Korespondenční **S**eminář **I**nspirovaný **Ch**emickou **T**ematikou

Ročník 25 (2026/2027)

Série 1

KSICHT je financován Ministerstvem školství, mládeže a tělovýchovy spolu se Sekcí chemie, Přírodovědecké fakulty Univerzity Karlovy.

Děkujeme za podporu!

Přírodovědecká fakulta
Univerzita Karlova

KSICHT probíhá pod záštitou Přírodovědecké fakulty Univerzity Karlovy

Chemie je všude: je ve vodě, je v půdě, je ve vzduchu a je i v nás samotných. Veškeré materiály jsou tvořeny chemickými látkami, chemické reakce nám každodenně pomáhají s tvarováním světa kolem nás a biochemické reakce nás vlastně utvářejí: katalytické reakce umožňují každodenní běh našich těl, neurotransmitery jsou nositeli našich emocí a naše DNA může dát vzniknout novým generacím. Avšak bez porozumění tajemným nebezpečenstvím s chemií spojeným jsme jí vydáni napospas, proto stojí za to ji poznat blíže a hlouběji, aby se stala naším dobrým sluhou a ne obávaným pánem.

Proč řešit KSICHT?

Milí řešitelé, KSICHT je zde již čtvrt století proto, aby Vám ukázal různá zákoutí chemie a přivedl Vás k jejich objevování. V průběhu školního roku k Vám doputují čtyři brožurky s úlohami z různých oblastí chemie, při jejichž řešení se naučíte mnoho nového a navíc si užijete kopu srandy, protože úkoly jsou mnohdy poněkud… neortodoxní. Prostřednictvím našeho seriálu se pak můžete seznámit s některými velkými chemickými tématy, která se Vám pokusíme předestřít stravitelně, zábavně a užitečně. V letošním ročníku to bude seriál s názvem *Analýza a vizualizace vědeckých dat*, který pro Vás budou psát Jakub Krieger a Jan Stoklasa. V neposlední řadě můžete v každé brožurce sledovat osudy skutečně neohroženého komiksového hrdiny, a sice Zajíčka chemika.

KSICHT pořádá v průběhu ročníku dva výlety, na kterých je možné se setkat s ostatními řešiteli, s organizátory a autory úloh. Celý ročník je zakončen týdenním soustředěním na Přírodovědecké fakultě UK, kde si mimo jiné vyzkoušíte práci v laboratořích a vyslechnete přednášky předních českých a světových vědců. Kapacitu tohoto soustředění máme pro 30 řešitelů, rozhodovat bude celkové umístění po 4. sérii. Vzhledem k požadavkům grantu, ze kterého je KSICHT spolufinancován, je počet účastníků ze Slovenska na závěrečném soustředění omezen na 8.

Mimo to úspěšní řešitelé mohou mít na vybraných vysokých školách odpuštěné přijímací zkoušky a získat motivační stipendium[1].

1 Více informací najdete na webových stránkách KSICHTu.

Termín pro odeslání řešení 1. série: 2. 11. 2026

Elektronicky (PDF)	Papírově
http://ksicht.natur.cuni.cz/odeslani-reseni	**KSICHT** **Přírodovědecká fakulta UK** **Hlavova 2030** **128 43, Praha 2**

Jak řešit KSICHT? **http://ksicht.natur.cuni.cz/**

V každé brožurce je pro Vás připraveno 5 úloh k vyřešení. Jsou mezi nimi zábavné hříčky i opravdové oříšky. Pokuste se poradit si s nimi, jak nejlépe umíte, ale pokud je nevyřešíte všechny, nic se nestane. Budeme rádi, když nám pošlete odpovědi, byť jen na část úkolů, které úloha obsahuje. Dbejte však, aby Vaše odpovědi byly srozumitelné a aby bylo zřejmé (zejména u výpočtů), jak jste k řešení dospěli.

Každou úlohu vypracujte **samostatně** na list formátu A4, na němž bude uvedeno **Vaše jméno**, **název a číslo úlohy**. V případě, že posíláte úlohy přes webový formulář (námi preferovaný způsob odeslání), uložte každou úlohu do samostatného souboru PDF[2]. Pro kreslení chemických vzorců doporučujeme používat programy dostupné zdarma: MDL ISIS/Draw, ChemSketch (freeware s povinnou registrací) nebo Chemtool.

Vypracované řešení úlohy odešlete organizátorům nejpozději do data uvedeného na následující stránce elektronicky nebo papírově (rozhoduje čas na serveru KSICHTu či datum poštovního razítka).

Autoři poté Vaše řešení opraví, ohodnotí je a pošlou Vám je zpět společně s následující brožurkou a dalšími úlohami k řešení. Řešitelé, kteří získají alespoň 50 % bodů z celého ročníku, obdrží certifikát o úspěšném absolvování semináře.

Vaše umístění ve výsledkové listině je také kritériem pro účast na závěrečném soustředění, detaily k přihlašování uvedeme v brožurce čtvrté série.

V případě jakýchkoliv dotazů se na nás neváhejte obrátit na e-mail ksicht@natur.cuni.cz nebo v případě dotazu ohledně úlohy napište autorovi úlohy na jmeno.prijmeni@ksicht.natur.cuni.cz.

2 Neposílejte naskenovaná řešení s výjimkou obrázků, text bývá špatně čitelný.

KSICHTí desatero řešení úloh

Vzhledem k tomu, že se opakovaně někteří řešitelé dopouští neodpustitelných či méně závažných prohřešků, kvůli kterým zbytečně přicházejí o body, vytvořili jsme pro Vás seznam zásad, kterých je dobré se držet.

1. Jen jeden KSICHT řešiti budeš.
2. Nebudeš si zoufat, že nevyřešíš všechno a správně.
3. Nebudeš se klanět Güghlu ni jiným vyhledávačům. Informaci svou si vždy ověříš[3].
4. Nezkopíruješ $W^iK^ip_ed_{ii}$ českou ni anglickou ni v jazyku jiném psanou[4].
5. Pamatuj na den odeslání, že ti má být svatý. Čtyři týdny řešiti budeš, dne (před)posledního odesláno míti budeš.
6. Rukopis vlastnoruční nenaskenuješ, ale do obálky vložíš a poštou odešleš.
7. Neudáš výsledku bez výpočtu.
8. Neopíšeš nadbytek číslic z kalkulátoru svého[5].
9. Nepožádáš o řešení bližního svého.
10. KSICHTí jméno důsledně šířiti budeš.

3 Smyslem korespondenčního semináře je také dát vám příležitost naučit se vyhledávat, třídit a kriticky vyhodnocovat dostupné informace. Proto můžete k řešení používat jakékoli tištěné i elektronické zdroje, se kterými je ale třeba správně zacházet – více v další poznámce.

4 Odevzdání textu získaného pomocí Ctrl+C, Ctrl+V není řešením úlohy. Tím má být vaše vlastní formulace odpovědí na otázky v úloze, kterou jste sestavili na základě informací dostupných klidně i na Wikipedii. Zejména u internetových zdrojů je třeba každý zdroj kriticky zhodnotit: zdaleka ne každá stránka, příspěvek na blogu či diskusním fóru obsahuje pravdivé informace.

5 Tzv. kalkulátorový syndrom: „Svět byl stvořen za 6,9999999999942 dní.“ Toto není ani správná, ani přesná hodnota.

KSICHTí pravidla používání AI

Umělá inteligence se stala standardním pracovním nástrojem. Chceme, aby byla přínosem i pro KSICHT. Domluvme se nejprve na následující trojici pravidel, která nám pomůže udržet KSICHT smysluplný.

1. Pište vlastními slovy

Nekopírujte text KSICHTí brožurky do AI a neodevzdávejte generovaný text jako odpověď.

Použití AI při řešení KSICHTu je povoleno: AI vám pomůže pochopit nové koncepty, najít dílčí informace nebo provést rešerši. Na Vás zůstává **vlastními slovy formulovat finální odpověď i se případně AI zeptat.**

Pokud odevzdáváte svou odpověď na základě AI, popište, jak jste ji použili – nemusíte kopírovat každý prompt dlouhé konverzace, ale vystihněte, o čem byla. Autorovi úlohy to umožní zjistit, jak přemýšlíte, a poskytnout Vám zpětnou vazbu.

2. Respektujte autory

Autor může specifikovat dílčí požadavky, jak si při řešení s AI počínat: ve své úloze může AI povolit a soustředit se na práci s ní, nebo ji naopak zcela omezit. **Respektujte je.**

Předmětem hodnocení bude i sdělnost Vašich řešení. Nevypisujte odstavce, ať už vaše vlastní, nebo generované, doufajíce, že si v tom najdeme správnou odpověď. Autor KSICHTí úlohy není detektiv.

3. Chraňme KSICHTí kulturu.

Chcete přemýšlet a učit se? Hledáte lidi, mezi které tím zapadnete? Potom je tu KSICHT pro Vás. Proč? KSICHT má právě takovou kulturu. Kulturu, ke které patří společná **radost z učení se**; kulturu, kde jsou **upřímnost**, **pravdivost** stejně jako **vlastní tvůrčí zápal** na vrcholu žebříčku hodnot. Z této kultury vyrostl KSICHT takový, jaký jsme jej poznali my, bývalí řešitelé, dnes už orgové. **Kulturu nemůžeme vymáhat**: chceme ji zachovat, předat dál. Pomozte nám!

- Pište odpovědi, které jsou opravdu Vaše.
- Pište odpovědi, které umíte obhájit a za kterými si stojíte.
- Dejte najevo zvídavost.
- Když už náhodou opravdu nevíte, nezachraňujte to bezhlavým generováním textu. Je v pořádku nevědět všechno.

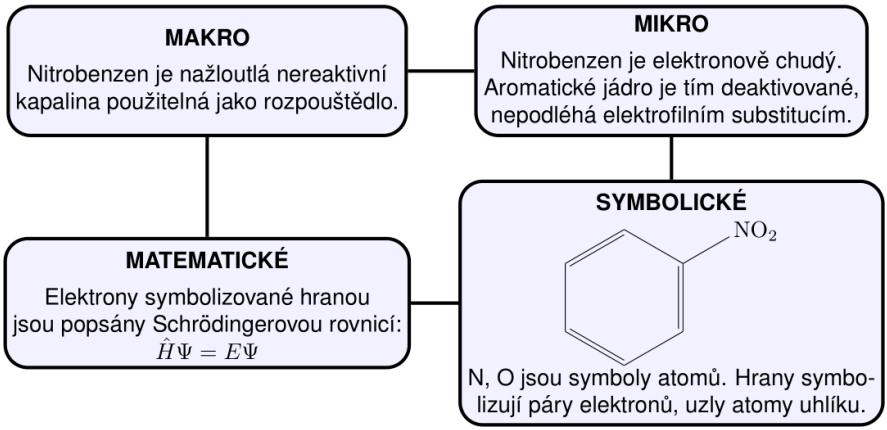
Jako autoři tuto **kulturu implicitně vkládáme do KSICHTích brožurek. Přejeme si dostávat to samé od Vás, řešitelů.** Ve většině případů se tak děje. Neudělejme z KSICHTu soutěž v psaní promptů. Díky!

Vlastními slovy i ve světě s AI

Už se jste někdy měli nápad, který jen co se dostal na papír, přestal dávat smysl nebo alespoň vypadal jinak, než jste jej měli v hlavě? Je to normální. Proces psaní, kreslení nebo výpočtu totiž funguje jako síto: Skrze gramatická či algebraická pravidla a grafomotoriku zapojuje řadu center lidského mozku. Pouze toto spojení umožňuje hlubokou úvahu. Samostatné vyjadřování vede ke komplexnímu myšlení.

Hlavním cílem KSICHTích úloh je pomoci Vám chápat chemii v souvislostech. Chemie je komplexní disciplína, vyžaduje neustálé propojování mnoha způsobů myšlení, viz schéma níže. Schéma bylo vytvořeno v souladu s trojicí pravidel.

Obr. 1: Chápání pojmu *nitrobenzen* znázorněné pomocí tzv. Johnsonova čtyřúhelníku. Pochopení vyžaduje propojení makroskopického, mikroskopického, symbolického a matematického vnímání. LaTeXový kód generován pomocí ChatGPT-5.0, content *bez AI*.



AI je efektivně schopna vyřešit většinu KSICHTích úloh, formulovat řešení na úrovni odborné publikace i detailně vysvětlit svůj postup. Pochopit řešení a zvládnout jej vypracovat však není totéž. S ohledem na to vyžadují KSICHTí pravidla vlastní formulaci řešení i případných promptů do AI. Ač řešení tímto způsobem zabere více času, úlohy tak lépe naplňují svůj cíl a KSICHT neztrácí smysl jako soutěž.

Úvodníček

Velevážené obecenstvo,

jsme poctěni, že jste dorazili na premiéru nejnovější série naší show Ksichťáci. Dvacet úloh, nespočet účastníků, ale jen jeden odejde z hradu Ksichtoklát s odměnou ve zlatých cihlách – pardon, režie mě právě upozorňuje, že z důvodu rozpočtových škrtů nabízíme místo každé cihly jeden atom zlata. Inu, nevadí, pojďme se raději podívat na to, jaké zrádné úlohy na naše soutěžící čekají.

Jako správné televizní drama začínáme zvolna, u čaje s hradním pánem Šimonem. Jeho *Čajová vášeň* je jistě sdílená mnohými z vás, a slibujeme, že čaj je 100% neškodný.

Pak už ale epizoda nabírá obrátky, jelikož soutěžící se k nám musí dostat až z daleké Itálie. *Toskánská* expertka Adéla si na ně nachystala několik tamních… specialit.

Další výkony ve fyzické výzvě pak už jistě povedou k tomu, že se naši soutěžící zapotí. Dojde na rozklad, eliminaci (soutěžících), možná i nějaké to kousání (potravy). A kdo nás provede *Bioenergetickou* povahou této úlohy lépe než mistr Tadeáš.

Následující epizodu jsme pak natáčeli na kolejích kolem hradu Ksichtoklát. V týmové úloze účastníci versus fyzikální zákony nebylo jasného vítěze, ale dívalo se na to dobře, jak nám tvrdí *Vlaková* vedoucí Júlia.

No, a na závěr nějaké to šokující odhalení – Jan reveals himself as a fluent English speaker! A plánuje chytit nic netušící účastníky do pasti svých Markovových řetězců. Zda jeho *Markov's DNA chains* polapí mnohé, nebo jeho plán selže, uvidíme až v příští epizodě.

Po tomto cliffhangeru už běží titulky a reklamy. I my jich tu pár máme! Třeba na náš výlet s KSICHTem, který proběhne tento podzim; detaily najdete v brožurce nebo na webu.

Užijte si řešení úloh!

Za všechny showrunnery KSICHTu

Tom Heger

Podzimní víkend s KSICHTem

Srdečně Vás zveme na Podzimní víkend s KSICHTem, který se uskuteční od 20. do 22. listopadu v Nymburce. Přihlásit se můžete na webu akce[6], kde najdete také všechny důležité informace. Těšíme se na Vás!

[6] https://ksicht.natur.cuni.cz/akce/19-podzimni-vikend-s-ksichtem/

Seriál: Analýza a vizualizace vědeckých dat

Autor: Jakub Krieger a Jan Stoklasa

1. díl: Chyby měření a trocha matematiky

Představte si Tadeáše, který nahlásí Honzovi do telefonu: „Přijdu do deskoherny kolem půl sedmý.“ V 18:30 v deskoherně čeká Honza, ale Tadeáš přijde až v 18:35. Když se potkají, pozdraví se a začnou konverzovat, a o Tadeášovu pozdním příchodu není ani zmínky. Nyní si představte Kubu, který nahlásí „Přijdu v 18:33.“ Reálně však přijde až v 18:35. Než ho Kuba stihne pozdravit, prohlásí Honza sarkasticky „Jdeš o dvě minuty pozdě!“ [7]

Asi intuitivně nahlédnete, proč si „přesnější“ Kuba vysloužil od Honzy komentář – na rozdíl od Tadeáše avizoval něco, co pak nesplnil. Přitom nehraje roli jenom čas, který nahlásil, ale také způsob, jakým to vyjádřil. V něm je totiž ukrytá informace o předpokládané přesnosti, respektive chybě, které se Kuba pravděpodobně dopouští, když uvádí časový údaj. Čas 18:33 totiž implikuje minutovou přesnost, a tu Kuba nedodržel.

Rozbor této zdánlivě banální situace z běžného života zde provádíme z toho důvodu, že něco velmi podobného funguje i v přírodních vědách. Kdykoliv něco ve vědě měříme, nebo obecně poznáváme, potýkáme se s tím, že věci nejsou poznatelné s libovolnou přesností. Vždy se dopustíme nějaké, byť sebemenší, chyby způsobené naším měřicím aparátem. A i kdybychom sestavili sebedokonalejší měřicí přístroj, narazíme na omezení, která nám klade kvantová mechanika. Zkrátka, nic nelze změřit dokonale přesně. Pokud chceme jako vědci podávat pravdivé informace, musíme proto udat nejen naměřenou hodnotu, ale i nějakým způsobem vyjádřit, jak je přesná.

Platné cifry

Pokud jste si četli KSICHTí desatero[8], určitě jste si všimli bodu číslo 8. „Neopíšeš nadbytek číslic z kalkulátoru svého“, včetně poznámky pod čarou „Tzv. kalkulátorový syndrom: „Svět byl stvořen za 6,9999999999942 dní.“ Toto není ani správná, ani přesná hodnota.“ Společně se nyní nad tímhle bodem desatera zamyslíme a trochu více jej rozvedeme, jak jsme to vlastně mysleli.

Bez újmy na obecnosti budeme hovořit o tom, jak daleko hodíte koulí (nebo jiným trojrozměrným konvexním tělesem). A budeme tvrdit, že jsme hodili do vzdálenosti 8 nebo 8,0 libovolných jednotek, třeba mNM (tisícina námořní míle). Jaký je mezi těmito dvěma hodnotami rozdíl, když přece obě jsou osm? Pomůžeme si Obrázkem 1. Představte si, že máte k dispozici jen číslice zobrazené na ose, tj. 0 – 1 – 2 – 3 atd. Jak se dovíte později, pravděpodobnost, že se trefíte

[7] Jakékoliv podobnosti postav s reálnými osobami jsou čistě náhodné.

[8] A pokud jste ho nečetli, měli byste se nad sebou vážně zamyslet a následně si ho přečíst.

přesně na nějakou hodnotu, kterou má vaše osa vyznačenou, je hodně nízká[9]. Ve většině případů se tedy stane, že vaše konvexní těleso skončí někde mezi značkami. Obrázek 1 zobrazuje třeba takovou situaci – pokud míček dopadne kamkoliv do oblasti vyznačené červenou[10] úsečkou, tak zahlásíte, že váš hod byl do vzdálenosti 8 tisícin námořní míle. Upusťme od toho, zdali je to hodně nebo málo[11].

Obrázek 1: Grafické znázornění hodnoty 8

Ve všech případech bychom tvrdili, že konvexní těleso dopadlo někde do vzdálenosti d a přitom:

$$d \in \langle 7{,}5\ ;\ 8{,}5 \rangle$$

Kdybychom měli jemnější osu, tedy dílky by třeba rostly po 0,1, tak bychom mohli tvrdit, že vzdálenost byla 8,0. V tomto případě se velikost toho intervalu výrazně zmenšila, nyní je jen:

$$d \in \langle 7{,}95\ ;\ 8{,}05 \rangle$$

Intuitivně cítíme, že s počtem platných cifer, které uvádíme, souvisí něco, co se běžně označuje jako nejistota měření – tedy to, jak moc si jsme jistí naší odpovědí.

Ilustrujme použití platných cifer na dalším příkladu. V pracovních postupech syntetické chemie se dočtete například: „Pomocí analytických vah odvažte na lodičce přibližně přesně 0,2 g analyzované látky.“ [1] Tím se myslí, že máte navážit hodnotu v okolí hodnoty 0,2 g, ale nemusí to být přesně 0,2000 g – to by se vám vzhledem k hmotnosti jednotlivých krystalků povedlo jen velmi stěží. Stačí, když navážíte hodnotu kolem 0,2 g, a pak si výsledek zapíšete s počtem platných cifer, který odpovídá přesnosti vašich vah, například 0,1967 g. Hodnota 0,2 g je orientační, hodnotou 0,1967 g si musí být jistí.

[9] Je fakt malá.

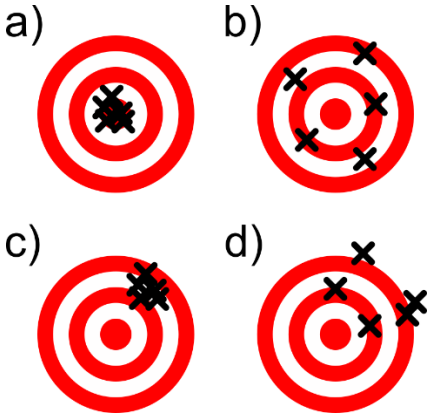
[10] V černobíle verzi brožurky myšleno šedou úsečkou.

[11] Tahle diskuse by vyžadovala zmínění dalších skutečností, například v jakém směru je měřena vzdálenost, protože lidé mají ve dvou ze třech os standardního 3D prostoru lepší schopnosti hodu než v té třetí ose.

Systematická vs. náhodná chyba, aneb střílení na terč

Měření ve vědě je jako střílení na terč. Na Obrázku 2 jsou vyobrazeny různé možnosti, jak takové střílení může dopadnout. Aktivní čtenáři si před čtením dalšího textu mohou zkusit sami přijít na význam obrázku.

Terče a) a c) sdílejí tu vlastnost, že zásahy, tedy měření, dopadly velmi blízko sebe. Na terči c) ovšem není průměrná poloha zásahů ve středu terče – zásahy jsou systematicky posunuté doprava nahoru. Tomu říkáme *systematická chyba*. Pokud se podíváme na terče a) a b), zjistíme, že mají podobnou průměrnou polohu zásahů, ale v případě b) jsou zásahy rozprostřené dál od sebe. Tomu říkáme *náhodná chyba*. V případě d) je vysoká jak náhodná, tak systematická chyba.



Obrázek 2: Možné výsledky měření ilustrované jako střílení na terč.

Reálné měření ve vědě má ale od střílení na terč jednu zásadní odlišnost. Terče, na které střílíme, nevidíme. Předem nevíme, jaká je ta opravdová správná hodnota – to by pak měření postrádalo smysl. Představte si obrázek 2 bez terčů, a snadno vám dojde, jaké to nese problémy.

Systematická chyba

Systematické chyby „zatěžují stejným způsobem výsledek každého nezávisle opakovaného měření“. [2]

Co může být zdrojem takové systematické chyby? Pokud na chvíli zabrousíme do chemické laboratoře, napadne nás třeba ulomená špička pipety, špatně používaná (nedomáčknutá) pipeta, zbytková kapalina v odměrném válci, špatně kalibrovaný (nebo nekalibrovaný) přístroj nebo třeba přetitrovaní (zastavení titrace až po dosažení bodu ekvivalence). Problém může být také v pracovním postupu. Pokud například budeme měřit vývoj pH v nějakém roztoku a umožníme přístup vzduchu, nevyhnutelně pH v čase poklesne kvůli rozpouštění CO_2 do tohoto

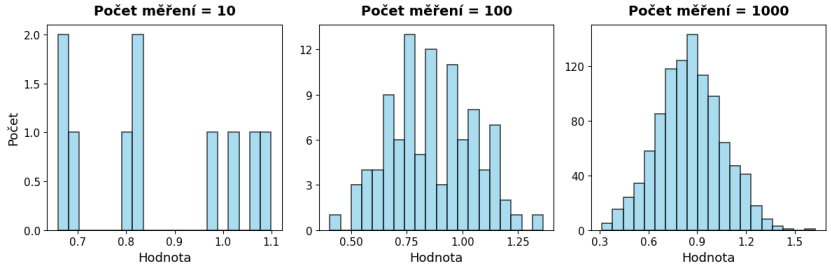
roztoku. Pokud si toho nejsme vědomi, budeme tento jev nesprávně interpretovat jako důsledek experimentu, a dopustíme se tak systematické chyby.

Jak můžeme velikost systematické chyby odhadnout? Na základě příkladů výše nás může napadnout použít jiný přístroj, jinou pipetu či nechat experiment nezávisle provést dalšího člověka. Ne vždy ovšem na systematickou chybu přijdeme, ve vědě je proto důležité dostatečně zdokumentovat náš způsob měření, aby někdo jiný mohl experiment zopakovat a případnou systematickou chybu odhalit.

Náhodná chyba

Představte si, že v laboratoři připravujete vícekrokovou syntézou nějakou látku a na konci určujete její hmotnost. Provedli jste postup pětkrát úplně stejně (alespoň podle vašich rozlišovacích schopností) a dostali jste výtěžky 0,86 g, 0,83 g, 0,86 g, 0,88 g a 0,85 g. Jelikož nedokážete ukázat na nějakou konkrétní příčinu, která by vaše hodnoty systematicky odchylovala, můžete odchylky považovat za náhodnou chybu. Lze ovšem nějak vyčíslit velikost této chyby? Ano, lze – s trochou matematiky!

Jak by to vypadalo, kdybychom počet opakování měření postupně zvyšovali? Obrázek 3 to ukazuje pomocí histogramů – grafů, které ukazují počet výskytů hodnot v daném intervalu.

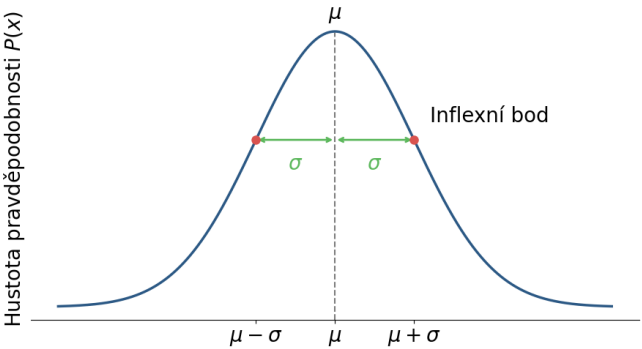


Obrázek 3: Příklady histogramů naměřených hodnot pro různý počet měření.

Pokud bychom počty měření navyšovali více a více, náš histogram by byl hladší a hladší a za jistých předpokladů[12] by se blížil tvaru na obrázku 4 – křivce Gaussovy funkce[13]. [3]

[12] Například že hodnota pozorované veličiny může nabývat libovolné hodnoty od $-\infty$ do $+\infty$.

[13] Lze to dokázat pomocí takzvané centrální limitní věty.



Obrázek 4: Gaussova funkce.

Gaussova funkce má matematický předpis [3]

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right),$$

kde x je pozorovaná veličina, μ je průměr (poloha středu „kopečku“) a σ směrodatná odchylka, která určuje šířku „kopečku“[14]. Ve statistice hovoříme o Gaussovu *rozdělení*, protože určuje, jak je *rozdělená* pravděpodobnost naměření různých hodnot. Gaussovo rozdělení není jediné možné rozdělení, existuje jich mnoho dalších[15]. Pravděpodobnost, že naměříme hodnotu v nějakém rozmezí hodnot, lze ale vždy určit jako plochu pod grafem v rozmezí těchto hodnot. Pro Gaussovo rozdělení je například pravděpodobnost, že hodnota bude v rozmezí $(\mu - \sigma, \mu + \sigma)$, cca 68 %.

Proč pomocí funkce průměr v Excelu nespočítáme průměr?

Ač vás to asi překvapí, opravdu tomu tak je. Začněme tím, že si řekneme, co je to průměr. Jak jste se v předchozí kapitole dozvěděli, průměr μ je jeden z parametrů, který společně se směrodatnou odchylkou σ popisují Gaussovo rozdělení. Pomocí těchto dvou parametrů a předpisu Gaussova rozdělení jsme schopni generovat hodnoty mající chtěnou statistickou distribuci.

[14] Pokud byste chtěli exaktnější definici, tak směrodatná odchylka je odmocnina z rozptylu (variance), pro který platí: $\sigma^2 = \frac{1}{n-1}\sum_i (x_i - \mu)^2$, kde $n-1$ je počet stupňů volnosti (v našem případě počet nezávislých měření).

[15] Můžete si například rozmyslet, jak by vypadalo rozdělení pravděpodobnosti hodnot při hodu férovou šestistrannou kostkou.

Uvažujme teď opačný případ: máme hodnoty, o kterých předpokládáme, že mají Gaussovo rozdělení[16] a budeme chtít dostat zpátky zmíněné parametry, pro naši otázku průměr. Pravděpodobně vás napadne použít Excel® nebo podobný nástroj a použít funkci *průměr*, respektive její anglickou variantu *mean*. Pokud si myslíte, že jste tím právě spočítali skutečný průměr, tak Vás musím zklamat, nespočítali. Data, která jste měli k dispozici, popisovala *nějakou* náhodnou veličinu[17]. Hodnota, kterou jste spočítali tím, že jste všechna čísla sečetli a vydělili jejich počtem, musí tedy taky být náhodná veličina – jinými slovy podle toho, jaká máte data, tak takovou hodnotu *průměru* dostanete. A jak tedy nazvat to, co jsme právě spočítali? Běžně se používá výraz ***odhad průměru***, anglicky ***mean estimate***, značí se $\hat{\mu}$ a má také svůj průměr a směrodatnou odchylku.

Rozdíl mezi průměrem a jeho odhadem by měl být tedy zřejmý. Průměr je jedna konkrétní (správná) hodnota, zatímco spočtený odhad průměru je náhodná veličina, která se jí v závislosti na množství a kvalitě dat ji blíží.

Matematické okénko: (vhodné pro zájemce)
Proč můžeme odhadnout průměr tak, že spočítáme odhad průměru?

Aneb, proč je počítání průměru tak populární a my lidé jej používáme i tehdy, když bychom neměli? Ukažme si to s trochou matematiky. K odhadu parametrů se často používá hodnota založená na maximalizaci věrohodnosti dat. Představte si, že máte data $X = \{x_1, x_2, \dots, x_n\}$, která jsou generována z Gaussova rozdělení s parametry μ a σ. Největší věrohodnost dosáhneme tím, že nastavíme takové parametry μ a σ, aby pravděpodobnost, že naše data X byla generována z právě tohoto rozdělení, byla co největší. Matematicky řečeno, odhad $\hat{\mu}$ mající maximální věrohodnost je taková hodnota možných parametrů μ, která maximalizuje věrohodností funkci $\mathcal{L}$:

$$\arg\max_{\mu} \mathcal{L} = \prod_i \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right)$$

po zlogaritmování dostaneme:

$$\arg\max_{\mu} \log \mathcal{L} = \frac{-1}{2\sigma^2} \sum_i (x_i - \mu)^2 - \log\sqrt{2\pi}\,\sigma$$

Z čehož za použití derivace[18] dostaneme, že

$$\hat{\mu} = \frac{1}{n} \sum_i x_i$$

[16] Tento předpoklad normálnosti můžeme také testovat, což bude pravděpodobně zmíněno v příštích dílech seriálu. Tím ověříme, že data skutečně mají *nějaké* normální rozdělení.
[17] Kvůli tomu, že je náhodná, ten *„průměr"* chceme spočítat.
[18] Aplikace je ponechána zkušenému čtenáři jako domácí cvičení.

Což je shodou náhod definice průměru. Co jsme to právě ukázali? Že způsob, jak s maximální věrohodností odhadnout skutečný průměr veličiny s **Gaussovým rozdělením,** je spočítat průměr známých dat. Ale pozor! Tohle platí pouze pro náhodné veličiny mající Gaussovo rozdělení, pro jiná statistická rozdělení dostaneme jiné odpovědi.

Literatura

[1] J. K. L. M. Vojtěch Kubíček, Anorganické praktikum, Univerzita Karlova.

[2] J. Čížek, „Úvod do praktické fyziky NOFY055,“ 2020. [Online]. Available: https://physics.mff.cuni.cz/kfnt/vyuka/upf/cizek/.

[3] G. Cowan, Statistical Data Analysis, Oxford: Clarendon Press, 1998.

Zadání úloh 4. série 24. ročníku KSICHTu

Úloha č. 1: Čajová **(8 bodov)**

Autor: Šimon Zuber

Čaj je hneď po vode najviac používaná kvapalina. Pijávam ho každé ráno už od detstva, no záľuba ku čaju prišla pred pár rokmi. Začalo sa to pestovaním a zbieraním bylín a prešlo to až ku čajom ako takým. Teraz by sa hodilo upresniť si, čo je vôbec čaj. Čaj je výluh z rastliny Camellia sinensis. Podľa tohto môžeme teda označiť všetko ostatné ako nečaje. Pravda je taká, že čajovinu, ktorá ide do vrecúška, kúpite vo Vietname napr. za 3 doláre za kilo (takže v jednom vrecúšku môže byť i len 0,05 centu čajoviny (The Global Economy, 2026)). Poďme sa ale teraz pozrieť ako taký proces výroby čaju vôbec vyzerá.

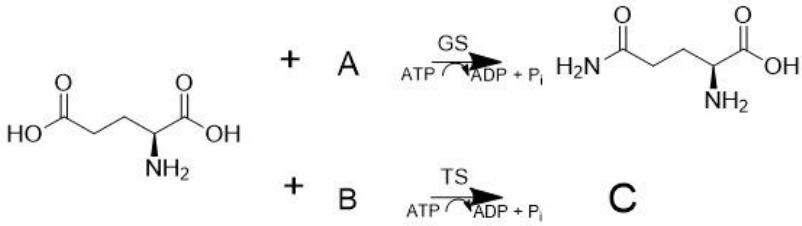
Časť A – Teoretická

1. Pri procese výroby nás hlavne zaujíma agitating (rozrušenie) čaju a fermentovanie (po chemickej správnosti oxidácia). Napíšte mi 5 základných typov čaju podľa fermentácie.

2. Napíšte aspoň jednu časť procesu, ktorú majú všetkých 5 typov spoločnú, a dve, ktoré sú rozdielne.

3. „Najjednoduchší“ na počet procesov je biely čaj. Na začiatku, po tom ako ich zoberú zo stromov, je *withering*, kedy sa listy rozložia na veľkú plochu a nechajú sa usušiť/zvädnúť. Napíšte 4 faktory, ktorými dokážeme ovplyvniť tento proces.

4. Pri tomto procese sa veľa toho mení, napr. cukry sa hydrolyzujú. Takisto sa látky chemicky premieňajú a mení sa ich chuť alebo iné senzorické vlastnosti, ako napríklad taký L-theanín(C) alebo L-glutamín. Doplňte reaktanty A, B, C a celé názvy pre GS a TS.



5. Z kyseliny glutámovej môžeme zosyntetizovať aj inú látku. Je po nej pomenovaný celý typ čaju. Má výraznú umami chuť. Listy sa po *witheringu* dajú do sudu, ktorý sa naplní dusíkom a ostanú tam po určitú dobu. Ako sa volá daná látka a prečo je pre telo prospešná?

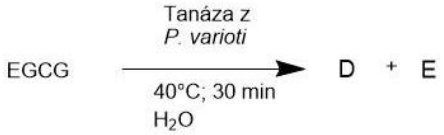
6. Pri *witheringu* takisto dochádza ku syntéze aromatických a prchavých molekúl. Pozrime sa preto na niektoré z nich. Doplňte skeletový vzorec, arómu alebo chuť a typ čaju, pre ktoré je to charakteristické (stačí jeden).

Molekula	**Vzorec**	**Aróma/chuť**	**Typ čaju**
(E)-2-oktenál			
2-fenylethanol			
Myricetín			
Trans-β-jonón			

7. Jedny z najvýznamnejších látok v čaji sú polyfenoly. Je to veľká skupina organických látok vyskytujúcich sa v prírode, ktoré obsahujú veľa (preto „poly“) fenolových skupín. Problém je, že sa ťažko, alebo len v malých množstvách dostanú do nášho tela. Pomenujte a nakreslite dva polyfenoly v listoch.

8. Kde a prečo sa používa v priemysle *Paecilomyces varioti tanáza?*

9. Doplňte nasledujúci mechanizmus a nakreslite a pomenujte EGCG, D, E.



10. Mohli ste si všimnúť, že pri „čínskom“ spôsobe čajovania sa nepoužívajú vrecúška, namiesto toho sa používa napríklad zhong alebo chahu. Jeden z dôvodov je lineárna chuť výluhu. Vysvetlite ako to súvisí s Fickovým zákonom. Pridajte aj zdroj.

11. Ďalším dôležitým faktorom je samotný materiál čajníku, môže byť napríklad zo striebra, skla, ale najviac používaný materiál je „*purple clay*.“ Medzi jej vlastnosti patrí dobré udržanie tepla a zníženie horkosti čaju. Čím je to? Nakreslite vertikálny prierez strany čajníku a vyznačte čo tam spôsobuje danú vlastnosť. (Aj ten prierez má svoj dôvod!)

12. Takisto ako pri víne, aj pri čaji je dôležitý *terroir* (súbor všetkých prírodných podmienok ktoré ovplyvňujú pestovanie plodiny). Chuťovo iné sú vysokohorské čaje (takisto aj množstvo čajoviny sa vyžaduje väčšie, pre množstvo pre pochopenie plnej komplexnej chute). Klimatické podmienky vo vyšších nadmorských výškach, najmä nedostatok priameho slnečného žiarenia a chlad, menia biochemickú rovnováhu v listoch. Ako tieto podmienky bránia premene L-theanínu na polyfenoly?

Časť B – Praktická

13. Ako sme sa dozvedeli v časti teoretickej, vytvorenie čajoviny môže byť pomerne jednoduchý spôsob, u nás sa to predsa robí už mnoho rokov s bylinkami. Pozrime sa ako by to vyzeralo keby sme pripravili bylinky práve čínskym spôsobom úpravy listov. Odtrhnite bylinku vášho výberu zo zeminy, napríklad v záhrade, alebo môže byť aj mäta z obchodu ktorá ale bude ešte zasadená. Rozdeľte vzorku na polovice a dajte ich do prostredia kde zmeníte iba jeden faktor na základe Vašej odpovede v úlohe 2a. Po 48 hodinách odoberte vzorky a porovnajte štruktúru listov a spravte výluh. Do odpovede mi prosím napíšte krátky report o rozdieloch v štruktúre pri daných prostrediach a rozdieloch vo výluhu. Posielajte fotku a report na mail simon.zuber@ksicht.natur.cuni.cz. Prajem krásne čajovanie.

Úloha č. 2: Toskánská **(9 bodů)**

Autorka: Adéla Hanousková

Nápad na tuhle úlohu vznikl během zdánlivě nekonečné cesty z Toskánska, kam se nyní společně se Zajíčkem Chemikem podíváme. Náš malý výlet může sloužit jako důkaz toho, že chemie nás provází na každém kroku. Stačí se jen pozorně dívat.

Je konec prosince a naše cesta začíná v malém obytném voze. Zajíček už se těší, až vyrazí, čeká ho však dlouhá cesta a nemůže podcenit přípravy. Jelikož chce ušetřit, rozhodl se, že si bude vařit na plynovém vařiči. K tomu potřebuje propan-butanovou tlakovou láhev.

1. Aby se Zajíček dostal do své cílové destinace, musí přejet Alpy, kde jsou hlášené mrazy okolo −6 °C. Pokud tam chce vařit, měl by si raději koupit láhev s poměrem propanu a butanu 6:4 nebo 4:6? Své rozhodnutí zdůvodněte.

Díky správnému výběru plynové láhve jsme hory překonali bez sebemenšího problému s hladem a nyní se dostáváme k první zastávce naší výpravy. Otevírá se nám dechberoucí pohled na renesanční Florencii. Zajíček si chce město pořádně prohlédnout, a tak zavítá do Galerie Akademie, kde se nachází slavné Michelangelovo dílo, socha Davida.

2. Zajíček je touto 517 cm vysokou a 5,5 t těžkou mramorovou sochou uchvácen. Neubrání se však své zvídavosti a přemýšlí, zda by šlo sochu, která je složena převážně z $CaCO_3$, rozpustit v kyselině chlorovodíkové. Představme si, že reakce probíhá za ideálních podmínek, s 37% HCl a mramorem, který je celý z bezvodého $CaCO_3$.

   a) Napište vyčíslenou rovnici reakce kyseliny chlorovodíkové s Davidem.

   b) Kolik litrů kyseliny by bylo potřeba na rozpuštění celé sochy, kdyby reakce měla 100% účinnost? Hustota HCl je 1,18 g/cm^3

   c) Stručně okomentujte, na jaké problémy bychom mohli narazit, kdybychom se tuto reakci skutečně pokoušeli provést.

David naštěstí stále stojí ve Florencii, ale Zajíčka to us ilovné přemýšlení úplně vyčerpalo. Rozhodl se tedy doplnit energii oblíbeným nápojem v blízké kavárně.

Za moment už má na stole šálek horkého italského espressa. Ochutná a nespokojeně se zamračí. Přemýšlí, proč má na jazyku místo blaženého pocitu hořkou pachuť.

3. a) Který chemický proces probíhající při pražení je zodpovědný za hořkou a spálenou chuť italské kávy?

   b) Stručně popište, k čemu během něj dochází. (1–2 věty)

   c) Takto pražená káva obsahuje velké množství pyridinu. Uveďte triviální název a strukturní vzorec látky, ze které při tomto procesu pyridin vzniká.

   d) Kromě pyridinu z této látky vzniká určitá kyselina, která je zároveň vitamínem. O jakou kyselinu jde? Jak se nazývá nemoc, ke které může vést nedostatek této kyseliny?

Ač ochuzen o chuťový prožitek, doplnil si Zajíček dostatečné množství energie, aby mohl pokračovat ve svém dobrodružství. Rozhodl se zpříjemnit si chladné prosincové odpoledne, oblékl si plavky a zamířil do přírodních termálních lázní městečka Saturnie.

4. Dříve než však mohl spatřit průzračnou modř a ponořit své znavené tělo do vody mající hřejivých 37,5 °C, udeřil ho do čumáčku hnilobný vaječný odér.

   a) Zajíček ihned identifikoval, se kterou sloučeninou má čest. Napište název sloučeniny a vysvětlete proč si návštěvníci lázní mohou bezpečně užít koupel navzdory vysoké toxicitě této látky.

Zápachu si nakonec přivykl, stejně jako autorka úlohy, která ale narozdíl od Zajíčka udělala amatérskou chybu a vběhla do vody se stříbrným řetízkem na krku. Ten brzy svůj lesk ztratil, nyní ho pokrývá matná tmavá vrstva čehosi. Zajíček si poklepal na čelo a pomohl autorce vysvětlit, co se právě stalo.

   b) Napište vyčíslenou rovnici reakce, která se na řetízku odehrála.

   c) Pomozte nešikovné autorce navrhnout, jak její řetízek zachránit. (uveďte stručný postup a vyčíslenou chemickou rovnici.)

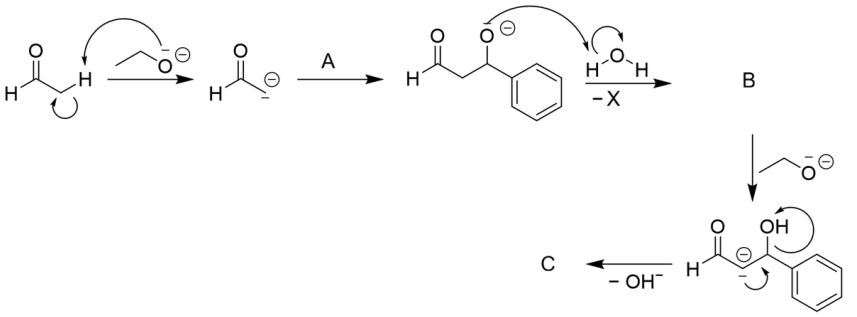
Při koupání Zajíčka zaujal ještě jeden barevný jev. Horniny omývané vodou z pramenů měly neobvyklou, téměř bílou barvu. Onen barevný film tvoří určitý druh mikroorganismů.

   d) Uveďte, o jaké organismy (obecně) jde a napište příklad rodu, který by teoreticky mohl být za tento jev odpovědný.

Vzhledem k daleké cestě, kterou Zajíček podnikl, nemohl slavit Vánoce tradičním způsobem. Rozhodl se tedy, že si alespoň syntetizuje své oblíbené vánoční vůně do malých lahviček, aby vánoční atmosféra mohla putovat s ním.

5. Identifikujte, jaké látky má Zajíček v lahvičkách:

   a) V reakčním mechanismu doplňte látky A a B, vedlejší produkt X a produkt C. U produktu uveďte jeho charakteristickou vůni a název.



   b) S další vonnou látkou se pojí její neodmyslitelná chuť. Jedním ze zdrojů pro tuto sloučeninu může být například polymerní lignin. Napište její vzorec a název.

   c) V poslední lahvičce nalezneme směs tří látek, které společně tvoří vůni borovice. Bez vánočního stromečku by se Zajíčkova výprava totiž neobešla. Jde o dva typické izomery monoterpenu a ester kyseliny octové. Uveďte vzorce a názvy všech tří látek.

Přestože byla zima v rozpuku, při pohledu na zářivý nápis Gelato Zajíček neodolal a nechal si nandat kelímek plný lahodné zmrzliny. Radoval se, že mu v té zimě alespoň neroztaje tak rychle a plastovou lžičkou úhledně opracovával jahodový kopeček.

6. Aby zmrzlina držela svou strukturu, přidávají se do ní dvě klíčové skupiny látek – emulgátory a stabilizátory.

   a) K oběma napište, jakou roli ve struktuře zmrzliny hrají.

   b) Napište 2 příklady emulgátoru a 2 příklady stabilizátoru, které se používají pro výrobu zmrzliny.

   c) Pokuste se schematicky nakreslit strukturu zmrzlinového kopečku. Schéma by mělo obsahovat emulgátory i stabilizátory. (Nápověda: nezapomeňte na proteiny).

Náš výlet se pomalu chýlí ke konci a zbývá jen jediné – přivézt domů italské víno. Pro toskánský region je typické víno Chianti a Zajíček se rozhodne podívat se na tento vyhlášený kousek zblízka.

7. Hned na první pohled ho zaujme nápis DOCG na hrdlu láhve.

   a) Co tato zkratka značí?

   b) Etiketa hlásá, že jde o suché víno. Zajíček tento pojem slyšel už mnohokrát, nikdy se však nezamýšlel nad tím, co přesně to o vínu vypovídá? Poraďte mu.

   V 2 dl Chianti s aciditou 5,9 g/l je přibližně 630 mg zbytkového cukru.

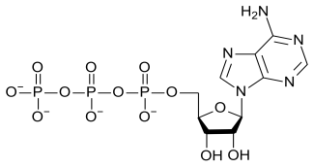
   c) Vysvětlete, proč jde o suché víno. (Budete potřebovat tabulku s obsahy zbytkových cukrů, kterou najděte na internetu. Uveďte zdroj vaší informace.)

   d) Může vinař označit své víno za suché, mělo-li by o 2 mg/cm^3 cukru víc?

Poslední údaj, na který se Zajíček podívá, je číslo na cenovce. Vypadá to, že si domů přiveze raději sušenky. :D

Úloha č. 3: Bioenergetická I **(12 bodů)**

Autor: Tadeáš Grabic



V lidském těle každou sekundu proběhne nespočet chemických procesů. Buňky vyrábějí nové proteiny, Na–K pumpy neúnavně obnovují membránový potenciál v neuronech a enzymy v jaterních buňkách pracují na odstraňování toxinů z krve. Valná většina těchto procesů vyžaduje dodání energie a my se v této dvoudílné úloze podíváme na to, kde se energie v těle bere.

Při řešení této úlohy ***prosím nepoužívejte umělou inteligenci****.*

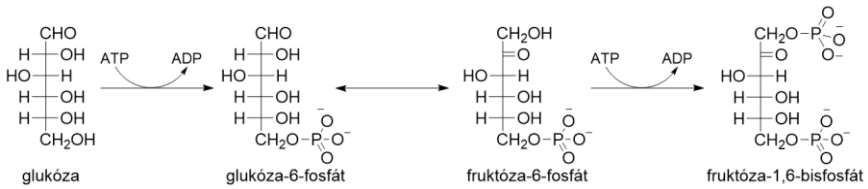
V buňce probíhají z termodynamického hlediska dva typy reakcí. Důležitou roli hraje veličina zvaná Gibbsova energie. Její změna má vliv na samovolnost reakce – je-li změna Gibbsovy energie záporná, reakce uvolňuje energii a probíhá samovolně. Takové reakce pak označujeme jako exergonické. Při kladné změně Gibbsovy energie reakce samovolně neprobíhají a označujeme je jako endergonické.

Aby mohly tyto reakce probíhat, tak je k jejich průběhu potřeba nějakou energii dodat. Toho je docíleno pomocí takzvaných spřažených reakcí – energeticky nevýhodná reakce je spojena s nějakou reakcí, která je natolik energeticky výhodná, že průběh potřebné reakce umožní. Častým „univerzálním platidlem“ je hydrolýza ATP, jehož struktura je ukázána na obrázku v úvodu úlohy.

ATP je získáváno pomocí katabolických reakcí – štěpením větších molekul na menší. Mezi nejdůležitější katabolické reakce patří glykolýza a β-oxidace, na které se podíváme v této úloze.

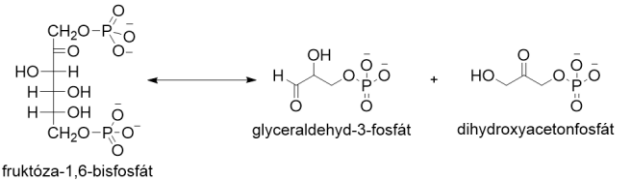
1. Jak je glykolýza nazývána jinak? Berte v potaz metabolické dráhy aktivní u člověka.

2. Prvním krokem glykolýzy je fosforylace glukózy za vzniku *glukózy-6-fosfátu*. Proč buňka tuto reakci provádí?



Obr. 1: Přípravná fáze glykolýzy.

Následuje izomerace *glukózy-6-fosfátu* na *fruktózu-6-fosfát*. Ta je pak za spotřeby druhé molekuly ATP přeměněna na *fruktózu-1,6-bisfosfát*, která je štěpena na dvě triózy – *glyceraldehyd-3-fosfát* a *dihydroxyacetonfosfát*. Do dalších reakcí míří pouze jedna – *glyceraldehyd-3-fosfát*.



Obr. 2: Štěpení fruktózy-1,6-bisfosfátu na dvě triózy.

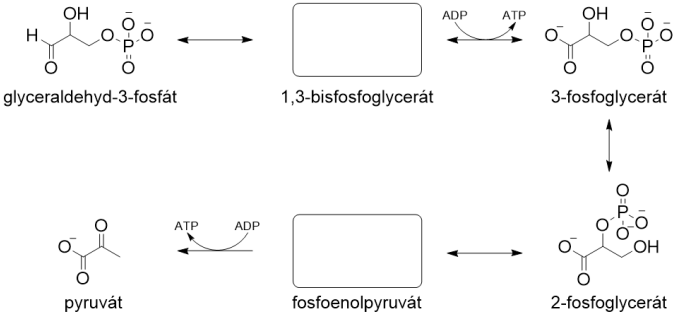
3. Druhá z trióz – *dihydroxyacetonfosfát* však nepřijde nazmar. Buňka totiž umí přeměňovat jednotlivé izomery sacharidů mezi sebou.

   a) Která třída enzymů obecně katalyzuje přeměnu izomerů mezi sebou a jak se tato třída nazývá?

   b) Kolik různých izomerů hexóz existuje?

   c) Umí lidská buňka využít všechny tyto izomery? Stručně zdůvodněte svou odpověď.

Glyceraldehyd-3-fosfát je poté oxidační reakcí přeměněn na *1,3-bisfosfoglycerát*. Makroergní povaha (schopnost uvolňovat velké množství energie při štěpení) této molekuly umožňuje substrátovou fosforylaci – fosfát z polohy 1 je přenesen na molekulu ADP, přičemž vznikne *3-fosfoglycerát* a ATP. Ten je izomerizován na *2-fosfoglycerát* a poté dehydratován za vzniku *fosfoenolpyruvátu*. Ten je opět makroergní, takže se může uskutečnit druhá substrátová fosforylace. Z *fosfoenolpyruvátu* vzniká *pyruvát* a fosfát je navázán na ADP za vzniku ATP.



Obr. 3: Oxidační fáze glykolýzy

4. V textu výše je zmíněna oxidační reakce. Která látka slouží v tomto případě jako akceptor elektronů a co z ní vzniká?

5. Nakreslete struktury *1,3-bisfosfoglycerátu* a *fosfoenolpyruvátu*. Vyznačte v nich makroergní část a stručně okomentujte, proč nese vyznačená část velké množství energie a umožňuje fosforylaci ADP na ATP.

Osud vzniklého *pyruvátu* závisí na podmínkách v buňce. V případě aerobních podmínek vstupuje do Krebsova cyklu, kde je přeměněn na CO_2. Za anaerobních podmínek je kvašením přeměněn na *laktát*, respektive *ethylalkohol*.

6. Proč v anaerobním prostředí probíhá kvašení a zpracování sacharidů nekončí u pyruvátu?

Při glykolýze tedy vzniknou 4 molekuly ATP, ale zároveň jsou dvě spotřebovány při aktivaci glukózy. Celkový zisk na jednu molekulu glukózy tedy činí dvě molekuly ATP… no, to není úplně moc, ale anaerobním bakteriím (a červeným krvinkám) to stačí.

7. Je ATP jedinou energeticky využitelnou molekulou vzniklou při glykolýze? Pokud ne, napište, jaké produkty glykolýzy lze dále využít k produkci energie a kde.

Dalším běžným zdrojem energie jsou triacylglyceroly (TAG). Ty jsou odbourávány pomocí β-oxidace.

8. Nakreslete strukturu 1,2-dipalmitoyl-3-stearoyl-glycerolu.

Při β-oxidaci jsou odbourávány pouze mastné kyseliny. V TAG se však kromě nich nachází glycerolová kostra.

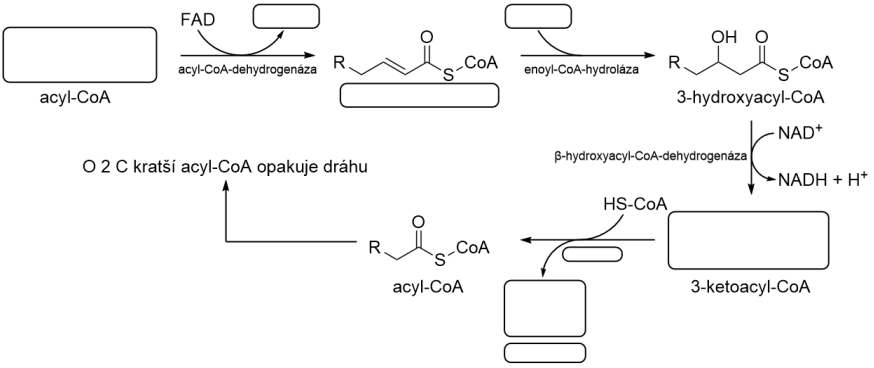
9. Do jaké metabolické dráhy se glycerol zapojuje a jak? Nakreslete schéma vstupu glycerolu do této metabolické dráhy včetně zapojených kofaktorů a energetických molekul, enzymy pojmenovávat nemusíte.

   Mastná kyselina musí být před vstupem do β-oxidace aktivována.

10. V čem tato aktivace spočívá? Napište, kolik ekvivalentů ATP tato aktivace buňku stojí.

Aktivovaná MK pak vstupuje do vlastní β-oxidace, jejíž schéma je na obrázku níže.

11. Do schématu β-oxidace doplňte chybějící vzorce, názvy či zkratky.



Produktem β-oxidace jsou tedy redukované formy koenzymů a *acetyl-CoA*. Ten poté vstupuje do Krebsova cyklu, který pro svůj průběh vyžaduje aerobní podmínky.

12. Může β-oxidace poskytovat energii za anaerobních podmínek? Vysvětlete vaši odpověď.

Úloha č. 4: Vlaková **(10 bodov)**

Autorka: Júlia Hodálová

Zajíček Chemik sa chystá na KSICHTí výlet vlakom. Pri kúpe lístka (s miestenkou) uvažuje, či si má sadnúť bližšie k rušňu (česky lokomotivě) alebo radšej nie. „Lokomotíva je motor vlaku. Ako hlučná bude v prvom vozni a v druhom vozni, kde budem mať väčší pokoj na vyriešenie tejto úlohy?“

„Lokomotíva je motorom celého vlaku. Spaľovaním paliva vyrába energiu, ktorá sa premení na teplo a to teplo sa premení na prácu, ktorá rozpohybuje vlak,“ uvažuje Zajíček.

1. a) Nájdite jednu chybu v Zajíčkovej úvahe a opravte ju.

   b) Zdôvodnite svoju odpoveď v otázke 1a). [max 2 vety]

   c) Aký zákon porušuje Zajíčkova úvaha?

Zajíček sa ale rýchlo uvedomí a pustí sa do hlbšieho porozumenia zákonov termodynamiky. V termodynamike sa svet delí na dve časti: pozorovanú sústavu (vymedzený priestor, ktorý skúmame, napríklad skúmavka, reaktor, vlak...) a okolie (všetko ostatné). Sústavy možno rozdeliť na základe vymieňania energie a hmoty s okolím.

2. a) Aký typ sústavy je vlak bez výmeny cestujúcich (nikto nenastupuje, nevystupuje, zmena hmotnosti paliva je zanedbateľná)?

   b) Aký typ sústavy je vlak bez cestujúcich uzavretý v dokonale zateplenom depe (zateplený ako kalorimeter)?

   c) Ktorý typ termodynamickej sústavy nikdy nemôže predstavovať idúci vlak (otvorenú, uzavretú alebo izolovanú)? Odpoveď zdôvodnite. [max 3 vety]

Prvý zákon termodynamiky pozná Zajíček zo školy:

$$\Delta U = Q + W$$

Slovne si ho môžeme interpretovať ako „zmena vnútornej energie sústavy (systému) je rovná súčtu tepla a práce, ktorú sústava prijala/odovzdala.“ Sústavou prijatá práca alebo teplo (z okolia) sa značí kladným znamienkom. Práca alebo teplo odovzdané sústavou do okolia sa označuje záporným znamienkom.

3. Aké bude znamienko tepla sústavy (vlaku) a prečo? Použite 1. zákon termodynamiky a odpovedajte $+ - -$ nedá sa určiť.

a) Ak je zmena vnútornej energie sústavy kladná a súčasne vlak konal prácu na okolí.

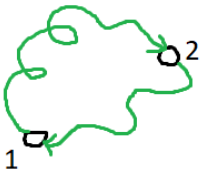
b) Ak je zmena vnútornej energie sústavy kladná a súčasne okolie konalo prácu na systéme.

V tejto úlohe nás (a Zajíčka) bude zaujímať najmä druhý zákon termodynamiky. Slovne ho môžeme definovať takto: „Nemožno zostrojiť taký stroj, ktorý dokáže všetko dodané teplo premeniť na prácu.“ Na konštrukciu stroja potrebujeme chladič, do ktorého je odovzdaná zvyšná energia, ktorá sa nedá premeniť na prácu. Prácou je myslená práca objemová:

$$W = F \cdot \Delta d = p \cdot S \cdot \Delta d = -p \cdot \Delta V.$$

Kde Δd je posunutie piestu, F je sila, S plocha piestu, p je konštantný tlak (ide o izobarický dej) a ΔV je objem vymedzený $S \cdot \Delta d$. Objemovú prácu koná plyn uzavretý pohyblivým piestom.

Ale ako zostrojiť takú lokomotívu (takú sústavu), ktorá by bola čo najefektívnejšia? Motor lokomotívy (tepelný stroj) pracuje cyklicky. To znamená, po dokončení cyklu[19] sa stavové funkcie „vrátia“ na pôvodné hodnoty (v ideálnom prípade), čiže zmena stavovej veličiny je **nulová**.



Obr. 1: Termodynamický cyklus z východiskového stavu 1, cez stav 2 a späť do stavu 1.

Pracujúce stroje možno opísať Carnotovým cyklom. Účinnosť cyklicky pracujúceho tepelného zdroja (η = „éta“) je daná pomerom vykonanej práce a prijatého tepla:

$$\eta = \frac{-W}{Q_h}.$$

[19] Cyklus nemá začiatok ani koniec. Vhodnejší výraz je „stav, kde začíname naše pozorovanie“ príde do „stavu, kde sme začínali naše pozorovanie“.

Význam $-W$ spočíva v našej dohode, že práca vykonaná (odovzdaná) systémom do okolia sa značí záporným znamienkom. Teda vo finále máme $-(-W)$, čo zaručuje kladnú účinnosť. Niektoré zdroje uvádzajú napr.:

$$\eta = \frac{|W|}{|Q_h|}$$

s absolútnymi hodnotami. Netreba sa toho báť. Účinnosť musí byť kladné číslo menšie ako 1, ide o ten istý zápis napísaný trochu inak.

4. V akom prípade by bol člen Q_h záporný? Uveďte jeden príklad takéhoto stroja.

   Nápoveda: Kedy je prijaté teplo zo zdroja záporné? (tj. teplo „prúdi“ odkiaľ kam?)

Vieme, že pre uzavretý systém platí, že po dokončení cyklu je zmena stavovej veličiny nulová. Takže prvý zákon termodynamiky môžeme prepísať:

$$\Delta U = Q_{celkové} + W \Rightarrow 0 = Q_{celkové} + W \Rightarrow -W = Q_{celkové}$$

Čiže práca, ktorú odovzdal systém okoliu, je rovná celkovému teplu. Z druhého termodynamického zákona vieme, že nemožno zostrojiť stroj, ktorý premieňa teplo z jedného tepelného zásobníka na prácu. Zásobníky tepla potrebujeme (najmenej) dva. Celkové teplo súčtom tepla prijatého a odovzdaného:

$$-W = Q_h + Q_c$$

Indexami sa myslí teplo ohrievača (h, *hot*) a chladiča (c, *cold*). Pre účinnosť získame vzťah:

$$\eta = \frac{Q_h + Q_c}{Q_h}.$$

Po detailnom rozpísaní fáz Carnotovho cyklu vieme účinnosť prepísať do tvaru:

$$\eta = \frac{T_h - T_c}{T_h}.$$

5. a) Sú parné lokomotívy teoreticky účinnejšie v zime alebo lete?

   b) Na základe a) formulujte predpoklad, zaručujúci čo najväčšiu účinnosť stroja.

6. Parné lokomotívy dosahovali vo všeobecnosti účinnosť 5 % až 8 %. Prácu konala para, ktorá dosahovala teplotu 350 °C. Chladičom bol okolitý vzduch (atmosféra).

   a) Vypočítajte teoretickú účinnosť parnej lokomotívy. Teplotu chladiča považujme napríklad priemernú teplotu v Česku za rok 2025, t. j. 11,8 °C (nezabudnite na postup).

b) Prečo sa vypočítaná hodnota líši od skutočnej? Uveďte aspoň 2 faktory.

7. Prácu koná plyn, ktorý sa buď rozpína ($\Delta V > 0 \Rightarrow W < 0$) alebo stlačuje ($\Delta V < 0 \Rightarrow W > 0$).

   a) Majme 2 rovnaké nádoby s piestom, ktoré majú počas celého nášho pokusu konštantnú teplotu T. V jednej nádobe je 1 mol Ne, v druhej 1 mol SO_2. Oba plyny necháme expandovať z rovnakého počiatočného objemu V_1 na rovnaký konečný objem V_2.

      i. Vykoná väčšiu prácu 1 mol Ne alebo 1 mol SO_2? Svoju odpoveď slovne zdôvodnite.

      ii. Uveďte a okomentujte 2 rovnice, ktoré možno použiť v i. Nezabudnite na koeficienty pre oba plyny.

   b) Teraz majme plyn, ktorý tvorí zmes 0,5 mol Ne a súčasne 0,5 mol SO_2 pri rovnakých podmienkach ako v a). Porovnajte vykonanú prácu s prácou čistého 1 mol SO_2 a 1 mol Ne (tj. bude <,> alebo =).

V dnešnej dobe možno vidieť parné lokomotívy len výnimočne. Súčasné vlaky majú spaľovacie motory. Modernejšie vlaky sú elektrické alebo výnimočnejšie vodíkové.

Vodíkové vlaky sú predmetom súčasnej vedy. Vlak poháňaný iba vodíkom nie je tak efektívny ako hybridný vlak poháňaný vodíkovým palivovým článkom a súčasne batériou[20]. V palivovom článku (*fuel cell*) reaguje vodík s kyslíkom, za vzniku vodnej pary, tepla, respektive energie (vo forme elektrickej energie). Elektrická energia poháňa vlak. Zvyšná energia sa ukladá do batérie.

8. V Nemecku (Sasku) premáva historicky prvý osobný hybridný vodíkovo batériový vlak[21] [Alstom Coradia iLint, 107 t (bez pasažierov), 150 miest na sedenie].

   a) Koľko energie sa uvoľní úplným zreagovaním 1 kg vodíka, keď vieme, že:

$$H_2(g) + 0{,}5\ O_2(g) \rightarrow H_2O(g)$$

$$\Delta H° = -241{,}8\ \text{kJ} \cdot \text{mol}^{-1}$$

   b) Železničná vzdialenosť medzi Prahou (hlavní nádraží) a Brnom (hlavní nádraží) je približne 190 km. Predpokladajme, že vlak ide konštantnou

[20] https://www.sciencedirect.com/science/article/pii/S0360319922047231

[21] https://www.alstom.com/solutions/rolling-stock/alstom-coradia-ilint-worlds-1st-hydrogen-powered-passenger-train

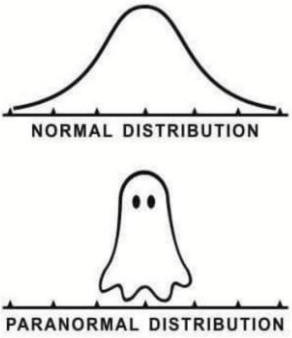
rýchlosťou a cesta trvá približne 2 h a 50 min. Priemerný výkon vlaku je 200 kW. Vypočítajte spotrebu vodíka (v kilogramoch) vlaku Alstom Coradia iLint, predpokladajte výkon 200 kW a celkovú účinnosť 40 %.

Nápoveda: niektoré údaje môžu byť navyše; nezabúdajte, že $\eta = \frac{W}{E}$.

Uf, Zajíček zavrie učebnicu, odloží kalkulačku a začne sa tešiť zo zaujímavých vecí a blížiaceho sa výletu.

Task n. 5: Markovs′ DNA chain **(11 points)**

Author: Jan Stoklasa



Statistics is not just an abstract field of mathematics; you will encounter it in places that are not obvious at first glance. One example of a simple model is the one introduced in the early 20th century by the Russian mathematician Andrey Andreyevich Markov.

Answers are accepted in English, Czech and Slovak, preferably in English.

A small free hint for you: using a computer program, *e.g.*, Excel or Python can significantly help you not only at the end of this task – feel free to use one but mention how you used it in your solution.

First, let me make one statement that is absolutely correct, but totally useless: *"Markov models are models that use Markov property".* I assume that does not help much. Let's try to explain it better.

Imagine you are on the 3rd floor of a tall building near the staircase. From here you have two actions you can take – go up one floor or down one floor. To put this into mathematical terms, we define the state space as a set of all possible floors where we can be, and the action space as a set of all possible actions we can take. Assume that the probability of taking an action "going down" is twice likely as taking an action "going up".

1. What are the probabilities that after taking one action you will be on the 1st, 2nd, 3rd or 4th floor?

What do we need to calculate the probabilities in question 1? We used the probabilities of taking individual actions (in this case two possible actions) and the information that we were on the 3rd floor. Mathematically speaking, we only need a state from the state space and an action from the action space. However, we do not need the information about how we got into that state. In other words: The future is independent of the past given the present.

Let's discover one more characteristic of Markov processes: the assumption of stationarity. This means that the probability of taking an action a in a state s is independent of time. Now, imagine our state space are not the floors in the building but three places that you visit during your regular day: an office, a restaurant and your home – and the actions are moving from one place to another;

2. Does this example follow or break the assumption of stationarity? Why?

For all our following examples, we will assume stationarity. And for now, let's bring in some chemistry.

DNA can be viewed as a very long string (text) over the alphabet of four letters (nucleobases) – A (adenine), C (cytosine), G (guanine) and T (thymine). Although these bases form pairs, only one strand of DNA is of interest to us.

How do we model DNA as a Markov model? First, there is the initial probability – the probability that the segment of DNA that was given to us begins with a specific base, more formally, the probability of moving from initial state (state of nothingness) to state with exactly one base. Since we have four nucleobases (letters) in our alphabet, we need four numbers describing the following:

$$\mathrm{P(A|begin)},\ \mathrm{P(G|begin)},\ \mathrm{P(C|begin)} \text{ and } \mathrm{P(T|begin)}$$[22]

Note: DNA is extremely long and the segments we analyze can randomly start anywhere along its whole length. More accurately, therefore, the above probabilities are the probabilities of a nucleobase's occurrence throughout the whole DNA strand. This may help you with question 5 below.

Furthermore, there are other probabilities, e.g., $\mathrm{P(T|A)}$ describing the probability that in any sequence a base A is followed by a base T. These probabilities are called transition probabilities.

3. How many distinct transition probabilities do we have in our DNA model?

Finally, we need the opposite of initial probabilities – terminal probabilities – for moving from any state of one base to the end state, e.g., $\mathrm{P(end|A)}$. These terminal probabilities determine the probability that a sequence terminates after a particular nucleotide, and therefore affect the distribution of sequence lengths generated by the model – higher values of terminal probabilities generally result in shorter sequences on average, while lower terminal probabilities result in longer

A graphical way to visualize our Markov model is depicted in Figure 1. Each arrow in the figure represents one possible action that has its own probability.

4. For the DNA model in Fig. 1, formally define the state space.

[22] $\mathrm{P(A|begin)}$ is a shorthand notation for $\mathrm{P}(s_i = \mathrm{A}|s_{i-1} = \mathrm{begin})$, *i.e.*, the probability that *i*-th state is A if the *i*–1-th state is "begin"

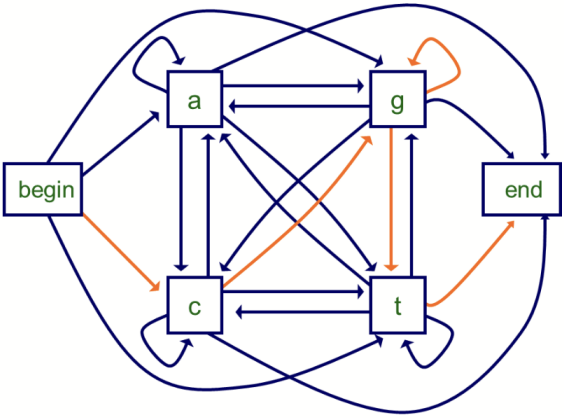


Figure 1: Schematic visualization of DNA sequence Markov model. The transitions corresponding to the sequence "cggt" are marked by orange colour (see text below question 4).

How do we calculate the probability of generating a sequence X of length L from this model? Since we use the key assumption of Markov chains, *i.e.*, that the probability of each x_i depends only on the value of x_{i-1}, the answer is the product of initial probabilities and all transition probabilities in the sequence[23]

$$\mathrm{P}(X) = \mathrm{P}(x_1|\mathrm{begin}) \prod_{i=2}^{L} \mathrm{P}(x_i|x_{i-1})$$

Let's show a particular example of a sequence X = "cggt"

$$\mathrm{P}(\mathrm{cggt}) = \mathrm{P}(\mathrm{c}|\mathrm{begin}) \cdot \mathrm{P}(\mathrm{g}|\mathrm{c}) \cdot \mathrm{P}(\mathrm{g}|\mathrm{g}) \cdot \mathrm{P}(\mathrm{t}|\mathrm{g}) \cdot \mathrm{P}(\mathrm{end}|\mathrm{t})$$

But how do we obtain these probabilities? The easiest way is to use the Maximum Likelihood Estimate (MLE): assume a given set of data D. Then we set the parameters Θ of our model to maximize the probability of generating all sequences in D, *i.e.*, to make the data D look as likely as possible under the model. Now, try it for yourself:

Let's have data $D = \{\mathrm{accgcgctta}, \mathrm{gcttagtgac}, \mathrm{tagccgttac}\}$

[23] Π is a symbol for product (https://en.wikipedia.org/wiki/Product_(mathematics)#Product_of_a_sequence)

MLE can be implemented by using relative frequencies in the given data. The initial probabilities can be calculated as follows[24]:

$$P(x|\text{begin}) = \frac{n_x}{\sum_{i\in\{\text{a,c,g,t}\}} n_i} \text{ where } n_x \text{ is count of base } x \text{ in data } D$$

5. For the given data D, calculate all initial probabilities as a relative frequency.

   Similarly, the transition probabilities can be calculated as follows:

   $P(x|y) = \frac{n_{yx}}{\sum_{i\in\{\text{a,c,g,t}\}} n_{yi}}$, where n_{yx} is count of the subsequence yx in data D

The probability of y being followed by x is the ratio of the counts of all pairs yx to the counts of all pairs containing y and any other nucleobase (a, c, g or t)

We previously included the terminal probabilities to describe their properties in the introduction above; however, for simplicity, we will exclude them from further calculations.

6. For the given dataset D, calculate all transition probabilities in our DNA model. Please ignore the terminal probabilities in this example.

   Note: The probabilities can often be organized into a structure called a *transition matrix* or *table*, but feel free to use any form that suits you. However, if you use the transition matrix (table), make sure to describe it properly to identify which values can be found in the specific columns and rows.

As you can see, some of your transition probabilities are equal to zero, quite a peculiar number if we are talking about probabilities… Let's see if that is a problem.

7. Compute the probabilities of generating the following two sequences $D_{\text{test}} = \{\text{acgcgct}, \text{acggcgc}\}$ from our model, using the results of questions 5 and 6.

As you can see, having a transition probability of zero is really a problem – if a pair that was not seen in the training data occurs in a sequence, the probability of the whole sequence becomes zero. However, that can be solved. Namely, so-called Laplace estimates represent a way of smoothing discrete variables[25]. It is a really

[24] $\sum$ is a symbol for summation (https://en.wikipedia.org/wiki/Summation)

[25] Mathematically, we maximize the posterior probability instead of the likelihood. Laplace estimate is one of the feasible approaches. What do we mean by "*maximizing posterior probability*"? Your prior assumption belief is that no natural DNA transition should have an absolute zero probability just because of a small sample size. The posterior is your updated belief which combines the actual observed data with your prior assumptions.

easy approach, in which we pretend that we have seen every possible combination of two nucleobases once before we start counting pairs in the data.

$$P(x|y) = \frac{n_{yx}+1}{\sum_{i\in\{a,c,g,t\}}(n_{yi}+1)}$$, where 1 is a pseudocount.

8. Recalculate the answer from question 5 by using the Laplace estimate (pseudocounts +1). All values should now be non-zero values.

In your answer to question 7 you have may noticed another problem. As the length of the tested sequence increases, the probability becomes smaller. This leads to two main results. First, we can only compare the probabilities of two sequences of the same length. Second, after a certain point, the probability is too small for your calculator or computer program to handle. However, mathematics provides a solution – negative logarithm. Usually, we are not interested in the exact value of the probability – we might just want to compare two values and select the larger one (just as you will do in the last question). Therefore, if we apply the negative logarithm[26] to obtain the negative log-probability of the sequence, we transform the product of probabilities

$$\mathrm{P}(X) = \mathrm{P}(x_1|\mathrm{begin}) \prod_{i=2}^{L} \mathrm{P}(x_i|x_{i-1})$$

into a sum of negative log-probabilities:

$$-\log \mathrm{P}(X) = -\log \mathrm{P}(x_1|\mathrm{begin}) + \sum_{i=2}^{L} -\log(x_i|x_{i-1})$$

This preserves the relative order of two different sequences or probabilities from two different models, but it makes the value reasonably small to fit our calculator or computer memory. However, due to multiplying the log-probability by −1, we have to switch from searching for the maximum probability to finding the minimum negative-log probability. This may be useful in the last question.

Let's apply this knowledge onto something more interesting.

CG-content is the percentage of C and G bases in the genome. For humans, this content is approximately 42%. For any pair of consecutive nucleotides (in short, any dinucleotide), we can calculate their probability using the CG-content. In

[26] The base of the logarithm can impact the properties described in following section. Make sure you do not use any extraordinary logarithm base without knowing how it can transform the logarithm function

DNA, not all dinucleotides are equally likely to occur. For example, C immediately followed by G is not as common as expected (4.4% in human).

The C nucleotide in CG dinucleotide often mutates to T. This mutation is called methylation. However, methylation usually does not occur in parts of the genome called CpG-islands. The CpG-islands are located around starts of genes.

Knowing that the CpG-islands usually precede genes, finding those regions might be used for classification of genes.

For this question, we will use the following matrix probabilities depicted in Figure 2.

$P(c \mid a)$

	a	*c*	*g*	*t*
a	.18	.27	.43	.12
c	.17	.37	.27	.19
g	.16	.34	.38	.12
t	.08	.36	.38	.18

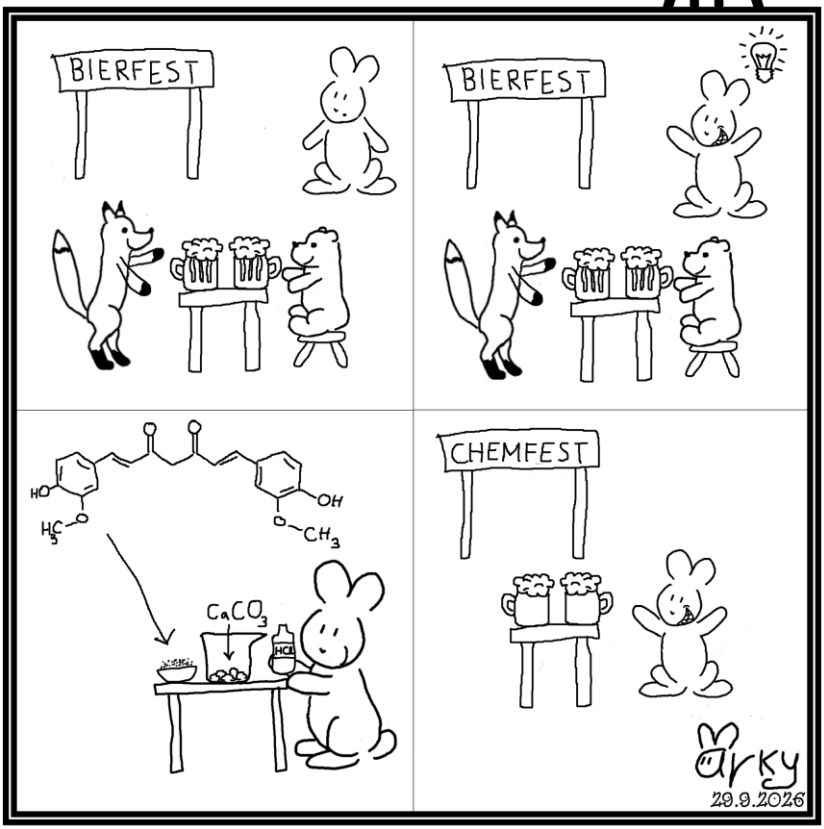
CpG

	a	*c*	*g*	*t*
a	.30	.21	.28	.21
c	.32	.30	.08	.30
g	.25	.24	.30	.21
t	.18	.24	.29	.29

null

Figure 2: Values of Transition probabilities of CpG and null DNA models

The CpG matrix shows the transition probabilities of sequences that are considered a CpG island; the null matrix shows the probabilities of other parts of DNA. We can classify the sequence as CpG or non-CpG based on the probabilities obtained from these two models defined by the matrices in Fig. 2. To do so we select the model that maximizes the probability of the sequence being generated by it.

9. For the following sequences, determine whether they are CpG island sequences or not. All sequences start with the nucleobase "a", so the initial probabilities can be ignored. Similarly, please ignore the terminal probabilities.
   c) aaattgaacgtccccgggcc
   d) aaggaacaaggacagaagga
   e) aggcgcccgagtgagacttttgcgtgccg
   f) atgccatgcggtgttttctagctcggggatgcgccgccat
   g) aaaagggcttagtgttacattcgattccaccctctccgaaatgctgaagacctgttctctgcatggtctggcg tcatcag

Zajíček chemik