NMAI059 Probability and statistics 1
Class 2

Robert Šámal

What we have learned

what happens — events we care about — prob.

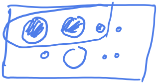
(1) $P(\Omega)=1$, $P(\emptyset)=0$

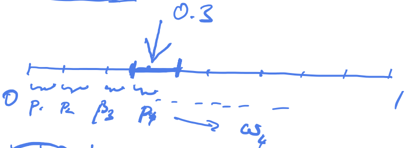
(2) $P(A_1 \cup A_2 \cup \dots) = P(A_1)+P(A_2)+\dots$ (disjoint sets)

- definition of a probability space $(\Omega, \mathcal{F}, P)$: two axioms
- **naive** probability space: Ω finite, $\mathcal{F} = \mathcal{P}(\Omega)$,
  $P(A) := |A|/|\Omega|$

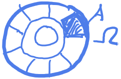
  $\frac{\#\text{ good}}{\#\text{ all}}$ (1) $P(\emptyset)=\frac{0}{|\Omega|}=0$, $P(\Omega)=\frac{|\Omega|}{|\Omega|}=1$ (2) $\frac{|A_1 \cup A_2 \cup \dots|}{|\Omega|} = \frac{|A_1|+|A_2|+\dots}{|\Omega|} = \frac{|A_1|}{|\Omega|}+\frac{|A_2|}{|\Omega|}+\dots$

- **discrete** probability space: $\Omega = \{\omega_1, \omega_2, \dots\}$, $\mathcal{F} = \mathcal{P}(\Omega)$,
  weights $\sum_i p_i = 1$
  $P(A) := \sum_{i:\omega_i \in A} p_i$





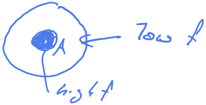
- **geometric** probability space:
  $\Omega \subseteq \mathbb{R}^d$ with a finite volume,
  $P(A) := V_d(A)/V_d(\Omega)$



- probability space **continuous with density**:
  weights
  $\Omega \subseteq \mathbb{R}^d$ with a function f, where $\int_\Omega f = 1$,
  $P(A) := \int_A f$

  $f: \Omega \to R$

  $P(A) = \int_A f$



What we have learned: Basic properties

In a probability space $(\Omega, \mathcal{F}, P)$ we have for $A, B \in \mathcal{F}$

- $P(A^c) = 1 - P(A) \qquad (A^c = \Omega \setminus A)$
- $A \subseteq B \Rightarrow P(A) \leq P(B)$ $\quad P(B \setminus A) = P(B) - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A_1 \cup A_2 \cup \dots) \leq \sum_i P(A_i)$ (subaditivity, Boole inequality)

- We define conditional probability (when $P(B) > 0$).

Prosecutor's fallacy

$P(\text{evidence} \mid \text{innocent})$ very low

$\neq$

$P(\text{innocent} \mid \text{evidence})$ can be still high

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

≤ 1, ≥ 0

cond. prob. is a prob.

- $Q(A) = P(A \mid B)$ satisfies the axioms of probability

$\{0,1\}^{\mathbb{N}}$

(1) $Q(\emptyset) = \frac{P(\emptyset \cap B)}{P(B)} = 0$, $Q(\Omega) = \frac{P(B)}{P(B)} = 1$ (2)

$A_1, A_2, \dots$ disj.: $Q(A_1 \cup A_2 \cup \dots) = \frac{P((A_1 \cup A_2 \cup \dots) \cap B)}{P(B)}$

$= \frac{P((A_1 \cap B) \cup (A_2 \cap B) \cup \dots)}{P(B)} = \frac{P(A_1 \cap B) + \dots}{P(B)}$ $= Q(A_1) + Q(A_2) + \dots$

disj.

Overview

Conditional probability

Discrete random variables

Examples of discrete r.v.'s

Chain rule

symm. def.

- $P(A \cap B) = P(B)P(A \mid B)$ $= P(A) \cdot P(B \mid A)$

Theorem

If $A_1, \ldots, A_n \in \mathcal{F}$ *and* $P(A_1 \cap \cdots \cap A_n) > 0$, *then*

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) =$$

$$P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \ldots P(A_n \mid \bigcap_{i=1}^{n-1} A_i)$$

n! theorems for the price of one

Proof

$$\cancel{P(A_1)} \cdot \frac{\cancel{P(A_1 \cap A_2)}}{\cancel{P(A_1)}} \cdot \frac{\cancel{P(A_1 \cap A_2 \cap A_3)}}{\cancel{P(A_1 \cap A_2)}} \cdots \boxed{\frac{P(A_1 \cap \cdots A_n)}{\cancel{P(A_1 \cap \cdots \cap A_{n-1})}}}$$

$$\frac{\#\text{good}}{\#\text{all}} = \frac{\binom{39}{3}}{\binom{52}{3}}$$

4.13

- Ex.: we pick 3 cards from a deck of 52. What is P(no heart)?

A_i: = "i-th card is not a heart"

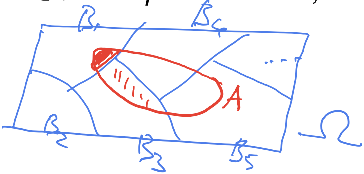
$$P(A_1 \cap A_2 \cap A_3) = P(A_1)\, P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2) = \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50}$$

Law of total probability

Definition

Countable family of sets $B_1, B_2, \ldots \in \mathcal{F}$ is a partition of Ω, if

- $B_i \cap B_j = \emptyset$ *for* $i \neq j$ *and*
- $\bigcup_i B_i = \Omega$.



Theorem

If $B_1, B_2, \ldots$ is a partition of Ω and $A \in \mathcal{F}$, then

$$P(A) = \sum_i P(A \mid B_i) P(B_i)$$

what if $P(B_i) = 0$?

undef $\cdot$ 0

(terms with $P(B_i) = 0$ are counted as 0).

$A = A \cap \Omega = (A \cap B_1) \cup (A \cap B_2) \cup \ldots$ pwise disj:

$P(A) = \sum_i P(A \cap B_i) = \sum_i P(B_i) \cdot P(A \mid B_i)$

Law of total probability – exhausting all possibilities

c_1 c_2 c_3

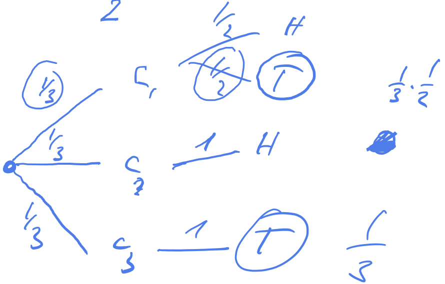
- Application 1. We have three coins: H+T, H+H, T+T, we choose from them at random. What is the probability that we toss a tail?

$$P(T) = P(C_1)\cdot P(T/C_1) + P(C_2)\cdot P(T/C_2) + P(C_3)P(T/C_3)$$

$$\frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot 0 + \frac{1}{3}\cdot 1$$

$$= \frac{1}{2}$$

$$P(T) = \frac{3}{6}$$



Law of total probability – “wishful thinking”

$P(\infty \text{ play}) = 0$

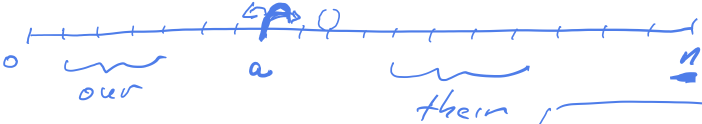
(Markov chain / random walk)

- Application 2. Gambler’s ruin. $P(\text{we lose}) = \frac{b}{a+b} = 1 - P(\text{we win})$

  We have a CZK (crowns), our opponent b CZK. We play repeatedly a fair game for 1 CZK, until someone loses all his/her money. What is the probability that we win?

$B_1 = \{\text{win } 1^{st} \text{ round}\}$

$B_2 = B_1^c$



$n = a+b$

$p_n = 1$, p_3, d_3, p_2, d_2, p_1, d_1, $p_0 = 0$

$p_a = P(\text{we win the game}) = ?$

$\boxed{p_0 = 0 \qquad p_n = 1}$

$\boxed{0 < a < n}$ $= P(\text{win} \mid \text{win 1. round}) \cdot P(\text{win } 1^{st} \text{ round})$

$+ \; P(\text{win} \mid \text{loose } 1^{st} \text{ round}) \cdot P(\text{loose } 1^{st} \text{ round})$

$\boxed{p_a = \frac{1}{2} \cdot p_{a+1} + \frac{1}{2} \, p_{a-1}}$

$n+1$ lin. eq's for $n+1$ var's

$2p_a = p_{a+1} + p_{a-1}$

$\Updownarrow$

$\underbrace{p_a - p_{a-1}}_{d_a} = \underbrace{p_{a+1} - p_a}_{d_{a+1}}$

$(n-1 \text{ eq's})$

$d_1 = d_2 = \dots = d_n$

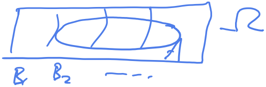
$\boxed{d_1 + d_2 + \dots + d_n} = \frac{1}{n}$

$(p_1 - p_0) + (p_2 - p_1) + \dots + (p_n - p_{n-1})$

$= p_n - p_0 = 1 - 0 = 1$

$\boxed{p_a = \frac{a}{n} = \frac{a}{a+b}}$

Bayes' rule



observation

Theorem

Let $B_1, B_2, \ldots$ be a partition of Ω, $A \in \mathcal{F}$ and $P(A), P(B_j) > 0$. Then

state of the world

$$\underbrace{P(B_j \mid A)}_{\text{updated theory, posterior}} \overset{(1)}{=} \frac{P(A \mid B_j)P(B_j)}{P(A)} \overset{(2)}{=} \frac{P(A \mid B_j)P(B_j)}{\sum_i \underbrace{P(A \mid B_i)}_{\text{the actual observation}}\underbrace{P(B_i)}_{\text{theory ab. world, prior}}}$$

(terms with $P(B_i) = 0$ are counted as 0).

Proof (1) def. of cond. prob.

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

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

(2) use LOTP for $P(A)$

Natural appl.

3 coins, at random, we toss T

C_1	TH
C_2	HH ×
C_3	TT

$$P(C_2 \mid T) = 0$$

$$P(C_1 \mid T) = \frac{P(T \mid C_1) \cdot P(C_1)}{P(T)} = \frac{\frac{1}{3} \cdot \frac{1}{2}}{\frac{1}{2}} = \frac{1}{3}$$

Bayes' rule

state of the world: D ... disease, D^c

$P(T/D)$ = sensit.

$P(T^c/D^c)$ = specif.

observ.: T ... we test +

$$P(D/T) = \frac{P(D) \cdot P(T/D)}{P(D)P(T/D) + P(D^c)P(T/D^c)} = \frac{p \cdot 0.8}{p \cdot 0.8 + (1-p) \cdot 0.01}$$

FPR

$p = 0.001$ .... 7 %

$p = 0.016$ .... 56 %

$p = 0.05$ .... 80 %

Independent events

Definition

Events $A, B \in \mathcal{F}$ are independent if $P(A \cap B) = P(A)P(B)$.

- Then we also have $P(A \mid B) = P(A)$, provided $P(B) > 0$.

$\Omega = \{T,H\} \times \{T,H\}$

Ex.: we toss a coin twice.

- $A = \{\omega \in \Omega : \omega_1 = H\} =$ "first toss was a head"
- $B = \{\omega \in \Omega : \omega_2 = H\} =$ "second toss was a head"
- $C = \{\omega \in \Omega : \omega_1 \neq \omega_2\} =$ "exactly one toss was a head"

$P(A \cap B) = \frac{1}{4}$ $=$

$P(A).P(B) = \cancel{\frac{1}{2} \cdot \frac{1}{2}}$ $\frac{2}{4} \cdot \frac{2}{4}$

$\boxed{A, B \text{ are indep.}}$ "by def."

$=$ $P(A \cap C) = P(B \cap C) = P(A).P(C) = P(B).P(C)$

$\boxed{\begin{array}{l} A, C \text{ are indep.} \\ B, C \text{ are indep.} \end{array}}$

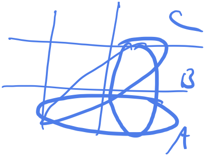
Mutually independent events

Definition

Events $\{A_i : i \in I\}$ are (mutually) independent if for every finite set $J \subseteq I$

$$\boxed{P(\bigcap_{i\in J} A_i) = \prod_{i\in J} P(A_i).}$$

If the condition is true only for sets J with $|J| = 2$, we call the collection $\{A_i\}$ pairwise independent.



A, B, C are pairwise indep.

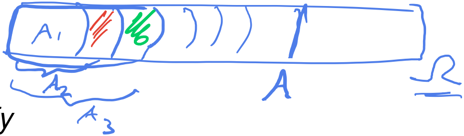
But not mutually indep.

$$P(A \cap B \cap C) = P(\emptyset) = 0$$

$$\neq$$

$$\frac{1}{8} = P(A) \cdot P(B) \cdot P(C)$$

Continuity of probability



Theorem

Suppose that events in $\mathcal{F}$ satisfy

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$$

and $A = \cup_{i=1}^{\infty} A_i$. Then we have

$$P(A) = \lim_{i\to\infty} P(A_i).$$

B

$A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \ldots$

$P(A) = P(A_1) + P(A_2 \setminus A_1) + P(A_3 \setminus A_2) + \ldots$

$= \lim_{i\to\infty} \underbrace{P(A_1) + \ldots + P(A_i \setminus A_{i-1})}_{P(A_i)}$

- $A_n \subset \{H, T\}^{\mathbb{N}}$, $A_n =$ in the first n tosses there was at least one tail.

$P(A_n) = 1 - \frac{1}{2^n}$

$A = \bigcup_{n=1}^{\infty} A_n =$ "at least one T in the whole ∞ seq."

$P(A) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} 1 - \frac{1}{2^n} = 1$

Overview

Conditional probability

Discrete random variables

Examples of discrete r.v.'s

Random variable

Often we are interested in a number given as a result of a random experiment.

- We throw a dart and measure the distance from the center of the dartboard.
- We roll a die untill we get a six, then count how many rolls it took.
- In a quicksort algorithm (with a random choice of pivot) we measure the number of operations.

Definition

Given a probability space $(\Omega, \mathcal{F}, P)$. *We call a function* $X : \Omega \to \mathbb{R}$ *a discrete random variable, if* $Im(X)$ *(range of* X*) is a countable set and if for every real* x *we have*

$$\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}.$$

PMF

Definition

Probability mass function, pmf of a discrete random variable X *is a function* $p_X : \mathbb{R} \to [0, 1]$ *such that*

$$p_X(x) = P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

- $\sum_{x \in Im(X)} p_X(x) =$?
- $S := Im(X) \qquad Q(A) := \sum_{x \in A} p_X(x)$
  $(S, \mathcal{P}(S), Q)$ is a discrete probability space.
- For $S = \{s_i : i \in I\}$ countable set of reals and $c_i \in [0, 1]$ satisfying $\sum_{i \in I} c_i = 1$ there is a probability space and a discrete r.v. X on it such that $p_X(s_i) = c_i$ for $i \in I$.

Overview

Conditional probability

Discrete random variables

Examples of discrete r.v.'s

Bernoulli/alternate distribution

- $X =$ number of tails in one toss of a coin (not necessary a fair one)
- We write $X \sim Bern(p)$. (Sometimes $Alt(p)$.)

- Given $p \in [0, 1]$.
- $p_X(1) = p$
- $p_X(0) = 1 - p$
- $p_X(k) = 0$ for $k \neq 0, 1$

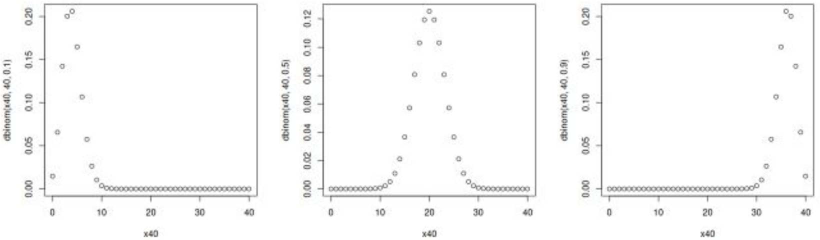
- For an event $A \in \mathcal{F}$ we define *indicator random variable* I_A:
- $I_A(\omega) = 1$ if $\omega \in A$, $I_A(\omega) = 0$ otherwise.
- $I_A \sim Bern(P(A))$

Binomial distribution

- $X =$ number of tails in n independent tosses of a loaded coin.
- We write $X \sim Bin(n, p)$.

- $X = \sum_{i=1}^{n} X_i$ for independent r.v.'s $X_1, \ldots, X_n \sim Bern(p)$.
- Given $p \in [0, 1]$.
- $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0, 1, \ldots, n\}$

Binomial distribution: pmf



Generated by the following code in R

```
x40 <- 0:40
plot(x40, dbinom(x40, 40, 0.1))
plot(x40, dbinom(x40, 40, 0.5))
plot(x40, dbinom(x40, 40, 0.9))
```