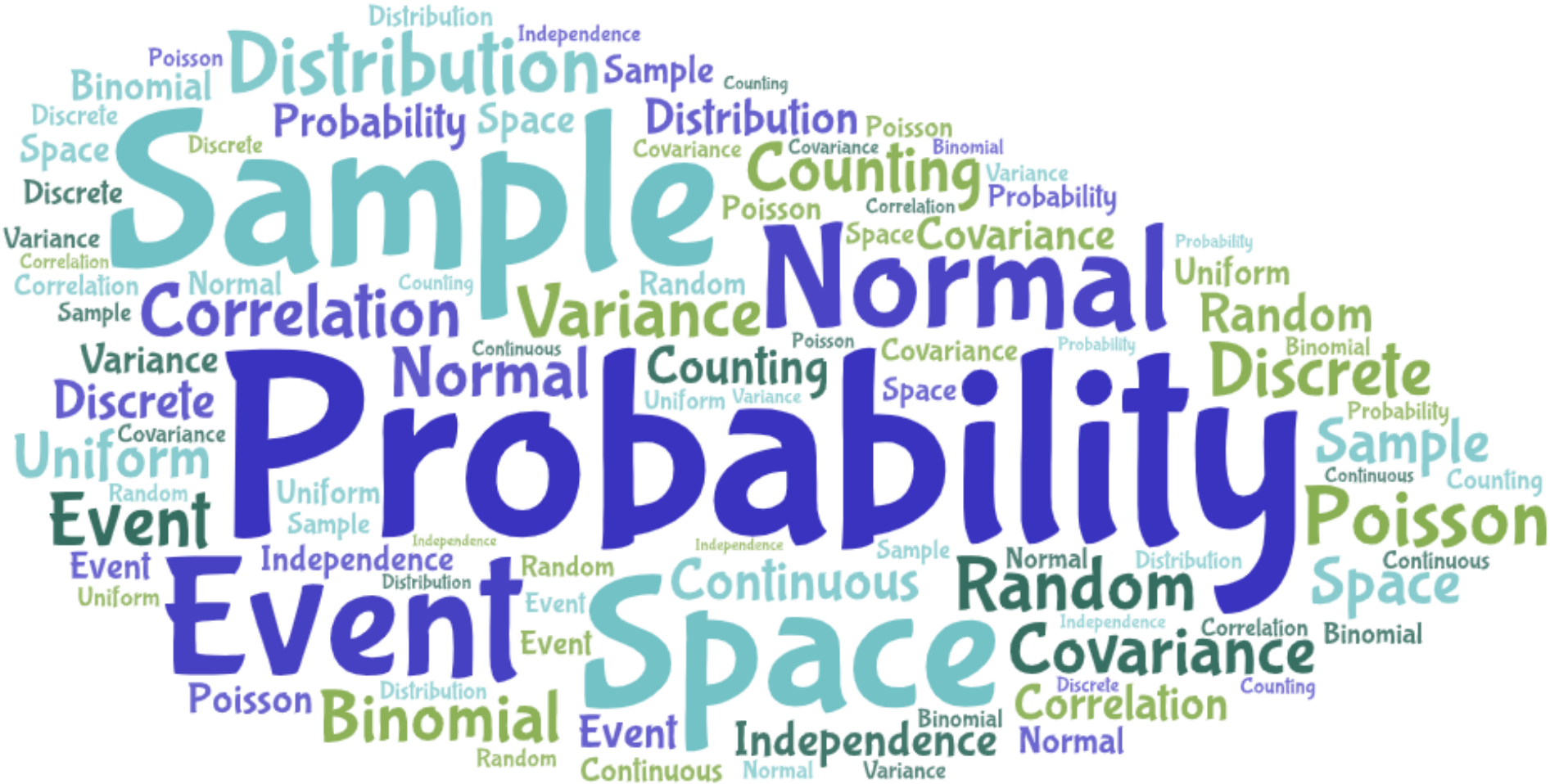
Discrete Mathematics and Probability

Week 7 and 8



Markulf Kohlweiss

Inclusion-exclusion

Proposition

For any events:

$$\begin{aligned}\mathbf{P}(E \cup F \cup G) = \mathbf{P}(E) + \mathbf{P}(F) + \mathbf{P}(G) \\ - \mathbf{P}(E \cap F) - \mathbf{P}(E \cap G) - \mathbf{P}(F \cap G) \\ + \mathbf{P}(E \cap F \cap G).\end{aligned}$$

Inclusion-exclusion

Proposition

For any events:

$$\begin{aligned}\mathbf{P}(E \cup F \cup G) = \mathbf{P}(E) + \mathbf{P}(F) + \mathbf{P}(G) \\ - \mathbf{P}(E \cap F) - \mathbf{P}(E \cap G) - \mathbf{P}(F \cap G) \\ + \mathbf{P}(E \cap F \cap G).\end{aligned}$$

$$\begin{aligned}\mathbf{P}(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{1 \le i \le n} \mathbf{P}(E_i) \\ - \sum_{1 \le i_1 < i_2 \le n} \mathbf{P}(E_{i_1} \cap E_{i_2}) \\ + \sum_{1 \le i_1 < i_2 < i_3 \le n} \mathbf{P}(E_{i_1} \cap E_{i_2} \cap E_{i_3}) \\ - \cdots \\ + (-1)^{n+1} \mathbf{P}(E_1 \cap E_2 \cap \cdots \cap E_n).\end{aligned}$$

Example

Example

In a sports club,

36 members play tennis, 22 play tennis and squash,
28 play squash, 12 play tennis and badminton,
18 play badminton, 9 play squash and badminton,
4 play tennis, squash and badminton.

How many play at least one of these games?

Topics

- Recap: examples with equally likely outcomes
- Conditional probability: how knowledge influences probability
- Bayes' theorem: link probabilities of related events

Equally likely outcomes

The return of counting

Finite sample space, $|\Omega| = N < \infty$, has special important case where each experiment outcome has *equal probability*:

$$\mathbf{P}(\omega) = \frac{1}{N} \qquad \text{for all } \omega \in \Omega$$

Definition

Outcomes $\omega \in \Omega$ are also called *elementary events*.

Example

Example

Rolling two dice, what is the probability that the sum of the numbers shown is 7?

What's wrong with this solution? "The number 7 is one out of the possible values $2, 3, \ldots, 12$ for the sum, and the answer is $\frac{1}{11}$."

Conditional probability

Conditional probability

Often you have *partial information* about the outcome of an experiment. This alters the likelihoods for various outcomes.

Example

Roll two dice. What is the probability that the sum of the numbers is 8? What if we know that the first die shows a 5?

Reduced sample space

We reduced our world to the event we were given:
$F = \{\text{first die shows } 5\} = \{(5, 1), (5, 2), \ldots, (5, 6)\}$

Definition

The event that is given to us is called a *reduced sample space*. We can simply work in this set to figure out the conditional probabilities given this event.

The event F has 6 equally likely outcomes. Only one of them, $(5, 3)$, provides a sum of 8. Hence the conditional probability is $\frac{1}{6}$.

Definition of conditional probability

The question can be reformulated.

$$E = \{\text{the sum is } 8\} = \{(2, 6), (3, 5), \ldots, (6, 2)\}$$

"In what proportion of cases in F will E also occur?"
"How does probability of 'E and F' compare to probability of F?"

Definition

Let F be an event with $\mathbf{P}(F) > 0$.
The *conditional probability of E given F* is:

$$\mathbf{P}(E \mid F) = \frac{\mathbf{P}(E \cap F)}{\mathbf{P}(F)}$$

Axioms

Proposition

Conditional probability $\mathbf{P}(\cdot \,|F)$ *satisfies the axioms of probability:*

1. *conditional probability is non-negative:* $\mathbf{P}(E|\,) \geq 0$;
2. *conditional probability of sample space is one:* $\mathbf{P}(\Omega \,|\, F) = 1$;
3. *for countably many mutually exclusive events* E_1, E_2, ...:

$$\mathbf{P}\Big(\bigcup_i E_i \,\Big|\, F\Big) = \sum_i \mathbf{P}(E_i \,|\, F)$$

How to compute conditional probabilities

Corollary

- $\mathbf{P}(E^c \mid F) = 1 - \mathbf{P}(E \mid F)$
- $\mathbf{P}(\emptyset \mid F) = 0$
- $\mathbf{P}(E \mid F\} = 1 - \mathbf{P}(E^c \mid F) \leq 1$
- $\mathbf{P}(E \cup G \mid F) = \mathbf{P}(E \mid F) + \mathbf{P}(G \mid F) - \mathbf{P}(E \cap G \mid F)$
- *If* $E \subseteq G$, *then* $\mathbf{P}(G - E \mid F) = \mathbf{P}(G \mid F) - \mathbf{P}(E \mid F)$
- *If* $E \subseteq G$, *then* $\mathbf{P}(E \mid F) \leq \mathbf{P}(G \mid F)$

BUT: Don't change the condition!
$\mathbf{P}\{E \mid F\}$ and $\mathbf{P}\{E \mid F^c\}$ have nothing to do with each other.

Multiplication rule

Proposition (*Multiplication rule*)

$$\mathbf{P}(E_1 \cap \cdots \cap E_n) = \mathbf{P}(E_1) \cdot \mathbf{P}(E_2 \mid E_1) \cdot \mathbf{P}(E_3 \mid E_1 \cap E_2) \cdots \mathbf{P}(E_n \mid E_1 \cap \cdots \cap E_{n-1})$$

Example again

Example

An urn contains 6 red and 5 blue balls. We draw three balls at random, at once (that is, without replacement). What is the chance of drawing one red and two blue balls?

Bayes' theorem

Bayes' Theorem

The aim is to say something about $\mathbf{P}(F \mid E)$, once we know $\mathbf{P}(E \mid F)$ (and other things...). This will be very useful, and serve as a fundamental tool in probability and statistics.

The Law of Total Probability

Theorem (*Partition Theorem*)

$$\mathbf{P}(E) = \mathbf{P}(E \mid F) \cdot \mathbf{P}(F) + \mathbf{P}(E \mid F^c) \cdot \mathbf{P}(F^c)$$

The Law of Total Probability

Theorem (*Partition Theorem*)

$$\mathbf{P}(E) = \mathbf{P}(E \mid F) \cdot \mathbf{P}(F) + \mathbf{P}(E \mid F^c) \cdot \mathbf{P}(F^c)$$

Definition

Countably many events $F_1, F_2, \ldots$form a *partition* of Ω if $F_i \cap F_j = \emptyset$ and $\bigcup_i F_i = \Omega$.

Theorem (*Partition Theorem*)

For any event E and any partition $F_1, F_2, \ldots$:

$$\mathbf{P}(E) = \sum_i \mathbf{P}(E \mid F_i) \cdot \mathbf{P}(F_i)$$

Example

Example

According to an insurance company:

- 30% of population are *accident-prone*:
  they will have an accident in any given year with 0.4 chance.
- 70% of population are *careful*:
  they have an accident in any given year with 0.2 chance.

How likely is a new customer to have an accident in 2023?

Bayes' Theorem

Theorem (Bayes' Theorem)

$$\mathbf{P}\{F \mid E\} = \frac{\mathbf{P}\{E \mid F\} \cdot \mathbf{P}\{F\}}{\mathbf{P}\{E \mid F\} \cdot \mathbf{P}\{F\} + \mathbf{P}\{E \mid F^c\} \cdot \mathbf{P}\{F^c\}}$$

If $\{F_i\}_i$ *partitions* Ω*, then:*

$$\mathbf{P}\{F_i \mid E\} = \frac{\mathbf{P}\{E \mid F_i\} \cdot \mathbf{P}\{F_i\}}{\sum_j \mathbf{P}\{E \mid F_j\} \cdot \mathbf{P}\{F_j\}}$$

Belief update

Example

Consider the insurance company again. Imagine it's now 2024. We learn that the new customer did have an accident in 2023. Now what is the chance that they are accident-prone?

Summary

- Probability: multiple ways to compute
- Conditional probability: reduced sample space, multiplication rule
- Bayes' theorem: partition theorem, belief update

Topics

- Independence: what information changes probability
- Random variables: when variables depend on chance
- Expectation: most likely outcomes of experiment
- Variance: how much the experiment can deviate

Independence

Independence

Sometimes partial information on an experiment does not change the likelihood of an event.

Definition

Events E and F are *independent* if $\mathbf{P}(E \mid F) = \mathbf{P}(E)$.
Equivalently: $\mathbf{P}(E \cap F) = \mathbf{P}(E) \cdot \mathbf{P}(F)$.
Equivalently: $\mathbf{P}(F \mid E) = \mathbf{P}(F)$.

Proposition

If E and F are independent events, then E and F^c are also independent.

Examples

Independence

Definition

Three events E, F, G are *(mutually) independent* if:

$$\begin{aligned}
\mathbf{P}\{E \cap F\} &= \mathbf{P}\{E\} \cdot \mathbf{P}\{F\}, \\
\mathbf{P}\{E \cap G\} &= \mathbf{P}\{E\} \cdot \mathbf{P}\{G\}, \\
\mathbf{P}\{F \cap G\} &= \mathbf{P}\{F\} \cdot \mathbf{P}\{G\}, \\
\mathbf{P}\{E \cap F \cap G\} &= \mathbf{P}\{E\} \cdot \mathbf{P}\{F\} \cdot \mathbf{P}\{G\}.
\end{aligned}$$

For more events the definition is that any (finite) subset of them have this factorisation property.

Examples

Random variables

Random variables

"Random variables $\approx$ random numbers". But *random* means that there must be some kind of experiment behind these numbers.

Definition

A *random variable* is a function from the sample space Ω to the real numbers $\mathbb{R}$.

Discrete random variables

Definition

A random variable X that can take on countably many possible values is called *discrete*.

Probability mass function

Definition

The *probability mass function (pmf)*, or *distribution* of a discrete random variable X gives the probabilities of its possible values:

$$\mathfrak{p}_X(x_i) = \mathbf{P}(X = x_i),$$

Proposition

$$\mathfrak{p}(x_i) \geq 0 \qquad \textit{and} \qquad \sum_i \mathfrak{p}(x_i) = 1$$

Examples

Cumulative distribution function

Definition

The *cumulative distribution function (cdf)* of a random variable X:

$$F\colon \mathbb{R} \to [0, 1], \qquad x \mapsto F(x) = \mathbf{P}(X \leq x).$$

Cumulative distribution function

Proposition

A cumulative distribution function F:

- *is non-decreasing: if $x \leq y$ then $F(x) \leq F(y)$*
- *has limit $\lim_{x \to -\infty} F(x) = 0$ on the left*
- *has limit $\lim_{x \to \infty} F(x) = 1$ on the right*

Expectation

Expectation

Once we have a random variable, we want to quantify its *typical* behaviour in some sense. Two of the most often used quantities for this are the *expectation* and the *variance*.

Definition

The *expectation* of a discrete random variable X is:

$$\mathbf{E}X = \sum_i x_i \cdot \mathfrak{p}(x_i)$$

provided the sum exists. Also called *mean*, or *expected value*.

Examples

Properties of expectation

Proposition (expectation of a function of a random variable)

If X is a discrete random variable, and $g\colon \mathbb{R} \to \mathbb{R}$ a function, then:

$$\mathbf{E}g(X) = \sum_i g(x_i) \cdot \mathfrak{p}(x_i) \qquad \textit{(if it exists)}$$

Corollary (expectation is linear)

If X is a discrete random variable, and a, b fixed real numbers:

$$\mathbf{E}(aX + b) = a \cdot \mathbf{E}X + b.$$

Moments

Definition (moments)

Let $n \in \mathbb{N}$. The n^{th} *moment* of a random variable X is:

$$\mathbf{E}X^n$$

The n^{th} *absolute moment* of X is:

$$\mathbf{E}|X|^n$$

Variance

Example

3. Variance

Definition (variance, standard deviation)

The *variance* and the *standard deviation* of a random variable are:

- $\mathbf{Var}\, X = \mathbf{E}(X - \mathbf{E}X)^2$.
- $\mathrm{SD}\, X = \sqrt{\mathbf{Var}\, X}$.

Example

Properties of the variance

Proposition (equivalent form of the variance)

$\mathbf{Var}\, X = \mathbf{E}X^2 - (\mathbf{E}X)^2$ *for any random variable* X.

Corollary

$\mathbf{E}X^2 \geq (\mathbf{E}X)^2$ *for any random variable* X,
with equality only if X *is constant.*

Examples

Properties of the variance

Proposition (variance is not linear)

Let X be a random variable, a and b fixed real numbers. Then:

$$\mathbf{Var}(aX + b) = a^2 \cdot \mathbf{Var}\, X$$

Summary

- Independence: what information changes probability
- Random variables: when variables depend on chance
- Expectation: most likely outcomes of experiment
- Variance: how much the experiment can deviate