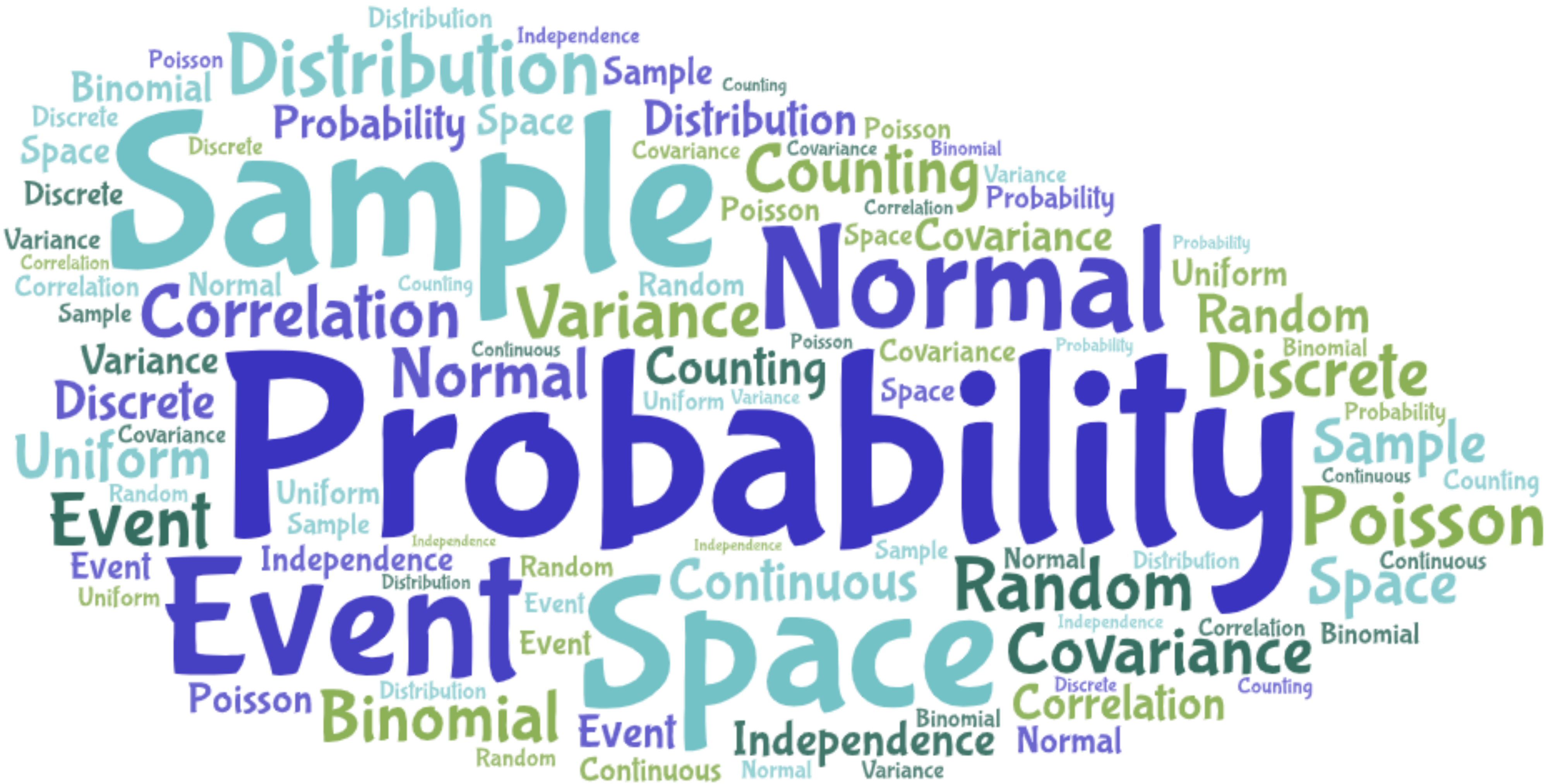
Discrete Mathematics and Probability

Week 8



Markulf ~~Chris Heunen~~ Kohlweiss

Multiplication rule

Proposition (*Multiplication rule*)

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

$= P(E_1) \cdot P(E_2) \cdots P(E_n)$

only if E_i are mutually independent

~~linear~~

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?

$$P(R_1 \wedge B_2 \wedge B_3) + P(B_1 \wedge R_2 \wedge B_3) + P(B_1 \wedge B_2 \wedge R_3)$$

$$= P(R_1) \cdot P(B_2 | R_1) \cdot P(B_3 | R_1 \wedge B_2)$$

$$+ P(B_1) \cdot P(R_2 | B_1) \cdot P(B_3 | B_1 \wedge R_2)$$

$$+ P(B_1) \cdot P(B_2 | B_1) \cdot P(R_3 | B_1 \wedge B_2)$$

$$= \frac{6}{11} \cdot \frac{5}{10} \cdot \frac{4}{9} + \frac{5}{11} \cdot \frac{6}{10} \cdot \frac{4}{9} + \frac{5}{11} \cdot \frac{4}{10} \cdot \frac{6}{9} = \frac{\cancel{3} \cdot \cancel{5} \cdot 4 \cdot \cancel{6}}{11 \cdot \cancel{10} \cdot \cancel{9}} = \frac{4}{11}$$

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*)

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

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 $F_1, F_2, \ldots$:

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

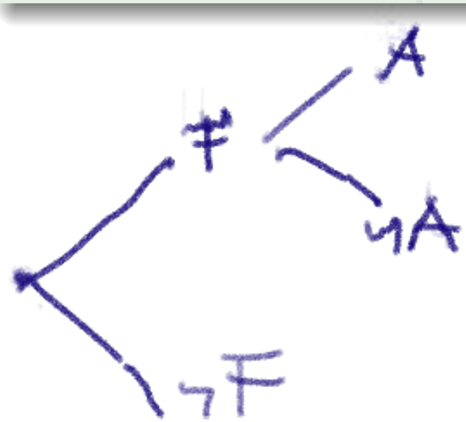
Example

Example

According to an insurance company:

- 30% of population are *accident-prone*: F
  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 A in 2023?



can be calculated using Law of total probability.

Example

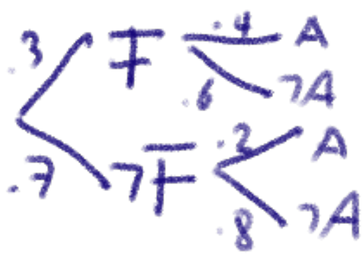
Example

According to an insurance company: F

- 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?

A



$P(A) = P(A|F)\cdot P(F) + P(A|F^c)\cdot P(F^c)$

$= 0.4 \cdot 0.3 + 0.2 \cdot 0.7 = 0.26$

Initial probability of new customer being accident prone is $P(F)=0.3$
what is probability after customer had an accident $P(F|A)$?

Bayes' Theorem

Theorem (Bayes' Theorem)

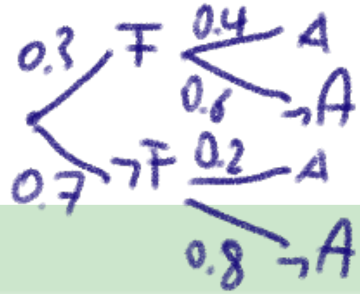
$$\mathbf{P}\{F \mid E\} = \frac{\mathbf{P}\{E \mid F\} \cdot \mathbf{P}\{F\}}{\mathbf{P}\{E \mid F\} \cdot \mathbf{P}\{F\} + \mathbf{P}\{E \mid F^c\} \cdot \mathbf{P}\{F^c\}}$$

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

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

Proof: Def of conditional prob + Law of total probability

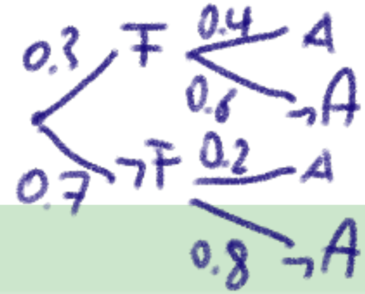
Belief update



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?

Belief update



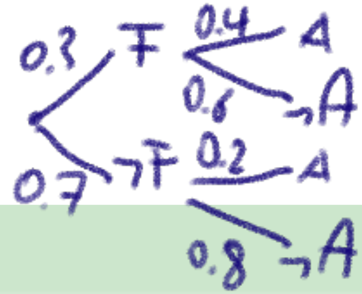
Example

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?

$$P(F \mid A) = \frac{P(A \mid F) \cdot P(F)}{P(A \mid F) \cdot P(F) + P(A \mid F^c) \cdot P(F^c)}$$

$$= \frac{0.4 \cdot 0.3}{0.4 \cdot 0.3 + 0.2 \cdot 0.7}$$

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?

$$P(F \mid A) = \frac{P(A \mid F) \cdot P(F)}{P(A \mid F) \cdot P(F) + P(A \mid F^c) \cdot P(F^c)}$$

$$= \frac{0.4 \cdot 0.3}{0.4 \cdot 0.3 + 0.2 \cdot 0.7} = 46\%$$

instead of

$P(F) = 0.3$ 30%

Summary

practice, otherwise it takes you too long, like when I calculate in this lecture.

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

Topics

Experiments with numerical outcome.

- 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

stretch goal

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.

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

Always give you some example propositions.

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

independent events $\neq$ mutually exclusive events

independence usually either trivial or tricky

Examples

$$\Omega = \text{two dice}$$

$$\left.\begin{aligned} E &= \{\text{sum is } 6\} \\ F &= \{\text{first is } 3\} \end{aligned}\right\} \text{not independent}$$

$$\frac{1}{36} = P(E \cap F) \neq P(E) \cdot P(F) = \frac{5}{36} \cdot \frac{1}{6}$$

Examples

Ω = two dice

$E = \{\text{sum is } 6\}$

$F = \{\text{first is } 3\}$

$\frac{1}{36} = P(E \cap F) \neq P(E) \cdot P(F) = \frac{5}{36} \cdot \frac{1}{6}$

not independent

$E' = \{\text{sum is } 7\}$

$F = \{\text{first is } 3\}$

$\frac{1}{36} = P(E' \cap F) = P(E') \cdot P(F) = \frac{\cancel{6}}{36} \cdot \frac{1}{\cancel{6}}$

independent!

Examples

Ω = two dice

$E = \{\text{sum is } 6\}$

$F = \{\text{first is } 3\}$

$\frac{1}{36} = P(E \cap F) \neq P(E) \cdot P(F) = \frac{5}{36} \cdot \frac{1}{6}$

not independent

$E' = \{\text{sum is } 7\}$

$F = \{\text{first is } 3\}$

$\frac{1}{36} = P(E' \cap F) = P(E') \cdot P(F) = \frac{\cancel{6}}{36} \cdot \frac{1}{\cancel{6}}$

independent!

$G = \{\text{second is } 4\}$

pairwise independent

$\frac{1}{36} = P(E' \cap G) = P(E') \cdot P(G) = \frac{6}{36} \cdot \frac{1}{6}$

$\frac{1}{36} = P(F \cap G) = P(F) \cdot P(G) = \frac{1}{6} \cdot \frac{1}{6}$

Examples

Ω = two dice

$E = \{\text{sum is } 6\}$

$F = \{\text{first is } 3\}$

$\frac{1}{36} = P(E \cap F) \neq P(E) \cdot P(F) = \frac{5}{36} \cdot \frac{1}{6}$

not independent

$E' = \{\text{sum is } 7\}$

$F = \{\text{first is } 3\}$

$\frac{1}{36} = P(E' \cap F) = P(E') \cdot P(F) = \frac{6}{36} \cdot \frac{1}{6}$

independent!

$G = \{\text{second is } 4\}$

pairwise independent

$\frac{1}{36} = P(E' \cap G) = P(E') \cdot P(G) = \frac{6}{36} \cdot \frac{1}{6}$

$\frac{1}{36} = P(F \cap G) = P(F) \cdot P(G) = \frac{1}{6} \cdot \frac{1}{6}$

But: $\frac{1}{36} = P(E' \cap F \cap G) \overset{?}{=} P(E') \cdot P(F) \cdot P(G) = \frac{6}{36} \cdot \frac{1}{6} \cdot \frac{1}{6}$

Kind of makes sense: degrees of freedom

Independence

Definition

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

$$\mathbf{P}\{E \cap F\} = \mathbf{P}\{E\} \cdot \mathbf{P}\{F\},$$
$$\mathbf{P}\{E \cap G\} = \mathbf{P}\{E\} \cdot \mathbf{P}\{G\},$$
$$\mathbf{P}\{F \cap G\} = \mathbf{P}\{F\} \cdot \mathbf{P}\{G\},$$
$$\mathbf{P}\{E \cap F \cap G\} = \mathbf{P}\{E\} \cdot \mathbf{P}\{F\} \cdot \mathbf{P}\{G\}.$$

For more events the definition is that any (finite) subset of them have this factorisation property. have to go all the way to n

Examples

Examples

Fix parameter $0 < p < 1$

n independent experiments, each succeeds with probability p

- chance that every one succeeds?

Examples

Fix parameter $0 < p < 1$

n independent experiments, each succeeds with probability p

- chance that every one succeeds?

$$p^n \xrightarrow[n \to \infty]{} 0$$

- chance that at least one succeeds?

$$1-(1-p)^n \xrightarrow[n \to \infty]{} 1$$

Murphy's Law

Examples

Fix parameter $0 < p < 1$

n independent experiments, each succeeds with probability p

- chance that every one succeeds?

$$p^n \xrightarrow[n \to \infty]{} 0$$

- chance that at least one succeeds?

$$1 - P(\text{each one fails}) = 1 - (1-p)^n \xrightarrow[n \to \infty]{} 1$$

Murphy's Law

- chance that exactly k succeed?

$$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

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 $\mathbb{R}$.

flip 3 coins $\quad \Omega = \{H,T\}^3 \quad$ X number of Hs

$X(TTT) = 0$

$X(HTT) = X(THT) = X(TTH) = 1$

$X(HHT) = X(HTH) = X(THH) = 2$

$X(HHH) = 3$

$P(X=1) = P(X^{-1}(1)) = P(\{HTT, THT, TTH\}) = \frac{3}{8}$

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 $\mathbb{R}$.

flip 3 coins X = random variable that counts heads

not between 0 or 1

$X : \Omega \to \mathbb{R}$

$X(TTT) = 0$

$X(HTT) = X(THT) = X(TTH) = 1$

$X(HHT) = X(HTH) = X(THH) = 2$

$X(HHH) = 3$

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 $\mathbb{R}$.

flip 3 coins X = random variable that counts heads

not between 0 or 1

$X : \Omega \to \mathbb{R}$

$X(TTT) = 0$

$X(HTT) = X(THH) = X(HHT) = 1$

$X(HHT) = X(HTH) = X(THH) = 2$

$X(HHH) = 3$

$P(X=1) = P(X^{-1}(1)) = P(\{HTT, THT, TTH\}) = \frac{3}{8}$

Discrete random variables

Definition

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

Discrete random variables

Definition

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

e.g. – nr. heads in 3 coin flips

– nr. of flips needed to ~~find~~ see a head

discrete things I can always count.
don't need to integrate

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

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

Vice versa: any function $f : \mathbb{N} \to \mathbb{R}$ s.t. $f(x) \geq 0$, $\Sigma f(x) = 1$ is always a pmf

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

Examples

Fix parameter $\lambda > 0$

Define $p(i) = c \cdot \frac{\lambda^i}{i!}$

which c makes this a pmf?

$p(i) \geqslant 0 \iff c \geqslant 0$

Examples

Fix parameter $\lambda > 0$

Define $p(i) = c \cdot \frac{\lambda^i}{i!}$

which c makes this a pmf?

$p(i) \geq 0 \iff c \geq 0$

$$\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{\infty} c \cdot \frac{\lambda^i}{i!} = c \cdot \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = c \cdot e^{\lambda} \iff c = e^{-\lambda}$$

so e.g. $P(X=0) = p(0) = e^{-\lambda} \cdot \frac{\lambda^0}{0!} = e^{-\lambda}$

$$\begin{aligned} P(X>2) &= 1 - P(X \leq 2) \\ &= 1 - P(X=0) - P(X=1) - P(X=2) \\ &= 1 - e^{-\lambda} - e^{-\lambda}\lambda - e^{-\lambda}\frac{\lambda^2}{2} \end{aligned}$$

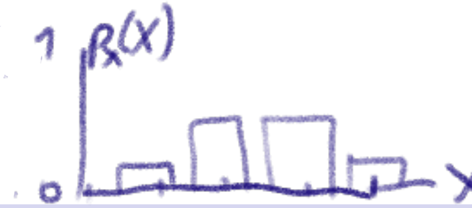
Cumulative distribution function

Definition

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

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

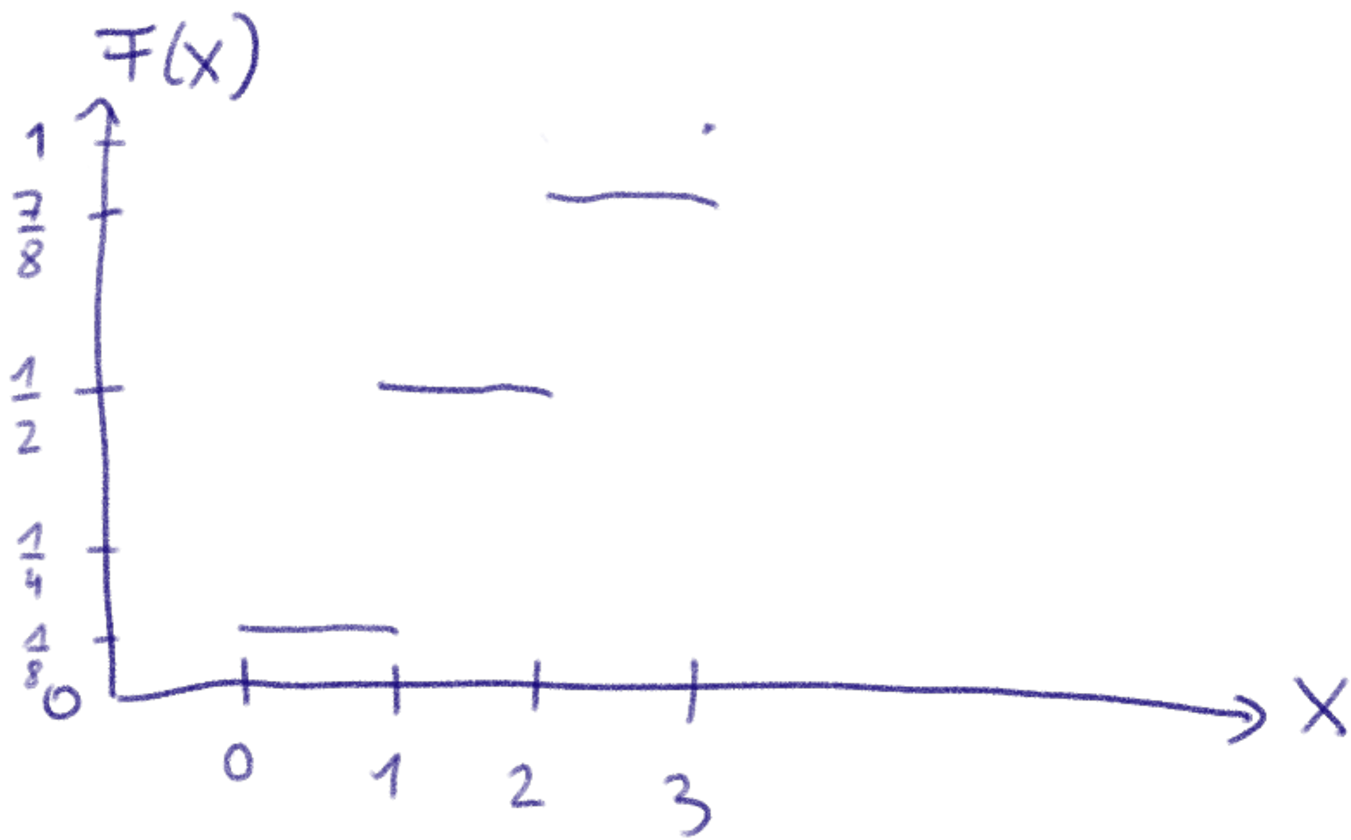
Cumulative distribution function



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

Will be very important when working with continuous random variables.

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*.

weighted average / center of mass

Examples

Examples

X = nr on fair die after roll

$$\mathbb{E}X = \sum_{i=1}^{6} x_i \cdot p(X=i) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} \dots + 6\frac{1}{6} = \frac{7}{2}$$

expectation need not be a possible value

Properties of expectation

Proposition (expectation of a function of a random variable)

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