mathsf{E}(X-\mathsf{E}\,X)(Y-\mathsf{E}\,Y) \\
&= \mathsf{Var}\,X+\mathsf{Var}\,Y+2\,\mathsf{E}(XY-X\,\mathsf{E}\,Y-Y\,\mathsf{E}\,X+\mathsf{E}\,X\cdot\mathsf{E}\,Y) \\
&= \mathsf{Var}\,X+\mathsf{Var}\,Y+2(\mathsf{E}\,XY-\mathsf{E}\,X\cdot\mathsf{E}\,Y)
\end{aligned}
$$

Variance of the sum

$$\begin{aligned}
\mathsf{Var}(X+Y) &= \mathsf{E}\left((X+Y)-\mathsf{E}(X+Y)\right)^2 \\
&= \mathsf{E}\left((X-\mathsf{E}\,X)+(Y-\mathsf{E}\,Y)\right)^2 \\
&= \mathsf{E}\left((X-\mathsf{E}\,X)^2+2\,(X-\mathsf{E}\,X)(Y-\mathsf{E}\,Y)+(Y-\mathsf{E}\,Y)^2\right) \\
&= \mathsf{E}(X-\mathsf{E}\,X)^2+\mathsf{E}(Y-\mathsf{E}\,Y)^2+2\,\mathsf{E}(X-\mathsf{E}\,X)(Y-\mathsf{E}\,Y) \\
&= \mathsf{Var}\,X+\mathsf{Var}\,Y+2\,\mathsf{E}(XY-X\,\mathsf{E}\,Y-Y\,\mathsf{E}\,X+\mathsf{E}\,X\cdot\mathsf{E}\,Y) \\
&= \mathsf{Var}\,X+\mathsf{Var}\,Y+2(\mathsf{E}\,XY-\mathsf{E}\,X\cdot\mathsf{E}\,Y)
\end{aligned}$$

The difference $\mathsf{E}\,XY-\mathsf{E}\,X\cdot\mathsf{E}\,Y$ is called the **covariance** of X and Y, denoted $\mathsf{Cov}(X,Y)$.

Independence of random variables

Recall that two events A and B are called **independent** if $\mathsf{P}(AB) = \mathsf{P}(A)\,\mathsf{P}(B)$.

Two (discrete) random variables X and Y are **independent** if for any values x and y the events $X = x$ and $Y = y$ are independent.

Independence of random variables, continued

Two (discrete) random variables X and Y are **independent** if for any values x and y the events $X = x$ and $Y = y$ are independent.

In particular, if X and Y are independent, then

$$\begin{aligned}
\mathsf{E}\,XY &= \sum_k k \cdot \mathsf{P}(XY = k) \\
&= \sum_k k \cdot \sum_{xy=k} \mathsf{P}(X = x, Y = y) \\
&= \sum_k k \cdot \sum_{xy=k} \mathsf{P}(X = x)\,\mathsf{P}(Y = y) \\
&= \sum_x x\,\mathsf{P}(X = x) \sum_y y\,\mathsf{P}(Y = y) \\
&= \left(\sum_x x\,\mathsf{P}(X = x)\right) \cdot \left(\sum_y y\,\mathsf{P}(Y = y)\right) = \mathsf{E}\,X \cdot \mathsf{E}\,Y
\end{aligned}$$

Independence of random variables, continued

Two (discrete) random variables X and Y are **independent** if for any values x and y the events $X = x$ and $Y = y$ are independent.

In particular, if X and Y are independent, then

$$\mathsf{E}\,XY = \mathsf{E}\,X \cdot \mathsf{E}\,Y$$

Independence of random variables, continued

Two (discrete) random variables X and Y are **independent** if for any values x and y the events $X = x$ and $Y = y$ are independent.

In particular, if X and Y are independent, then

$$\begin{aligned} \mathsf{E}\, XY &= \mathsf{E}\, X \cdot \mathsf{E}\, Y \\ \mathsf{Var}(X+Y) &= \mathsf{Var}\, X + \mathsf{Var}\, Y + 2(\mathsf{E}\, XY - \mathsf{E}\, X \cdot \mathsf{E}\, Y) \\ &= \mathsf{Var}\, X + \mathsf{Var}\, Y \end{aligned}$$

Independence of random variables, continued

Two (discrete) random variables X and Y are **independent** if for any values x and y the events $X = x$ and $Y = y$ are independent.

In particular, if X and Y are independent, then

$$\mathsf{Var}(X + Y) = \mathsf{Var}\, X + \mathsf{Var}\, Y.$$

Independence of random variables, continued

Two (discrete) random variables X and Y are **independent** if for any values x and y the events $X = x$ and $Y = y$ are independent.

In particular, if X and Y are independent, then

$$\mathsf{Var}(X + Y) = \mathsf{Var}\, X + \mathsf{Var}\, Y.$$

In general,

$$\mathsf{Var}(X + Y) = \mathsf{Var}\, X + \mathsf{Var}\, Y + 2\, \mathsf{Cov}(X, Y).$$

Variance: Summary

$$\mathsf{Var}\, X = \mathsf{E}(X - \mathsf{E}\, X)^2 \geq 0$$
$$\mathsf{Var}\, X = \mathsf{E}\, X^2 - (\mathsf{E}\, X)^2$$

$\mathsf{Var}\, X = 0$ iff there exists a constant c such that $\mathsf{P}(X = c) = 1$.

For all c we have $\mathsf{Var}(cX) = c^2 \cdot \mathsf{Var}\, X$ and $\mathsf{Var}(X + c) = \mathsf{Var}\, X$.

$$\mathsf{Var}(X + Y) = \mathsf{Var}\, X + \mathsf{Var}\, Y + 2\, \mathsf{Cov}(X, Y)$$
$$(\mathsf{Cov}(X, Y) = 0 \text{ if } X \text{ and } Y \text{ are independent})$$

Why variance?

Chebyshev inequality:

$$\mathsf{P}\big(\,|X - \mathsf{E}\,X| \geq t\,\big) \leq \frac{\mathsf{Var}\,X}{t^2}$$

Geometric distribution

Let $X_1, X_2, \ldots, X_n, \ldots$ be independent Bernoulli random variables with parameter p.
Call trial i a success if $X_i = 1$ and a failure otherwise.
Denote $q = 1 - p = \mathsf{P}(X_i = 0)$.

Geometric distribution

Let $X_1, X_2, \dots, X_n, \dots$ be independent Bernoulli random variables with parameter p.
Call trial i a success if $X_i = 1$ and a failure otherwise.
Denote $q = 1 - p = \mathsf{P}(X_i = 0)$.

Let Y denote the number of failures before the first success.
Random variable Y is said to have **geometric distribution** with parameter p.

y	0	1	2	$\dots$	n	$\dots$
$\mathsf{P}(Y = y)$	p	pq	pq^2	$\dots$	pq^n	$\dots$

Geometric distribution: Properties

y	0	1	2	$\ldots$	n	$\ldots$
$\mathsf{P}(Y = y)$	p	pq	pq^2	$\ldots$	pq^n	$\ldots$

$$\mathsf{E}\,X = \frac{q}{p}$$

$$\mathsf{Var}\,X = \frac{q}{p^2}$$

Summary of today's lecture

- Computational complexity:
  Decision and counting problems, complexity classes $\mathbf{P}$, $\mathbf{NP}$, and $\#\mathbf{P}$
- Probability theory: Measures, events, random variables, probability distributions

Agenda

Tuesday	computational complexity, probability theory
Wednesday	randomized algorithms, Monte Carlo methods
Thursday	hashing-based approach to model counting
Friday	from discrete to continuous model counting