Slezská univerzita v Opavě
Matematický ústav v Opavě

Aplikovaná statistika

Tomáš Kopf

Opava 2013

evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
Slezská univerzita v Opavě
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Hrazeno z prostředků projektu OPVK CZ.1.07/2.2.00/15.0174
Inovace bakalářských studijních oborů se zaměřením na spolupráci s praxí

Dítko moje,
nápady Tvoje
jsou jak
hbité, skryté, nečekané
příšery, potvory s raptory,
ach, teď zas' rozbilo's náhodné generátory.

Obsah

Předmluva

Toto je úvod do Bayesovské statistiky v rozsahu jednoho semestru.

Bayesovská statistika vychází ze subjektivně chápaných pravděpodobností, které jsou upravovány podle dostupných informací a používány pro praktické rozhodování na základě ztrátových funkcí resp. utilit. Její úvahy jsou tak zakotveny v konkrétních datech a v dohadech o dané situaci. Jejím cílem je doporučit nejlepší rozhodnutí pro tuto situaci.

Dějiny Bayesovské statistiky sahají daleko do minulosti. Tento přístup ke statistice je pojmenován po presbyteriánském knězi a matematikovi Thomasovi Bayesovi (přibl. 1701 – 7.dubna 1761), jehož posmrtně publikovaná práce [**3**] obsahuje jeden z prvních argumentů, které bychom dnes mohli označit za Bayesovské. Velmi brzy se ovšem objevila celá řada podstatných prací od Pierrea-Simona Laplace (23. března 1749 – 5. března 1827).

Že se tato praktická teorie v prostředí ovládaným konkurenční frekventistickou statistikou výrazněji prosadila až v závěru 20. století, má několik příčin.

Jednou příčinou byl jistě ostych pracovat se subjektivními pravděpodobnostmi, což vystavuje teorii útokům očerňujícím jí jako neobjektivní a nevědeckou. Opak je však pravdou: tato teorie zachází s pravděpodobnostmi zcela racionálně. Zjevným způsobem vyloží na stůl, co v je rozhodovacím procesu založeno na dohadech a co je založeno na ověřených faktech. To je jistě lepší, než subjektivní vstupy zastírat, nebo se dokonce vyhnout zodpovědnosti za důsledky statistické analýzy tím, že se statistika vyhne konkrétní vazbě na teorii rozhodování.

Je též nutno (byť bez přiměřené analýzy, která by šla za rámec tohoto textu) vzpomenout, že fundamentální podporu nachází použití subjektivních pravděpodobností v kvantové teorii, kterou případná závislost na hledisku subjektu pronásleduje od dob Schrödingerovy kočky [**20**].

Obrázek 0.1. **Schrödingerova kočka.** V tomto hypotetickém pokusu je kočka smrtícím mechanismem spouštěným kvantově-mechanickým radioaktivním rozpadem uvedena do superpozice živé a mrtvé kočky.

Druhou příčinou, která brzdila rozmach Bayesovského přístupu, je složitost výpočtů, které lze jen zřídka provést analyticky či analytickými aproximacemi. Tuto situaci zcela zásadním způsobem změnila kombinace vhodných numerických algoritmů s vzniklou možností je provést na dostatečně výkonných počítačích.

Text má za cíl zpřístupnit studentu Bayesovskou statistiku do té míry, aby jí v principu dokázal použít na praktickém problému ve své bakalářské či diplomové práci. Zabývá se následujícími tématy:

V Kapitole 1 je poskytnuta Bayesovská interpretace pravděpodobnosti, která jednak celou teorii zakotvuje v reálném světě, ale má i dopad na její formální podobu, která se svými cíli odlišuje od jiných přístupů k situacím s částečnou nejistotou.

V Kapitolách 2 a 3 je zavedeno a formalizováno rozdělení pravděpodobnosti pomocí teorie míry. Cílem není probrat teorii míry ve vší obecnosti a hloubce, ale trvat na základech této teorie, aby případná speciální přednáška z teorie míry a integrálu byla v jasném kontaktu s textem a aby byla připravena scéna pro sofistikovanější přístupy ze stochastických procesů.

Následně je v Kapitolách 4,5 a 6 diskutováno generování vzorků podle rozdělení pravděpodobnosti specializovanými i obecnými metodami se zvláštním důrazem na generování vzorků pomocí Markovových řetězců Metropolisovým-Hastingsovým algoritmem.

Kapitoly 7, 8 a 9 předkládají základní schéma Bayesovských statistických modelů, jejich zadání pomocí Bayesovských sítí a jejich řešení, v přímé vazbě na teorii rozhodování.

Pro praktické, numerické řešení složitějších úloh je v Kapitole 10 zavedena a ilustrována metoda Monte Carlo s Markovovými řetězci (MCMC) a uzavírá tak první cestu od pojmových a matematických základů k praktickým výpočtům.

Samostatné, fundamentální celky, které lze do značné míry pochopit bez vazby na vše další, jsou

- Matematické základy: Kapitoly 2 a 3,
- Bayesovské základy: Kapitoly 1,7 (bez Sekce 7.4) a 9.

Obsah ostatních pěti kapitol se týká metod uchopení a zvládání výpočtů v praktických situacích.

Veškeré zde uvedené grafy a výpočty byly provedeny pomocí programu Maple 15.

Pro lineárně postupujícího čtenáře je text mimo příkladů ilustrujících výklad proložen kontrolními otázkami, které mají dát příležitost k pozastavení se nad prezentovanou látkou. Ze stejného důvodu je každá kapitola ukončena Shrnutím. Úlohy potom slouží pro vyžití ve výpočtech.

Další cvičení k textu lze nalézt na síti na

`http://maplikace.math.slu.cz/AplikovanaStatistika/`.

Ta jsou většinou interaktivní. Některá vyžadují pouze novější prohlížeč, jiná vyžadují možnost použít program Maple 15.

Čtenáři, který nemíní respektovat řazení textu, je k dispozici seznam použitých symbolů a rejstřík na konci textu. Pro snadnější kontakt s odbornou literaturou jsou v rejstříku uvedeny za hesly i jejich anglické podoby.

Autor děkuji za upozornění na některé chyby H. Baranovi a posluchačům přednášek z Aplikované statistiky konaných v zimního semestru 2012.

Autor bude vděčný za poznámky k obsahu a upozornění na případné další chyby, které můžete zaslat na adresu

`Tomas.Kopf@math.slu.cz`

aby mohly být použity k vylepšení asymptotiky případných příštích verzi tohoto textu.

Ve Štítině, 31. srpna 2012

Tomáš Kopf

KAPITOLA 1

Náhoda a pravděpodobnost

Klíčová slova: Pravděpodobnost, sázka, Holandská kniha, podmíněná pravděpodobnost, Bayesův vzorec.

Abstrakt: Jistota, že nastane nějaký jev, se vyjadřuje jeho pravděpodobností. Tu lze kvantitativně zjistit podle toho, kolik nejvíce jsme na daný jev ochotni vsadit. Pokud přidělíme různým jevům pravděpodobnosti nevhodným způsobem, povede vhodný rozpis sázek sestavený protivníkem k tomu, že určitě prohrajeme. Z požadavku nepřidělovat pravděpodobnosti takto nevhodným, neracionálním způsobem plynou doporučení pro určení pravděpodobností. Doporučení pro určení pravděpodobností jsou základní a závazná pro tento předmět.

1.1. Úvod

Pravděpodobnost je základním pojmem pro kvantitativní úvahy v situaci, kdy si nemůžeme být jisti, zda něco nastane či nikoliv. Disciplína vystavěná na tomto pojmu se nazývá **statistikou**. Hned na začátku však narazíme na problém, že existují různé představy o tom, co pravděpodobnost vlastně je a tyto různé představy mají navíc dopad na celou další teorii. Je tedy třeba zpřesnit, co se pravděpodobností myslí.

V tomto textu se budeme zabývat tzv. **Bayesovskou statistikou** (viz [**18, 10**], zde diskutovaný pohled pochází od Jeffreye [**12, 21**]), která chápe pravděpodobnost jako subjektivní míru očekávání toho, že nastane nějaký jev.

Podobně je často míněno vyjádření pravděpodobnosti v hovorové řeči:

„Na 99 procent tam přijdu."
„Já tak padesát na padesát."

Očekávání se mohou člověk od člověka lišit a Bayesovská statistika nemá za cíl volně volitelná očekávání jakkoliv předepisovat. Ať si každý nastaví svá očekávání, že různé jevy nastanou, jak myslí. To však neznamená, že očekávání mají být nahodilá, neuvážená a znovu nám hovorové užití pravděpodobnosti poskytne návod, jak stanovení pravděpodobnosti propůjčit závažnost, totiž propojením s důsledky:

„Zítra to Opava na sto procent projede, to je jasné."
„Tak to nemáš pravdu! Vsadím své boty, že vyhrajou."

Sázka je jednou z možností, jak zaručit, že pravděpodobnosti jsou stanovovány zodpovědně.

Osnova této kapitoly je následující: V první sekci je pravděpodobnost kvantitativně zavedena pomocí sázky. V další sekci je ukázáno, že bez jakýchkoliv předsudků o přidělování pravděpodobností různým jevům se některé možnosti jeví jako nepřiměřené, nevhodné, mající za následek jistou prohru určitele pravděpodobností při konfrontaci s šikovně zvoleným systémem sázek, tzv. **Holandskou knihou**. Z požadavku, že k danému určení pravděpodobností nemá být možno sestavit Holandskou knihu, plynou **doporučení Bayesovské statistiky** pro určení pravděpodobností. Ta budeme v dalším považovat za závazná. V poslední sekci je potom zavedena **podmíněná pravděpodobnost** pomocí podmíněné sázky a i pro její určení lze nalézt doporučení.

Poznámka 1. Tzv. **frekventistická statistika** vychází z toho, že pravděpodobnost $P(A)$ je zlomek všech předepsaných situací, kdy uvažovaný jev A (v předepsané situaci) nastal:

$$P(A) = \frac{\text{počet předepsaných situací, kdy nastal jev } A}{\text{počet všech předepsaných situací}} \tag{1}$$

Pokud je (v určité idealizaci) předepsaných situací nekonečně mnoho, je třeba výše uvedený vztah pro pravděpodobnost $P(A)$ nahradit vhodnou limitou z výrazu na pravé straně.

Můžeme tento přístup chápat s božským odstupem tak, že všechny situace ve vesmíru, ať v minulosti, současnosti či budoucnosti, nějak dopadnou, takže nakonec je v každé situaci přesně určeno, jestli v ní jev A nastal, či nikoliv a i pravděpodobnost $P(A)$ je přesně a jednoznačně určena.

Zdrojem naší nejistoty je potom pouze to, že v konfrontaci s jednou z předepsaných situací ve své lidské omezenosti nevíme, jestli právě máme co do činění se situací, kdy jev A nastane, nebo se situací, kdy jev A nenastane. Jakou z nich jsme právě vybrali? V tomto kontextu je přirozené, že ve frekventistické statistice hrají důležitou roli **náhodné výběry** z předepsaných situací.

Poznámka 2. Vedle frekventistické a Bayesovské statistiky existují další přístupy k nakládání s nejistotou či náhodou, viz např. **teorii možnosti** [**24, 7**].

1.2. Pravděpodobnost

Sázka je jednou z možností, jak zaručit, že pravděpodobnosti jsou stanovovány zodpovědně. Budeme mít za to, že dotyčný přisuzuje jevu A pravděpodobnost $P(A)$, pokud je ochoten prodat i koupit následující směnku[1] (sázku) za €$P(A)$:

$$\boxed{\begin{array}{c}\text{Tato směnka má hodnotu}\\ \textbf{€1},\\ \textbf{pokud nastane } A.\end{array}} \tag{2}$$

[1]Pro jednoduchost předpokládáme, že hodnota peněz je zcela stálá a že nezáleží na tom, kdy k výplatě výher dojde.

Ochotu prodat i koupit bychom mohli nahradit tím, že dotyčný má stanovit hodnotu té směnky a protihráč se potom bude rozhodovat, jestli mu ji prodá, nebo ji od něho koupí.

Ať tak či tak, dotyčný se vystavuje možné ztrátě, pokud pravděpodobnost stanoví nesprávně. Pokud jí stanoví příliš vysoko, mohlo by se mu stát, že nakoupí takovou směnku příliš draho, zatímco jev A spíše nenastane. Pokud stanoví pravděpodobnost příliš nízko, prodá takovou směnku pod cenou, zatímco ona docela jistě poskytne €1. Toto dilema je osobní záležitostí onoho dotyčného a různí lidé dojdou k různým pravděpodobnostem.

KONTROLNÍ OTÁZKA 1.1. *Provoz družice po jejím vynesení na oběžnou dráhu je spolehlivý a slibuje v penězích korigovaných na inflaci 120 000 000,-€. Sdružení investorů je za její navržení, výrobu a umístění ochotno zaplatit maximálně 70 000 000,-€. Jaká je dle mínění sdružení nejvýše pravděpodobnost, že raketa dopravující družici na oběžnou dráhu v průběhu (nepojištěného) letu exploduje?*

Bayesovská statistika tedy připouští, aby jevům, které mohou nastat, každý přisuzoval různé pravděpodobnosti, jak myslí. Zároveň však propojením s důsledky formuluje normy, které ten či onen nemusí dodržovat, ale když je bude porušovat, se zlou se potáže.

KONTROLNÍ OTÁZKA 1.2. *Jev A nastane s pravděpodobností $P(A) = 90\%$. Spolehli jste se na něj, ale on nenastal a Vy jste se se zlou potázali. Chování dle pravděpodobností Vás přeci před ztrátou mělo chránit, a přesto jste ztrátu utržili. Není to v rozporu?*

1.3. Doporučení Bayesovské statistiky

DOPORUČENÍ 1. *Pravděpodobnost $P(A)$ libovolného jevu A má splňovat*

$$0 \leq P(A) \leq 1 \tag{3}$$

DŮKAZ. Doporučení není něco vynutitelného. Není automatické, že bude vždy dodrženo. Někdo může klidně určitému jevu přisoudit zápornou pravděpodobnost. Důkazem doporučení tedy rozumíme pouze důkaz toho, že jeho nedodržením se dotyčný vystavuje zaručené ztrátě.

Pokud by tedy někdo přiřadil jevu A zápornou pravděpodobnost $P(A) < 0$, znamenalo by to, že je ochoten někomu dát směnku uvedenou a ještě mu k tomu zaplatit €$\mid P(A) \mid$, čímž přijde zcela jistě o €$\mid P(A) \mid$ a navíc případně o €1, pokud by jev A nastal.

Pokud by někdo přiřadil jevu A pravděpodobnost $P(A) > 1$, koupil by si tedy směnku za €$P(A)$, přičemž směnka má hodnotu maximálně €1 (když totiž A nastane), takže jistě přijde alespoň o €$P(A)$-€1.

KONTROLNÍ OTÁZKA 1.3. *Lze „odstát“ nebezpečný jev tím, že se sdruží s jevem se stejně velkou, ale zápornou pravděpodobností?*

DOPORUČENÍ 2. *Jistý jev A má pravděpodobnost $P(A) = 1$ a nemožný jev B má pravděpodobnost $P(B) = 0$.*

DŮKAZ. Protože jev A zaručeně nastane, má směnka () hodnotu €1. Obdobně směnka pro nemožný jev nikdy nepřinese zisk a má tedy nulovou hodnotu.

DOPORUČENÍ 3. *Pokud A a B jsou vzájemně se vylučující jevy, potom má pro jejich pravděpodobnosti $P(A)$, $P(B)$ a pro pravděpodobnost $P(A \vee B)$, že nastane A nebo B, platit:*

$$P(A \vee B) = P(A) + P(B). \tag{4}$$

DŮKAZ. Směnky odpovídající pravděpodobnostem vystupujícím v tomto doporučení tedy jsou:

Tato směnka má hodnotu
€1,
pokud nastane A **nebo** B. (5)

Tato směnka má hodnotu
€1,
pokud nastane A.

Tato směnka má hodnotu
€1,
pokud nastane B. (6)

Pokud by dotyčný přiřadil příslušné pravděpodobnosti tak, že by platilo $P(A \vee B) < P(A)+P(B)$, koupil by dvě směnky (6) za €$P(A)$+€$P(B)$ a prodal směnku (5) za €$P(A \vee B)$. Tím celkově utratil víc, než dostal, takže na penězích tratil:

$$€P(A \vee B) - €P(A) - €P(B) < €0 \tag{7}$$

Zato ale má jiné směnky. Pokud nastane A, získá z první ze směnek (6) částku €1, ale zároveň prodejem směnky (5) o €1 přišel, takže na směnkách ani nezískal, ani netratil. Obdobné by to bylo, pokud by nastal jev B. Pokud nenastane ani A, ani B, jsou všechny uvažované směnky bezcenné. Na směnkách tedy dotyčný ve všech případech nic netratil ani nezískal. Celkově tedy tratil. Podobně lze vyšetřit případ $P(A \vee B) > P(A) + P(B)$.

KONTROLNÍ OTÁZKA 1.4. *Musí pravděpodobnosti*

- *P(chlapec), že Vaše první pravnouče bude chlapec,*
- *P(dívka), že Vaše první pravnouče bude dívka,*

splňovat $P(\text{chlapec}) + P(\text{dívka}) = 1$?

DOPORUČENÍ 4. *Pokud m, n jsou přirozená čísla, potom směnka*

Tato směnka má hodnotu
€$\frac{m}{n}$,
pokud nastane A. (8)

má mít hodnotu €$\frac{m}{n}P(A)$.

DŮKAZ. n těchto směnek musí mít stejnou hodnotu jako m směnek (1).

ÚLOHA 1. **Mistrovství světa v kuličkách.**

V roce 2015 se Mistrovství světa v kuličkách v Greyhound public house v Tinsley Green zúčastní týmy Anglie, Německa, USA, Číny, Nepálu a Francie. V místní sázkové kanceláři můžete koupit nebo prodat sázku na výhru v daném kurzu, libovolného objemu. Pouze jeden tým může soutěž vyhrát, není možná situace, kdy je výherců více.

Sázka na výhru	**Kurz**
USA	*10*
Anglie	*4*
Německo	*4*
Francie	*10*
Nepál	*5*
Čína	*10*
Asijský tým	*3*
Evropský tým	*2*
Favorité (Anglie nebo Německo)	*2*
Outsideři (USA, Francie nebo Nepál)	*3*

Tabulka 1.1. Kurzy možných sázek

Tak například sázka na výhru asijského týmu za 10 liber vyhrává 3x10 liber = 30 liber, pokud vyhraje Čína nebo Nepál a prohrává (tj. je bezcenná), pokud vyhraje jiný tým.

1. Jaké pravděpodobnosti přisuzuje sázková kancelář jednotlivým výhrám?

2. Sestavte Holandskou knihu (zaručeně vyhrávající rozpis sázek) pro sázky na tomto mistrovství.

3. Za předpokladu, že každá jednotlivá sázka libovolného objemu na možnost v tabulce je zpoplatněna částkou 1,- libry, kolik nejméně potřebujete peněz, abyste podle své Holandské knihy zaručeně získali?

Nápověda: *Pokud určené pravděpodobnosti přesně nesplňují doporučení Bayesovské statistiky, obsahují důkazy těchto doporučení návod jak sestavit Holandskou knihu.*

Úloha je dostupná v interaktivní podobě jako

Cvičení na síti 1. **Mistrovství světa v kuličkách.**.

1.4. Podmíněná pravděpodobnost

Podmíněnou pravděpodobnost $P(B \mid A)$ lze zjišťovat na základě podmíněné sázky. Budeme mít za to, že dotyčný přisuzuje jevu B, který může nastat poté, co bude uskutečněn jev A, podmíněnou pravděpodobnost $P(B \mid A)$, pokud je ochoten prodat i koupit následující směnku za €$P(B \mid A)$:

Tato směnka má hodnotu **€1**, **pokud nastane zároveň** A **i** B **se zárukou,** že její kupní cena bude navrácena, pokud A nenastane.

(9)

Doporučení 5. **(tzv. Bayesův vzorec).** *Nechť $P(A \wedge B)$ je pravděpodobnost, že nastane zároveň jev A i jev B. Potom se doporučuje, aby platilo:*

$$P(A \wedge B) = P(B \mid A)P(A) \tag{10}$$

Důkaz. Ve směnce (9) můžeme přímo dosadit hodnotu, kterou má, pokud A nenastane, bez toho, aby se její hodnota pro dotyčného změnila. Dostaneme směnku,

Tato směnka má hodnotu **€1**, **pokud nastane zároveň A i B** a hodnotu **€**$P(B \mid A)$, **pokud nenastane A.**

(11)

která ovšem zase má mít stejnou cenu jako směnky

Tato směnka má hodnotu **€1**, **pokud nastane zároveň A i B**	Tato směnka má hodnotu **€**$P(B \mid A)$, **pokud nenastane A.**

(12)

Mělo by tedy platit (s použitím Doporučení 4, 2 a 3):

$$€P(B \mid A) = €P(A \wedge B) + €(P(B \mid A)P(\neg A)), \tag{13}$$

kde $P(\neg A)$ je pravděpodobnost, že nenastane jev A, pro kterou by ovšem mělo platit, že $P(\neg A) = 1 - P(A)$. Dostaneme tedy z (13):

$$€P(B \mid A) = €P(A \wedge B) + €(P(B \mid A)(1 - P(A))) \tag{14}$$

$$P(B \mid A) = P(A \wedge B) + (P(B \mid A)(1 - P(A))) \tag{15}$$

$$0 = P(A \wedge B) - P(B \mid A)P(A) \tag{16}$$

$$P(A \wedge B) = P(B \mid A)P(A) \tag{17}$$

Definice 1. *Jevy A, B jsou nezávislé, pokud platí:*

$$P(A \mid B) = P(A) \tag{18}$$

Poznámka 3. Dle (212) máme prohozením jevů A a B též

$$P(A \mid B)P(B) = P(A \wedge B) = P(B \mid A)P(A), \tag{19}$$

protože $A \wedge B = B \wedge A$. Pokud platí (18), platí potom též

$$P(B \mid A) = P(B) \tag{20}$$

Nezávislost je tedy symetrický vztah (symetrická relace) mezi jevy A a B.

OBRÁZEK 1.1. **Souhra náhod.** Souběžné jevy nemusí být statisticky nezávislé.

KONTROLNÍ OTÁZKA 1.5. *Albert půjde do zahradní hospody „U plastového kelímku“ s pravděpodobností $P(A) = 20\%$. Šance, že potká Bohouše, pokud tam Albert půjde, je $P(B \mid A) = 40\%$. Šance, že Bohouš potká Alberta, pokud Bohouš do hospody půjde, je $P(A \mid B) = 10\%$. Jaká je pravděpodobnost $P(B)$, že Bohouš zaujme své místo „U plastového kelímku“?*

SHRNUTÍ 1. *Pravděpodobnost $P(A)$ jevu A vyjadřuje subjektivní jistotu, že jev A nastane. Lze jí kvantitativně stanovit pomocí sázky, kterou dotyčný bude považovat za vyváženou.*

Pokud se dotyčný rozhoduje racionálně, musí platit následující pravidla:

$$0 \leq P(A) \leq 1, \tag{21}$$

$$P(A) = 1 \qquad \text{pro jistý jev } A, \tag{22}$$

$$P(A) = 0 \qquad \text{pro nemožný jev } A, \tag{23}$$

$$P(A \vee B) = P(A) + P(B) \qquad \text{pro neslučitelné jevy } A,\ B, \tag{24}$$

$$P(A \wedge B) = P(B \mid A)P(A) \tag{25}$$

Splnění těchto pravidel budeme od nynějška vždy požadovat a předpokládat.

KAPITOLA 2

Pravděpodobnost a teorie míry

Klíčová slova: Teorie míry, Kolmogorovovy axiomy pravděpodobnosti, rozdělení pravděpodobnosti.

Abstrakt: Exaktní matematickou formalizací pravděpodobnosti je teorie míry. Její základní pojmy jsou zde shrnuty ve vztahu k pravděpodobnosti.

2.1. Úvod

Na základě rozboru z předchozí kapitoly můžeme nyní zodpovědně přidělovat pravděpodobnosti různým jevům. Například bychom mohli zvážit, s jakou pravděpodobností bude na příští Zlaté tretře v Ostravě překonán světový rekord na 100 m. Jaká je však pravděpodobnost, že bude světový rekord přesně vyrovnán? Že některý běžec bude tak rychlý, že ohrozí světový rekord, má malou pravděpodobnost, ale že by se navíc ještě přesně trefil na určený čas?

Ve skutečnosti vyrovnání světového rekordu až tak nemožné není: Podle pravidel Mezinárodní asociace atletických federací se čas neměří přesněji než na setiny sekundy a nejlepší časy se liší právě jen o setiny.

V technických možnostech by však bylo měřit závod o mnoho řádů přesněji, například laserovým sledováním umělé pihy na nose každého ze závodníků a tím by se přesné vyrovnání světového rekordu značně komplikovalo. V idealizaci zcela přesného měření bychom přesnému vyrovnání světového rekordu mohli přiřadit nulovou pravděpodobnost. Nulovou pravděpodobnost, tedy stejnou jako pravděpodobnost nemožného jevu, budeme pak ale patrně chtít přisoudit libovolnému zvolenému výsledku. S jistotou tedy nastane jev, který má nulovou pravděpodobnost? Znamená to potom, že pravděpodobnost, že cokoliv nastane, je nulová? To je znepokojující.

Znepokojení je oprávněné, poukazuje však na jednoduchou věc: Zacházení s teoriemi o pravděpodobnosti vyžaduje přesnou matematickou formalizaci. Bez přesné matematické formulace se rychle dostaneme do úzkých, budeme bádat nad domnělými paradoxy, a to zbytečně, protože přesná matematická formulace je k dispozici: **teorie míry**.

Teorie míry bere v úvahu, že uvažované možnosti lze vyjádřit pomocí prostoru (množiny) Ω, jehož jednotlivé body (prvky) $x \in \Omega$ představují **elementární jevy**. V případě možného času závodníka si můžeme pod prostorem Ω představit reálnou, časovou osu $\mathbb{R}$. (Nebudeme se zdržovat rozjímáním nad tím, že záporného času by závodník mohl dosáhnout během v tratě v protisměru a byl tak dříve v cíli než na startu. Daný pohled je pochopitelně určitou idealizací.) Body $x \in \mathbb{R}$ reálné osy potom představují možné, naprosto přesně změřené časy. Teorie míry hned od počátku bere v úvahu, že přiřazovat pravděpodobnosti jednotlivým bodům (elementárním jevům) nemusí být nejšťastnější, jak bylo vysvětleno výše a přiřazuje namísto toho pravděpodobnosti $P(A)$ podmnožinám, tzv. **jevům** $A \subset \Omega$. Příkladem může být pravděpodobnost překonání světového rekordu, kdy uvažovanou podmnožinou A reálné osy $\mathbb{R}$ jsou všechny časy menší, než současný světový rekord t_S:

$$A = \{t \in \mathbb{R} \text{ taková, že } 0 < t < t_S\} \tag{26}$$

To nevylučuje možnost přiřadit i jednotlivým bodům x pravděpodobnosti $P(x)$ oklikou, totiž přiřazením pravděpodobnosti podmnožině $\{x\}$, která obsahuje právě onen bod $x \in \Omega$:

$$P(x) \equiv p(\{x\}) \tag{27}$$

Shrnuto: Teorie míry vychází z množiny elementárních jevů Ω, jejíž podmnožinám $A \subset \Omega$ bude přiřazovat jejích pravděpodobnosti (nazývané též **míry**) $P(A)$.

V následující sekci zavedeme základní pojmy teorie míry a vysvětlíme jejích vztah k pravděpodobnosti. V další sekci se potom budeme zabývat rozděleními pravděpodobnosti, které se používají k popisu náhodných jevů.

2.2. Teorie míry

Fundamentální matematické myšlenky se projevují tím, že se objevují v různých kontextech znovu a znovu, až do omrzení. Teorie míry je jednou z nich. Není proto divné, že původně nevznikla jako teorie pravděpodobnosti, ale za jiným účelem, totiž jako teorie umožňující definovat, **měřit** velikosti podmnožin nějakého prostoru, například obsahy podmnožin v rovině. Že se tatáž věc hodí i na matematickou formalizaci teorie pravděpodobnosti je zjištěním Andreje Nikolajeviče Kolmogorova [**14**].

2.2.1. Měřitelné prostory. Jednou technickou překážkou, která činí teorii míry netriviální, je snaha měřit libovolné podmnožiny A prostoru Ω, což ovšem často není slučitelné s dalšími požadavky, které bychom na různé modely této teorie chtěli klást. Řešením tohoto problému je trochu slevit ze zamýšleného cíle měřit všechny podmnožiny prostoru Ω, ale omezit se jen na některé, tzv. **měřitelné**. Množina $\mathcal{F}$ všech měřitelných množin se však neurčuje zcela nahodile, ale tak, aby splňovala následující definici:

DEFINICE 2. **σ-algebra** $(\Omega, \mathcal{F})$ *je neprázdný systém podmnožin $\mathcal{F}$ množiny Ω obsahující prázdnou množinu a uzavřený vůči doplňkům a spočetným sjednocením.* $(\Omega, \mathcal{F})$ *se potom nazývá* **měřitelným prostorem**. *Podmnožina $A \subset \Omega$ se označuje jako* **měřitelná**, *pokud $A \in \mathcal{F}$.*

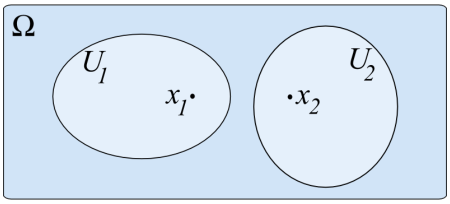
2.2.2. Vztah k topologickým prostorům. V běžných případech je prostor Ω, na kterém pracujeme, topologickým prostorem - tedy je opatřen pojmem toho, co jsou otevřené množiny. Podobně, jako σ-algebra, i topologický prostor je zadán systémem podmnožin množiny Ω:

DEFINICE 3. **Topologický prostor** $(\Omega, \mathcal{T})$ *je neprázdný systém podmnožin $\mathcal{T}$ množiny Ω obsahující prázdnou množinu a též samotnou množinu Ω a uzavřený vůči konečným průnikům a libovolným sjednocením. Říkáme, že prostor Ω je opatřen topologií $\mathcal{T}$. Podmnožina*

$U \subset \Omega$ se označuje jako **otevřená**, *pokud $U \in \mathcal{T}$. Podmnožina $U \subset \Omega$ obsahující bod x se označuje jako* **otevřené okolí** *bodu x.*

Obvykle budeme chtít, aby tyto dvě struktury na Ω, tedy struktura σ-algebry $(\Omega, \mathcal{F})$ a struktura topologického prostoru $(\Omega, \mathcal{T})$, byly v souladu. Obecné topologie však mohou mít řadu komplikujících vlastností, a proto se omezíme pouze na takové topologické prostory, které jsou navíc lokálně kompaktní a Hausdorffovy, jak je definováno v následujícím:

DEFINICE 4. *Topologický prostor $(\Omega, \mathcal{T})$ je* **Haussdorffův**, *pokud libovolné dva jeho body $x_1, x_2 \in \Omega$ lze oddělit nepřekrývajícími se otevřenými množinami. Tj. existují otevřené množiny $U_1, U_2 \in \mathcal{T}$ tak, že $U_1 \cap U_2 = \emptyset$ a že $x_1 \in U_1$ a $x_2 \in U_2$. Jinými slovy: Existují neprotínající se (disjunktní) otevřená okolí bodů x_1, x_2.*



OBRÁZEK 2.1. **Hausdorffův topologický prostor.** Libovolné dva body $x_1 \neq x_2$ lze od sebe oddělit disjunktními otevřenými okolími.

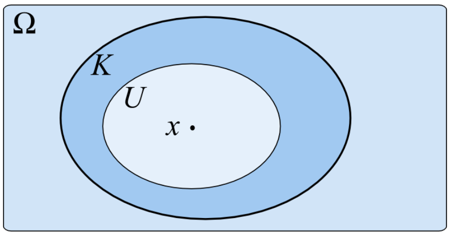
Množina $K \subset \Omega$ je **kompaktní**, *pokud z jejího libovolného* **pokrytí** *$\bigcup_{i \in I} U_i$ otevřenými množinami U_i (množina I, ze které je index i může být nekonečná, třeba $\mathbb{N}$):*

$$K \subset \bigcup_{i \in I} U_i \tag{28}$$

lze vybrat konečné pokrytí

$$K \subset \bigcup_{i_1, i_2, \ldots, i_n} U_i \tag{29}$$

Topologický prostor $(\Omega, \mathcal{T})$ je **lokálně kompaktní**, *pokud každý jeho bod $x \in \Omega$ je spolu s některým svým otevřeným okolím U obsažen v některé kompaktní podmnožině K prostoru Ω, tj. $x \in U \subset K \subset \Omega$.*



OBRÁZEK 2.2. **Lokálně kompaktní topologický prostor.** Libovolný bod $x \in \Omega$ má otevřené okolí U obsažené v některé kompaktní množině K: $x \in U \subset K \subset \Omega$.

Poznámka 4. Výše uvedené dodatečné podmínky na topologický prostor (lokální kompaktnost, Hausdorffovskost) jsou dosti abstraktní a technické. Na první pohled mohou vypadat jako nahodilé a nepochopitelné. Je to tím, že v sobě skrývají desetiletí vývoje, kdy se matematici snažili dokázat věty, které by spolu skloubily různé matematické teorie. Tohoto skloubení však nelze dosáhnout s obecnou Definicí 3, což se ukázalo při snaze matematiků dokázat příslušné věty. Byly potřebné doplňující požadavky. A právě ty jsou obsaženy v Definici 4.

Běžné topologické prostory tyto podmínky splňují, např. $\mathbb{R}^n$ s topologií generovanou kartézskými součiny intervalů.

Ostatní topologické prostory, které nejsou Hausdorffovy lokálně kompaktní, zde budeme mít za výtržníky a dále se jim nebudeme věnovat.

Pro podrobnější seznámení viz specializovanou přednášku o míře a integrálu nebo literaturu, např. [**22**], [**5**].

Definice 5. *Nechť je dán prostor Ω opatřený topologií $\mathcal{T}$ tak, že je lokálně kompaktním Hausdorffovým prostorem. Potom jeho* **Borelova algebra** $(\Omega, \mathcal{B})$ *je nejmenší σ-algebra obsahující $\mathcal{T}$, tj. nejmenší σ-algebra na Ω, pro kterou platí $\mathcal{T} \subset \mathcal{B}$.*

Poznámka 5. Borelova algebra $(\Omega, \mathcal{B})$, nebo její vhodné rozšíření $(\Omega, \mathcal{L})$ se σ-algebrou tzv. Lebesgueovsky měřitelných množin, $\mathcal{B} \subset \mathcal{L}$ popsanou v další kapitole, neobsahuje všechny podmnožiny dané množiny Ω. Jako zajímavý protipříklad viz Vitaliho množiny, Příklad 4. I tento protipříklad ukazuje, že sestrojit množinu, která není v tomto kontextu měřitelná, není snadné. Takže zatímco z důvodu logické správnosti musíme otázku měřitelnosti podmnožin prostoru Ω s sebou vláčet celou teorií, je užitečné si podržet intuici, že „všechny rozumné“podmnožiny $A \in \Omega$ jsou měřitelné a že v praktických příkladech nebude zpravidla třeba měřitelnost podmnožin prověřovat, protože ji obecné výsledky zaručují.

2.2.3. Měřitelná zobrazení.

Definice 6. *Funkce $f : \Omega_1 \to \Omega_2$ z měřitelného prostoru $(\Omega_1, \mathcal{F}_1)$ do měřitelného prostoru $(\Omega_2, \mathcal{F}_2)$ je* **měřitelná**, *pokud vzorem každé měřitelné množiny je měřitelná množina.*

Pokud máme co do činění s reálnou funkcí f a předpokládáme na oboru hodnot $\mathbb{R}$ strukturu měřitelného prostoru generovanou otevřenými intervaly, můžeme se spokojit s konkrétnější definicí:

Definice 7. **(Speciální případ).** *Funkce $f : \Omega \to \mathbb{R}$ je* **měřitelná**, *pokud pro každé $r \in \mathbb{R}$ platí:*

$$\{\omega \in \Omega \mid f(\omega) < r\} \in \mathcal{F},$$

tj., pokud vzorem libovolného otevřeného intervalu v $\mathbb{R}$ je měřitelná množina.

Jedno snadné kritérium pro měřitelnost zobrazení je následující:

Tvrzení 1. *Pokud $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$ jsou Borelovy algebry, potom každé spojité zobrazení $f : \Omega_1 \to \Omega_2$ je měřitelné.*

Kontrolní otázka 2.1. *Proč platí Tvrzení 1?*

Měřitelná funkce $X : \Omega \to \mathbb{R}^n$ z měřitelného prostoru $(\Omega, \mathcal{F})$ do měřitelného prostoru $(\mathbb{R}^n, \mathcal{B})$ s Borelovskou σ-algebrou $\mathcal{B}$ generovanou otevřenými množinami (nebo, v obecnějším smyslu, do jakéhokoliv měřitelného prostoru) se nazývá **náhodnou proměnnou**, či **náhodnou veličinou**.

Náhodné proměnné $X : \Omega \to \mathbb{R}^n$ lze chápat jako souřadnicové funkce na prostoru Ω. Tyto souřadnicové funkce však nemusí nutně představovat souřadný systém umožňující popisu Ω do všech detailů, ale vystihující jen některé uvažované rysy přiřazované elementárním jevům v Ω.

OBRÁZEK 2.3. **Body mass index (BMI)** je složením měřitelného zobrazení $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, které index spočte na základě hmotnosti m a výšky těla h, a měřitelného zobrazení $f : \Omega \to \mathbb{R} \times \mathbb{R}$, které každému člověku $x \in \Omega$ přiřadí jeho hmotnost a výšku.

PŘÍKLAD 1. *Nechť Ω obsahuje lidi jako elementární jevy a nechť náhodná proměnná $BMI : \Omega \to \mathbb{R}$ přiřadí každému člověku $x \in \Omega$ jeho tzv.* **Body Mass Index**[ι]*:*

$$BMI(x) = \frac{\textit{tělesná hmotnost v kilogramech}(x)}{(\textit{tělesná výška v metrech}(x))^2} \tag{30}$$

Samotný prostor $\mathbb{R}$ hodnot náhodné veličiny BMI je měřitelným prostorem a jeho měřitelné podmnožiny tak představují jevy. Jev $A = (-\infty, 16.5)$ se označuje jako těžká podvýživa, jev $B = [16.5, 18.5)$ jako podváha, jev $C = [18.5, 25]$ jako ideální váha, jev $D = (25, 30]$ jako nadváha a jev $E = (30, \infty)$ jako obezita. Tyto jevy jsou zvoleny jako vzájemně se vylučující a vyčerpávají $\mathbb{R}$:

$$A \cup B \cup C \cup D \cup E = \mathbb{R}.$$

Při jiné volbě jevů A, B, C, D, E by tomu ale tak nemuselo být.

Náhodná proměnná BMI představuje pro některé účely užitečnou souřadnici na prostoru Ω, v žádném případě však nevystihuje v detailu identitu jeho bodů $x \in \Omega$ a není tedy souřadným systémem.

2.2.4. Míra.

DEFINICE 8. **Mírou** *μ na měřitelném prostoru $(\Omega, \mathcal{F})$ se nazývá zobrazení*

$$\mu : \mathcal{F} \to [0, +\infty) \cup \{+\infty\},$$

pokud platí

(1) $\mu(\emptyset) = 0$,
(2) $\mu(\cup_i A_i) = \sum_i \mu(A_i)$ *pro spočetně mnoho libovolných po dvou disjunktních množin $A_i \in \mathcal{F}$.*[1]

DEFINICE 9. *Libovolná míra μ zadaná na Borelově algebře $(\Omega, \mathcal{B})$ se nazývá Borelovou mírou.*

2.2.5. Invariantní míry. Na některých prostorech Ω jsou dány symetrie. Symetrie se zadávají grupou symmetrií G a působením každého prvku $g \in G$ na prostor Ω zobrazením $\alpha_g : \Omega \to \Omega$, přičemž má pro všechna $g_1, g_2 \in G$ a jejích součin $g_1 \cdot g_2$ v grupě G platit

$$\alpha_{g_1}(\alpha_{g_2}(x)) = \alpha_{g_1 \cdot g_2}(x) \qquad \text{pro všechna } x \in \Omega. \tag{31}$$

PŘÍKLAD 2. *Posunutí v $\Omega = \mathbb{R}^n$ jsou symetrie. Grupa symetrií je v tomto případě $G = \mathbb{R}^n$ s grupovou operací danou součtem příslušných souřadnic. Působení posunutí $g = [t_1, t_2, \dots t_n] \in \mathbb{R}^n$ na bod $x = [x_1, x_2, \dots x_n] \in \mathbb{R}^n$ je dáno součty příslušných souřadnic jako:*

$$\alpha_g(x) = \alpha_{[t_1, t_2, \dots t_n]}([x_1, x_2, \dots x_n]) = [t_1 + x_1, t_2 + x_2, \dots t_n + x_n] \tag{32}$$

Jednoduchost důležitého předchozího příkladu může být matoucí tím, že jak prostor Ω, tak grupa symetrií G byly zadány kopiemi stejného prostoru $\mathbb{R}^n$. Proto je užitečný i následující příklad:

PŘÍKLAD 3. *Otočení okolo počátku v $\Omega = \mathbb{R}^n$ jsou symetrií. Grupa symetrií je v tom případě grupa speciálních ortonormálních transformací $G = SO(n)$.*

Na symetrických prostorech je přirozené požadovat, aby jak teorie míry, tak topologie tyto symetrie respektovaly. To znamená, že působení symetrie dané zobrazením $\alpha_g : \Omega \to \Omega$ je měřitelné a spojité a že uvažovaná Borelova míra μ se při zobrazení zachovává. Říkáme, že daná míra je **invariantní**. To znamená, že pokud obrazem měřitelné množiny $A \subset \Omega$ je měřitelná množina $\alpha_g(A)$, potom platí:

$$\mu(A) = \mu(\alpha_g(A)). \tag{33}$$

DEFINICE 10. *Borelova míra (v užším smyslu slova) na $\Omega = \mathbb{R}^n$ je Borelovou mírou v obecnějším smyslu (viz Definici 9) spolu s následujícími předpoklady:*

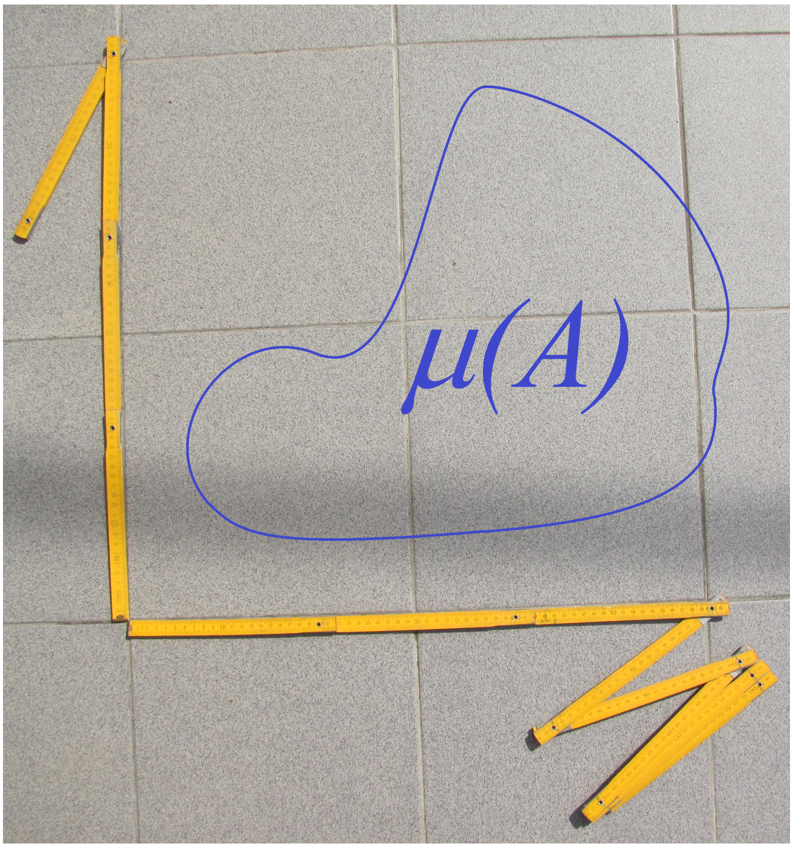
(1) *Topologie na $\Omega = \mathbb{R}^n$ je obvyklou topologií používanou na tomto prostoru. Je generovaná otevřenými hyper-krychlemi, tj. součiny otevřených intervalů*

$$(a_1, b_1) \times (a_2, b_2) \times \dots \times (a_n, b_n) \subset \underbrace{\mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}}_{n\text{-krát}}. \tag{34}$$

[1]Předpokládáme zde, že sčítání je rozšířeno přirozeným způsobem na hodnotu $+\infty$ tak, že $+\infty + r = +\infty$ pro libovolné $r \in [0, +\infty) \cup \{+\infty\}$

(2) *Míra hyper-krychle* $(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n)$ *je*

$$\mu\left((a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n)\right) = (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n) \tag{35}$$



OBRÁZEK 2.4. **Obsah.** Běžný obsah lze chápat jako Borelovu míru na $\mathbb{R}^2$.

TVRZENÍ 2. *Míra v předchozí Definici 10 je invariantní mírou vůči symetrii danou grupou posunutí v* $\mathbb{R}^n$ *a je určena jednoznačně.*

Tato míra je dokonce invariantní vzhledem k vyšší symetrii, totiž vůči grupě Euklidovských transformací generovaných posunutími a otočeními v $\mathbb{R}^n$.

DŮKAZ. Tvrzení lze dokázat na základě toho, že míra hyper-krychle je invariantní vůči Euklidovským transformacím. Je přitom třeba zohlednit, že pokud Euklidovská transformace není čistým posunutím, není možno přímočaře použít vzorec (35) pro výpočet míry transformované hyper-krychle.

CVIČENÍ NA SÍTI 2. **Borelova algebra a Borelova míra**.

POZNÁMKA 6. V definici Borelovy míry v užším smyslu jsme se omezili na prostory $\Omega = \mathbb{R}^n$ a grupu posunutí $G = \mathbb{R}^n$. Obdobnou míru lze sestrojit pro jiné prostory s Borelovou algebrou na lokálně kompaktních prostorech. Nazývá se **(levou) Haarovou mírou**.

POZNÁMKA 7. Mlčky zde předpokládáme, že míry, které definujeme, také existují. Pro důkaz těchto skutečností viz příslušnou literaturu, např. [**22**], [**5**], [**4**].

PŘÍKLAD 4. *Tento příklad ukazuje, že žádná tzv. Vitaliho množina není měřitelná vzhledem k Borelově algebře na $\mathbb{R}$ s Borelovou mírou nebo vůči některému jejímu translačně invariantnímu rozšíření. Příklad je trochu složitější a lze jej přeskočit, pokud uvěříte, že neměřitelné množiny v tomto případě existují. Zvídavý čtenář by se však neuvedením tohoto příkladu cítil nutně odbytý.*

V prvním kroku sestrojíme Vitaliho množinu. K tomu zavedeme na $\mathbb{R}$ relaci ekvivalence $\sim$ tak, že reálná čísla v_1 , v_2 jsou ekvivalentní, pokud se liší o racionální číslo:

$$v_1 \sim v_2 \Leftrightarrow v_2 - v_1 \in \mathbb{Q} \tag{36}$$

*Reálná čísla se potom rozpadají do nekonečně mnoho tříd ekvivalence $[v]$, $v \in \mathbb{R}$, z nichž každá je hustá v $\mathbb{R}$ (tj, k libovolnému bodu $r \in \mathbb{R}$ existuje libovolně blízko bod $w \in [v]$ z dané třídy $[v]$). Každá z tříd ekvivalence má tedy bod (i nekonečně mnoho bodů) v intervalu $[0,1]$.***Vitaliho množinu** *V (jednu z mnoha možných) obdržíme tak, že do ní z každé z různých tříd ekvivalence vybereme jeden její reprezentant ležící v intervalu $[0,1]$. Že je takový výběr možný, je zaručeno* **axiomem výběru**, *o který se musíme opřít.*

Nyní ukážeme sporem, že Vitaliho V množina nemůže být měřitelná. Sestrojíme nejdříve její posunutí V_q o racionální číslo q z intervalu $[-1,1]$:

$$V_q = \{v + q \mid v \in V\}, \qquad\qquad q \in \mathbb{Q} \cap [-1,1]. \tag{37}$$

Platí:

$$[0,1] \subset \bigcup_{q \in \mathbb{Q} \cap [-1,1]} V_q \subset [-1,2] \tag{38}$$

Protože Borelova algebra je translačně invariantní, budou množiny V_q měřitelné, pokud V je měřitelná. Měřitelné množiny jsou dle definice uzavřené vůči spočetným sjednocením, a protože racionálních čísel (i jejích průnik s $[-1,1]$) je spočetně mnoho, je sjednocení v (38) spočetné a měřitelné. Má tedy v důsledku invariance míry, $\mu(V_q) = \mu(V)$ míru

$$\mu\left(\bigcup_{q \in \mathbb{Q} \cap [-1,1]} V_q\right) = \sum_{q \in \mathbb{Q} \cap [-1,1]} \mu(V_q) = \sum_{q \in \mathbb{Q} \cap [-1,1]} \mu(V) \tag{39}$$

a platí dle (38):

$$1 = \mu([0,1]) \leq \sum_{q \in \mathbb{Q} \cap [-1,1]} \mu(V) \leq \mu([-1,2]) = 3 \tag{40}$$

Jenže to je ve sporu s tím, že prostřední výraz v (40) může nabývat pouze hodnoty 0 a $+\infty$:

$$\sum_{q \in \mathbb{Q} \cap [-1,1]} \mu(V) = \begin{cases} 0 & \text{pro } \mu(V) = 0, \\ \infty & \text{pro } \mu(V) > 0. \end{cases} \tag{41}$$

Předpoklad, že V je měřitelná, tedy nemůže být pravdivý.

2.2.6. Rozdělení pravděpodobnosti. V teorii pravděpodobnosti budeme míru značit P a nazývat **rozdělením pravděpodobnosti** na Ω. Jedinou odchylkou Kolmogorovovy teorie pravděpodobnosti od obecné teorie míry je předpoklad, že celková míra prostoru Ω všech elementárních jevů je jednotková. To předpokládáme zpravidla i zde:

$$P(\Omega) = 1. \tag{42}$$

Důvodem tohoto omezení je předpoklad, že prostor Ω obsahuje veškeré možné elementární jevy. Jev určený tím, že nastane některý elementární jev z Ω, je tedy jistým jevem a musí proto podle poznatků z předchozí kapitoly mít jednotkovou pravděpodobnost. Při této příležitosti je též dobré si všimnout, že vztahy v definici míry souhlasí s příslušnými dalšími doporučeními Bayesovské statistiky zdůvodněnými v předchozí kapitole a zapadají tak do zdůvodněných požadavků teorie pravděpodobnosti.

$(\Omega, \mathcal{F}, P)$ se potom nazývá **pravděpodobnostním prostorem**. V tomto kontextu se definiční vztahy míry spolu s (42) nazývají Kolmogorovovými axiomy.

SHRNUTÍ 2. *Rozdělení pravděpodobnosti P přiřazuje měřitelným podmnožinám A prostoru Ω elementárních jevů jejich pravděpodobnost $P(A)$. P je mírou na Ω s dodatečnou podmínkou, že*

$$P(\Omega) = 1. \tag{43}$$

Míra P splňuje dle definice náledující:

Pro prázdnou množinu $\emptyset$ platí

$$P(\emptyset) = 0. \tag{44}$$

Pro sjednocení spočetně mnoha měřitelných podmnožin $A_1, A_2, \ldots \in \Omega$, kde se žádné dvě různé podmnožiny nepřekrývají (tj. $A_i \cap A_j = \emptyset$ pro $i \neq j$), platí

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i). \tag{45}$$

Vztahy (43), (44), (45) se nazývají Kolmogorovovými axiomy pro rozdělení pravděpodobnosti P.

KAPITOLA 3

Rozdělení pravděpodobnosti

Klíčová slova: Lebesgueův integrál, hustota pravděpodobnosti, střední hodnota, rozptyl

Abstrakt: Míra či rozdělení pravděpodobnosti umožňují zavést integrál. Rozdělení pravděpodobnosti lze s pomocí integrálu přes Borelovu míru zadat hustotou pravděpodobnosti a částečně charakterizovat střední hodnotou a rozptylem.

3.1. Úvod

Rozdělení pravděpodobnosti P na prostoru elementárních jevů Ω nám umožňuje jevu $A \subset \Omega$ přiřadit pravděpodobnost $P(A)$. To však bezprostředně vede na možnost počítat vážené průměry z vhodných funkcí $f : \Omega \to \mathbb{R}$. Nejjednodušší v tomto případu diskutované funkce jsou zavedeny v následující definici:

DEFINICE 11. *Funkce $f : \Omega \to \mathbb{R}$ je* **jednoduchá**, *pokud nabývá pouze konečně mnoha různých hodnot $\{y_i\}_{i=1}^n$ tak, že vzorem $A_i = f^{-1}(y_i)$ každé z hodnot y_i je měřitelná množina.*

Střední hodnotou $E(f)$ funkce f nazýváme potom průměr z hodnot y_i vážených jejích pravděpodobnostmi, tj. pravděpodobnostmi $P(A_i)$ jejích vzorů a označujeme jej alternativně jako integrál funkce f:

$$E(f) \equiv \int_\Omega f\,dP \equiv \int_\Omega f(x)dP(x) = \sum_{i=1}^n f_i P(A_i) \tag{46}$$

Stejným výrazem definujeme tzv. **Lebesgueův integrál** jednoduché funkce $f : \Omega \to \mathbb{R}$ pro obecnou míru μ na Ω:

$$\int_\Omega f\,d\mu \equiv \int_\Omega f(x)d\mu(x) = \sum_{i=1}^n y_i \mu(A_i) \tag{47}$$

TVRZENÍ 3. **(Linearita Lebesgueova integrálu).** *Pokud $f, g : \Omega \to \mathbb{R}$ jsou jednoduché funkce potom je jednoduchá i jejich lineární kombinace $af + bg$, s libovolnými koeficienty $a, b \in \mathbb{R}$ a platí*

$$\int_\Omega af + bg\,d\mu = a\int_\Omega f\,d\mu + b\int_\Omega g\,d\mu \tag{48}$$

První sekce této kapitoly rozšíří Lebesgueův integrál na mnohem širší třídu funkcí.

Další sekce se věnuje důležitému speciálnímu případu: Zatímco pro daný měřitelný prostor zpravidla existuje mnoho různých měr, je v R^n na vhodné σ-algebře jedna význačná míra, tzv. Lebesgueova míra, která navíc na funkcích majících jak Lebesgueův, tak Riemannův integrál dává pro oba integrály stejný výsledek. Kdo umí počítat Riemannovy integrály, ten se tedy nemusí učit nový kalkulus pro integrály Lebesgueovy.

Ve třetí sekci je ukázáno, že mnohá rozdělení pravděpodobnosti P na $\mathbb{R}^n$ lze zadat ve vztahu k význačné Lebesgueově míře pomocí nezáporné funkce $p(x)$ na Ω, tzv. **hustoty pravděpodobnosti**.

Poslední sekce se zabývá některými často užívanými rozděleními pravděpodobnosti na $\mathbb{R}$ a jejích charakterizací.

Pro podrobnější seznámení se s tématem viz např. [**22**], [**5**].

POZNÁMKA 8. Speciálním případem jednoduché funkce je **charakteristická funkce** $\chi_A : \Omega \to \mathbb{R}$ pro měřitelnou množinu A:

$$\chi_A(x) = \begin{cases} 1 & \text{pro } x \in A, \\ 0 & \text{jinak.} \end{cases} \tag{49}$$

S pomocí charakteristické funkce lze míru $\mu(A)$ množiny A vyjádřit jako integrál z její charakteristické funkce:

$$\mu(A) = \int_\Omega \chi_A d\mu, \tag{50}$$

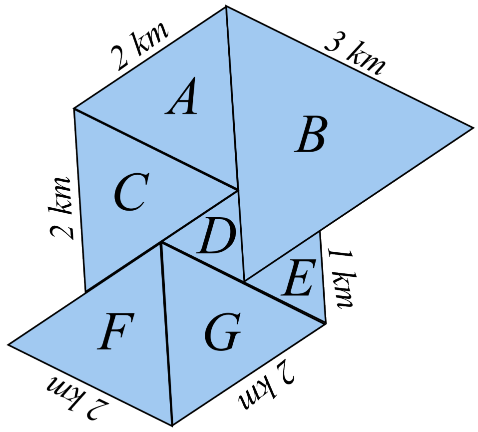
a tedy i pro rozdělení pravděpodobnosti

$$P(A) = \int_\Omega \chi_A dP. \tag{51}$$

Každou jednoduchou funkci f s hodnotami $\{y_i\}_{i=1}^n$ a vzory hodnot $A_i = f^{-1}(y_i)$ lze napsat jako lineární kombinaci charakteristických funkcí:

$$f(x) = \sum_{i=1}^n y_i \chi_{A_i}(x). \tag{52}$$

KONTROLNÍ OTÁZKA 3.1. *Rovnoměrně zabydlená oblast je obsluhována vysílači s oblastmi působnosti ve tvaru rovnostranných trojúhelníků dle Obrázku 3.1. Cena pořízení a provozu vysílačů přepočtena na jedno spojení je uvedená v Tabulce 3.1. Jaké jsou průměrné náklady na jedno spojení?*



OBRÁZEK 3.1. **Poloha vysílačů.** Vysílače obsluhují oblasti ve tvoru rovnostranných trojúhelníků.

Vysílač	*A*	*B*	*C*	*D*	*E*	*F*	*G*
1 spojení [Kč]	*0.011*	*0.008*	*0.013*	*0.021*	*0.024*	*0.010*	*0.012*

TABULKA 3.1. Náklady na pořízení a provoz vysílačů přepočteny na jedno spojení

3.2. Lebesgueův integrál pro obecnou míru

DEFINICE 12. *Pro funkci $f : \Omega \to \mathbb{R}$ je f^+, f^- její* **kladná část**, *resp.* **záporná část***:*

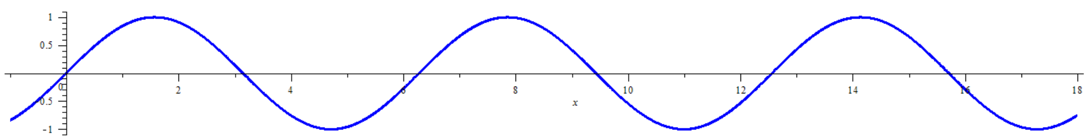
$$f^+(x) = \max(f(x), 0), \qquad f^-(x) = -\min(f(x), 0). \tag{53}$$

LEMMA 4. *Pro funkci $f : \Omega \to \mathbb{R}$ jsou f^+, f^- nezáporné funkce a platí*

$$f(x) = f^+(x) - f^-(x) \tag{54}$$

DŮKAZ. Rozbor případů $f(x) \geq 0$ a $f(x) < 0$.

KONTROLNÍ OTÁZKA 3.2. *Načrtněte kladnou čás* $\sin^+(x)$ *a zápornou část* $\sin^-(x)$ *funkce* $\sin(x)$*,*



OBRÁZEK 3.2. **Graf funkce** $\sin(x)$**.**

LEMMA 5. *Nechť $f : \Omega \to \mathbb{R}$ je měřitelná. Potom jsou měřitelné i funkce f^+, f^-.*

DŮKAZ. Viz definice měřitelnosti: Definice 7.

DEFINICE 13. *Nechť* $f^+ : \Omega \to \mathbb{R}$ *je nezáporná měřitelná funkce na měřitelném prostoru* $(\Omega, \mathcal{F})$ *s mírou* μ. *Potom* **Lebesgueův integrál** *funkce* f^+ *je definován jako*

$$\int_\Omega f^+ \, d\mu = \sup \left\{ \int_\Omega g \, d\mu, \; kde \; g \leq f^+ \; je \; jednoduchá \; funkce \; na \; \Omega \right\}. \tag{55}$$

Měřitelná funkce $f : \Omega \to \mathbb{R}$ *je* **Lebesgueovsky integrovatelná**, *pokud alespoň jeden z Lebesgueových integrálů její kladné části* f^+, *resp. záporné části* f^- *je konečný.* **Lebesgueův integrál** *funkce* f *je potom definován jako*

$$\int_\Omega f \, d\mu = \int_\Omega f^+ \, d\mu - \int_\Omega f^- \, d\mu. \tag{56}$$

POZNÁMKA 9. Takto koncipovaná definice připouští jako hodnoty Lebesgueova integrálu $+\infty$, $-\infty$.

POZNÁMKA 10. Linearita Lebesgueova integrálu (48) platí nejen pro jednoduchou ale i pro obecnou Lebesgueovsky integrovatelnou funkci. Důkaz však vyžaduje trochu práce z důvodu suprema vyskytujícího se v Definici 13.

3.3. Speciální případ: Lebesgueova míra

Lebesgueova míra μ_L je dána rozšířením Borelovy míry na $\mathbb{R}^n$ tak, že k Borelově algebře jsou přidány všechny podmnožiny N množin s nulovou Borelovou mírou a vygenerována nejmenší σ-algebra je obsahující, σ-algebra $(\mathbb{R}^n, \mathcal{L})$ tzv. **Lebesgueovsky měřitelných množin**. Integrál založený na této míře je tzv. Lebesgueův integrál v užším slova smyslu a značí se symbolem dx v integrálu:

$$\int_{\mathbb{R}^n} \bullet d\mu_L \equiv \int_{\mathbb{R}^n} \bullet dx \tag{57}$$

Technicky užitečné je následné tvrzení umožňující počítat Lebesgueův integrál, pokud známe kalkulus pro Riemannův integrál:

TVRZENÍ 6. *Pokud funkce* $f : \mathbb{R} \to \mathbb{R}$ *je na* $\mathbb{R}^n$ *na intervalu* $[a, b]$ *Riemannovsky integrovatelná, potom je na* $[a, b]$ *i Lebesgueovsky integrovatelná a oba integrály jsou si rovny.*

DŮKAZ. Viz [**4**], 1.díl, strana 138.

3.4. Hustota pravděpodobnosti

Je běžné setkávat se ve statistice s různými rozděleními pravděpodobnosti P na stejném měřitelném prostoru $(\Omega, \mathcal{F})$. Zde se omezíme na případ $\Omega = \mathbb{R}^n$ se σ-algebrou $\mathcal{F} = \mathcal{L}$ Lebesgueovsky měřitelných množin. Mnohé z rozdělení pravděpodobnosti lze potom vyjádřit skrz Lebesgueovu míru pomocí tzv. hustoty pravděpodobnosti:

DEFINICE 14. *Spojitá, nezáporná funkce* $p : \mathbb{R}^n \to \mathbb{R}$ *je splňující*

$$\int_{\mathbb{R}^n} p(x) dx = 1 \tag{58}$$

se nazývá **hustotou pravděpodobnosti** *pro pravděpodobnostní míru* P *s Lebesgueovým integrálem*

$$\int_{\mathbb{R}^n} f(x)dP(x) = \int_{\mathbb{R}^n} f(x)p(x)dx \tag{59}$$

Tento vztah mezi P a $p(x)$ je využíván tak často a automaticky, že za cenu určité terminologické nepřesnosti i $p(x)$ označíme jako „**rozdělení pravděpodobnosti** $p(x)$“místo delšího (ale správnějšího) „ rozdělení pravděpodobnosti s hustotou pravděpodobnosti $p(x)$“, jak je tomu zvykem i ve velké části literatury.

Poznámka 11. Zde uvedená definice není nejobecnější možná. Vyhýbá se z důvodu stručnosti řadě případných technických obtíží a též otázce, kdy lze dvě obecné míry porovnat pomocí zobecněné hustoty pravděpodobnosti. Hlubší vhled lze nalézt např. v [**5**].

3.5. Některá rozdělení pravděpodobnosti a jejích charakteristiky

V této sekci uvedeme pomocí spojité hustoty pravděpodobnosti několik potřebných rozdělení pravděpodobností na $\mathbb{R}$.

Na $\mathbb{R}$ existuje navíc význačná funkce x, která každému bodu $x \in \mathbb{R}$ přiřadí příslušné číslo $x \in \mathbb{R}$ a která umožňuje systematicky rozdělení pravděpodobnosti $p(x)$ charakterizovat pomocí jeho **momentů**

$$E(x^n) = \int_{\mathbb{R}} x^n p(x)dx \qquad \text{pro } n \in \mathbb{N}. \tag{60}$$

V praxi se nejčastěji uplatňuje **střední hodnota** $E(x)$ rozdělení pravděpodobnosti $p(x)$

$$E(x) = \int_{\mathbb{R}} xp(x)dx \tag{61}$$

a **rozptyl** $Var(x)$ rozdělení pravděpodobnosti $p(x)$ sestavený z prvních dvou momentů:

$$Var(x) = \int_{\mathbb{R}} (x - E(x))^2 p(x)dx = E(x^2) - E(x)^2 \tag{62}$$

Cvičení na síti 3. **Střední hodnota a rozptyl**.

Rozptyl $Var(x)$ umožňuje omezit pravděpodobnost, že se hodnota x náhodného vzorku odchýlí o vice než danou mez ϵ od střední hodnoty $E(x)$:

Věta 1. **Čebyševova nerovnost.** *Nechť rozdělení pravděpodobnosti* $p(x)$ *na* $\mathbb{R}$ *má konečnou střední hodnotu* $E(x)$ *a konečný rozptyl* $Var(x)$ *a nechť* x_1 *je pozorovaný vzorek. Potom platí:*

$$P(\mid x_1 - E(x) \mid < \epsilon) \geq 1 - \frac{Var(x)}{\epsilon^2} \tag{63}$$

Důkaz. Pravděpodobnost $P(\mid x_1 - E(x) \mid < \epsilon)$ je zapsána poněkud úsporně jako pravděpodobnost, že je splněna podmínka $\mid x_1 - E(x) \mid < \epsilon$. Podrobněji jí vyjádříme tak, že je to míra $P(A)$ množiny A všech bodů x, ve kterých je podmínka $\mid x_1 - E(x) \mid < \epsilon$ splněna:

$$A = \{x_1 \in \mathbb{R} \| \mid x_1 - E(x) \mid < \epsilon\} \tag{64}$$

$$P(\mid x_1 - E(x) \mid < \epsilon) \equiv P(A). \tag{65}$$

Podívejme se na pravděpodobnost $P(\mathbb{R} \setminus A)$ doplňku $\mathbb{R} \setminus A$ k množině A, na kterém musí platit opak podmínky $\mid x_1 - E(x) \mid < \epsilon$, tedy

$$\mid x - E(x) \mid \geq \epsilon \qquad \text{na } \mathbb{R} \setminus A. \tag{66}$$

což lze ekvivalentně vyjádřit jako

$$\frac{\mid x - E(x) \mid^2}{\epsilon^2} \geq 1 \qquad \text{na } \mathbb{R} \setminus A. \tag{67}$$

Podle (51) lze napsat pravděpodobnost množiny $\mathbb{R} \setminus A$ jako integrál z její charakteristické funkce $\chi_{\mathbb{R} \setminus A}$:

$$P(\mathbb{R} \setminus A) = \int_{\mathbb{R}} \chi_{\mathbb{R} \setminus A} dP. \tag{68}$$

Platí ovšem podle (67) na $\mathbb{R} \setminus A$ a podle $\frac{|x-E(x)|^2}{\epsilon^2}$ na $\mathbb{R}$ (a tedy i na $A \in \mathbb{R}$):

$$\chi_{\mathbb{R} \setminus A} \leq \frac{\mid x - E(x) \mid^2}{\epsilon^2} \tag{69}$$

a proto platí

$$P(\mathbb{R} \setminus A) = \int_{\mathbb{R}} \chi_{\mathbb{R} \setminus A} dP \leq \int_{\mathbb{R}} \frac{\mid x - E(x) \mid^2}{\epsilon^2} dP = \frac{Var(x)}{\epsilon^2}, \tag{70}$$

$$P(A) = 1 - P(\mathbb{R} \setminus A) = 1 - \frac{Var(x)}{\epsilon^2}. \tag{71}$$

3.5.1. Normální (Gaussovo) rozdělení. Toto rozdělení na $\Omega = \mathbb{R}$

$$N(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{72}$$

je určeno dvěma parametry $\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}^+$.

Jeho střední hodnota a rozptyl jsou

$$E(x) = \mu, \qquad Var(x) = \sigma^2. \tag{73}$$

3.5.2. Bernoulliho rozdělení. Rozdělení je dáno na dvoubodovém prostoru elementárních jevů $\Omega = \{0, 1\}$ (které lze často interpretovat jako uskutečnění či neuskutečnění sledovaného jevu) jako

$$P(x \mid p) = \begin{cases} p & \text{pro } x = 1, \\ 1 - p & \text{pro } x = 0. \end{cases} \tag{74}$$

a je určeno jedním parametrem $p \in [0, 1]$.

Jeho střední hodnota a rozptyl jsou

$$E(x) = p, \qquad Var(x) = p(1 - p). \tag{75}$$

3.5.3. Rovnoměrné rozdělení na intervalu $[0, 1]$. Rozdělení $U_{[0,1]}$ je dáno na intervalu $\Omega = [0, 1]$ hustotou pravděpodobnosti $u(x) = 1$ nebo formálně na $\mathbb{R}$ jako

$$u(x) = \begin{cases} 1 & \text{pro } x \in [0, 1], \\ 0 & \text{jinak}, \end{cases} \tag{76}$$

což souhlasí s tzv. **charakteristickou funkcí** intervalu $[0, 1]$

Jeho střední hodnota a rozptyl jsou

$$E(x) = \frac{1}{2}, \qquad Var(x) = \frac{1}{12}. \tag{77}$$

ÚLOHA 2. *Ověřte výpočtem uvedené střední hodnoty a rozptyly uvedené v (73), (75) resp. (77) pro rozdělení (72), (74) resp. (76).*

SHRNUTÍ 3. *Na základě pojmu míry na prostoru Ω je možno definovat integrál funkce f přes tuto míru μ,*

$$\int_\Omega f d\mu \tag{78}$$

což je pro normalizovanou míru vážený průměr hodnot funkce s váhou danou pravděpodobností.

Pomocí hustoty pravděpodobnosti $p(x)$ lze vyjádřit některá rozdělení pravděpodobnosti (se symbolem integrace dP) na $\mathbb{R}^n$ ve vztahu k rovnoměrné Lebesgueově míře (se symbolem integrace dx):

$$\int_{\mathbb{R}^n} f dP = \int_{\mathbb{R}^n} f(x)p(x)dx \tag{79}$$

Rozdělení pravděpodobnosti na $\mathbb{R}$ lze charakterizovat pomocí jejích momentů, zejména střední hodnoty $E(x)$ a rozptylu $Var(x)$:

$$E(x) = \int_{\mathbb{R}} xp(x)dx, \qquad Var(x) = \int_{\mathbb{R}} (x - E(x))^2 p(x)dx. \tag{80}$$

KAPITOLA 4

Generátory pseudonáhodných čísel

Klíčová slova: Generátory pseudonáhodných čísel, zobrazení náhodné veličiny, Boxův-Mullerův algoritmus, zamítací (accept-reject) algoritmus.

Abstrakt: Simulace a statistické výpočty vyžadují vytváření vzorků náhodných veličin s určeným rozdělením pravděpodobnosti. Jsou diskutovány generátory pseudonáhodných čísel s rovnoměrným rozdělením a metody jak na jejích základě vytvářet vzorky s jiným rozdělením pravděpodobnosti.

4.1. Úvod

K simulaci náhodných procesů, ale i k přibližným výpočtům je třeba vytvářet vzorky náhodné veličiny x (zde budeme zpravidla uvažovat $x \in \mathbb{R}$ nebo obecněji $x \in \mathbb{R}^n$) s předepsaným rozdělením pravděpodobnosti $p(x)$. Ty bychom si mohli opatřit na základě náhodných fyzikálních jevů, jakými jsou házení mincí či kostkou nebo radioaktivní rozpad. Pro generování většího množství vzorků jsou však tyto metody nepraktické, protože jsou pomalé, případně těžko kontrolovatelné nebo drahé. (To by se zásadně změnilo zavedením dostupných kvantových počítačů). Proto se náhodná čísla v praxi generují pomocí deterministických algoritmů, podle kterých může postupovat běžný počítač.

Postup je následující. Vychází se z počáteční hodnoty $x_0 \in \Omega$ v prostoru možných hodnot Ω, na kterou se aplikuje pevně stanovené zobrazení $D : \Omega \to \Omega$, čímž vznikne další hodnota $x_1 = Dx_0$. Další hodnoty se potom generují obdobně:

$$x_{n+1} = D(x_n) = D^{n+1}(x_0). \tag{81}$$

Lze namítnout, že se v tom případě vůbec nejedná o náhodné vzorky a to je jistě pravda. Z toho důvodu se taky označují jako **pseudonáhodné** vzorky. Při vhodném výběru zobrazení D je však posloupnost vzorků nerozlišitelná běžnými testy od skutečně náhodné posloupnosti vzorků s daným rozdělením pravděpodobnosti. Předpokládá se potom, že ani postupy, za jejichž účelem jsme si pseudonáhodné vzorky opatřili, je nebudou schopny rozlišit od náhodných vzorků.

Zda je nebo není možno pseudonáhodné a náhodné vzorky rozlišit na základě běžných testů, závisí samozřejmě na volbě daných testů a tato volba je poněkud nahodilá a diskutovaná odborníky. V běžné praxi se však tato nejasnost zpravidla neprojeví.

V praxi se zpravidla přímo generují jen pseudonáhodná čísla z intervalu $[0, 1]$ s rovnoměrným rozdělením pravděpodobnosti. Ta se též podrobují přísným testům. Následující sekce je věnována právě takovým generátorům pseudonáhodných čísel. Další sekce se potom věnují metodám, jak pomocí rovnoměrně rozložených pseudonáhodných čísel na jednotkovém intervalu vytvořit vzorky na jiných množinách a s jiným rozdělením pravděpodobnosti. Jedná se o metodu zobrazení náhodné veličiny a o zamítací metodu (rejection sampling, accept-reject algorithm).

4.2. Generátory pseudonáhodných čísel s rovnoměrným rozdělením pravděpodobnosti na jednotkovém intervalu

Numerické výpočty v počítačových programech zpravidla probíhají s určitou pevně zvolenou přesností. Číslo z intervalu $[0,1]$ s přesností $\frac{1}{N}$ se proto dá získat vygenerováním celého čísla $n \in \{1,2,\dots,N\}$ a výsledek je potom dán jako $x = \frac{n}{N}$. Budeme se proto v dalším zabývat pouze případem

$$\Omega = \{1,2,\dots,N\} \qquad \text{pro nějaké } N. \tag{82}$$

Příslušný generátor náhodných čísel je potom dán zobrazením

$$D : \{1,2,\dots,N\} \to \{1,2,\dots,N\} \tag{83}$$

Takto generovaná náhodná čísla se musí nejpozději po N hodnotách začít opakovat, což ovšem v praxi pro vhodné D není překážkou, pokud je N dostatečně velké.

PŘÍKLAD 5. *Jednoduchou volbou zobrazení D je*

$$D(n) = an + b \mod N \qquad \textit{pro } a,b \in \mathbb{Z} \tag{84}$$

a odpovídající generátor náhodných čísel je takzvaný **lineární kongruentní generátor**.

KONTROLNÍ OTÁZKA 4.1. *Pro $N = 100$ najděte konstanty $a,b \in \mathbb{N}$, $a,b > 1$ a počáteční hodnotu $n_0 \in \{0,1,2,\dots,99\}$ tak, že se po 15 krocích žádné z generovaných čísel neopakuje.*

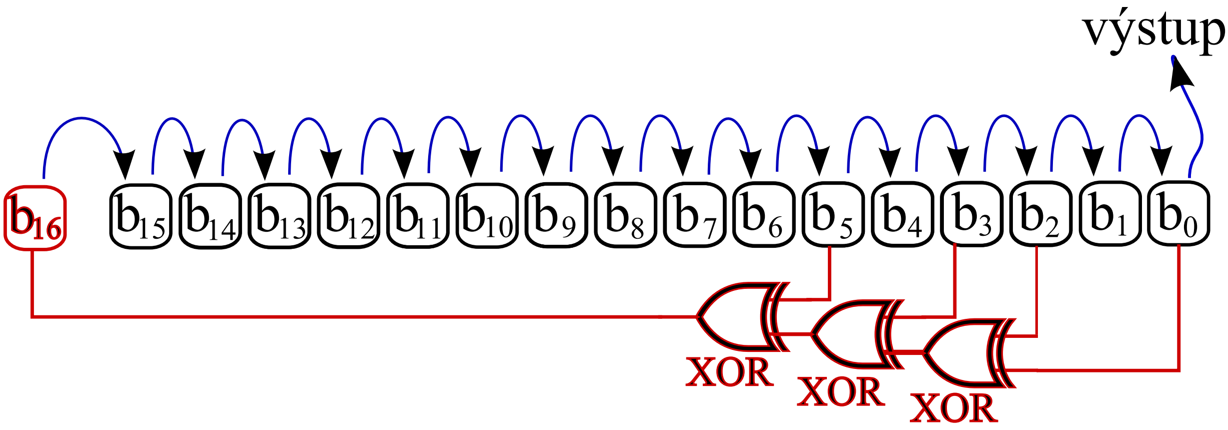
Další jednoduchý postup předpokládá čísla s určitým pevně zvoleným počtem k číslic. Předpokládáme, že číslo, které by ve svém zápisu mělo méně než k číslic je doplněno nulami na k číslic. Zobrazení D je určeno posunutím všech číslic o jednu pozici doprava, přičemž první číslice je nahrazena číslicí vypočtenou pomocí zpětnovazebné funkce f z některých číslic předchozího čísla. Poslední číslice je smazána. Odpovídající generátor náhodných čísel je takzvaný **lineární posuvný registr se zpětnou vazbou**.

Prakticky se v počítači tento postup používá v dvojkové (binární) číselné soustavě, kde možné číslice jsou pouze 0 a 1. Zpětnovazebná funkce f se zpravidla složením několika kopií funkce XOR:

$$\mathrm{XOR}(x,y) = \begin{cases} 0 & \text{pro } [x,y] = [0,0] \text{ nebo } [x,y] = [1,1], \\ 1 & \text{pro } [x,y] = [1,0] \text{ nebo } [x,y] = [0,1]. \end{cases} \tag{85}$$

PŘÍKLAD 6. *Příkladem je 16-bitový Fibonaccího lineární posuvný registr se zpětnou vazbou, viz Obrázek 4.1. Zpětnovazebná funkce f nabývá na čísle b s číslicemi $b_{15}, b_{14}, \dots, b_2, b_1, b_0$ hodnoty*

$$f(b) = XOR(b_5, XOR(b_3, XOR(b_2, b_0))) \tag{86}$$



OBRÁZEK 4.1. **Fibonaccího lineární posuvný registr se zpětnou vazbou.** Registr pracuje ve dvou krocích: Nejdříve je spočten pomocný bit b_{16} (výpočty označeny červeně) na základě obsahu registru v bitech b_0 až b_{15} a následně dojde k posunutí bitů registru (vyznačeno modrými šipkami), čímž se vysune výstupní bit a zároveň uvolní pomocný bit b_{16}.

Předchozí generátory jsou jednoduché a rychlé. Statistické vlastnosti jimi generovaných vzorků však ještě nejsou zcela vyhovující, ukazují však, jak je možno v principu postupovat. Pro moderní implementaci viz např. generátor zvaný **Mersene twister** [**15**], který je používaný v řadě programů.

POZNÁMKA 12. Přirozenou myšlenkou je využít v generování pseudonáhodných čísel z $[0,1]$ diskretizaci některého zobrazení často diskutovaného ve zkoumání chaosu, např. tzv **tent map** $T : [0,1] \to [0,1]$:

$$T(x) = \begin{cases} 2x & \text{pro } x \in [0, \frac{1}{2}] \ , \\ 2(1-x) & \text{pro } x \in (\frac{1}{2}, 1], \end{cases} \tag{87}$$

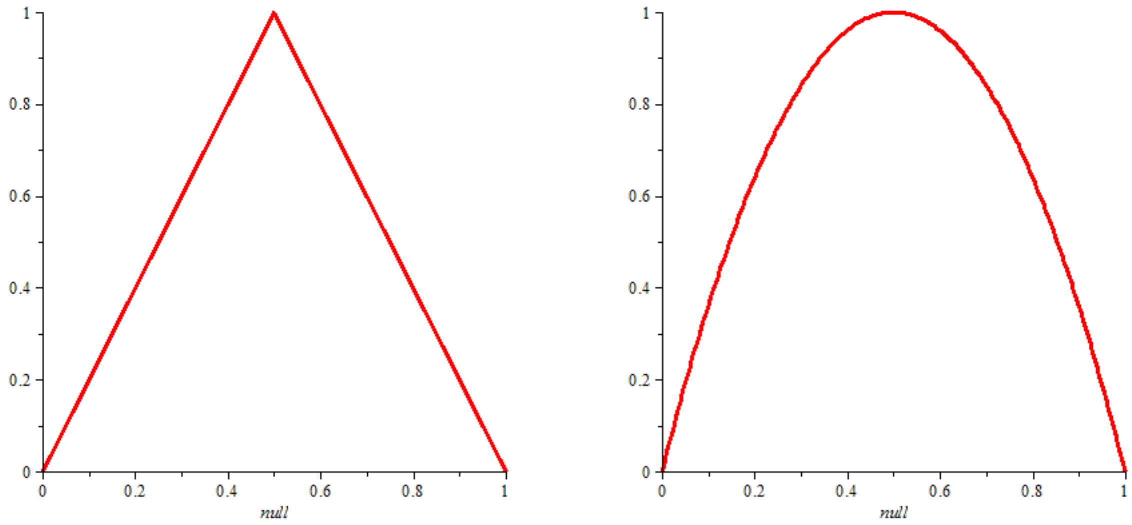
nebo logistické zobrazení L_α:

$$L_\alpha(x) = \alpha x(1-x) \tag{88}$$

s teoretickým rozdělením na invariantním intervalu $[0,1]$ pro $\alpha = 2$:

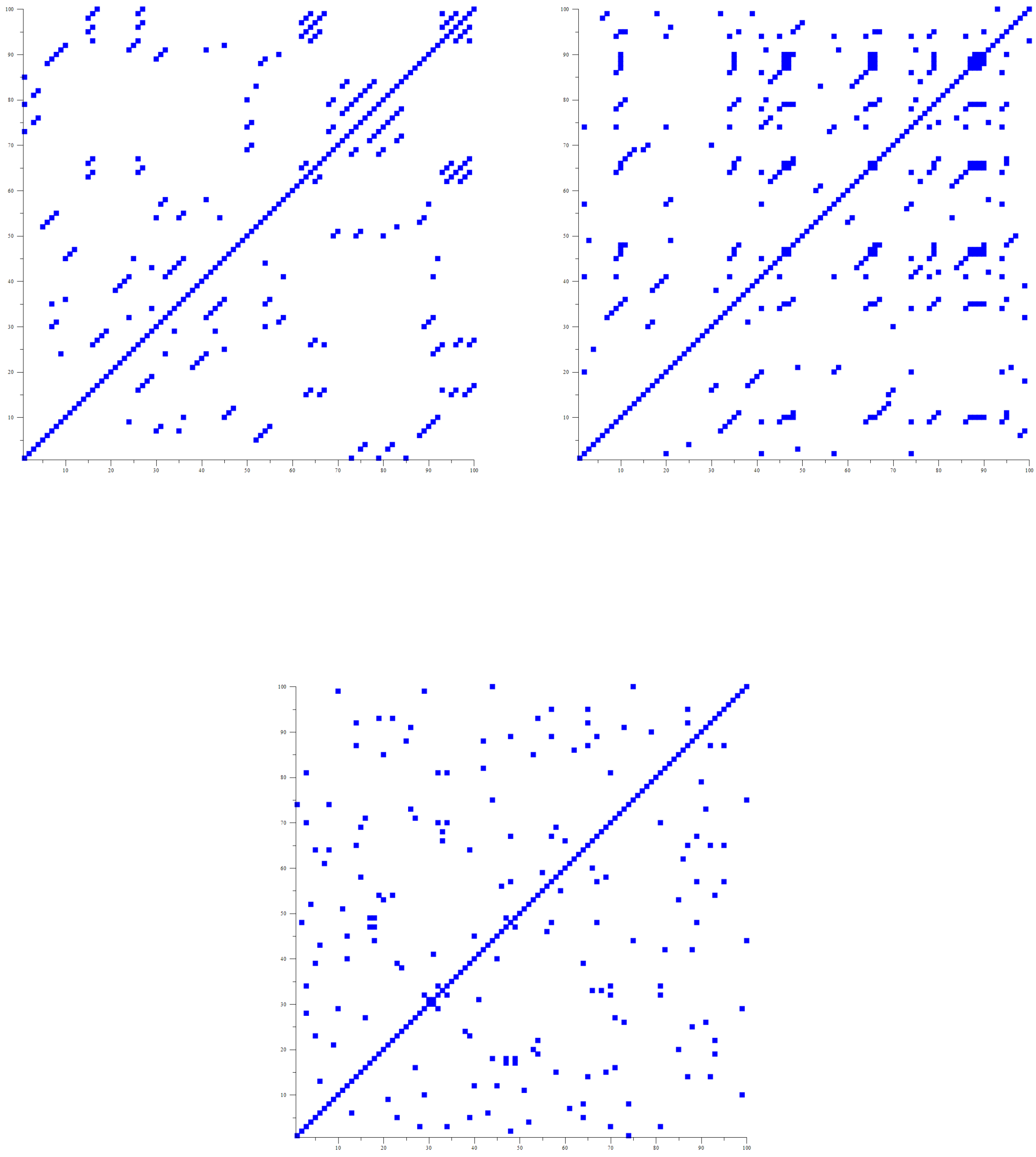
$$p(x) = \frac{1}{\pi}\sqrt{x(1-x)}. \tag{89}$$

(viz [**19**], str. 37)



OBRÁZEK 4.2. **Tent map a logistické zobrazení.** Logistické zobrazení (vpravo) je vykresleno pro $\alpha = 4$.

Ukazuje se však, že tato myšlenka, přinejmenším v této jednoduché podobě, nezaručuje jako výsledek použitelný generátor pseudonáhodných čísel (viz též Obrázek 4.2) a není navíc numericky příliš efektivní.



Obrázek 4.3. **Rekurenční grafy.** Jednou z možností získání rychlého vhledu do chování diskrétního dynamického systému generujícího posloupnost hodnot $x_0, x_1, x_2, \ldots$ jsou rekurenční grafy [**8**]. Na osách rekurenčního grafu se vynáší indexy $i, j = 1, 2, \ldots$ a jejich průsečík je zabarven, pokud se x_j dostalo k x_i blíže než ϵ. Na grafech jsou v následujícím pořadí rekurenční grafy pro tent map, logistické zobrazení a běžný generátor náhodných čísel (Mersene twister). Diagonála je automaticky součástí rekurenčního grafu. Zatímco pro generátor náhodných čísel nelze rozeznat další strukturu, jsou v prvních dvou případech zřetelné obrazce. Použité hodnoty v grafech byly $x_0 = 0.1234567890$, $\epsilon = 0.01$.

POZNÁMKA 13. Z pohledu běžných simulací ve statistice je např. Mersene twister dobrou volbou pro generátor pseudonáhodných čísel. To však neplatí v případě použití pseudonáhodných čísel v kryptografii, kdy je třeba zaručit, že generovaná čísla jsou nejen nerozlišitelná od vzorku generovaného rovnoměrným rozdělením pomocí statistických testů, ale že je i velmi náročné, technicky dostupnými prostředky takřka nemožné pochopit vztahy mezi generovanými hodnotami.

POZNÁMKA 14. Vzorku x rovnoměrného rozdělení na intervalu $[0, 1]$ lze v simulacích použít k uskutečnění události A s pravděpodobností $P(A)$:

- Pokud $x \leq P(A)$, jev A se uskuteční,
- Pokud $x > P(A)$, jev A se neuskuteční.

KONTROLNÍ OTÁZKA 4.2. *Jak můžete na základě vzorku x rovnoměrného rozdělení na intervalu* $[0, 1]$ *simulovat hod kostkou?*

CVIČENÍ NA SÍTI 4. **Generátor číslic 0 a 1**.

4.3. Simulace obecných rozdělení pravděpodobnosti inverzním zobrazením ze stejnoměrného rozdělení

TVRZENÍ 7. *Předpokládejme, že je dána měřitelná funkce $f : X \to Y$ a že na prostoru X je dána míra μ. Potom je zobrazením f jednoznačně určena míra na Y, kterou budeme označovat $f^*\mu$, a to vztahem:*

$$f^*\mu(B) = \mu(f^{-1}(B)) \qquad \textit{pro každou měřitelnou množinu } B \in Y. \tag{90}$$

DŮKAZ. Připomeňme nejdříve, že o měřitelnosti funkce $f : X \to Y$ lze mluvit pouze pokud X a Y jsou měřitelné prostory. Na každém z nich musí tedy být dána σ-algebra, kterou označíme $\mathcal{P}$, resp. $\mathcal{Q}$. Že tyto σ-algebry jsou předpokládány, ale nikoliv zmíněny v textu tvrzení, se může jevit jako zlomyslnost, která je však v matematických textech častou zkratkou. Zkrátka, stručnost je vykoupená předpokladem, že si čtenář nutné a nevyhnutné detaily dokáže sám doplnit. S tímto doplněním tedy předpokládáme měřitelné prostory $(X, \mathcal{P})$, $(Y, \mathcal{Q})$.

Definiční vlastností měřitelnosti funkce f je skutečnost, že vzorem libovolné měřitelné množiny $B \in \mathcal{Q}$ je měřitelná množina $B^{-1} \in \mathcal{P}$. Protože B^{-1} je měřitelná, lze stanovit její míru $\mu(f^{-1}(B))$ a pravá strana vztahu (90) proto dává smysl a lze jej použít na definici levé strany vztahu (90). K ověření definičních vlastností míry si je třeba uvědomit, že pokud množiny $B_1, B_2 \in \mathcal{Q}$ jsou disjunktní, tak potom totéž platí pro jejich vzory $B_1^{-1}, B_2^{-1} \in \mathcal{P}$.

POZNÁMKA 15. Pokud uvažovanou mírou μ v předchozím Tvrzení 7 je rozdělení pravděpodobnosti, potom i míra $\mu(f^{-1}(B))$ je rozdělením pravděpodobnosti, pokud $\mu(f(X)) = 1$, což je nutně splněno, pokud zobrazení f je surjektivní (je "na").

Zaměříme se nyní na měřitelná zobrazení, která jsou nejen surjektivní, ale zároveň také injektivní ("prostá"), tj. zaměříme se na bijekce. V takovém případě jsou vztahy mezi odpovídajícími vzorky generovanými podle příslušných rozdělení pravděpodobností obzvláště jednoduché.

TVRZENÍ 8. *Nechť je $f : X \to Y$ bijekce měřitelných prostorů a nechť P je rozdělení pravděpodobnosti na X. Pokud $y \in Y$ je vzorek generovaný podle rozdělení pravděpodobnosti f^*P na Y, potom $f^{-1}(y) \in X$ je vzorek generovaný podle rozdělení pravděpodobnosti P na X.*

DŮKAZ. Protože f je injektivní (prosté), je vzorem jednotlivého vzorku $y \in Y$ jednotlivý vzorek $f^{-1}(y) \in X$. Vztah odpovídajících rozdělení pravděpodobností je dán Tvrzením 7.

DŮSLEDEK 9. *Nechť je dáno rozdělení pravděpodobnosti P s hustotou pravděpodobnosti $p(x)$ na $\mathbb{R}$ takové, že jeho tzv.* **distribuční funkce**

$$F(x) = \int_{-\infty}^{x} p(t)dt \tag{91}$$

je bijekcí $F : \mathbb{R} \to [0, 1]$.

Potom F^*P je rovnoměrné rozdělení pravděpodobnosti a podle předchozího Tvrzení 8 je možno vzorek podle pravděpodobnostního rozdělení P generovat jako $F^{-1}(y)$, kde $y \in Y = [0, 1]$ je vzorek generovaný podle rovnoměrného rozdělení na jednotkovém intervalu $[0, 1]$.

DŮKAZ. V tomto případě je mlčky předpokládáno, že příslušné σ-algebry jsou generovány intervaly. Postačí tedy prokázat vztah (90) na intervalech. Nechť tedy je dán interval $[a, b] \subset \mathbb{R}$ a jeho obraz $B = [F(a), F(b)] \in [0, 1]$, přičemž $[a, b] = F^{-1}B = F^{-1}[F(a), F(b)]$. Musíme prokázat, že

$$F^*P([a, b]) = P(F^{-1}(B)) \tag{92}$$

odpovídá rovnoměrnému rozdělení. To je ovšem pravda, protože platí

$$P(F^{-1}(B)) = P([a, b]) = \int_a^b p(x)dx = F(b) - F(a) = \int_{F(a)}^{F(b)} 1\, dy, \tag{93}$$

kde 1 v posledním integrálu představuje hustotu pravděpodobnosti rovnoměrného pravděpodobnostního rozdělení na jednotkovém intervalu.

KONTROLNÍ OTÁZKA 4.3. *Exponenciální rozdělení s parametrem α je dáno hustotou pravděpodobnosti*

$$p(t \mid \alpha) = \begin{cases} \alpha e^{-\alpha t} & \textit{for } t \geq 0, \\ 0 & \textit{for } t < 0. \end{cases} \tag{94}$$

Jak lze generovat vzorek x tohoto rozdělení, pokud máme k dispozici náhodné číslo y generované dle rovnoměrného rozdělení na jednotkovém intervalu $[0, 1]$?

Výše uvedené metody lze rozšířit na případ více dimenzí a funkce, které nejsou bijekcemi. Jako příklad uveďme bez důkazu následující, užitečný algoritmus pro generování dvou nezávislých reálných náhodných veličin x_1, x_2 s normálním rozdělením pravděpodobnosti:

ALGORITMUS 1. **(Boxův-Mullerův algoritmus).** *Nechť jsou* $u_1, u_2 \in [0,1]$ *nezávislé náhodné veličiny s rovnoměrným rozdělením nad jednotkovým intervalem* $[0,1]$ *a nechť je dáno měřitelné zobrazení* $f : [0,1] \times [0,1] \to \mathbb{R} \times \mathbb{R}$ *předpisem*

$$x_1 = \sqrt{-2 \log u_1} \cos 2\pi u_2 \tag{95}$$

$$x_2 = \sqrt{-2 \log u_1} \sin 2\pi u_2 \tag{96}$$

Potom jsou $x_1 \in \mathbb{R}$, $x_2 \in \mathbb{R}$ *nezávislé náhodné veličiny s normálními rozděleními*

$$p(x_1) = N(x_1 \mid 0,1) \qquad\qquad p(x_2) = N(x_2 \mid 0,1) \tag{97}$$

4.4. Zamítací metoda (rejection sampling, accept-reject algorithm)

VĚTA 2. **(Fundamentální věta simulace).** *Nechť je hustota pravděpodobnosti* $p(x)$ *na* Ω *omezena hodnotou* m*. Generovat hodnoty* $x \in \Omega$ *podle pravděpodobnostního rozdělení* $p(x)$ *je potom ekvivalentní generování* $(x,u) \in \Omega \times [0,m]$ *s rovnoměrným rozdělením na* $\Omega \times [0,m]$ *s podmínkou přijetí hodnoty* (x,u)*:*

$$u \leq p(x). \tag{98}$$

DŮKAZ. Rozdělení pravděpodobnosti $p(x,u)$ na $\Omega \times [0,m]$ lze s pomocí rovnoměrných rozdělení pravděpodobnosti $U_\Omega(x)$ na Ω resp. $U_{[0,m]}(u)$ na $[0,m]$ formálně zapsat jako

$$p(x,u) = \frac{1}{N} U_\Omega(x) U_{[0,m]}(u) H(p(x) - u) \tag{99}$$

s normalizační konstantou N:

$$N = \int_{\Omega \times [0,m]} U_\Omega(x) U_{[0,m]}(u) H(p(x) - u) d\Omega\, du \tag{100}$$

a kde H je Heavisideova funkce:

$$H(t) = \begin{cases} 0 & \text{pro } t < 0, \\ 1 & \text{pro } t \geq 0, \end{cases} \tag{101}$$

Výsledné rozdělení pravděpodobnosti na Ω je potom

$$\int_{[0,m]} p(x,u) du = \int_{[0,m]} \frac{1}{N} U_\Omega(x) U_{[0,m]}(u) H(p(x) - u)\, du = \tag{102}$$

$$= \frac{1}{N} U_\Omega(x) \int_{[0,p(x)]} U_{[0,m]}(u) H(p(x) - u)\, du = \tag{103}$$

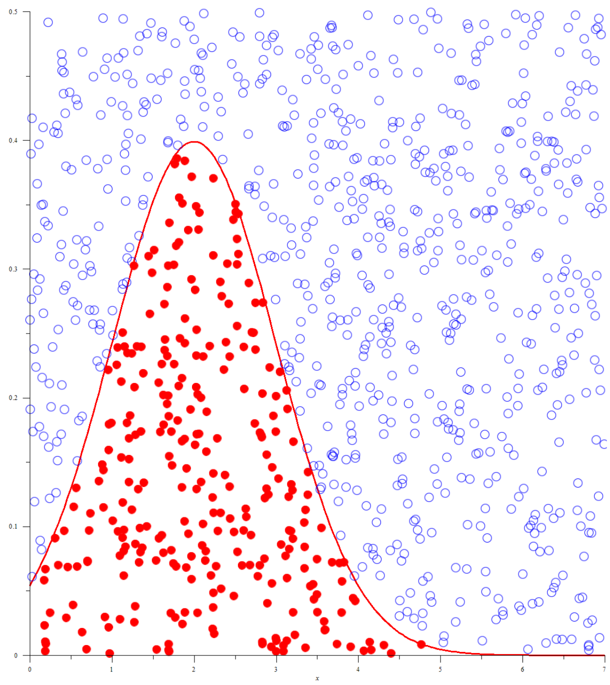
$$= \frac{1}{N} U_\Omega(x) \frac{p(x)}{m} = p(x) \tag{104}$$

kde poslední rovnost plyne z toho, že $\frac{1}{Nm} U_\Omega(x)$ je konstantní na Ω a $p(x)$ je normalizováno na Ω.

Algoritmus 2. **Zamítací metoda (accept-reject).** *Prakticky lze potom generovat* (x,u) *s rovnoměrným pravděpodobnostním rozdělením na* $\Omega \times [0,m]$*, kde* $m \geq \sup_{x\in\Omega} p(x)$ *hodnotu* x *přijmout při* $u \leq p(x)$ *nebo odmítnout při* $u > p(x)$ *a generovat další hodnoty* (x,u)*. Přijaté hodnoty* x *jsou potom generovány dle rozdělení* $p(x)$*.*

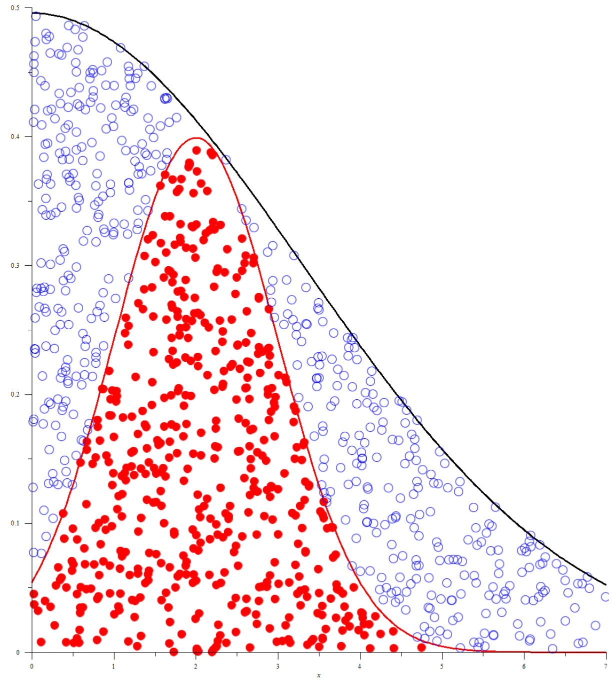
Poznámka 16. Výsledky Algoritmu 2 se nezmění, pokud by namísto hustoty $p(x)$ byla použita obecně nenormalizovaná hustota $cp(x)$ pro konstantu $c > 0$. Při zadávání hustoty $p(x)$ proto není třeba obtěžovat se s její normalizací.

Z Obrázku 4.4 je zřejmý princip algoritmu: obdélník $\Omega \times [0,m]$ je rovnoměrně pokryt generovanými body a hodnoty nad grafem $p(x)$ jsou zamítnuty. Metoda je neefektivní v tom smyslu, že zamítnutých bodů nad grafem může být nutné vygenerovat mnoho, přičemž do výsledného vzorku nevcházejí, byly vygenerovány zbytečně. Tomu lze zamezit generováním podle jiného rozdělení $q(x)$, které (po případném přenásobením vhodnou kladnou konstantou) dokáže shora oříznout generované body a snížit tak podíl zamítnutých bodů, viz Obrázek 4.5. To je obzvlášť důležité, pokud má být zahrnuta asymptotická oblast rozdělení.



Obrázek 4.4. **Zamítací metoda.** Obrázek zachycuje 1000 pseudonáhodných vzorků generovaných s rovnoměrným rozdělením na $[0,7] \times [0,0.5]$. Jen vzorky (plné, červené kroužky) ležící pod křivkou $p(x)$ jsou přijaty a představují výsledné vzorky. Body nad křivkou (prázdné, modré kroužky) jsou zamítnuty a dále se nepoužijí. Křivka v tomto grafu představuje nenormalizovanou hustotu pravděpodobnosti. Výsledky je tedy třeba normalizovat.

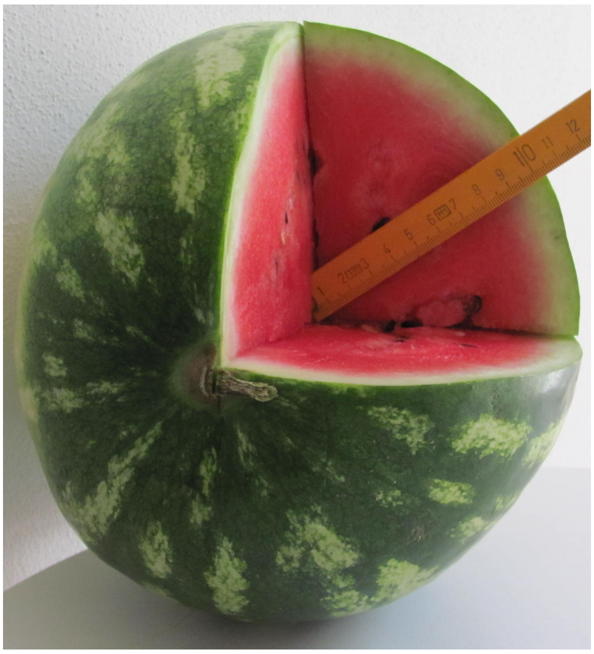
Algoritmus 3. **Vzorkování dle důležitosti (importance sampling).** *Nechť pro rozdělení pravděpodobnosti* $q(x)$ *na* Ω*, ke kterému je možno generovat vzorky, existuje konstanta* $M > 0$ *tak, že pro rozdělení pravděpodobnosti* $p(x)$ *na* Ω *platí* $p(x) < M\,q(x)$*. Prakticky lze potom generovat* (x, u) *s pravděpodobnostním rozdělením* $q(x) \times U[0, M\,q(x)]$*, hodnotu* x *přijmout při* $u \leq p(x)$ *nebo odmítnout při* $u > p(x)$ *a generovat další hodnoty* (x, u)*. Přijaté hodnoty* x *jsou potom generovány dle rozdělení* $p(x)$*.*



Obrázek 4.5. **Vzorkování dle důležitosti.** Obrázek zachycuje 1000 pseudonáhodných vzorků generovaných podle funkce důležitosti $q(x)$ na $\Omega = [0, 7]$. Z generovaných bodů ležících pod M-násobkem $Mq(x)$ funkce důležitosti (černá, horní křivka) jsou přijaty jen vzorky (plné, červené kroužky) ležící pod křivkou $p(x)$ (červená křivka). Body nad křivkou (prázdné, modré kroužky) jsou zamítnuty a dále se neuvažují. Křivky v tomto grafu představují nenormalizované hustoty pravděpodobnosti. Výsledky je tedy třeba normalizovat.

Z porovnání Obrázků 4.4 a 4.5 je zřejmé, že v případě generování dle důležitosti došlo při vygenerování stejného počtu 1000 bodů k vygenerování více použitelných vzorků. Z grafů není tak zřejmé, že ovšem přetrvává následující obtíž: Při práci ve vysoké dimenzi získají oblasti ve větších vzdálenostech od počátku velkou váhu (viz Úloha 3) a jakákoliv odchylka mezi funkci $Mq(x)$ (M-násobku funkce důležitosti) a rozdělením pravděpodobnosti $q(x)$ povede k značnému podílu odmítnutých případů. Abychom se tomu vyhnuli, je třeba ve vysokých dimenzích velmi pečlivě sladit asymptotické chování $Mq(x)$ a $p(x)$, čímž metoda ovšem ztrácí na obecné použitelnosti. Těmto obtížím ve vysokých dimenzích se vyhne generování vzorků pomocí Markovových řetězců, které je popsáno v dalších dvou kapitolách.

ÚLOHA 3. *Dokažte, že ve vysokých dimenzích jsou melouny samá slupka.*



OBRÁZEK 4.6. Meloun v $\mathbb{R}^3$

Předpokládejte, že meloun má tvar koule s poloměrem R, přičemž na slupku připadá vnějších 10% poloměru. Nechť je $V^n_{vnitřek}$ n-rozměrným objemem jedlého vnitřku melounu a V^n_{slupka} n-rozměrným objemem šlupky. Spočtěte podíl $\frac{V^n_{slupka}}{V^n_{vnitřek}+V^n_{slupka}}$ objemu slupky na celkovém objemu melounu v několika malých dimenzích n a dokažte, že

$$\lim_{n\to\infty} \frac{V^n_{slupka}}{V^n_{vnitřek} + V^n_{slupka}} = 100\%. \tag{105}$$

Pomůcka: *Jednotkový n-rozměrný objem v $\mathbb{R}^n$ představuje jednotková krychle $[0,1]^n$. n-rozměrný objem n-rozměrné koule $B^n(R)$ s poloměrem R je*

$$V(B^n(R)) = \frac{\pi^{n/2}}{\Gamma(1+n/2)}R^n = \begin{cases} \frac{\pi^{n/2}}{(n/2)!}R^n & \text{pro } n \in \mathbb{N} \text{ sudé,} \\ \frac{(2\pi)^{(n-1)/2}}{n!!}R^n & \text{pro } n \in \mathbb{N} \text{ liché.} \end{cases} \tag{106}$$

kde dvojitý faktoriál je $n!! = 1 \cdot 3 \cdot 5 \cdot 7 \cdot \cdots \cdot (n-2) \cdot n$ pro n liché.

Shrnutí 4. *Pseudonáhodná čísla na jednotkovém intervalu lze generovat s rovnoměrným rozdělením na počítači dle standardních algoritmů. Na jejich základě lze generovat vzorky podle speciálních rozdělení nebo zamítací metodou i pro obecná, těžko analyticky zpracovatelná rozdělení.*

Obecné metody se ovšem mohou dostat do úzkých při snaze věrně zachytit asymptotiku rozdělení pravděpodobnosti v nekonečnu, či v případě vysokých dimenzí.

KAPITOLA 5

Markovovy řetězce

Klíčová slova: Konečné a spojité Markovovy řetězce, Markovova vlastnost, matice přechodu, stacionární stav, ergodicita.

Abstrakt: Markovovy řetězce jsou diskrétní dynamické systémy, které mají za vhodných okolností jednoznačný stacionární stav (rozdělení pravděpodobnosti), který je v pravděpodobnosti limitou jejích časového vývoje. Dynamikou Markovova řetězce generované vzorky nejsou nezávislé, ale dokáží nezávislé vzorky z důvodu ergodicity asymptoticky zastoupit při výpočtu středních hodnot.

5.1. Úvod

V předchozí kapitole jsme se seznámili s některými postupy generování vzorků dle daného rozdělení pravděpodobnosti. Vedle nich však existuje další možnost generování vzorků, vycházející z určitého druhu diskrétního stochastického dynamického systému, tzv. **Markovova řetězce**. Tato metoda umožňuje efektivně generovat vzorky i pro rozdělení pravděpodobnosti bez zvláštních vlastností a na případně vysoko-rozměrném prostoru a je proto základem řady moderních numerických metod ve statistice.

Fakt, že Markovův řetězec je diskrétní stochastický dynamický systém, znamená, že hodnoty x_t, kterých nabývá, se mění v diskrétních časových krocích (tj. je to diskrétní dynamický systém; čas t nabývá jen hodnot v přirozených číslech a nikoliv v libovolných reálných číslech, jak by odpovídalo plynulému toku času[1]) a hodnoty v jednotlivých časech nejsou pevně, deterministicky určeny, ale dány náhodným (stochastickým) procesem.

Definice 15. *Posloupnost náhodných veličin $x_0, x_1, x_2, \ldots$ je* **Markovovým řetězcem**, *pokud platí tzv.* **Markovova vlastnost** *pro podmíněné pravděpodobnosti:*

$$P(x_n \mid x_0, x_1, \ldots x_{n-1}) = P(x_n \mid x_{n-1}) \qquad \text{pro libovolné } n \in \mathbb{N}. \tag{107}$$

Markovův řetězec $x_0, x_1, x_2, \ldots$ je **konečný**, *pokud x_n nabývá jen konečně mnoha hodnot pro každé $n \in \mathbb{N}_0$. Markovův řetězec je* **časově homogenní** *(nebo též* **stacionární***), pokud x_n nabývá pro každé $n \in \mathbb{N}_0$ hodnot ve stejné, časově neproměnné množině a platí:*

$$P(x_{n+1} \mid x_n) = P(x_n \mid x_{n-1}) \qquad \text{pro libovolné } n \in \mathbb{N}. \tag{108}$$

Pro konečný Markovův řetězec tvoří pravděpodobnosti

$$M_{x_{n,n-1}} = P(x_n \mid x_{n-1}) \tag{109}$$

[1]Existuje ovšem obdoba Markovových řetězců pro spojitou časovou proměnnou, tzv. **Markovovy procesy**

tzv. **matici pravděpobnosti přechodu**.

V dalším budeme předpokládat, že všechny Markovovy řetězce jsou časově homogenní.

Markovova vlastnost (107) znamená, slovně vyjádřeno, že pokud známe hodnoty $x_0, x_1, \ldots x_n$, je pravděpodobnost následné hodnoty x_n určena *pouze* předcházející hodnotou x_{n-1} a nezávisí na ostatních předcházejících členech.

Alternativně se dá říct, že k vygenerování náhodné hodnoty x_n stačí znát x_{n-1}.

Markovovy řetězce mají za vhodných podmínek následující vlastnost: rozdělení pravděpodobnosti, které se na nich v čase vyvíjí, se asymptoticky blíží k rovnovážnému stavu s určitým rozdělením pravděpodobnosti P. Vzorky generované daným časovým vývojem jsou potom asymptoticky vzorky rovnovážného rozdělení pravděpodobnosti P.

Princip si v příští sekci ujasníme na případu konečných Markovových řetězců. V poslední sekci zmíníme situaci technicky náročnějšího, ale v praxi užívanějšího případu Markovova řetězce se spojitým prostorem stavů. V tom případě je třeba podmíněné pravděpodobnosti $P(x_n \mid \ldots)$ v (107) chápat jako hustoty pravděpodobnosti v x_n. Pro podrobnější informace ovšem odkážeme na specializovanou odbornou literaturu, viz [**19**], [**17**].

Otevřená zůstane otázka, zda pro dané rozdělení pravděpodobnosti P lze nalézt vhodný Markovův řetězec tak, aby se s jeho pomoci daly asymptoticky generovat vzorky pro rozdělení P. Kladnou odpověď poskytne následující kapitola v podobě Metropolisova-Hastingsova algoritmu.

5.2. Konečné Markovovy řetězce

Označme stavy konečného Markovova řetězce přirozenými čísly:

$$x_n \in \{1, \ldots, N\} \tag{110}$$

Matice pravděpodobností přechodu konečného, časově homogenního Markovova řetězce jsou rovny konstantní matici M. Matice M typu $N \times N$ je tzv. stochastickou maticí, tj. každý její sloupec představuje pravděpodobnostní rozdělení na N bodech:

$$M_{ij} \geq 0, \qquad \sum_{i=1}^{N} M_{ij} = 1. \tag{111}$$

Pravděpodobnost přechodu ze stavu x_0 v čase 0 do stavu x_n v čase n je dána opakovaným použitím Bayesova vzorce (212) a sesčítáním přes všechny možnosti $x_1, \ldots, x_{n-1}$:

$$P(x_n \mid x_0) = \sum_{x_1, \ldots, x_{n-1}} P(x_n \mid x_{n-1}) P(x_{n-1} \mid x_{n-2}) \ldots P(x_1 \mid x_0) = (M^n)_{x_n x_0}, \tag{112}$$

kde poslední výraz je zapsán pomocí maticové mocniny matice pravděpodobnosti přechodu M definované v (109).

Následuje řada definic. Přibližují nám, čeho si lze na Markovových řetězcích všimnout a zachycují některé jejích rysy, které jsou technicky užitečné buď přímo, nebo v jejich obdobě pro spojité Markovovy řetězce. Klíčové však je, že tyto definice zároveň poskytují pojmovou strukturu pro zavedení ergodicity (Definice 21), která nakonec zaručuje, že lze (asymptoticky) jediným časovým vývojem $x_0, x_1, x_2, \ldots$ z libovolně zvoleného počátečního stavu x_0 vyčerpávajícím způsobem vystihnout vlastnosti rovnovážného, stacionárního stavu Markovova řetězce zmíněného v Úvodu.

DEFINICE 16. *Stav $j \in \{1, \dots, N\}$ časově homogenního Markovova řetězce je* **dosažitelný** *ze stavu $i \in \{1, \dots, N\}$ (což značíme $i \to j$), pokud pravděpodobnost přechodu $P(x_n \mid x_0)$ je nenulová (a tedy kladná) pro nějaký čas $n \geq 0$.*

Stavy i, j spolu **komunikují** *(značíme $i \leftrightarrow j$) pokud i je dosažitelný z j a j je dosažitelný z i, tj.*

$$i \leftrightarrow j, \text{ pokud } i \to j \text{ a } j \to i. \tag{113}$$

Markovův řetězec je **ireducibilní**, *pokud všechny jeho stavy vzájemně komunikují.*

DEFINICE 17. *Stav i je* **přechodný**, *pokud pravděpodobnost, že se v žádném dalším čase již do i nevrátí, je nenulová, tj. pravděpodobnost návratu do i je menší než 1:*

$$\sum_{n=1}^{\infty} P(x_n = i \mid x_0 = i) \equiv \sum_{n=1}^{\infty} (M^n)_{ii} < 1 \tag{114}$$

Stav je **rekurentní**, *pokud není přechodný.*

Návrat do původního stavu nemusí být možný v libovolném počtu kroků. Následující definice především charakterizuje, v kterém počtu kroků se nelze vrátit do původního stavu $x_0 = i$:

DEFINICE 18. *Stav i má* **periodu** *k, jestliže k je největší společný dělitel počtů kroků, kdy návrat má nenulovou pravděpodobnost:*

$$\{n \in \mathbb{N} \mid P(x_n = i \mid x_0 = i) > 0\} \tag{115}$$

Stav i je **aperiodický**, *pokud má periodu $k = 1$.*

KONTROLNÍ OTÁZKA 5.1. *Mějme dány následující stochastické 3×3-matice určující odpovídající Markovovy řetězce na množině $\{1, 2, 3\}$.*

$$M_A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad M_B = \begin{pmatrix} 1/2 & 1/3 & 0 \\ 1/2 & 1/3 & 1/2 \\ 0 & 1/3 & 1/2 \end{pmatrix}, \quad M_C = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{116}$$

Pro každý Markovův řetězec rozhodněte pro $i, j \in \{1, 2, 3\}$

(1) *Které j je dosažitelné z i?*
(2) *Která i, j spolu komunikují?*
(3) *Je daný Markovův řetězec ireducibilní?*
(4) *Který stav i je přechodný a který rekurentní?*
(5) *Který stav i je periodický (S jakou periodou?) a který je aperiodický?*

Pro konkretní průběh $x_0, x_1, x_2, \dots$ lze definovat následující:

DEFINICE 19. **Doba prvního návratu** *T_i stavu i je nejmenší počet kroků $n \geq 1$, po kterém se stav Markovova řetězce začínající v $x_0 = i$ znovu vrátí do i, tj. $x_n = i$.*

Dobu prvního návratu lze chápat jako náhodnou veličinu, ale nikoliv na prostoru $\{1,\ldots,N\}$, ale na prostoru $\{1,\ldots,N\}^{\mathbb{N}}$ všech realizací dynamiky $x_0, x_1, x_2, \ldots$. Potom má smysl mluvit o střední hodnotě této náhodné veličiny:

DEFINICE 20. **Střední doba prvního návratu** *je definována jako*

$$E(T_i) = \sum_{n=1}^{\infty} nP(T_i = n \mid x_0 = i) \tag{117}$$

Stav i je **kladně rekurentní***, pokud $E(T_i) < \infty$.*

DEFINICE 21. *Stav i časově homogenního Markovova řetězce je* **ergodický***, pokud je kladně rekurentní a aperiodický. Markovův řetězec je ergodický, pokud je ireducibilní a všechny jeho stavy jsou ergodické.*

DEFINICE 22. *Pravděpodobnostní rozdělení*

$$P = \begin{pmatrix} p(1) \\ \vdots \\ p(N) \end{pmatrix} \tag{118}$$

je **stacionárním stavem** *pro stochastickou $N \times N$-matici M, pokud platí:*

$$P = MP \tag{119}$$

TVRZENÍ 10. *Každá stochastická matice M má stacionární stav, který odpovídá její největší vlastní hodnotě. Takových vektorů může být více.*

Pokud navíc platí, že

$$M_{ij} > 0 \qquad \text{pro } i, j \in \{1,\ldots,N\} \tag{120}$$

potom je stacionární stav jednoznačně určený a platí

$$\lim_{n\to\infty} (M^n)_{ij} = p_i \qquad \text{pro } j \in \{1,\ldots,N\} \tag{121}$$

DŮKAZ. (viz. Perronova-Frobeniova věta a Kap. 11 v [**11**].)

KONTROLNÍ OTÁZKA 5.2. *Markovův řetězec na $\{1,2,\ldots,9\}$ je dán maticí $M = \frac{1}{45}S$, kde S je 9×9-matice vzniklá řešením následujícího sudoku:*

5	3			7				
6			1	9	5			
	9	8					6	
8				6				3
4			8		3			1
7				2				6
	6					2	8	
			4	1	9			5
				8			7	9

(122)

(1) *Ověřte, že M je stochastická matice.*
(2) *Spočtěte stacionární stav Markovova řetězce.*

OBRÁZEK 5.1. **Zákaznické karty** jsou jedním z pokusů ovlivnit věrnost zákazníků.

ÚLOHA 4. *Předpokládejte čtvrť se supermarkety* **A**, **B** *a s populací 12000 nakupujících obyvatel. Předpokládejte týdenní útratu na obyvatele při rodinném nákupu 1000,-Kč na osobu a týden. Zisk z obratu předpokládejte 10%.*

- *Nakupující v supermarketu* **A** *se do něj vrátí s pravděpodobností 90% a s pravděpodobností 10% půjde příští týden raději do supermarketu* **B**.
- *Nakupující v supermarketu* **B** *se do něj vrátí s pravděpodobností 80% a s pravděpodobností 20% půjde příští týden raději do supermarketu* **A**.

V supermarketu **B** *chtějí investovat do opatření, kterým by se i u jejich zákazníků dosáhla věrnost 90%. Jaký roční zisk (ze kterého by se potom též dalo financovat ono opatření) by toto opatření přineslo. Modelujte tuto situaci pomocí Markovových řetězců.*

CVIČENÍ NA SÍTI 5. **Generování pseudotextu**.

5.3. Spojité Markovovy řetězce

Ve spojitém případě lze zavést pojmy obdobné pojmům užívaným pro konečné Markovovy řetězce a tak v předchozí sekci diskutované konečné Markovovy řetězce dávají dobrou představu o podstatě věci. Ve spojitém případě nutná technická prohloubení teorie však jdou za rámec tohoto textu (viz např. [**19**], [**17**]) a omezíme se zde proto pouze na několik poznámek.

I v spojitém případě definujeme stacionární stav Markovova řetězce:

DEFINICE 23. *Rozdělení pravděpodobnosti* π *je* **stacionárním stavem** *pro Markovův řetězec s pravděpodobností přechodu* $P(x_n \mid x_{n-1})$*, pokud platí:*

$$\pi(x_n) = \int_{x_{n-1}} P(x_n \mid x_{n-1})\pi(x_{n-1})dx_{n-1} \tag{123}$$

Ireducibilita se definuje vůči danému rozdělení pravděpodobnosti $f(x)$:

DEFINICE 24. *Markovův řetězec je* f**-ireducibilní***, pokud pro každou měřitelnou množinu* A *s kladnou mírou vůči* f*, tj.* $f(A) = \int_A f(x)dx > 0$*, platí:*

Existuje čas $n \in \mathbb{N}$*, že pro libovolný počáteční bod* $x_0 \in \Omega$ *je pravděpodobnost výskytu v* A *v čase* n *nenulová, tj.*

$$\int_A P(x_n \mid x_0)dx_n > 0 \tag{124}$$

Výše popsané skutečnosti umožňují asymptoticky generovat vzorky podle stacionárního rozdělení pravděpodobnosti Markovova řetězce. Přitom jsme však zcela opomenuli otázku, po kolika krocích jsme již naším vzorkem „dostatečně blízko“ stacionárního rozdělení. Zde se spokojíme s pragmatickou odpovědí, že musíme jít „dostatečně daleko“, čímž je myšleno, že lze v praxi dosáhnout potřebné přesnosti zpravidla i na běžném počítači. Přesnější kritéria a analýzy jsou ale pochopitelně velmi důležité a jsou předmětem specializované odborné literatury na toto téma, viz např. [**17, 16, 19**].

PŘÍKLAD 7. **Náhodná procházka.** *Náhodná procházka na* $\mathbb{Z}^n$ *resp. na* $\mathbb{R}^n$ *je Markovovým řetězcem s podmíněnou pravděpodobností*

$$P(x_n \mid x_{n-1}) = q(x_n - x_{n-1})\mu_L \tag{125}$$

s daným tvarem funkce $q(\bullet)$ *a s Lebesgueovou mírou . Tato podmíněná pravděpodobnost je automaticky invariantní vůči posunutí* $a \in \mathbb{Z}^n$*, resp.* $a \in \mathbb{R}^n$*:*

$$x_{n-1} \to x_{n-1} + a \tag{126}$$

$$x_n \to x_n + a \tag{127}$$

Podmíněnou pravděpodobnost $P(x_n \mid x_{n-1})$ *je třeba chápat jako míru v prvním argumentu, která je zadaná pomocí hustoty* $q(x_n - x_{n-1})$*. Jak diskrétní, ale nekonečný případ* $\mathbb{Z}^n$*, tak i spojitý případ* $\mathbb{R}^n$ *se opakovaně vyskytují v řadě situací.*

ÚLOHA 5. *Simulujte na počítači opakovaně náhodnou procházku v* $\mathbb{Z}$ *s podmíněnou pravděpodobností danou funkcí* q*:*

$$q(z) = \begin{cases} 1 & \text{pro } z \in \{+1, -1\}, \\ 0 & \text{jinak.} \end{cases} \tag{128}$$

Popište charakter výsledku $P(x_n \mid x_0)$ *po* n *krocích na základě získaného vzorku.*
Má tento Markovův řetězec stacionární stav daný limitou $\lim_{n\to\infty} P(x_n \mid x_0)$*?*

SHRNUTÍ 5. **Markovův řetězec** *je dynamický systém, jehož vývoj probíhá nikoliv v spojitém čase, ale v jednotlivých diskrétních krocích od náhodné veličiny* x_n *v čase* $n \in \mathbb{N}$ *k náhodné veličině* x_{n+1}*, přičemž stav po určitém kroku je podmíněn pouze stavem před daným krokem a nezávisí na žádných dalších informacích z předcházejících kroků (to je tzv.* **Markovova vlastnost***). Změna stavu v jednom kroku je proto úplně charakterizována* **pravděpodobností přechodu** $P(x_{n+1} \mid x_n)$*.*

Za vhodných podmínek má Markovův řetězec jednoznačně určený **stacionární stav***, tj. rozdělení pravděpodobnosti, které se vývojem Markovova řetězce nemění. Za vhodných podmínek je Markovův řetězec* **ergodický***, tj. jednotlivý vývoj z libovolného počátečního bodu* x_0 *skoro vždy asymptoticky probíhá (při nezohlednění pořadí hodnot, pouze jejích výskytů) hodnoty tak, jako by byly generovány ze stacionárního stavu Markovova řetězce.*

KAPITOLA 6

Metropolisův-Hastingsův algoritmus

Klíčová slova: Metropolisův-Hastingsův algoritmus, náhodné procházky, asymptotické chování.

Abstrakt: Metropolisův-Hastingsův algoritmus je obecná konstrukce ergodického Markovova řetězce se stacionárním stavem rovným cílovému rozdělení pravděpodobnosti. Podmíněná pravděpodobnost vyskytující se v konstrukci je libovolná, má však vliv na rychlost konvergence. Postup je ilustrován na jednoduchých příkladech.

6.1. Úvod

Podle předchozí kapitoly lze Markovovy řetězce využít ke generování vzorků, které budou mít asymptoticky rozdělení odpovídající stacionárnímu rozdělení daného řetězce. Je však možné najít vhodný Markovův řetězec, který by měl předepsané stacionární rozdělení s **cílovou hustotou pravděpodobnosti** $p(\theta)$? Obecné a efektivní řešení této úlohy poskytuje tzv. **Metropolisův-Hastingsův algoritmus**:

Algoritmus 4. *Nechť je dána cílová hustota pravděpodobnosti $p(\theta)$ a podmíněná hustota pravděpodobnosti $q(\theta' \mid \theta)$. Potom lze zadat Markovův řetězec následujícím postupem:*

(1) *Z m-té hodnoty $\theta^{(m)}$ lze generovat mezivýsledek ξ podle hustoty pravděpodobnosti $q(\xi \mid \theta)$.*
(2) *Novou, $(m+1)$-ní hodnotou $\theta^{(m+1)}$ bude potom buď, s pravděpodobností ρ získaná mezihodnota ξ, nebo jinak předchozí hodnota $\theta^{(m)}$:*

$$\rho = \min\left(\frac{p(\xi)q(\theta^{(m)} \mid \xi)}{p(\theta^{(m)})q(\xi \mid \theta^{(m)})}, 1\right) \tag{129}$$

$$\theta^{(m+1)} = \begin{cases} \xi & \textit{s pravděpodobností } \rho, \\ \theta^{(m)} & \textit{jinak} \end{cases} \tag{130}$$

6.2. Asymptotické chování

Že Metropolisův-Hastingsův Markovův řetězec má vhodné asymptotické vlastnosti, lze ověřit pomocí následující věty (viz [**19**], str. 274):

Věta 3. *Nechť nosič* supp f *rozdělení pravděpodobnosti f je souvislý a nechť f je omezené a kladné na každé kompaktní množině svého nosiče. Pokud existují $\epsilon, \delta > 0$ tak, že*

$$q(\theta \mid \xi) > \epsilon \qquad \textit{pro } \mid \theta - \xi \mid < \delta, \tag{131}$$

potom platí:

(1) *Markovův řetězec je f-ireducibilní.*
(2) *Pokud pro funkci $h(\theta)$ platí $\int \mid h(\theta) \mid f(\theta) d\theta < \infty$, potom*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} h(\theta^{(t)}) = \int h(\theta)\, f(\theta)\, d\theta \tag{132}$$

(3) *Pro libovolnou počáteční rozdělení pravděpodobnosti $\mu(\theta_0)$*

$$\lim_{n\to\infty} \sup_{A\in\mathcal{F}} \int_A \left(\int P(\theta_n \mid \theta_0) d\mu(\theta_0) - f(\theta_n) \right) d\theta_n = 0 \tag{133}$$

6.3. Metropolisův-Hastingsův algoritmus s náhodnými procházkami

Jednou z jednoduchých možností, jak zvolit podmíněnou hustotu pravděpodobnosti $q(\theta' \mid \theta)$ tak, aby byly splněny předpoklady Věty 3 je zadat jí předpisem

$$q(\theta' \mid \theta) = q(\mid \theta - \xi \mid), \tag{134}$$

který sám o sobě zadává náhodnou procházku. Předpokládáme, že funke q je větší než ϵ pro nějaké δ-okolí nuly. Volí se tak, aby se podle ní daly snadno generovat vzorky. Příkladem je normální rozdělení $q(x) = N(x \mid 0, \delta)$ nebo rovnoměrné rozdělení $q(x) = U_{B_{0,\delta}(x)}$ na kouli $B_{0,\delta}$ s poloměrem δ v počátku. Symetrie (134) vůči záměně θ a ξ poněkud zjednoduší zápis Metropolisova-Hastingsova algoritmu:

ALGORITMUS 5. *Nechť je dána cílová hustota pravděpodobnosti $p(\theta)$ a podmíněná hustota pravděpodobnosti $q(\theta' \mid \theta)$. Potom lze zadat Markovův řetězec následujícím postupem:*

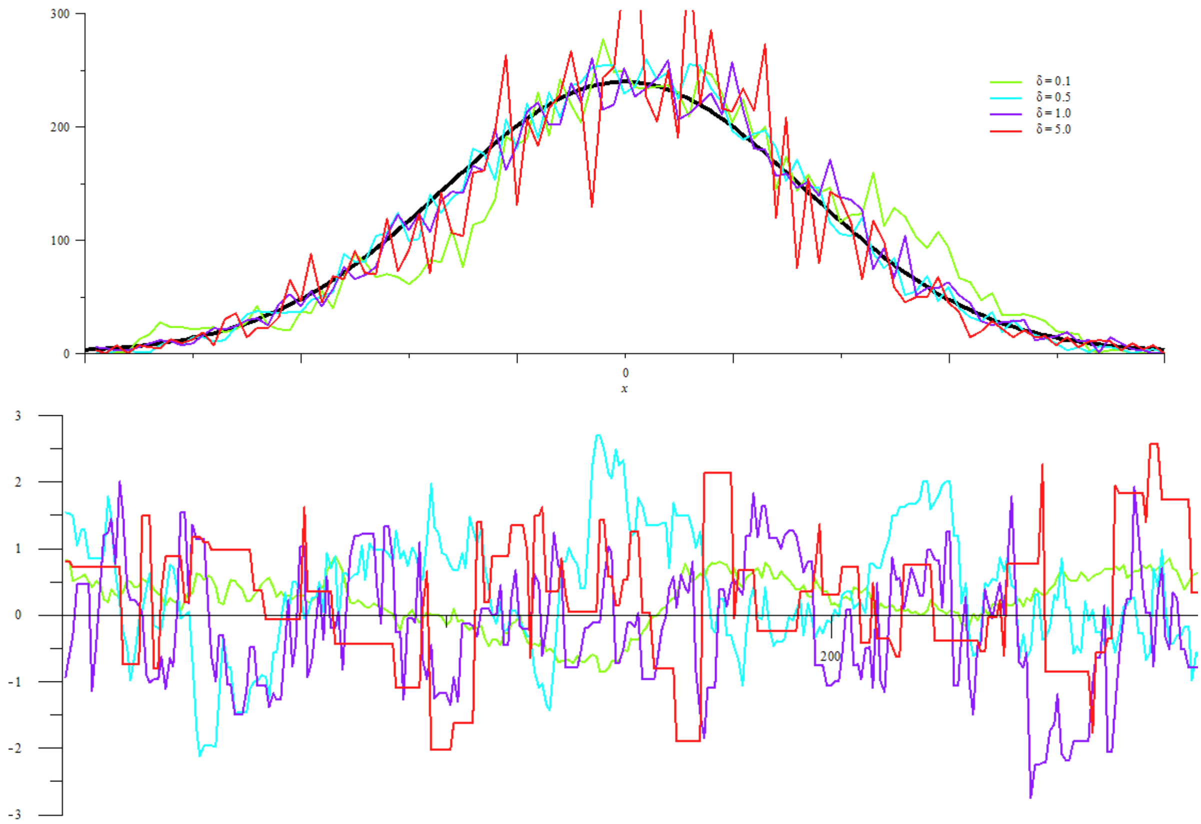
(1) *Z m-té hodnoty $\theta^{(m)}$ lze generovat mezivýsledek ξ podle hustoty pravděpodobnosti $q(\mid \theta - \xi \mid)$.*
(2) *Novou, $(m+1)$-ní hodnotou $\theta^{(m+1)}$ bude potom buď, s pravděpodobností ρ získaná mezihodnota ξ, nebo jinak předchozí hodnota $\theta^{(m)}$:*

$$\rho = \min\left(\frac{p(\xi)}{p(\theta^{(m)}))}, 1 \right) \tag{135}$$

$$\theta^{(m+1)} = \begin{cases} \xi & \textit{s pravděpodobností } \rho, \\ \theta^{(m)} & \textit{jinak} \end{cases} \tag{136}$$

Následující dva příklady mají názorně předvést fungování Metropolisova-Hastingsova algoritmu. Jedná se o 1- nebo 2-rozměrné případy, kdy by bylo patrně účinnější použít např. některou z variant zamítací metody. Je to tak proto, že se třicetirozměrné grafy špatně kreslí. Tak v těchto příkladech uvidíme princip algoritmu, nikoliv však dalekosáhlé výhody metod založených na Markovových řetězcích.

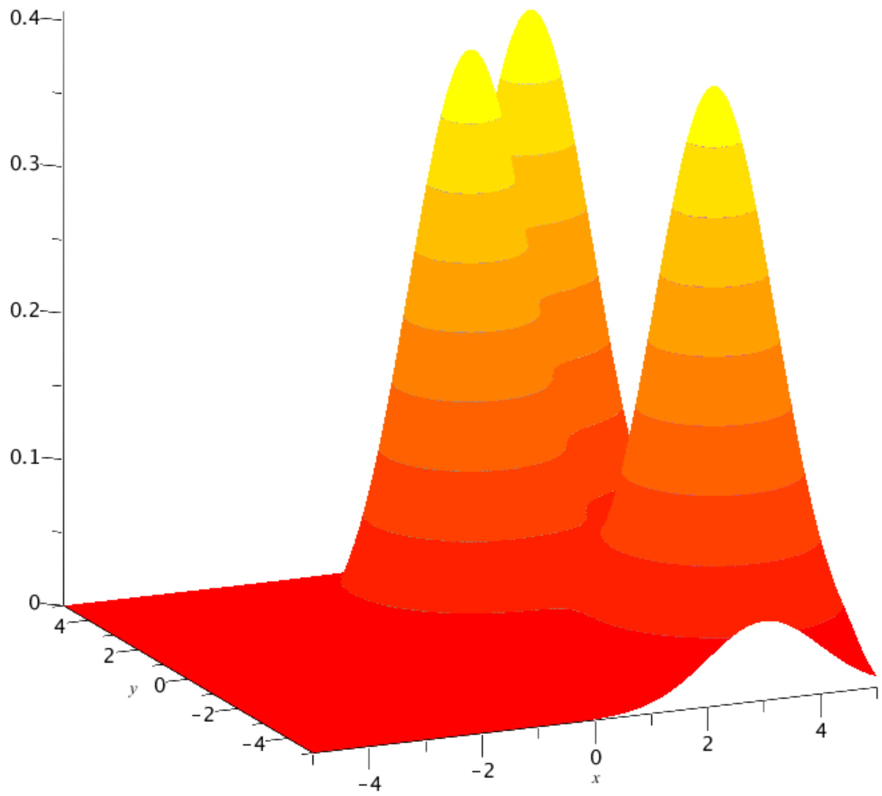
Příklad 8. **(Normální rozdělení na $\mathbb{R}$).**



Obrázek 6.1. **Normální rozdělení na $\mathbb{R}$.** Cílovým rozdělením pravděpodobnosti je normální rozdělení $N(\theta \mid 0, 1)$. Vývoj Markovova řetězce dle Metropolisova-Hastingsova algoritmu s náhodnou procházkou: $q(x) = N(x \mid 0, \delta)$ s $\delta = 0.1$, $\delta = 0.5$, $\delta = 1$, $\delta = 5$, 10000 kroků. Horní graf má zkoumaný interval $[-3, 3]$ rozdělený do 100 stejně velkých podintervalů a pro každý podinterval je vynesen počet bodů v tomto intervalu. Spodní graf ukazuje detail prvních stovek generovaných bodů pro každou hodnotu δ

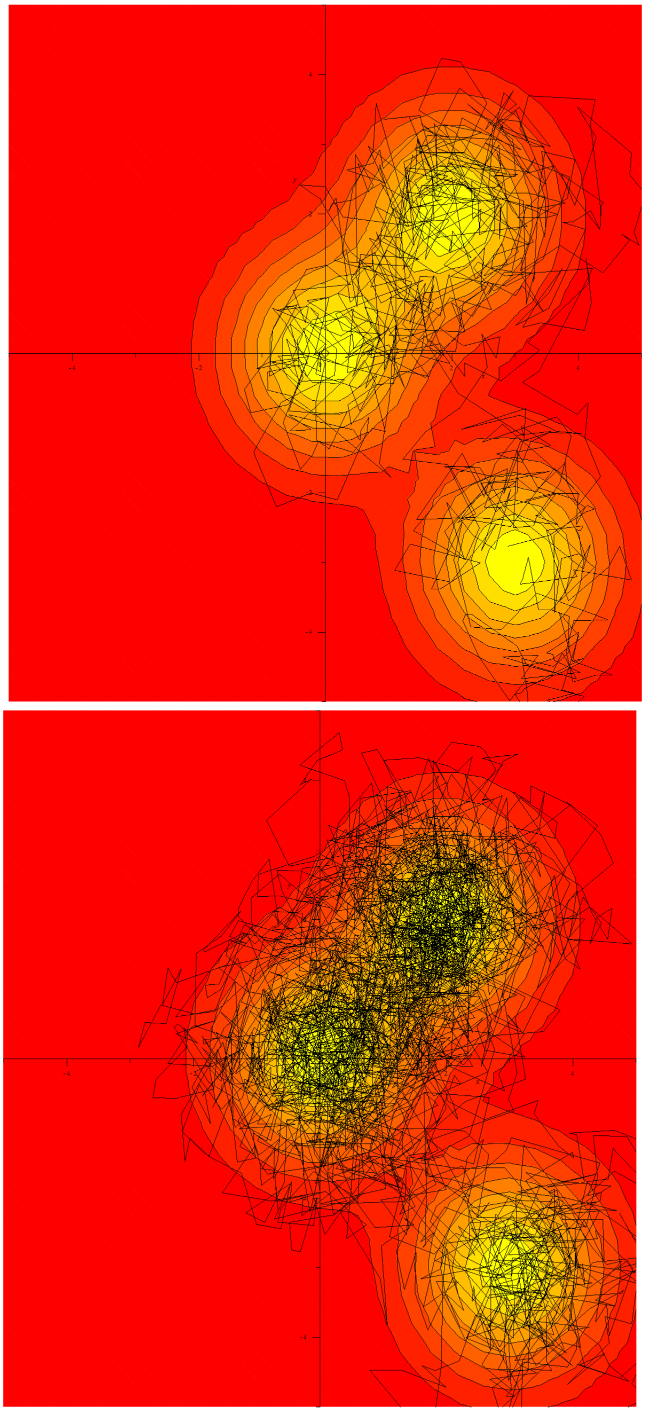
Kontrolní otázka 6.1. *První graf Příkladu 8, srovnávající rozvržení vzorků z Markovova řetězce ukazuje, že hodnoty vystihují původně zadané rozdělení (černá křivka) pro všechny volby δ. Je však patrné, že nejmenší a největší zvolená hodnota vedou k poněkud horším výsledkům. Vysvětlete tyto důvody.*

PŘÍKLAD 9. **(Směs normálních rozdělení na $\mathbb{R}^2$).**

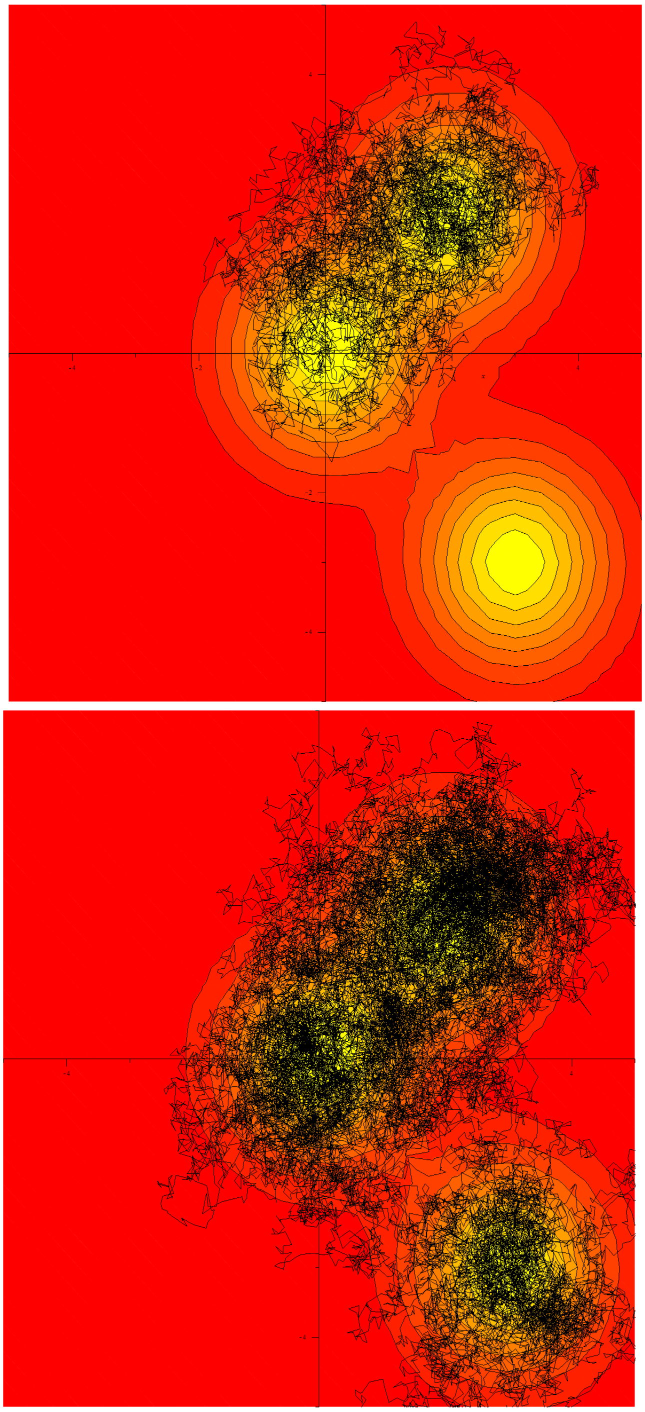


OBRÁZEK 6.2. **Směs normálních rozdělení na** $\mathbb{R}^2$. Cílové rozdělení pravděpodobnosti je rovnoměrnou směsí tří normálních rozdělení $N(\theta \mid [0,0],1)$, $N(\theta \mid [2,2],1)$, $N(\theta \mid [3,-3],1)$.

Následné simulace postupují dle Algoritmu 5 pro dvě různé volby hustoty pravděpodobnosti $q(\mid \theta - \xi \mid)$ (Obě jsou ve stejném tvaru $q(x) = N(x \mid 0, \delta)$, ovšem s jiným parametrem, $\delta = 0.5$ a $\delta = 0.1$.). Zatímco kterákoliv z voleb asymptoticky zaručuje vzorek vystihující cílové rozdělení, a po teoretické stránce na konkrétní volbě q zdánlivě nezáleží, má konkrétní volba rozhodující vliv na to, jak rychle (po jak mnoha krocích) simulace poskytne vzorek uspokojivě vystihující podstatné rysy cílového rozdělení pravděpodobnosti $p(\theta)$.



Obrázek 6.3. **Směs normálních rozdělení na** $\mathbb{R}^2$. Cílové rozdělení pravděpodobnosti je rovnoměrnou směsí tří normálních rozdělení $N(\theta \mid [0,0],1)$, $N(\theta \mid [2,2],1)$, $N(\theta \mid [3,-3],1)$. Vývoj Markovova řetězce dle Metropolisova-Hastingsova algoritmu s náhodnou procházkou: $q(x) = N(x \mid 0, \delta)$ s $\delta = 0.5$, 1000 a 4000 kroků. Při 4000 krocích vzorek rozumně pokrývá nejpodstatnější část rozdělení pravděpodobnosti.



OBRÁZEK 6.4. **Směs normálních rozdělení na** $\mathbb{R}^2$. Cílové rozdělení pravděpodobnosti je rovnoměrnou směsí tří normálních rozdělení $N(\theta \mid [0,0],1)$, $N(\theta \mid [2,2],1)$, $N(\theta \mid [3,-3],1)$. Vývoj Markovova řetězce dle Metropolisova-Hastingsova algoritmu s náhodnou procházkou: $q(x) = N(x \mid 0,\delta)$ s $\delta = 0.1$, 10000 a 40000 kroků. Počty kroků jsou oproti předchozímu obrázku desetinásobné. Přesto vzorek při 10000 krocích nevystihuje podstatnou část, tj normální rozdělení $N(\theta \mid [3,-3],1)$, ze směsi normálních rozdělení.

CVIČENÍ NA SÍTI 6. **2D Markovův řetězec daný Metropolisovým-Hastingsovým algoritmem**.

SHRNUTÍ 6. **Metropolisův-Hastingsův algoritmus**: *Nechť je dána cílová hustota pravděpodobnosti $p(\theta)$ a podmíněná hustota pravděpodobnosti $q(\theta' \mid \theta)$. Potom lze zadat Markovův řetězec následujícím postupem:*

(1) *Z m-té hodnoty $\theta^{(m)}$ lze generovat mezivýsledek ξ podle hustoty pravděpodobnosti $q(\xi \mid \theta)$.*

(2) *Novou, $(m+1)$-ní hodnotou $\theta^{(m+1)}$ bude potom buď, s pravděpodobností ρ získaná mezihodnota ξ, nebo jinak předchozí hodnota $\theta^{(m)}$:*

$$\rho = \min\left(\frac{p(\xi)q(\theta^{(m)} \mid \xi)}{p(\theta^{(m)})q(\xi \mid \theta^{(m)})}, 1\right) \tag{137}$$

$$\theta^{(m+1)} = \begin{cases} \xi & \text{s pravděpodobností } \rho, \\ \theta^{(m)} & \text{jinak} \end{cases} \tag{138}$$

Při vhodném použití dává tento algoritmus ergodický Markovův řetězec se stacionárním stavem rovným cílovému rozdělení pravděpodobnosti $p(x)$.

KAPITOLA 7

Statistický model a jeho aktualizace. Oblast spolehlivosti.

Klíčová slova: Statistický model, inverze a aktualizace statistického modelu, α-uvěřitelná oblast spolehlivosti.

Abstrakt: Předpoklady o rozdělení pravděpodobnosti lze vyjádřit statistickým modelem. Parametry modelu, předurčeny apriorními rozděleními pravděpodobnosti, jsou aktualizovány na základě získaných vzorků. Oblast očekávaného výskytu parametru modelu lze znázornit oblastí spolehlivosti.

7.1. Úvod

Pravděpodobnost zavedená v Kapitole 1 charakterizuje subjektivní názor na to, zda se určitý jev stane. Takto chápaná pravděpodobnost se jeví jako dosti libovolná. V této kapitole tuto libovůli omezíme dvěma způsoby:

(1) Pro uvažovaná rozdělení pravděpodobnosti navrhneme pevně určený tvar, ve kterém je libovůle již jen v několika (konečně mnoha) **parametrech**. Hodnoty parametrů nefixujeme, ale vyjádříme názor na jejích hodnotu pomocí tzv. **apriorního rozdělení pravděpodobnosti**. Tyto údaje tvoří **statistický model**. Pro jeho konkrétní podobu můžeme mít různé důvody - pokud nejsou dost pádné, vznikají a musí se prošetřit pochybnosti o tom, zda stanovený model odpovídá situaci. Na druhou stranu umožní přijetí modelu omezit náročnost problému a zachytit základní rysy situace pomocí několika málo konstant. Statistický problém tak může stále vyžadovat rozsáhlé numerické výpočty, ale stává se schůdnějším.

(2) Libovůle v rozdělení pravděpodobnosti byla v Kapitole 1 omezena pouze doporučeními Bayesovské statistiky. Ty se zdají být dosti slabým omezením. Situace se ovšem podstatně změní, pokud spolu se zjednodušením na základě volby statistického modelu získáme o situaci pozorovaná **data** (vzorky). V tom případě se dají pomocí Bayesova vzorce (212) na základě získaných dat upravit (aktualizovat) dohady o hodnotách parametrů modelu. To má za následek, že při velmi odlišných počátečních předpokladech jednotlivých subjektů o hodnotách parametrů se pod tíhou pozorování jejích předpoklady budou sbližovat. Tím celá teorie ztrácí na libovolnosti a získává na objektivitě, a to právě do té míry, do které to umožňuje kvalita dostupných dat.

Na základě dostupných dat aktualizované rozdělení pravděpodobnosti pro parametry statistického modelu, tzv. **aposteriorní rozdělení pravděpodobnosti** představuje úplný výsledek statistického usuzování v Bayesovském rámci. Do jisté míry lze rysy aposteriorního rozdělení pravděpodobnosti názorně zachytit pomocí tzv. **oblasti spolehlivosti**. Ta je též určitým protějškem pojmu „intervalu spolehlivosti“ ve frekventistické statistice, kde hraje prominentní roli. Přes všechnu názornost je však třeba vidět, že se v použití oblasti spolehlivosti ztrácí část informace obsažené v aposteriorním rozdělení pravděpodobnosti a že by se jako skutečný výsledek Bayesovské analýzy mělo vždy žádat aposteriorní rozdělení pravděpodobnosti (nebo jeho důsledky v teorii rozhodování, jak budeme diskutovat v Kapitole 9).

V Sekci 7.2 je zaveden statistický model. Aktualizace statistického modelu je v obecnosti i v konkrétním případě názorné analýzy hodů mince diskutována v Sekci 7.3. V Sekci 7.4 je potom zavedena a ilustrována oblast spolehlivosti.

7.2. Rozdělení pravděpodobnosti s parametry

Často se setkáme s případem, že rozdělení pravděpodobnosti sice neznáme, ale předpokládáme v určitém tvaru $p(x \mid \theta)$, kde θ je neznámý parametr nebo soubor parametrů.

PŘÍKLAD 10. *Normální rozdělen je dáno dvojicí parametrů $\theta = (\mu, \sigma)$:*

$$p(x \mid \theta) = N(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{139}$$

PŘÍKLAD 11. *Bernoulliho rozdělení pravděpodobnosti na náhodné veličině x, která může nabývat pouze hodnot z dvouprvkové množiny $\{rub, líc\}$ a modeluje hod mincí, má jeden parametr $\theta \in [0, 1]$:*

$$p(rub) = \theta \qquad\qquad p(líc) = 1 - \theta \tag{140}$$

Pokud by byla mince spravedlivá, platilo by $\theta = p(rub) = p(líc) = \frac{1}{2}$.

Pokud tedy parametr θ přesně neznáme, nejsme si jisti jeho hodnotou, přísluší mu rozdělení pravděpodobnosti $p(\theta)$, $\theta \in \Theta$, kde Θ je prostor všech možných hodnot parametru θ. O tvaru rozdělení pravděpodobnosti $p(\theta)$ se můžeme různě dohadovat. Jeho subjektivní volbu, učiněnou s nejlepším svědomím označujeme $\pi(\theta)$ a nazýváme **apriorním rozdělením pravděpodobnosti** pro parametr $\theta \in \Theta$. Potom máme dle Bayesova vzorce (212):

$$p(x, \theta) = p(x \mid \theta)\pi(\theta) \tag{141}$$

a rozdělení pravděpodobnosti $p(x)$ je potom

$$p(x) = \int_{\theta\in\Theta} p(x, \theta) d\theta = \int_{\theta\in\Theta} p(x \mid \theta)\pi(\theta) d\theta \tag{142}$$

DEFINICE 25. **Statistickým modelem** *pro náhodnou veličinu x je soubor uvažovaných rozdělení $\mathcal{P} = \{p_I(x)\}$ spolu s apriorním rozdělením pravděpodobnosti $p(I)$ na prostoru uvažovaných rozdělení pravděpodobnosti.*

Pokud množina $\mathcal{P}$ je parametrizována konečně-rozměrným parametrem θ, tj. $\mathcal{P} = \{p(x \mid \theta) \mid \theta \in \mathbb{R}^n\}$ s apriorním rozdělením $p(\theta)$, nazývá se **parametrickým statistickým modelem**.

Situace popsaná vzorci (141), (142) zadává jeden z nejjednodušších a nejběžnějších druhů parametrického statistického modelu. Zadání složitějších modelů pomocí tzv. Bayesovských sítí bude věnována další kapitola.

7.3. Aktualizace apriorního rozdělení pravdopodobnosti

Předpokládejme, že máme k dispozici naše ohodnocení nejistoty situace ohledně náhodné veličiny x a parametru θ podle (141) a dozvíme se, že při uskutečnění pokusu nastala hodnota $x = x_1$. Potom ovšem musíme v světle této informace upravit určené pravděpodobnosti parametru θ. Opět podle Bayesova vzorce máme

$$p(\theta \mid x_1) = \frac{p(x_1, \theta)}{p(x_1)} = \frac{p(x_1 \mid \theta)\pi(\theta)}{\int_{\theta\in\Theta} p(x_1 \mid \theta)\pi(\theta)d\theta} \tag{143}$$

Rozdělení pravděpodobnosti $p(\theta \mid x_1)$ se nazývá aposteriorním rozdělením pravděpodobnosti pro θ a liší se od apriorního rozdělení $\pi(\theta)$ tím, že bere jediným způsobem slučitelným s doporučeními Bayesovské statistiky do úvahy informaci o hodnotě x_1, která při stanovení apriorního rozdělení nebyla známa. Došlo tedy k **aktualizaci** rozdělení pravděpodobnosti pro parametr θ. Od toho okamžiku bychom tedy namísto $\pi(\theta)$ měli použít $p(\theta \mid x_1)$.

Předpokládejme, že nyní dojde k dalšímu pokusu s výsledkem $x = x_1$. Musíme tedy znovu, stejným způsobem aktualizovat rozdělení pravděpodobnosti pro parametr θ z $p(\theta \mid x_1)$ na výsledné $p(\theta \mid x_1, x_2)$.

Po plném zvážení n výsledků $x_1, x_2, \dots, x_n$ nezávislých pokusů je aktualizované rozdělení nakonec

$$p(\theta \mid x_1, x_2, \dots, x_n) = \frac{p(x_n \mid \theta)p(\theta \mid x_1, x_2, \dots, x_{n-1})}{p(x_n \mid x_1, x_2, \dots, x_{n-1})} = \tag{144}$$

$$= \frac{p(x_n \mid \theta)\dots p(x_2 \mid \theta)p(x_1 \mid \theta)\pi(\theta)}{p(x_n \mid x_1, x_2, \dots, x_{n-1})\dots p(x_2 \mid x_1)p(x_1)}. \tag{145}$$

OBRÁZEK 7.1. **Jablka.** Rozdělení pravděpodobnosti jejich průměru lze přibližně vyjádřit jako normální. Normální rozdělení je definováno na celé reálné ose, zatímco záporné průměry jablek nemají valného smyslu. Pokud pravděpodobnost záporného průměru je dle použitého normálního rozdělení mizivá, prakticky nulová, nepovede tato okolnost k vážnějším problémům.

Kontrolní otázka 7.1. *Jablka dané odrůdy mají průměr v centimetrech idealizovaně daný normálním rozdělením $N(x \mid \mu, 1)$ s apriorním rozdělením pravděpodobnosti pro parametr μ též daným normálním rozdělením:*

$$\pi(\mu) = N(\mu \mid 10, 3) \tag{146}$$

Jaké je aposteriorní rozdělení pravděpodobnosti pro střední průměr jablek v úrodě, pokud ve vzorku tří náhodně vybraných jablek jsou jejích průměry 10.5, 11.2 a 12.1?

Úloha 6. *Při hodu 1 Kč padne buď koruna nebo lev. Tendenci, že padne koruna, budeme označovat k, přičemž $k \in [0, 1]$ a $P(koruna \mid k) = k$. Předpokládejte rovnoměrné apriorní rozdělení pro tendenci k, tj. $p(k) = 1$. Proveďte několik hodů mincí a jim odpovídající Bayesovské aktualizace. Popište slovně výsledné, aposteriorní rozdělení pravděpodobnosti pro tendenci k.*

Obrázek 7.2. **Hod mince.** Přímý hod mince lze manipulovat jen obtížně. Pokud se ovšem mince namísto hodu roztočí svisle, lze jednu ze stran preferovat nenápadným opracováním hrany mince.

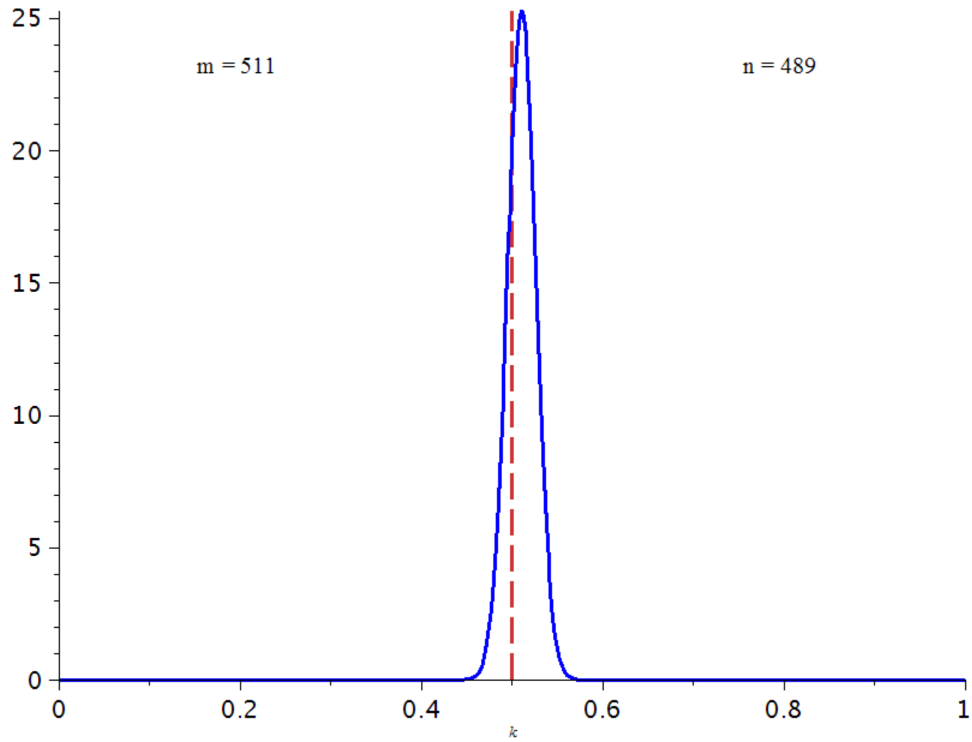
Řešení. Podle (144)-(145) je aktualizované, aposteriorní rozdělení pro parametr k po m hodech koruny a n hodech lva je

$$p(k \mid m, n) = \frac{k^m(1-k)^n}{\int_0^1 k^m(1-k)^n dk} = \frac{(m+n+1)!}{m!\,n!} k^m(1-k)^n, \tag{147}$$

protože

$$\int_0^1 k^m(1-k)^n dk = \frac{m!\,n!}{(m+n)!}\frac{1}{m+n+1}. \tag{148}$$

Podle vzorce na pořadí, ve kterém jsou jednotlivé výsledky hozeny, nezáleží. To je pochopitelné, protože se jedná o oddělené, nezávislé pokusy. Možné výsledky jsou znázorněny v Obrázku 7.3. Ukazuje se, že zvyšujícím se počtem hodů se rozdělení stává více lokalizované, až pro pravděpodobnou hodnotu parametru k nenechává příliš volnosti okolo hodnoty $\frac{m}{m+n}$. To odpovídá intuitivní představě, že jsme si po opakování většího počtu pokusů jistější, jaká je hodnota parametru k.

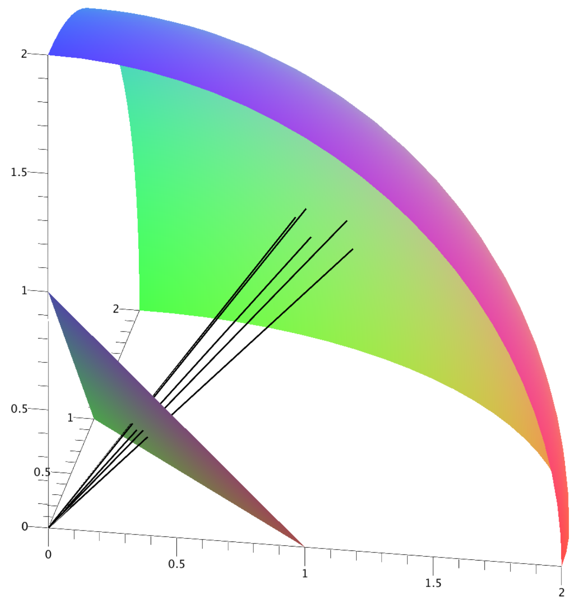


Obrázek 7.3. **Aposteriorní rozdělení pravděpodobnosti pro parametr k po 1000 hodech mincí.**

Poznámka 17. Mohlo by se zdát, že v Úloze 6 uvažované apriorní rozdělení pravděpodobnosti $p(k) = 1$ parametru k vyjadřuje naprostou neznalost parametru k, neboť je rovnoměrné v k. To je však omyl, protože parametrizace Bernoulliho rozdělení je svým způsobem nejednoznačná. Řešením tohoto problému je tzv. **Jeffreyovo neinformativní apriorní rozdělení** [**2**]. Rovnoměrnost tohoto rozdělení se neopírá o konkretní parametrizaci, ale o geometrií prostoru všech rozdělení pravděpodobnosti. V plné obecnosti jde tento pojem za rámec tohoto textu, Pro případ rozdělení pravděpodobnosti na konečném počtu n možností však lze názorně vyjádřit výsledek takovýchto úvah: Rozdělení pravděpodobností na n možnostech je dáno n pravděpodobnostmi $P_1, P_2, \ldots, P_n$ splňujícím

$$0 \leq P_i \leq 1 \qquad \text{pro } i \in \{1, 2, \ldots, n\}, \qquad \sum_{i=1}^{n} P_i = 1. \tag{149}$$

Každé rozdělení pravděpodobnosti lze tedy geometricky chápat jako bod v standardním simplexu v $\mathbb{R}^n$. Jeffreyovo neinformativní apriorní rozdělení je potom dáno tak, že není rovnoměrné vůči translacím v části roviny tvořenou daným simplexem, ale je rovnoměrné, pokud by se promítlo na část S^{n-1} sféry v $\mathbb{R}^n$ (viz Obrázek 7.4).



OBRÁZEK 7.4. **Jeffreyovo neinformativní apriorní rozdělení pravděpodobnosti na rozděleních pravděpodobnosti na třech možnostech.** Rozdělení pravděpodobnosti se třemi složkami $[P_1, P_2, P_3]$ leží v standardním symplexu (trojúhelníku). Jeffreyovo neinformativní apriorní rozdělení na tomto simplexu je rovnoměrné, pokud se promítne na sféru se středem v počátku, vůči rotacím sféry.

CVIČENÍ NA SÍTI 7. **Bayesovská aktualizace: Jeden ze tří**.

7.4. Oblast spolehlivosti

Aktualizované, aposteriorní rozdělení pravděpodobnosti $p(\theta \mid x)$ obsahuje veškerou informaci, kterou o hodnotě parametru θ na základě pozorování x máme k dispozici a považuje se za úplný a (bez získání dalších faktů či vazeb na teorii rozhodování - viz Kapitola 9) nevylepšitelný výsledek Bayesovského usuzování.

Někdy se však pro účely znázornění hodí vyjádřit úsudek o hodnotě parametru $\theta \in \Theta$ se ztrátou části dostupné informace, pomocí oblasti $R \subset \Theta$, která určuje pravděpodobný výskyt parametru θ:

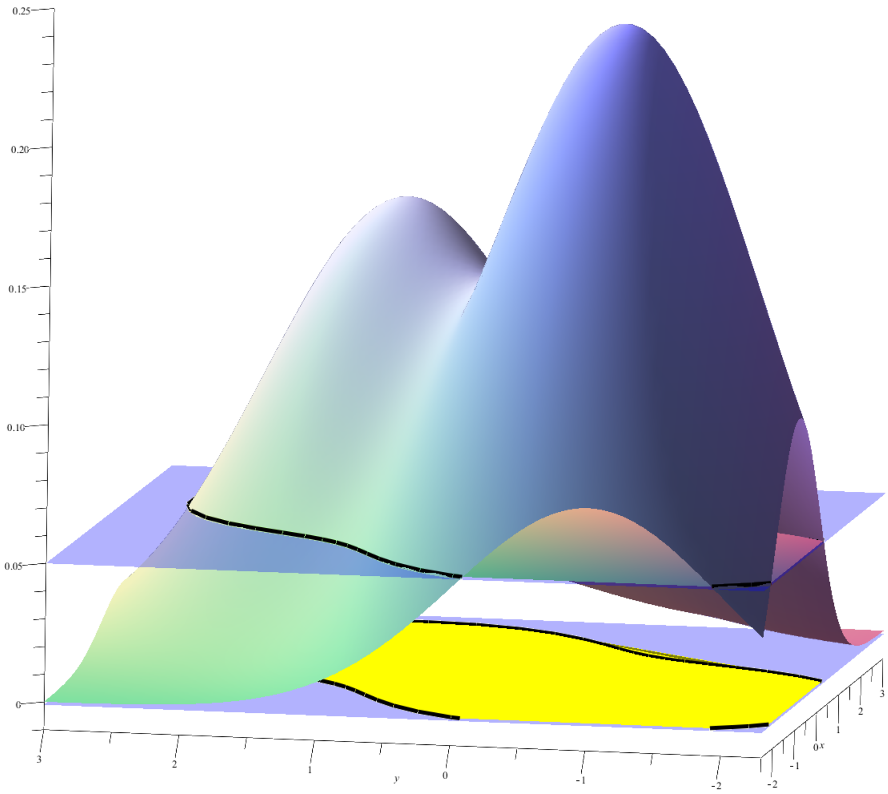
DEFINITION 1. *Oblast $R \subset \Theta$ je α-***uvěřitelnou oblastí spolehlivosti** *v prostoru Θ parametrů $\theta \in \Theta$, pokud platí*

$$P(R \mid x) \geq 1 - \alpha. \tag{150}$$

*Tato oblast R je α-***uvěřitelnou oblastí spolehlivosti pro nejvyšší aposteriorní hustotu***, pokud jí lze napsat pomocí oblastí $S(r)$ následovně:*

$$S(r) = \{\theta \in \Theta \mid p(\theta \mid x) \geq r\} \qquad \textit{pro } S(r) \subset \Theta \textit{ a } r \in [0,1], \tag{151}$$

$$R = \bigcap_{P(S(r)) \geq 1-\alpha} S(r). \tag{152}$$



Obrázek 7.5. α-**uvěřitelná oblast spolehlivosti pro nejvyšší aposteriorní hustotu**. Oblast R (žlutě) pod grafem hustoty pravděpodobnosti je určena tím, že na ní hustota pravděpodobnosti přesahuje hraniční hodnotu znázorněnou vodorovnou rovinou, a to tak, že R je nejmenší možná, pro kterou celková pravděpodobnost neklesne pod $1-\alpha$. Pravděpodobnost ocitnutí se mimo oblast R je menší než α.

Úloha 7. *Předpokládejme rozdělení $N(x^1, x^2 \mid \mu^1, \mu^2)$ pravděpodobnosti na $\mathbb{R}^2$ (Pro složky v $\mathbb{R}^2$ používáme horní indexy, které by se neměly omylem zaměnit za exponenty) dané jako součin normálních rozdělení (139) s tím, že parametr σ je známý, s pevně fixovanou hodnotou $\sigma = 1$ pro oba činitele:*

$$N(x^1, x^2 \mid \mu^1, \mu^2) = N(x^1 \mid \mu^1, 1) N(x^2 \mid \mu^2, 1) \tag{153}$$

předpokládejme dále, že apriorní rozdělení pravděpodobnosti $\pi(\theta)$ pro 2-rozměrný parametr $\theta = (\mu^1, \mu^2)$ je dáno jako rovnoměrné rozdělení a množině $[-1/2, +1/2] \times [-1/2, +1/2]$:

$$p(\theta) \equiv p((\mu^1, \mu^2)) \qquad = \begin{cases} 1 & \text{pro } (\mu^1, \mu^2) \in [-1/2, +1/2] \times [-1/2, +1/2], \\ 0 & \text{jinak.} \end{cases} \tag{154}$$

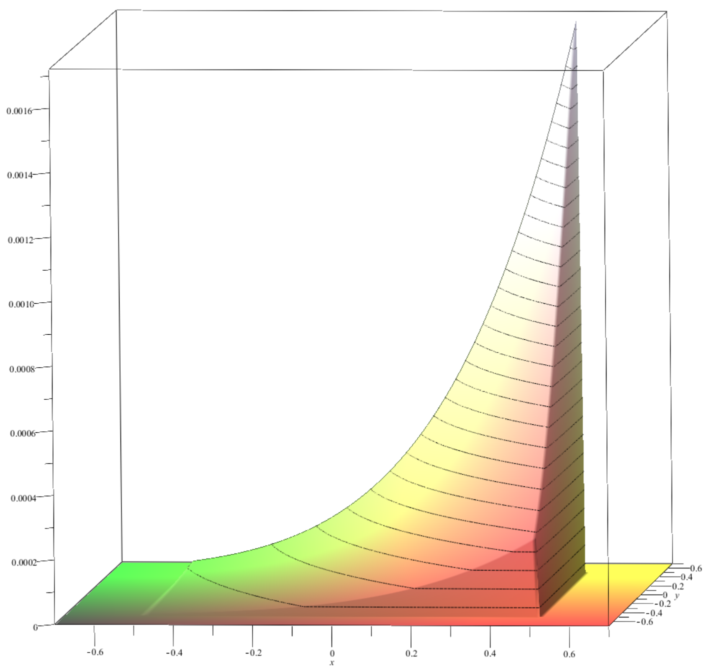
Za předpokladu, že byly získány dva datové body $x_1 = [1, 0]$ a $x_2 = [3, 2]$, určete 5%-uvěřitelnou oblast spolehlivosti parametru $\theta = (\mu^1, \mu^2) \in \mathbb{R}^2$, v případě nutnosti numerickou aproximací.

Řešení: Podle (144)-(145) je aktualizované, aposteriorní rozdělení pro parametr $\theta = (\mu^1, \mu^2)$ dáno jako

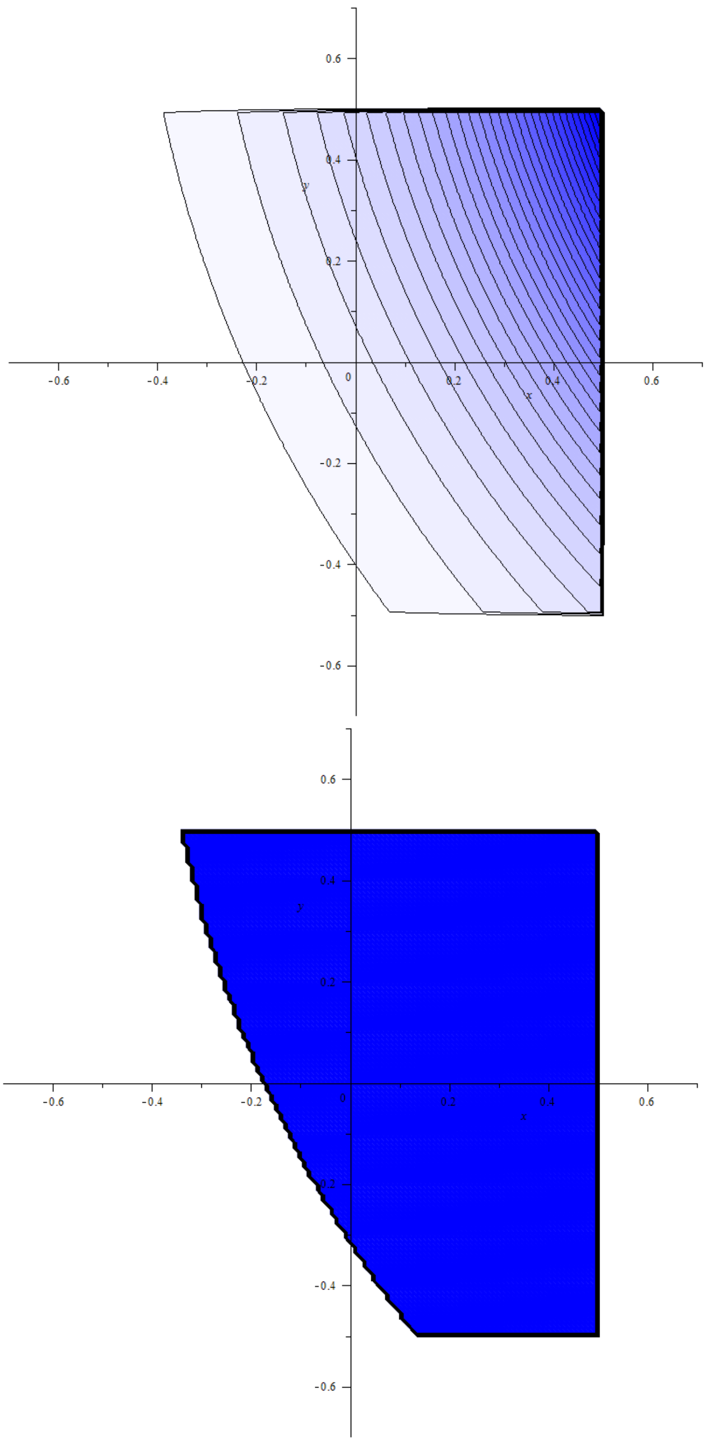
$$p(\theta \mid x_1, x_2) = \frac{N(x_1^1, x_1^2 \mid \mu^1, \mu^2) N(x_2^1, x_2^2 \mid \mu^1, \mu^2) \pi(\mu^1, \mu^2)}{I}, \tag{155}$$

kde

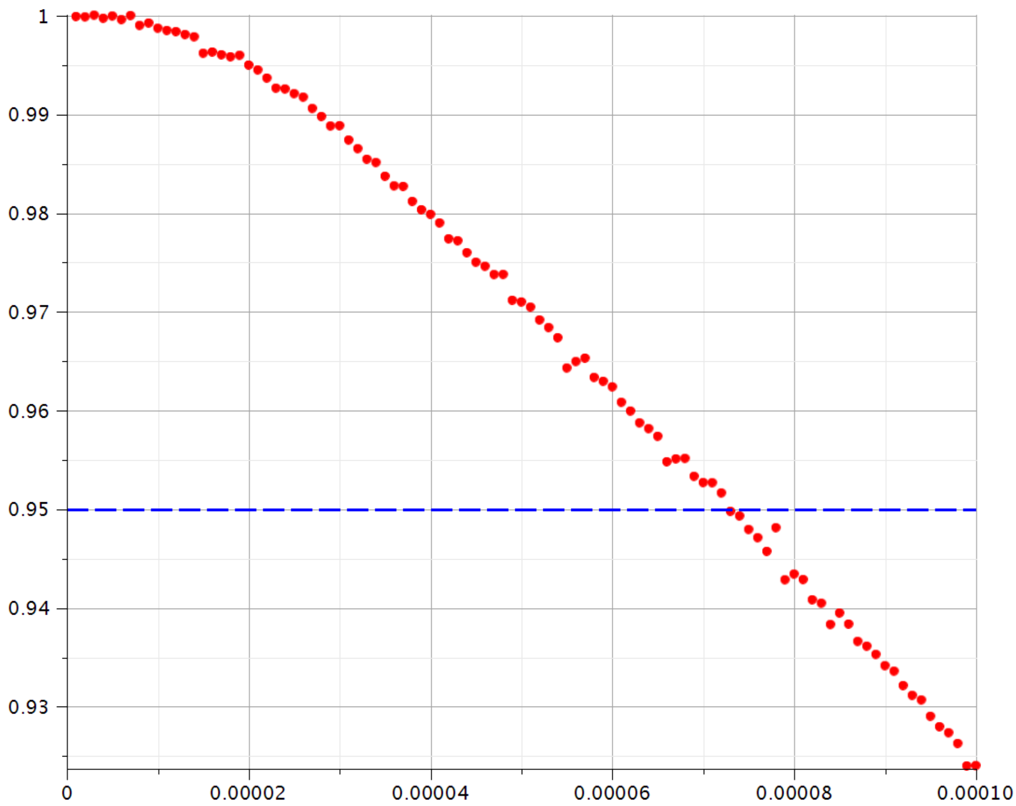
$$I = \int_{-1/2}^{+1/2} \int_{-1/2}^{+1/2} N(x_1^1, x_1^2 \mid \mu^1, \mu^2) N(x_2^1, x_2^2 \mid \mu^1, \mu^2) d\mu^1 d\mu^2 \tag{156}$$



Obrázek 7.6. **Aposteriorní hustota pravděpodobnosti.** Aposteriorní hustota pravděpodobnosti je nulová v místech, kde apriorní rozdělení pravděpodobnosti je nulové, tj. vně čtverce $[-1/2, +1/2] \times [-1/2, +1/2]$.



Obrázek 7.7. 5%**-uvěřitelná oblast spolehlivosti.** Na prvním obrázku jsou vykresleny vrstevnice hustoty pravděpodobnosti. Na druhém obrázku je zvolena vrstevnice tak (viz Obrázek 7.8), aby oblast uvnitř byla právě 5%-uvěřitelnou oblastí spolehlivosti.



OBRÁZEK 7.8. 5%-**uvěřitelná oblast spolehlivosti.** Volba správné vrstevnice v Obrázku je provedena integrací přes oblasti určené vrstevnicemi. V tomto grafu je pak vynesena pravděpodobnostní míra oblasti proti hodnotě vrstevnice. Příhodná hodnota příslušné vrstevnice je potom přibližně 0.0007. Graf není zcela hladký z důvodu použití numerické integrace metodou Monte Carlo, která tuto integraci zvládne. Podrobněji je tato metoda diskutována v Kapitole 10.

POZNÁMKA 18. Volba apriorního rozdělení pravděpodobnosti (154) má za následek, že jakékoliv aposteriorní rozdělení pravděpodobnosti bude nulové v místech, kde apriorní rozdělení pravděpodobnosti je nulové, tj. vně čtverce $[-1/2,+1/2]\times[-1/2,+1/2]$ (viz též Obrázek 7.6). Pokud máme nezpochybnitelné důvody, že mimo čtverec $[-1/2,+1/2]\times[-1/2,+1/2]$ hodnoty parametrů nesmí ležet, je to v pořádku. Pokud však ty důvody nejsou nezpochybnitelné, měli bychom se vyhnout tomu nastavit apriorní rozdělení pravděpodobnosti kdekoliv jako přesně nulové a raději mu přiřadit kladnou, byť nepatrnou hodnotu. To umožní, aby v případě našeho omylu naměřená data měla šanci při drtivé váze důkazů přebít naši domněnku, že v dané oblasti parametry nebudou.

SHRNUTÍ 7. **Statistickým modelem** *pro náhodnou veličinu* x *je rozdělení pravděpodobnosti* $p(x \mid \theta)$ *s konečně-rozměrným parametrem* θ *spolu s* **apriorním rozdělením pravděpodobnosti** $p(\theta)$*, které obsahuje subjektivní úvahy o tom, jaké hodnoty parametru* θ *lze čekat.*

Rozdělení pravděpodobnosti pro hodnotu parametru θ *je třeba* **aktualizovat** *podle zjištěných vzorků* $x_1, x_2, \ldots, x_n$*, čímž obdržíme* **aposteriorní rozdělení pravděpodobnosti** *pro parametr* θ*:*

$$p(\theta \mid x_1, x_2, \ldots, x_n) = \frac{p(x_n \mid \theta) \ldots p(x_2 \mid \theta) p(x_1 \mid \theta) p(\theta)}{p(x_n \mid x_1, x_2, \ldots, x_{n-1}) \ldots p(x_2 \mid x_1) p(x_1)}. \tag{157}$$

Typicky se rozdělení pravděpodobnosti $p(\theta \mid x_1, x_2, \ldots, x_n)$*, které se pro různé lidi liší v důsledku jejích různé volby apriorního rozdělení* $p(\theta)$*, začnou přibližovat vlivem aktualizací na základě dat* $x_1, x_2, \ldots, x_n$*, které sdílejí. Je to exaktní vyjádření toho, že fakta nutí lidi s různými počátečními názory svá stanoviska sbližovat, byť stále nejsou identicky stejná (pokud ovšem dodržují doporučení Bayesovské statistiky).*

Zásadním výsledkem ohledně hodnoty parametru θ*, vedle kterého není z hlediska Bayesovské statistiky nic dalšího třeba, je aposteriorní rozdělení pravděpodobnosti pro tento parametr. Pro názornost ho lze se ztrátou informace vystihnout pomocí tzv.* **oblasti spolehlivost***, což je zvolené okolí s mírou dostatečně blízkou k hodnotě* 1 *(k jistotě). Mimo snahu o názornost by se však vždy jako výsledek mělo považovat aposteriorní rozdělení pravděpodobnosti.*

KAPITOLA 8

Bayesův vzorec. Bayesovské sítě

Klíčová slova: Orientovaný acyklický graf, Bayesovská síť.

Abstrakt: Statistické souvislosti mezi částečně závislými jevy vyžadují celkové rozdělení pravděpodobnosti. Vhodným nástrojem pro sestavení takového rozdělení pravděpodobnosti je struktura Bayesovské sítě. Toto je velmi stručné shrnutí Bayesovských sítí.

8.1. Úvod

Podmíněná pravděpodobnost $P(A \mid B)$ umožňuje omezit a zohlednit kontext B pro pravděpodobnost jevu A. Pravděpodobnost $P(A \mid B)$ má často názorný význam, pokud B zapřičiňuje A a v tom případě je příhodnou vstupní informací pro vyšetřování složitějších souvislostí. Jinak ovšem muže zjištění podmíněných pravděpodobností na základě doporučení Bayesovské statistiky a zejména pak na základě Bayesova vzorce vést k rozsáhlejším výpočtům a případně k překvapivým výsledkům. To je ilustrováno zadáním a řešením následující Úlohy 8:

ÚLOHA 8. *Cestující prochází na letišti běžnou bezpečnostní kontrolou, která má mimo jiné zjistit, zda některý cestující u sebe nemá střelnou zbraň. Předpokládejte následující:*

- *Pravděpodobnost, že některý cestující má střelnou zbraň je*
$$P(\textit{má střelnou zbraň}) = 0.0001\%$$
.
- *Pravděpodobnost, že běžná prohlídka odhalí střelnou zbraň je*
$$P(\textit{zbraň je nalezena} \mid \textit{má střelnou zbraň}) = 80\%$$
.
- *Bezpečnostní pracovníci vybírají dle své intuice některé cestující k následné, důkladnější prohlídce s pravděpodobností*
$$P(\textit{důkladná prohlídka}) = 1\%$$
.
- *Intuice bezpečnostních pracovníků je velmi dobrá, takže pokud cestující má zbraň, je s vybrán k důkladnější prohlídce pravděpodobností*
$$P(\textit{důkladná prohlídka} \mid \textit{má střelnou zbraň}) = 90\%$$
.

Zjistěte:

(1) *Cestující vedle Vás byl vybrán k důkladnější prohlídce. Jaká je pravděpodobnost, že má u sebe střelnou zbraň?*
(2) *Cestující vedle Vás byl vybrán k důkladnější prohlídce a prošel s Vámi do letadla. Jaká je pravděpodobnost, že má u sebe střelnou zbraň?*
(3) *Po příchodu do letadla si k Vám přisedne cestující, o kterém nevíte, zda byl na důkladnější prohlídce. Jaká je pravděpodobnost, že má u sebe střelnou zbraň?*

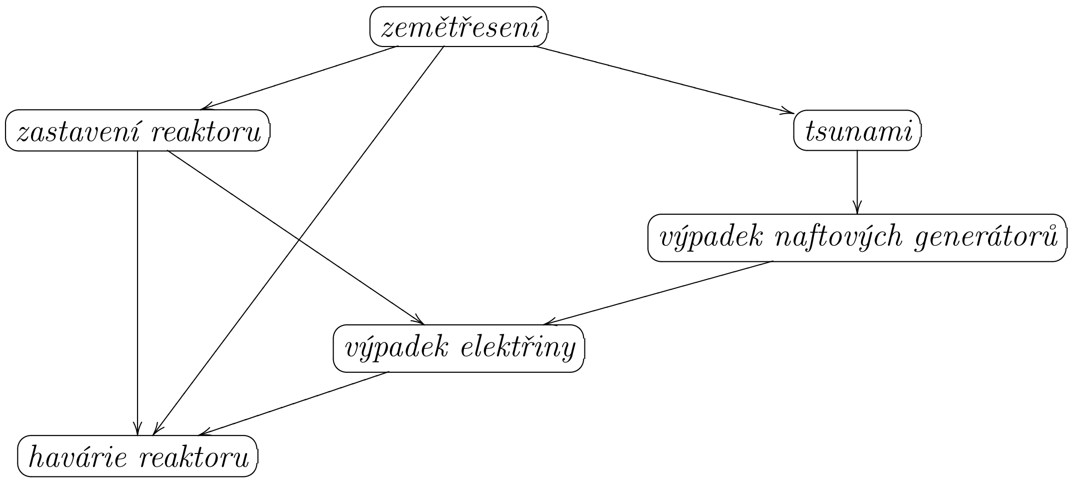
Řešení:
Systematický přístup k zadání a řešení komplikovaných situací zadaných podmíněnými pravděpodobnostmi, jak je ukázáno v dalších odsecích této kapitoly.

8.2. Orientované acyklické grafy

Bayesovská síť je grafický nástroj na modelování složitých rozdělení pravděpodobnosti. Náhodné proměnné jsou znázorněny jako vrcholy orientovaného acyklického grafu [**1**].

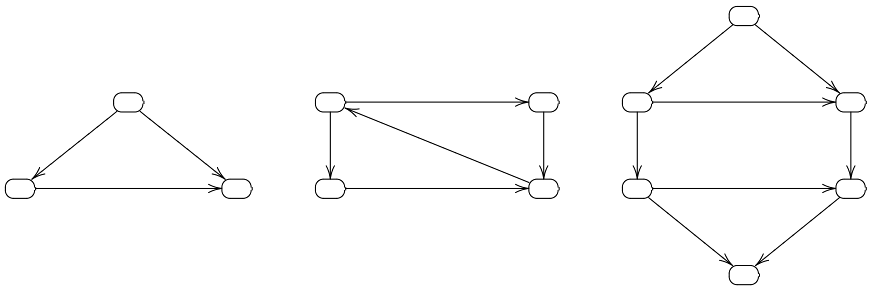
Definice 26. **Orientovaný acyklický graf** *sestává z vrcholů a orientovaných hran (šipek mezi vrcholy), přičemž libovolným procházením mezi vrcholy podél šipek, ve směru šipek se nelze vrátit do vrcholu, ze kterého jsme vyšli. (Graf neobsahuje kolečko z šipek).*

Příklad 12. *V následujícím grafu jsou ve vrcholech náhodné veličiny, které mají jen dvě možné hodnoty (nastalo/nenastalo). Obecně mohou samozřejmě náhodné veličiny nabývat libovolných jiných hodnot, pokud to odpovídá problému.*



Poznámka 19. Je názorné si představovat, že šipky v daném grafu představují zapříčinění, a tak je i vytvořen graf předchozího příkladu. Výsledné celkové rozdělení pravděpodobnosti však informaci o zapříčinění neobsahuje a proto existují příklady grafů, které dávají rozumné rozdělení pravděpodobnosti bez toho, aby odpovídaly naší představě o zapříčinění. V sestavování modelů se však zpravidla opíráme právě o pochopení příčinných souvislostí.

Kontrolní otázka 8.1. *Rozhodněte, který z následujících grafů je acyklický:*



(a) (b) (c)

Definice 27. *Nechť je dán vrchol v orientovaného acyklického grafu G.*

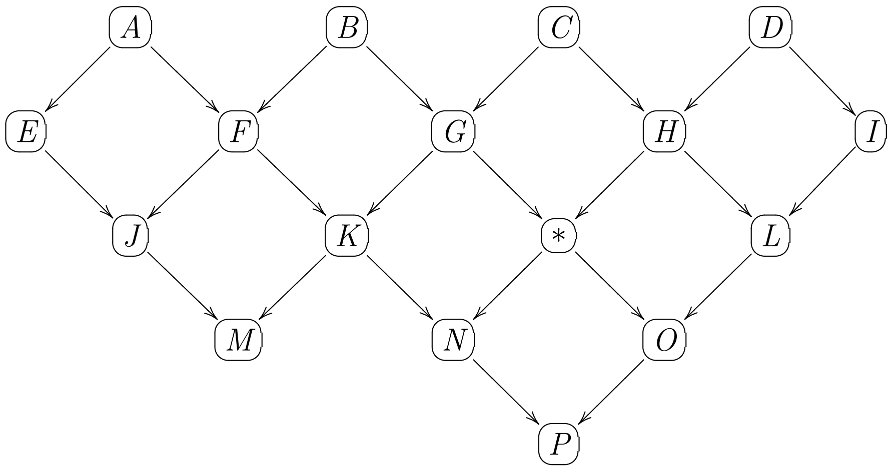
Vrchol r grafu G je **rodičem** *vrcholu v, pokud z r vede šipka do v. Množinu všech rodičů vrcholu v označujeme jako Parents(v).*

Vrchol d grafu G je **dítětem** *vrcholu v, pokud z v vede šipka do d. Množinu všech dětí vrcholu v označujeme jako Children(v).*

Vrchol p grafu G je **potomkem** *vrcholu v, pokud z v vede cesta z šipek do p. Množinu všech potomků vrcholu v označujeme jako Descendants(v).*

Vrchol n grafu G **není ve vztahu přímého potomstva** *s vrcholem v, pokud z v nevede cesta z šipek do n ani z n nevede cesta z šipek do v. Množinu všech vrcholů, které nejsou ve vztahu přímého potomstva s vrcholem v, označujeme jako Non-Descendants(v).*

Kontrolní otázka 8.2. *Najděte pro následující acyklický graf množiny Parents$(*)$, Children$(*)$, Descendants$(*)$ a Non-Descendants$(*)$. Najděte všechny vrcholy, pro které platí*



Najděte všechny vrcholy v, pro které platí, že nejsou ve vztahu přímého potomstva s žádným vrcholem w, který by byl ve vztahu přímého potomstva s vrcholem $*$.

8.3. Výpočet rozdělení pravděpodobnosti z Bayesovské sítě

Konstrukce celkového rozdělení pravděpodobnosti $p(v_1, \dots, v_n)$ pro náhodné veličiny $v_1, \dots, v_n$ odpovídající všem vrcholům $v_1, \dots, v_n$ orientovaného acyklického grafu G funguje následujícím způsobem:

TVRZENÍ 11. *Předpokládáme, že, pro orientovaný acyklický graf G,*

(1) *všechny podmíněné pravděpodobnosti $p(v \mid Parents(v))$ jsou nezávislé od proměnných $Non\text{-}Descendants(v)$.*

(2) *jsou zadány tzv.* **tabulky podmíněných pravděpodobností**

$$p(v \mid Parents(v))$$

pro každý vrchol v grafu G.

Výše uvedené údaje se nazývají **Bayesovskou sítí***.*

Potom existuje jednoznačně určené celkové rozdělení pravděpodobnosti $p(v_1, \dots, v_n)$ splňující předpoklady a je dáno výrazem

$$p(v_1, \dots, v_n) = \prod_{v \in G} p(v \mid Parents(v)) \tag{158}$$

POZNÁMKA 20. Pokud se omezíme jen na malou síť, jsou přímočaré výpočty zvládnutelné s pomocí běžného počítače a není ničeho dalšího třeba. Na libovolnou otázku ohledně systému lze potom odpovědět na základě zkonstruovaného celkového rozdělení pravděpodobnosti.

V praxi mohou Bayesovské sítě mít tisíce vrcholů a existují efektivní algoritmy pro zacházení s nimi [**6**]. Zde se specializovanými algoritmy a nutností jejích použití nebudeme zabývat v plné obecnosti ale omezíme se v Sekci 8.6 na speciální případ Bayesovských sítí, na skryté Markovovy řetězce.

PŘÍKLAD 13. *Zde uvedeme pro ilustraci specifikaci dvou z tabulek podmíněných pravděpodobností pro graf z Příkladu 12. Hodnoty jsou zde smyšlené, ilustrativní. Pravděpodobnosti pro zvolené podmínky se musí vždy sčítat na jedničku.*

- *Vrchol* zemětřesení *nemá žádné rodiče a jeho tabulka podmíněných pravděpodobností proto neobsahuje žádné podmínky:*

zemětřesení	*pravděpodobnost (denní)*
nastalo	*0.0001*
nenastalo	*0.9999*

- *Vrchol* zastavení reaktoru *má jednoho rodiče a jeho tabulka podmíněných pravděpodobností proto obsahuje jednu podmínku:*

zastavení reaktoru	*zemětřesení*	*pravděpodobnost*
nastalo	*nastalo*	*0.995*
nenastalo	*nastalo*	*0.005*
nastalo	*nenastalo*	*0.002*
nenastalo	*nenastalo*	*0.998*

- *Vrchol* *tsunami* *má jednoho rodiče a jeho tabulka podmíněných pravděpodobností proto obsahuje jednu podmínku:*

tsunami	*zemětřesení*	*pravděpodobnost*
nastalo	*nastalo*	*0.1*
nenastalo	*nastalo*	*0.9*
nastalo	*nenastalo*	*0.00001*
nenastalo	*nenastalo*	*0.99999*

- *Vrchol* *výpadek naftových generátorů* *má jednoho rodiče a jeho tabulka podmíněných pravděpodobností proto obsahuje jednu podmínku:*

výpadek naftových generátorů	*tsunami*	*pravděpodobnost*
nastalo	*nastalo*	*0.5*
nenastalo	*nastalo*	*0.5*
nastalo	*nenastalo*	*0.999*
nenastalo	*nenastalo*	*0.001*

- *Vrchol* *výpadek elektřiny* *má dva rodiče a jeho tabulka podmíněných pravděpodobností proto obsahuje dvě podmínky:*

výpadek elektřiny	*zastavení reaktoru*	*výpadek naftových generátorů*	*pravděpodobnost*
nastalo	*nastalo*	*nastalo*	*0.95*
nenastalo	*nastalo*	*nastalo*	*0.05*
nastalo	*nenastalo*	*nastalo*	*0.01*
nenastalo	*nenastalo*	*nastalo*	*0.99*
nastalo	*nastalo*	*nenastalo*	*0.01*
nenastalo	*nastalo*	*nenastalo*	*0.99*
nastalo	*nenastalo*	*nenastalo*	*0.000001*
nenastalo	*nenastalo*	*nenastalo*	*0.999999*

- *Vrchol* (*havárie reaktoru*) *má tři rodiče a jeho tabulka podmíněných pravděpodobností proto obsahuje tři podmínky:*

havárie reaktoru	*zastavení r.*	*zemětřesení*	*výpadek el.*	*pravděpodobnost*
nastala	*nastalo*	*nastalo*	*nastal*	*0.9*
nenastala	*nastalo*	*nastalo*	*nastal*	*0.1*
nastala	*nenastalo*	*nastalo*	*nastal*	*0.8*
nenastala	*nenastalo*	*nastalo*	*nastal*	*0.2*
nastala	*nastalo*	*nenastalo*	*nastal*	*0.8*
nenastala	*nastalo*	*nenastalo*	*nastal*	*0.2*
nastala	*nenastalo*	*nenastalo*	*nastal*	*0.7*
nenastala	*nenastalo*	*nenastalo*	*nastal*	*0.3*
nastala	*nastalo*	*nastalo*	*nenastal*	*0.001*
nenastala	*nastalo*	*nastalo*	*nenastal*	*0.999*
nastala	*nenastalo*	*nastalo*	*nenastal*	*0.01*
nenastala	*nenastalo*	*nastalo*	*nenastal*	*0.99*
nastala	*nastalo*	*nenastalo*	*nenastal*	*0.000001*
nenastala	*nastalo*	*nenastalo*	*nenastal*	*0.999999*
nastala	*nenastalo*	*nenastalo*	*nenastal*	*0.0000001*
nenastala	*nenastalo*	*nenastalo*	*nenastal*	*0.9999999*

Kontrolní otázka 8.3. *Pro Bayesovskou síť zadanou v Příkladech 12 a 13 vypočtěte následující:*

(1) *Spočtěte pravděpodobnost, že dojde k havárii reaktoru.*
(2) *Spočtěte pravděpodobnost, že dojde k havárii reaktoru za předpokladu, že došlo k zemětřesení.*
(3) *Spočtěte pravděpodobnost, že dojde k havárii reaktoru za předpokladu, že došlo k zemětřesení, ale nevznikla vlna tsunami.*

8.4. Generování vzorků a simulace na Bayesovských sítích.

Vygenerovat vzorek na základě rozdělení pravděpodobnosti zadaného Bayesovskou sítí je snadné, za předpokladu, že umíme v každém vrcholu vygenerovat hodnotu příslušné náhodné veličiny na základě příslušné tabulky podmíněných pravděpodobností, pokud jsou známy hodnoty náhodných veličin vzorku u všech rodičů vrcholu. Lze potom nejdříve vygenerovat hodnoty příslušející k vrcholům bez rodičů a postupně generovat další hodnoty ve vrcholech, jejíchž rodiče už jsou známy.

8.4.1. Simulace. Veškeré otázky ohledně rozdělení pravděpodobnosti jevů na Bayesovské síti lze v principu odpovědět exaktně. V praxi však může být síť velmi rozsáhlá a přímý výpočet nemusí být možný ani s použitím výkonných počítačů. Jednoduchou alternativou k sofistikovaným specializovaným algoritmům je simulace. Ta poskytne jen přibližný výsledek, jejiž výsledek lze ve vhodných případech získat rychle a jednoduše. Pokud bychom chtěli například zjistit pravděpodobnost hodnot příslušejících k několika málo vrcholům, můžeme si nechat vygenerovat dostatečný počet vzorků a příslušné pravděpodobnosti aproximovat relativními četnostmi jednotlivých kombinací hodnot v získaném souboru vzorků.

ÚLOHA 9. *Pro Bayesovskou síť zadanou v Příkladech 12 a 13 vypočtěte následující:*
(1) *Vygenerujte vzorek podle rozdělení pravděpodobnosti zadaného Bayesovskou sítí.*
(2) *Spočtěte simulací pravděpodobnost havárie reaktoru a porovnejte výsledek s přesným výpočtem.*

8.5. Markovovy řetězce a skryté Markovovy řetězce jako Bayesovské sítě

Speciálními případy Bayesovských sítí jsou **Markovovy řetězce**,

$$\boxed{x_0} \to \boxed{x_1} \to \boxed{x_2} \to \dots \to \boxed{x_n} \to \dots \tag{159}$$

a tzv. **skryté Markovovy řetězce**,

$$\begin{tikzcd} \boxed{x_0} \arrow[r] \arrow[d] & \boxed{x_1} \arrow[r] \arrow[d] & \boxed{x_2} \arrow[r] \arrow[d] & \dots \arrow[r] & \boxed{x_{n-1}} \arrow[r] \arrow[d] & \boxed{x_n} \arrow[r] \arrow[d] & \dots \\ \boxed{y_0} & \boxed{y_1} & \boxed{y_2} & \dots & \boxed{y_{n-1}} & \boxed{y_n} & \dots \end{tikzcd} \tag{160}$$

Skryté Markovovy systémy jsou příhodným případem, kdy se specializovanými algoritmy pracujícími na Bayesovské síti dá efektivně dosáhnout efektních výsledků.

8.6. Specializované algoritmy pro skryté Markovovy řetězce

DEFINICE 28. *Skrytým Markovovým řetězcem rozumíme konečný, časově homogenní Markovův řetězec s možnými hodnotami $x_n \in \{1, 2, \dots, N\}$ pro všechna $n \in \mathbb{N}_0$, který však není pozorován přímo, ale skrze náhodné veličiny $y_0, y_1, \dots$, nabývajících hodnot v konečné množině $\{1, 2, \dots, M\}$, přičemž jednotlivé hodnoty y_n pozorovaných náhodných veličin jsou určeny příslušnými nepozorovanými hodnotami x_n skrze stochastickou matici $O_{y_n x_n} = p(y_n \mid x_n)$ a jsou vzájemně podmíněně nezávislé:*

$$p(y_0, y_1, \dots, y_n \mid x_0, x_1, \dots, x_n) = \prod_{i=0}^{n} p(y_i \mid x_i) \tag{161}$$

Základní úlohy. Struktura skrytých Markovových řetězců vede na několik základních úloh [v závorkách jsou pro informaci uvedeny názvy algoritmů na řešení těchto úloh]:
(1) **Učení parametrů modelu.** Na základě jedné posloupnosti nebo konečné množiny posloupností pozorovatelných $y_0, y_1, \dots, y_k$ je třeba odhadnout parametry modelu, tj. matice M, O a počáteční stav π udávající pravděpodobnosti $\pi(x_0) = p(x_0)$ pro hodnoty x_0. Pro tento problém existuje efektivní řešení založeno na
   (a) maximální věrohodnosti [Baumův-Welchův algoritmus].
   (b) vzájemné informaci [Baldiův-Chauvinův algoritmus].
(2) **Úsudky na základě známých parametrů modelu.**
   (a) Pravděpodobnost pozorované posloupnosti: problém výpočtu $p(y_0, y_1, \dots, y_n)$ [dopředný algoritmus].

(b) Filtrování: problém nalezení $p(x_n \mid y_0, y_1, \ldots, y_n)$ [dopředný algoritmus].
(c) Shlazení: problém nalezení $p(x_k \mid y_0, y_1, \ldots, y_n)$ pro $0 \leq k < n$ [dopředný-zpětný algoritmus].
(d) Nejvěrohodnější vysvětlení: Problém nalezení konečné posloupnosti $x_0, x_1, \ldots x_n$ k pozorováním $y_0, y_1, \ldots, y_n$, pro kterou je $p(x_0, x_1, \ldots x_n \mid y_0, y_1, \ldots, y_n)$ maximální [Viterbiho algoritmus].

Naivní řešení úloh skrytých Markovových řetězců. Uvedené úlohy se mohou zdát přímočaré a snadné. Problém však spočívá v tom, že řada přímočarých řešení je v principu správná, ale co do výpočetní náročnosti exponenciálně rostoucí s délkou pozorování Markovova řetězce, a proto prakticky nepoužitelná. To je důvodem, proč na jednotlivé úlohy existují speciální algoritmy, které jsou podstatně efektivnější.

PŘÍKLAD 14. *Ukažme si potíž s přímočarými řešeními na problému nalezení pravděpodobnosti pozorované posloupnosti. Pravděpodobnost $p(y_0, y_1, \ldots, y_n)$ lze napsat následovně:*

$$p(y_0, y_1, \ldots, y_n) = \sum_{x_0=1}^{N} \sum_{x_1=1}^{N} \cdots \sum_{x_n=1}^{N} p(x_0, x_1, \ldots, x_n, y_0, y_1, \ldots, y_n) = \tag{162}$$

$$= \sum_{x_0=1}^{N} \sum_{x_1=1}^{N} \cdots \sum_{x_n=1}^{N} p(y_n \mid x_n) \ldots p(y_1 \mid x_1) p(y_0 \mid x_0) \times \tag{163}$$

$$\times\, p(x_n \mid x_{n-1}) \ldots p(x_1 \mid x_0) \pi(x_0), \tag{164}$$

což je výraz s N^n členy, z nichž každý vyžaduje výpočet $2n$ součinů.

8.7. Algoritmy pro úsudky na základě známých parametrů modelu.

Základní pomůckou při řešení úloh Markovových řetězců jsou tzv. dopředné veličiny $\alpha_k(x_k)$ a zpětné veličiny $\beta_k(x_k)$:

$$\alpha_k(x_k) = p(y_0, y_1, \ldots, y_k, x_k), \tag{165}$$

$$\beta_k(x_k) = p(y_{k+1}, y_{k+2}, \ldots, y_n \mid x_k), \tag{166}$$

efektivně vypočítatelné rekurzemi

$$\alpha_{k+1}(x_{k+1}) = O_{y_{k+1} x_{k+1}} \sum_{x_k=1}^{N} M_{x_{k+1} x_k} \alpha_k(x_k), \tag{167}$$

$$\beta_{k-1} x_{k-1} = \sum_{x_k=1}^{N} O_{y_k x_k} \beta_k(x_k) M_{x_k x_{k-1}}, \tag{168}$$

s počáteční, resp. koncovou hodnotou

$$\alpha_0(x_0) = O_{y_0 x_0} \pi(x_0), \tag{169}$$

$$\beta_n(x_n) = 1. \tag{170}$$

Provedení rekurze se nazývá **dopředným** resp. **zpětným algoritmem**. S pomocí těchto rekurzivně počítaných veličin lze řešit následující problémy s výsledky:

Pravděpodobnost pozorované posloupnosti: dopředný algoritmus.

$$p(y_0, y_1, \ldots, y_n) = \sum_{x_n=1}^{N} \alpha_n(x_n) \tag{171}$$

Filtrování: dopředný algoritmus.

$$p(x_n \mid y_0, y_1, \ldots, y_n) = \frac{\alpha_n(x_n)}{p(y_0, y_1, \ldots, y_n)} \tag{172}$$

Shlazení: dopředný-zpětný algoritmus. Pro $0 \leq k \leq n$:

$$p(x_k \mid y_0, y_1, \ldots, y_n) = \frac{\alpha_k(x_k)\beta_k(x_k)}{p(y_0, y_1, \ldots, y_n)} \tag{173}$$

Nejvěrohodnější vysvětlení: Viterbiho algoritmus.

$$\delta_k(x_k) = \max_{x_0, x_1, \ldots, x_{k-1}} p(y_0, y_1, \ldots, y_k, x_0, x_1, \ldots, x_k) \tag{174}$$

lze spočítat rekurzí

$$\delta_{k+1}(x_{k+1}) = O_{y_{k+1}x_{k+1}} \max_{x_k} M_{x_{k+1}x_k} \delta_k(x_k) \tag{175}$$

s počáteční hodnotou

$$\delta_0(x_0) = O_{y_0x_0}\pi(x_0). \tag{176}$$

SHRNUTÍ 8. *Bayesovská sít umožňují prakticky sestavit rozdělení pravděpodobnosti pro rozsáhlý systémy na základě orientovaného acyklického grafu vystihujícího, jehož vrcholy odpovídají náhodným veličinám systému, a tabulek podmíněných pravděpodobností pro jednotlivé náhodné veličiny.*

Bayesovské sítě se hodí především pro případy, kdy graf neobsahuje příliš mnoho ze všech šipek mezi vrcholy, které by byly možné. To je v praxi často rozumný předpoklad, kdy šipky často vyjadřují přímá kauzální působení mezi jednotlivými náhodnými veličinami. Neexistence šipky mezi dvěma vrcholy znamená, že příslušné náhodné veličiny jsou podmíněně nezávislé.

Bayesovská síť se ovšem hodí nejen pro sestavení rozdělení pravděpodobnosti pro celý systém. Lze jí s výhodou použít pro generování vzorků nebo pro nasazení speciálních algoritmů na otázky o systému, které by byly na základě rozdělení pravděpodobnosti systému sice v principu, ale nikoliv prakticky řešitelné.

KAPITOLA 9

Bayesovská teorie rozhodování. Ztrátová funkce, utilita. Klasické ztrátové funkce

Klíčová slova: Ztrátová funkce, utilita, rozhodnutí.

Abstrakt: Bayesovská statistika nachází nejlepší rozhodnutí jako minimalizující aposteriorní očekávanou ztrátu.

9.1. Úvod

Konečným cílem schématu Bayesovské statistiky užívající statistický model není stanovení aposteriorního rozdělení pravděpodobnosti pro parametry modelu (jak diskutováno v Kapitole 7), ale **rozhodnutí**, které na základě získaných poznatků učiníme. Touto volbou se zabývá **teorie rozhodování**. Její zcela přímočará formulace založená na **ztrátové funkci** je podána v Sekci 9.2. V praxi ovšem vyžaduje často značnou práci důkladně sestavit a zdůvodnit používanou ztrátovou funkci či jen znaménkem se od ní lišící **utilitu**. Je někdy těžké vědět, co vlastně chceme, co je náš cíl, či která ztráta z několika možných je větší než jiná. Problému existence a volby utility je věnována Sekce 9.3.

9.2. Bayesovská teorie rozhodování

Bayesovská teorie rozhodování je zadána

(1) třídou rozdělení pravděpodobností $p(x \mid \theta)$ danou předepsaným tvarem a možnými hodnotami parametru $\theta \in \Theta$,
(2) apriorním rozdělením $\pi(\theta)$ pro parametry θ,
(3) prostorem rozhodnutí $\mathcal{D}$ a ztrátovou funkcí $L : \Theta \times \mathcal{D} \to [0, \infty]$.

Užitečnost (utilita) rozhodnutí $d \in \mathcal{D}$ za hodnoty parametru θ je dána jako

$$U(\theta, d) = -L(\theta, d). \tag{177}$$

Obvykle nelze ztrátovou funkci volbou jednoho rozhodnutí d minimalizovat pro libovolnou hodnotu θ, s čímž je třeba se nějak vypořádat. [1]

[1] Frekventistické řešení tohoto problému je minimalizovat průměrnou ztrátu (frekventistické riziko)

$$R(\theta, \delta) = \int L(\theta, \delta(x)) p(x \mid \theta) dx,$$

kde $\delta : \mathcal{X} \to \mathcal{D}$ je rozhodovací pravidlo pro výsledek $x \in \mathcal{X}$.

To má řadu závad:

Bayesovské řešení minimalizuje aposteriorní očekávanou ztrátu ρ, která bere do úvahy známou, naměřenou hodnotu x vznikne integrací přes neznámý parametr θ:

$$\rho(\pi, d \mid x) = \int_{\Theta} L(\theta, d)\pi(\theta \mid x) d\theta \tag{178}$$

KONTROLNÍ OTÁZKA 9.1. *Komín přirozené tlakové nádoby upouští tlak s pravděpodobností q podle Bernoulliho rozdělení. Pokud parametr θ klesne pod hodnotu 0.1 je třeba očekávat nevyhnutný výbuch tlakové nádoby se škodami $100M€$. Apriorní odhad parametru θ je dán jako*

$$\pi(\theta) = 30\theta^2(1-\theta)^2. \tag{179}$$

Náklady na odstranění tlakové nádoby včetně ztrát na příjmech z turismu jsou $20M€$. Ztrátová funkce tedy je

$$L(\theta, d) = \begin{cases} -20M€ & \text{pro } d = \text{„nádoba odstraněna“}. \\ -100M€ & \text{pro } d = \text{„nádoba neodstraněna“ a } \theta < 0.1. \\ 0M€ & \text{pro } d = \text{„nádoba neodstraněna“ a } \theta \geq 0.1. \end{cases} \tag{180}$$

Pětkrát za sebou nedošlo k upuštění tlaku. Jak je třeba rozhodnout? Odstranit nebezpečí nebo zachovat přírodní úkaz?

DEFINICE 29. *Za předpokladu, že $\mathcal{D} = \Theta$, je $\delta^\pi : \mathcal{X} \to \mathcal{D}$* **Bayesovským odhadem** *parametru θ, pokud*

$$\delta^\pi(x) = \arg\min_{d \in \mathcal{D}} \rho(\pi, d \mid x) \tag{181}$$

V kontextu odhadů, kdy $\mathcal{D} = \Theta$ je lineární prostor, jsou běžnými ztrátovými funkcemi

$$L_1(\theta, d) = \mid \theta - d \mid \tag{182}$$

$$L_2(\theta, d) = (\theta - d)^2 \tag{183}$$

DEFINICE 30. **Integrovaným rizikem** *rozumíme*

$$r(\pi, \delta) = \int_{\mathcal{X}} \rho(\pi, d \mid x) m(x) dx = \int_{\Theta} \int_{\mathcal{X}} L(\theta, d) p(x \mid \theta) dx \pi(\theta) d\theta \tag{184}$$

Bayesovské riziko *je potom výraz*

$$r(\pi) = r(\pi, \delta^\pi) \tag{185}$$

VĚTA 4. *Odhad $\delta : \mathcal{X} \to \mathcal{D}$ minimalizuje integrované riziko $r(\pi, \delta)$, pokud je Bayesovským odhadem, tj. $\delta(x) = \delta^\pi(x)$.*

- Skutečně naměřená hodnota x se vlastně dál nevyužije, středuje se přes všechna $x \in \mathcal{X}$, která většinou vůbec nenastala.
- Frekventistická analýza předpokladů, že se věc bude znovu a znovu opakovat, což nemusí být pravda.
- Riziko $R(\theta, \delta)$ je závislé na parametru θ a nemusí tedy poskytnout úplné uspořádání rozhodovacích pravidel δ.

Rozhodovací pravidlo $\delta : \mathcal{X} \to \mathcal{D}$ lze při $\mathcal{D} = \Theta$ chápat jako **odhad** parametru δ.

Důkaz. viz [**18**].

Cvičení na síti 8. **Lov zlatých sršňů**.

9.3. Existence a volba utility

Výše uvedené úvahy se zdají být přímočaré, vycházejí však z netriviálního předpokladu, že máme k dispozici ztrátovou funkci či utilitu. V rámci zde používané subjektivní interpretace pravděpodobnosti je přirozené chápat též utilitu jako subjektivní hodnocení užitečnosti různých výsledků. Musí ale utilita, která vyjadřuje užitečnost rozhodnutí v předpokládaném kontextu parametru θ vůbec existovat? A pokud existuje, jak je možno stanovit její průběh?

Utilita se zpravidla sestavuje tak, že se nejdříve vyjasní faktický výsledek (též nazývaný odměna) r z prostoru všech možných odměn $\mathcal{R}$ a těmto výsledkům jsou potom přiřazeny hodnoty utility pomocí funce $U : \mathcal{R} \to \mathbb{R}$. Předpokládáme, že prostor odměn (výsledků) $\mathcal{R}$ je **úplně uspořádaná množina**, tj. na $\mathcal{R}$ je dána relace $\preceq$ (vyjadřující preferenci pro pravou stranu) tak, že

(1) $r_1 \preceq r_2$ nebo $r_2 \preceq r_1$ pro všechna $r_1, r_2 \in \mathcal{R}$ (porovnatelnost libovolných odměn),
(2) pokud $r_1 \preceq r_2$ a $r_2 \preceq r_3$, potom platí $r_1 \preceq r_3$ (tranzitivita).

Z tohoto pohledu je například zřejmé, že lákavě jednoduché hodnocení užitku penězi jako lineární stupnicí je nedostatečné: malý, pevně stanovený obnos je mnohem užitečnější chudému než bohatému. Peněžní zisk r je tedy lépe chápat jako výsledek v prostoru $\mathcal{R} = \mathbb{R}$, který by se však neměl zaměňovat za hodnotu utility $U(r)$. Přírůstek užitku, utility v důsledku přírůstku peněz je při zvyšující se celkové částce sice zpravidla[2] vnímán jako kladný, ale čím dál menší. To motivuje častou volbu utility $U(r)$ jako rostoucí konkávní omezenou funkci.

V kontextu statistiky je navíc typická situace, kdy výsledek není určen jednoznačně a je třeba přiřazovat utilitu nikoliv jednotlivým výsledkům (tedy $U(r)$, $r \in \mathcal{R}$), ale rozdělením pravděpodobnosti P nad výsledky $\mathcal{R}$ (které slibují různé možné výsledky s danými pravděpodobnostmi) jako $U(P)$.

Že utilita $U(P)$ existuje, lze dokázat, pokud přijmeme následující předpoklady [**23**]:

(1) Prostor $\mathcal{P}$ rozdělení pravděpodobností na $\mathcal{R}$ je úplně uspořádaná množina, tj. na $\mathcal{P}$ je dána relace $\preceq$ tak že platí
  (a) (úplnost) $P_1 \preceq P_2$ nebo $P_2 \preceq P_1$ pro všechna $P_1, P_2 \in \mathcal{P}$,
  (b) (tranzitivita) pokud $P_1 \preceq P_2$ a $P_2 \preceq P_3$, potom platí $P_1 \preceq P_3$.

---

[2] I s tím je ovšem někdy polemizováno, viz. např. odpověď R. P. Feynmana [**9**] na nabídku lépe placeného místa:

„Když jsem se dozvěděl plat, rozhodl jsem se, že ho nemohu přijmout. Důvod, proč ho musím odmítnout, spočívá v tom, že bych s ním mohl udělat všechno, co jsem vždycky udělat chtěl - najít si báječnou milenku, zařídit jí byt, kupovat jí hezké věci... S tím platem, co mi nabízíte, bych to mohl opravdu udělat a vím, jak by to se mnou dopadlo. Měl bych plnou hlavu toho, co milenka dělá; doma bych se hádal a tak dále. Tohle všechno by mi šlo tak na nervy, že bych byl nespokojený a nešťastný. Nemohl bych pořádně dělat fyziku a byl by z toho pěkný malér. To, po čem jsem vždycky toužil, by mi špatně posloužilo, a tak jsem se rozhodl, že vaši nabídku nemohu přijmout."

(2) (nezávislost) Pokud $P_1 \preceq P_2$, potom pro libovolné $\alpha \in [0,1]$ a libovolné $P \in \mathcal{P}$ platí

$$\alpha P_1 + (1-\alpha)P \preceq \alpha P_2 + (1-\alpha)P \tag{186}$$

(3) (Archimedova vlastnost) Pokud $P_1 \preceq P_2 \preceq P_3$, potom existují $\alpha, \beta \in (0,1)$ tak, že

$$\alpha P_1 + (1-\alpha)P_3 \preceq P_2 \preceq \beta P_1 + (1-\beta)P_3. \tag{187}$$

Platí potom, že existuje $U : \mathcal{R} \to \mathbb{R}$ tak, že

$$P_1 \preceq P_2 \Leftrightarrow \int_{\mathcal{R}} U(r)dP_1 \leq \int_{\mathcal{R}} U(r)dP_2. \tag{188}$$

Pro situaci statistických modelů, jak bylo diskutováno v předchozí sekci, máme $\mathcal{R} = \Theta$.

Úloha 10. **Petrohradský paradox.** *Případnému zájemci je nabídnuto, že po zaplacení vstupného €X bude moci házet mincí tak dlouho, dokud nepadne poprvé líc. V závislosti na čísle hodu n, kdy poprvé padl líc, je výplata ze hry 3^n.*

(1) *Jaké nejvyšší vstupné €X by zájemce měl být ochoten zaplatit, aby se hry zúčastnil, pokud pouze požaduje, aby ve hře dosáhl průměrného zisku?*
(2) *Diskutujte Vaší osobní utilitní funkci.*
(3) *Určete na základě Vaší osobní utility, kolik byste byli ochotní zaplatit, abyste si zahráli hru Petrohradského paradoxu.*
(4) *Předpokládejte, že pravidla jsou změněna tak, že pokud padne hlava prohráváte a nedostanete nic, zatímco, pokud odstoupíte, než padne první hlava, v n-tém kroku, dostanete obnos 2^{n-1}. Určete, v kterém kroku hry byste na základě Vaší osobní utility měli odstoupit, pokud jste se až do toho kroku hry dostali.*

Shrnutí 9. *Bayesovská teorie rozhodování je zadána*

(1) *statistickým modelem $p(x \mid \theta)$ s parametrem $\theta \in \Theta$ s apriorním rozdělením $\pi(\theta)$,*
(2) *prostorem rozhodnutí $\mathcal{D}$ a ztrátovou funkcí $L : \Theta \times \mathcal{D} \to [0, \infty]$.*

Sestavení ztrátové funkce $L(\theta, d)$ nebo ekvivalentně utility $U(\theta, d) = -L(\theta, d)$ je často složité.

Nejlepší rozhodnutí $d^{\pi}_{nejlepší}(x) \in \mathcal{D}$ je takové, které minimalizuje aposteriorní ztrátu ρ:

$$d^{\pi}_{nejlepší}(x) = \arg\min_{d \in \mathcal{D}} \rho(\pi, d \mid x), \tag{189}$$

kde

$$\rho(\pi, d \mid x) = \int_{\Theta} L(\theta, d)\pi(\theta \mid x)d\theta. \tag{190}$$

Bayesova statistika tak dává přímý návod na volbu rozhodnutí v dané situaci.

KAPITOLA 10

Monte Carlo a MCMC

Klíčová slova: metoda Monte Carlo, MCMC, numerické řešení statistických modelů.

Abstrakt: Metoda MCMC spojuje výhodu numerické integrace metodou Monte Carlo s efektivitou generování vzorků pomocí Markovových řetězců.

10.1. Úvod

Úlohy Bayesovské statistiky vedou zpravidla na integrály některé funkce $f(\theta)$ přes míru P s hustotou pravděpodobnosti $p(\theta)$ (viz např. (142)):

$$E(f) = \int_{\Omega} f(\theta) dP = \int_{\Omega} f(\theta) p(\theta) d\theta. \tag{191}$$

Tyto integrály je často nutno vyčíslit numericky. Numerickou integraci umožňuje například Simpsonova metoda. Tato a obdobné metody ovšem selhávají především proto, že v praktických případech je nutno integrovat přes mnoho-dimenzionální integrační proměnné, a numerická chyba těchto metod zpravidla roste exponenciálně s počtem dimenzí (tzv. **prokletí rozměrnosti**). Existuje však metoda nazývaná **integrací Monte Carlo**, která sice v malém počtu dimenzí konverguje pomaleji než jiné metody, ale která netrpí na jejích neduhy při vyšším počtu dimenzí.

Metoda Monte Carlo je založená na vygenerování náhodných vzorků podle míry P s hustotou pravděpodobnosti $p(\theta)$ a odhadem integrálu z funkce $f(\theta)$ přes tuto míru aritmetickým průměrem hodnot funkce $f(\theta_i)$ přes vygenerované vzorky $\{\theta_i\}_{i=1}^m$. Slabinou tohoto postupu pro obecnou míru ovšem je, že i generování vzorků v mnoha dimenzích naráží na značné obtíže.

Vzorky lze však i v mnoha rozměrech generovat asymptoticky pomocí Markovových řetězců zadaných tak, že jejích stacionární rozdělení pravděpodobnosti bude mít hustotu pravděpodobnosti právě $p(\theta)$, jak to umožňuje například Metropolisův-Hastingsův algoritmus. Takto získané vzorky lze potom využít v metodě Monte Carlo. Tato metoda využívající spojení Markovových řetězců a Monte Carla se nazývá anglicky **Markov Chain Monte Carlo** a označuje se zaužívanou zkratkou **MCMC**. Zde se omezíme na vysvětlení principu této v praxi nesmírně úspěšné metody, zatímco pro podrobné kvantitativní vyšetření jejích často pozoruhodných konvergenčních vlastností odkazujeme na specializovanou literaturu (viz např. [**19**]).

Tato a některé s MCMC příbuzné metody se mohou na první pohled jevit jako drobná technická vylepšení. Otevřely však v posledních letech široké možnosti v reálných aplikacích statistických metod. Pokud nám tedy záleží na praktickém uplatnění, jsou to metody zcela klíčové.

10.2. Integrování metodou Monte Carlo

Integrál funkce $f(x)$ přes míru P s hustotou pravděpodobnosti $p(x)$ označíme jako

$$E(f) = \int_\Omega f(x)dP = \int_\Omega f(x)p(x)dx \tag{192}$$

Tento integrál můžeme přibližně vyčíslit pomocí průměru $\bar{f}_m$ spočteném na náhodném výběru m náhodně s rozdělením pravděpodobnosti $p(x)$ vygenerovaných vzorků $x_1, x_2, \ldots, x_m$:

$$\bar{f}_m = \frac{1}{m}\sum_{i=1}^{m} f(x_i) \tag{193}$$

Kontrolní otázka 10.1. *Spočtěte metodou Monte Carlo integrál*

$$\int_{-\infty}^{+\infty} \mid x \mid e^{-\frac{x^2}{2}} dx \tag{194}$$

a porovnejte výsledek tohoto přiblížení s přesně spočtenou hodnotou.

Je ovšem důležité vědět, jak přesné toto přiblížení je. Platí následující skutečnosti:

Věta 5. **Zákon velkých čísel.** *Nechť $\bar{f}_m$ je průměr z m hodnot $f(x_i)$, $i \in \{1, 2, \ldots, m\}$ spočtený na m nezávislých vzorcích $x_1, x_2, ..., x_m \in \Omega$ podle pravděpodobnostního rozdělení $p(x)$ a nechť je $E(f)$ střední hodnota funkce f na Ω. Potom platí pro libovolné $\epsilon > 0$:*

$$\lim_{m\to\infty} P(\bar{f}_m = E(f)) = 1 \tag{195}$$

Důkaz. Průměr $\bar{f}_m$ má následující charakteristiky:

$$E(\bar{f}_m) = E(\frac{1}{m}\sum_{i=1}^{m} f(x_i)) = \frac{1}{m}\sum_{i=1}^{m} E(f(x_i)) = E(f) \tag{196}$$

$$Var(\bar{f}_m) = Var(\frac{1}{m}\sum_{i=1}^{m} f(x_i)) = \frac{1}{m^2}Var(\sum_{i=1}^{m} f(x_i)) = \frac{1}{m}Var(f) \tag{197}$$

Použitím Čebyševovy nerovnosti (63) na $\bar{f}_m$ dostaneme

$$P(\mid (\bar{f}_m - E(f) \mid < \epsilon) \geq 1 - \frac{Var(f)}{m\epsilon^2}, \tag{198}$$

z čehož plyne tvrzení.

Podle silného zákona velkých čísel tedy platí:

$$\tfrac{1}{m}\textstyle\sum_{i=1}^{m} f(x_i) \xrightarrow[\text{vždy}]{\text{skoro}} \int_\Omega f(x)p(x)dx \tag{199}$$

Poznámka 21. Předchozí je tzv. silná forma zákona velkých čísel. Jeho slabá forma (slabší tvrzení) říká, že

$$\lim_{m\to\infty} P(\bar{f}_m - E(f) \leq \epsilon) = 1 \tag{200}$$

Věta 6. **Centrální limitní věta (Lindebergova-Lévyho).** *Nechť jsou $x_1, \ldots, x_n$ náhodné veličiny se stejným rozdělením pravděpodobnosti s konečnou střední hodnotou $E(x_i) = \mu$ a konečným rozptylem $Var(x_i) = \sigma^2$. Pak platí:*

$$\lim_{n\to\infty} P\left(\frac{x - n\mu}{\sigma\sqrt{n}} \leq x\right) = p(x \mid 0,1), \tag{201}$$

kde $p(x \mid 0,1)$ je normální (Gaussovo) rozdělení se střední hodnotou 0 a rozptylem 1.

Důkaz. Viz [**13**].

Pokud $E(f^2) < \infty$, lze tedy dle centrální limitní věty odhadnout rozptyl $\bar{f}_m$

$$Var(\bar{f}_m) = \frac{1}{m}\int_\Omega (f(x) - E(f))^2 p(x)dx \tag{202}$$

pomocí výběrového rozptylu

$$v_m = \frac{1}{m^2}\sum_{i=1}^{m}(f(x_i) - \bar{f}_m)^2 \tag{203}$$

a $\bar{f}_m$ má potom přibližně normální rozdělení:

$$\frac{\bar{f}_m - E(f)}{\sqrt{v_m}} \approx N(\bar{f}_m \mid 0,1) \tag{204}$$

Kontrolní otázka 10.2. *Spočtěte opakovaně metodou Monte Carlo integrál z Kontrolní otázky 10.1 a diskutujte dosažené výsledky.*

Cvičení na síti 9. **Integrace metodou Monte Carlo**.

10.3. Markov Chain Monte Carlo

Metoda MCMC se liší od metody Monte Carlo vysvětlené v předchozí sekci tím, že ke generování vzorků není používán generátor náhodných veličin, jehož nezávislé realizace by určovaly posloupnost vzorků, ale Markovův řetězec. Po sobě jdoucí vzorky generované Markovovým řetězcem například na základě Metropolisova-Hastingsova algoritmu s náhodnou procházkou nejsou rozhodně nezávislé. To je zřejmé z blízkosti po sobě jdoucích kroků, pokud rozdělení q určující jednotlivé kroky je krátkého dosahu.

Přesto však posloupnost vzorků počínaje libovolnou počáteční hodnotou $\theta^{(0)}$ asymptoticky věrně reprezentuje cílové rozdělení pravděpodobnosti z důvodu *ergodicity* daného Markovova řetězce. Že vskutku dosazením posloupnosti vzorků generovaných Markovovým řetězcem namísto nezávisle generovaných vzorků cílového rozdělení do vzorců metody Monte Carlo dostaneme asymptoticky hodnotu příslušného integrálu, je přesně obsahem tvrzení (132) Věty 3.

Principem metody MCMC je tedy přiblížení

$$\int_\Omega f(x)p(x)dx \approx \frac{1}{m}\sum_{\{x_i\}} f(x_i) \tag{205}$$

kde $\{x_i\}$ je posloupnost vzorků Markovova řetězce se stacionárním rozdělenim pravděpodobnosti $p(x)$.

10.4. Aplikace MCMC

V následující úloze, aposteriorní rozdělení pravděpodobnosti pro parametry modelu vyžaduje numerickou integraci a vysoká dimenze prostoru parametrů neumožňuje použít běžná numerická přiblížení jako např. Simpsonovu metodu.

Úloha 11. *Statistický model určuje rozmístění N označených bodů $x_i = [x_i^1, x_i^2, \ldots, x_i^n] \in \mathbb{R}^n$ v n-rozměrném prostoru podle n-rozměrného normálního rozdělení s konstantním rozptylem $\sigma^2 = 1$ ve všech souřadnicích a s časově neměnnou střední hodnotou $\mu = [\mu^1, \mu^2, \ldots, \mu^n]$, která může být v každé souřadnici jiná. S běžným skalárním součinem $a \cdot b$ pro dva vektory*

$$a = [a^1, a^2, \ldots, a^n] \in \mathbb{R}^n, \tag{206}$$

$$b = [b^1, b^2, \ldots, b^n] \in \mathbb{R}^n \tag{207}$$

definovaný jako

$$a \cdot b = a^1b^1 + a^2b^2 + \ldots + a^nb^n, \tag{208}$$

lze toto normální rozdělení napsat jako

$$N(x_i \mid \mu, 1) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}(x_i-\mu)\cdot(x_i-\mu)}. \tag{209}$$

Apriorní rozdělení pravděpodobnosti pro parametry modelu $\mu = [\mu^1, \mu^2, \ldots, \mu^n]$ je přitom dáno n-rozměrným Cauchyho rozdělením definovaným jako

$$C(x_i \mid \mu) = \frac{\Gamma\left(\frac{1+n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\pi^{n/2}} \frac{1}{(1 + (x_i-\mu)\cdot(x_i-\mu))^{(1+n)/2}}. \tag{210}$$

Pro případ $n = 4$ bylo obdrženo měření poloh pro $N = 5$ realizací:

bod	**poloha**
x_1	*[0.03,0.12,0.45,0.26]*
x_2	*[-0.12,0.07,0.23,0.17]*
x_3	*[0.24,0.05,0.57,0.05]*
x_4	*[0.32,0.28,0.13,-0.01]*
x_5	*[0.19,0.18,0.72,0.31]*

Tabulka 10.1. Získané hodnoty.

Spočtěte aposteriorní rozdělení pravděpodobnosti pro parametry $\mu = [\mu^1, \mu^2, \ldots, \mu^n]$ *na základě údajů* $x_i = [x_i^1, x_i^2, x_i^3, x_i^4]$, $i = 1, 2, \ldots, 5$.

SHRNUTÍ 10. *Běžné postupy numerické integrace zpravidla selhávají ve vysokých dimenzích, nikoliv však* **metoda Monte Carlo***, která počítá integrál přes rozdělení pravděpodobnosti* $p(x)$ *aproximací*

$$\int_\Omega f(x)p(x)dx \approx \frac{1}{m}\sum_{i=1}^{m} f(x_i) \tag{211}$$

spočtenou na m *náhodně s rozdělením pravděpodobnosti* $p(x)$ *vygenerovaných vzorcích* $x_1, x_2, \ldots, x_m$.

I generování vzorků je ve vyšších dimenzích problematické, dá se však úspěšně řešit generováním přes vhodný ergodický Markovův řetězec (např. zkonstruovaný Metropolisovým-Hastingsovým algoritmem), jehož stacionární stav je právě rozdělení pravděpodobnosti $p(x)$*. Tímto vylepšením vzniká metoda* **MCMC** *vhodná pro numerické řešení řady náročných úloh.*

Odpovědi na kontrolní otázky

Kapitola 1. Náhoda a pravděpodobnost.

Kontrolní otázka 1.1. Pravděpodobnost je nejvýše 0.583.

Kontrolní otázka 1.2. Ztráta není v přímém, logickém rozporu s předpokladem, který jen vyjadřoval určitý stupeň důvěry. Na druhou stranu situace ale nevrhá dobré světlo na předpoklad, že jev nastane s 90% pravděpodobností, která byla předložena jako fakt.

Kontrolní otázka 1.3. Otázka je do jisté míry slovní hříčka. Jevy se neodstanou, ať jim subjektivně přiřadíme pravděpodobnosti, jaké chceme. To, co lze vyrušit, by byly finanční ztráty dané nebezpečím, a to koupí sázky za zápornou cenu, která by představovala určité pojištění - nejspíše ale nebude snadné získat prodejce, který vedle vydání směnky ještě daruje částku, která efektivně pojišťuje proti nebezpečí.

Kontrolní otázka 1.4. Pokud jste naprosto přesvědčeni, že budete mít pravnoučata (chlapce nebo dívky - kdo ví, jaké možností obsahuje budoucnost), je volba pravděpodobností s $P(\text{chlapec}) + P(\text{dívka}) = 1$ jistě možná. Pokud ale připustíte, že z nějakého důvodu by se mohlo stát, že například žádná pravnoučata mít nebudete, pak splnění podmínky $P(\text{chlapec}) + P(\text{dívka}) = 1$ **není nutné**. Disjunktní jevy nemusí vyčerpávat veškeré možnosti.

Kontrolní otázka 1.5. Podle Doporučení 5 platí

$$P(A \mid B)P(B) = P(A \wedge B) = P(B \mid A)P(A) \tag{212}$$

a tedy pravděpodobnost, že Bohouš zaujme své místo „U plastového kelímku“, je

$$P(B) = \frac{P(B \mid A)P(A)}{P(A \mid B)} = \frac{40\% \cdot 20\%}{10\%} = 80\%. \tag{213}$$

(K zamyšlení: Proč Bohouš není v hospodě s pravděpodobností 40%, když ho tam ale Albert s takovou pravděpodobností najde? Poskytoval by např. společný, ale poněkud se vylučující zájem obou o tutéž slečnu Cecilii možný mechanismus pro vysvětlení této situace?)

Kapitola 2. Pravděpodobnost a teorie míry.

Kontrolní otázka 2.1. Podle Definice 6 je třeba ověřit, že vzorem měřitelné množiny je měřitelná množina. Každou měřitelnou množinu z Ω_2 lze získat opakovanou aplikací operace doplňku a operace spočetného sjednocení z otevřených množin. Tutéž operaci můžeme provést na vzorech příslušných otevřených množin, přičemž vzorem otevřené množiny je otevřená množina. Platí

$$f^{-1}(\Omega_2 \setminus A) = \Omega_1 \setminus f^{-1}(A) \tag{214}$$

$$f^{-1}\left(\bigcup_{i=1}^{\infty} U_i\right) = \bigcup_{i=1}^{\infty} f^{-1}(U_i) \tag{215}$$

Můžeme tedy libovolný vzor $A_1 = f^{-1}(A_2)$ měřitelné množiny A_2 vygenerovat danými operacemi z otevřených množin přesně tak, jak bychom vygenerovali měřitelnou množinu A_2 z otevřených množin. A_1 musí tedy být měřitelná.

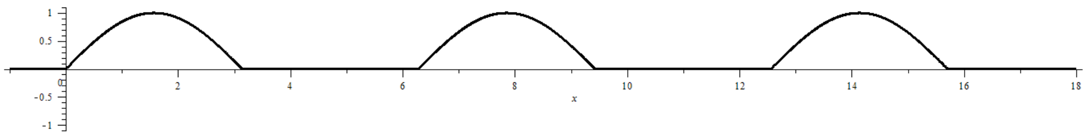
Kapitola 3. Rozdělení pravděpodobnosti.

Kontrolní otázka 3.1. Budeme vycházet z Borelovy míry na $\mathbb{R}^{\not{k}}$, která odpovídá běžnému chápání obsahu. Borelovou mírou každého rovnostranného trojúhelníku $\Delta(s)$ se stranou s je tedy jeho obsah $\mu(\Delta(s)) = \frac{\sqrt{3}}{4}s^2$. Míru je ovšem třeba pro výpočet střední hodnoty na obsluhované oblasti $\Omega = \Delta_A \cup \Delta_B \cup \Delta_C \cup \Delta_D \cup \Delta_E \cup \Delta_F \cup \Delta_G$ normalizovat, čímž získáme míru P:

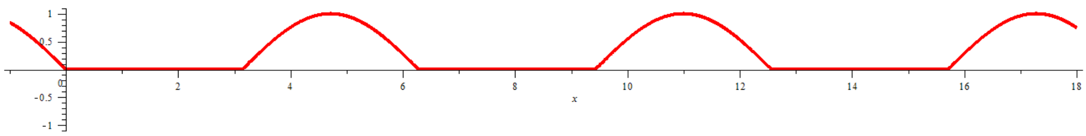
$$P := \frac{\mu}{\mu(\Delta_A) + \mu(\Delta_B) + \mu(\Delta_C) + \mu(\Delta_D) + \mu(\Delta_E) + \mu(\Delta_F) + \mu(\Delta_G)} \tag{216}$$

Průměrné náklady na jedno spojení jsou 0.0111 (což se liší od aritmetického průměr 0.0141 údajů pro jednotlivé vysílače).

Kontrolní otázka 3.2. Viz Obrázky .1, .2



Obrázek .1. **Graf funkce** $\sin^+(x)$**.**



Obrázek .2. **Graf funkce** $\sin^-(x)$**.**

Kapitola 4. Generátory pseudonáhodných čísel.

Kontrolní otázka 4.1. Příkladem řešení je volba

$$a = 2 \qquad b = 3 \qquad n_0 = 0 \tag{217}$$

což vygeneruje jako výsledek posloupnost začínající následovně:

$$0, 3, \mathbf{9}, 21, 45, 93, 89, 81, 65, 33, 69, 41, 85, 73, 49, 1, 5, 13, 29, 61, 25, 53, \mathbf{9}, \ldots \tag{218}$$

Až 23. člen této posloupnosti se rovná některému předchozímu.

Kontrolní otázka 4.2.

$$\operatorname{hod}(x) = \begin{cases} \text{⚀} & \text{pro } x \in [0, 1/6], \\ \text{⚁} & \text{pro } x \in (1/6, 2/6], \\ \text{⚂} & \text{pro } x \in (2/6, 3/6], \\ \text{⚃} & \text{pro } x \in (3/6, 4/6], \\ \text{⚄} & \text{pro } x \in (4/6, 5/6], \\ \text{⚅} & \text{pro } x \in (5/6, 1]. \end{cases} \tag{219}$$

Kontrolní otázka 4.3. Lze postupovat dle Důsledku 9 s jedinou úpravou: distribuční funkce $F(x \mid \alpha)$ je žádanou bijekcí pouze v případě, že se omezíme na nezáporná čísla:

$$F(x \mid \alpha) : [0, +\infty) \to [0, 1) \tag{220}$$

$$F(x \mid \alpha) = \int_0^x p(t \mid \alpha) dt = \int_0^x \alpha e^{-\alpha t} dt = 1 - e^{-\alpha x} \tag{221}$$

Pro inverzi distribuční funkce je třeba řešit rovnici

$$y = 1 - e^{-\alpha x} \tag{222}$$

což dá výsledek

$$x = -\frac{1}{\alpha} \ln{(1 - y)}. \tag{223}$$

Kapitola 5. Markovovy řetězce.
Kontrolní otázka 5.1. **Pro matici M_A:**

(1) $i \to j$ pro všechna $i, j \in \{1, 2, 3\}$ (libovolné j je dosažitelné z libovolného i).
(2) $i \leftrightarrow j$ pro všechna $i, j \in \{1, 2, 3\}$ (všechna i, j spolu komunikují)
(3) M_A je ireducibilní.
(4) Všechny stavy $i \in \{1, 2, 3\}$ jsou rekurentní.
(5) Všechny stavy $i \in \{1, 2, 3\}$ jsou periodické s periodou 3.

Pro matici M_B:

(1) $i \to j$ pro všechna $i, j \in \{1, 2, 3\}$ (libovolné j je dosažitelné z libovolného i).
(2) $i \leftrightarrow j$ pro všechna $i, j \in \{1, 2, 3\}$ (všechna i, j spolu komunikují)
(3) M_B je ireducibilní.
(4) Všechny stavy $i \in \{1, 2, 3\}$ jsou rekurentní.
(5) Všechny stavy $i \in \{1, 2, 3\}$ jsou aperiodické.

Pro matici M_C:

(1) $1 \to 1$, $1 \to 2$, $2 \to 1$, $2 \to 2$, $3 \to 3$.
(2) $1 \leftrightarrow 1$, $1 \leftrightarrow 2$, $2 \leftrightarrow 2$,$3 \leftrightarrow 3$,
(3) M_C není ireducibilní.
(4) Všechny stavy $i \in \{1, 2, 3\}$ jsou rekurentní.
(5) Všechny stavy $i \in \{1, 2, 3\}$ jsou aperiodické.

Kontrolní otázka 5.2. Všechny sloupce matice S mají dle pravidel sudoku součet roven součtu všech číslic $1, 2, \ldots, 9$, tedy 45. Matice M je tedy stochastická. Stacionární stav je dán jako rozdělení $P_i = 1/9$, $i \in \{1, 2, \ldots, 9\}$.

Kapitola 6. Metropolisův-Hastingsův algoritmus.

Kontrolní otázka 6.1. Pokud je parametr popisující délku kroku δ příliš malý, nedokáže Markovův řetězec dostatečně vykrýt potřebný prostor pod křivkou za tak malý počet iterací, protože jakékoliv přesuny jsou pomalé.

Při příliš velkém δ zkouší Metropolisův-Hastingsův algoritmus často velmi vzdálené kandidáty na další krok. Ve velké vzdálenosti je však normální rozdělení pravděpodobnosti mizivé, což vede k zamrznutí stavu Markovova řetězce ve stávající poloze.

Vysvětlení pro oba extrémní případy je podpořeno grafy Příkladu 8.

Kapitola 7. Statistický model a jeho aktualizace. Oblast spolehlivosti.

Kontrolní otázka 7.1. Aposteriorní rozdělení pravděpodobnosti pro střední průměr jablek v úrodě je

$$p(\mu) = \frac{N(10.5 \mid \mu, 1)N(11.2 \mid \mu, 1)N(12.5 \mid \mu, 1)N(\mu \mid 10, 3)}{\int_{-\infty}^{+\infty} N(10.5 \mid \mu, 1)N(11.2 \mid \mu, 1)N(12.5 \mid \mu, 1)N(\mu \mid 10, 3)d\mu} \tag{224}$$

Kapitola 8. Bayesův vzorec. Bayesovské sítě.

Kontrolní otázka 8.1. Acyklické jsou grafy (a) a (c).

Kontrolní otázka 8.2.

$$\text{Parents}(*) = \{\text{G}, \text{H}\}, \tag{225}$$

$$\text{Children}(*) = \{\text{N}, \text{O}\}, \tag{226}$$

$$\text{Descendants}(*) = \{\text{N}, \text{O}, \text{P}\}, \tag{227}$$

$$\text{Non-Descendants}(*) = \{\text{A}, \text{E}, \text{F}, \text{I}, \text{J}, \text{K}, \text{L}, \text{M}\}. \tag{228}$$

Vrcholy v, pro které platí, že nejsou ve vztahu přímého potomstva s žádným vrcholem w, který by byl ve vztahu přímého potomstva s vrcholem $*$: E.

Kontrolní otázka 8.3. Návod: Vypočtěte na počítači rozdělení pravděpodobnosti (158).

Kapitola 9. Bayesovská teorie rozhodování. Ztrátová funkce, utilita. Klasické ztrátové funkce.

Kontrolní otázka 9.1. Porovnáním s (147) lze usoudit, že aposteriorní rozdělení pravděpodobnosti bude

$$p(\theta) = 360 \cdot \theta^2 (1 - theta)^7 \tag{229}$$

Aposteriorní očekávaná ztráta tedy je:

$$\rho(\text{„nádoba odstraněna“}) = -20M€, \tag{230}$$

$$\rho(\text{„nádoba neodstraněna“}) = \int_0^{0.1} -100M€ \cdot 360 \cdot \theta^2 (1 - theta)^7 d\theta = -7.02. \tag{231}$$

Je rozumné nádobu neodstranit.

Kapitola 10. Monte Carlo a MCMC.

Kontrolní otázka 10.1. Pro řešení metodou Monte Carlo je nejdříve třeba integrál upravit, aby v něm vystupovala hustota generovatelné rozdělení pravděpodobnosti:

$$\int_{-\infty}^{+\infty} \mid x \mid e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{+\infty} \sqrt{2\pi} \mid x \mid \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}_{N(x|0,1)} dx. \tag{232}$$

Vzorky x_i, $i = 1, 2, \ldots, n$ normálního rozdělení $N(x \mid 0, 1)$ lze generovat Boxovým-Mullerovým algoritmem (Algoritmus 1) a výsledek je potom

$$\int_{-\infty}^{+\infty} \mid x \mid e^{-\frac{x^2}{2}} dx \approx \frac{1}{n} \sum_{i=1} n\sqrt{2\pi} \mid x_i \mid. \tag{233}$$

V tomto případě je pro porovnání možné spočítat přesné řešení (s použitím substituce $y = x^2/2$):

$$\int_{-\infty}^{+\infty} \mid x \mid e^{-\frac{x^2}{2}} dx = 2 \int_{0}^{+\infty} \mid x \mid e^{-\frac{x^2}{2}} dx = \tag{234}$$

$$= 2 \int_{0}^{+\infty} e^{-y} dy = 2 \left[-e^{-y} \right]_0^{\infty} = 2. \tag{235}$$

Kontrolní otázka 10.2. Výsledky integrace metodou Monte Carlo jsou rozmístěny jako vzorky normálního rozdělení.

Použité symboly

$\mathbb{R}$	množina reálných čísel
$\mathbb{N}$	množina přirozených čísel
Ω	množina elementárních jevů
$\mathcal{F}$	systém měřitelných množin
$\mathcal{B}$	systém měřitelných množin Borelovy algebry
$\mathcal{L}$	systém Lebesgueovsky měřitelných množin
$\mathcal{T}$	systém otevřených množin, topologie
(a, b)	otevřený interval
$[a, b]$	uzavřený interval
U	otevřená množina
K	kompaktní množina
$P(A)$	pravděpodobnost jevu A, pravděpodobnostní míra množiny $A \subset \Omega$
€$P(A)$	hodnota $P(A)$ v peněžních jednotkách
$P(A \mid B)$	podmíněná pravděpodobnost jevu A, pravděpodobnostní míra množiny $A \subset \Omega$ za předpokladu jevu B
μ	míra
dx	Lebesgueova míra na $\mathbb{R}^n$
$p(x)$	hustota pravděpodobnosti (Radonův-Nikodýmův faktor)
$p(x \mid y)$	podmíněná hustota pravděpodobnosti
U_M	rovnoměrné rozdělení pravděpodobnosti na množině M
$N(x \mid \mu, \sigma)$	Normální (Gaussovo) rozdělení se střední hodnotou μ a rozptylem σ^2
$C(x_i \mid \mu)$	Cauchyho rozdělení pravděpodobnosti
$\int f d\mu$	Lebesgueův integrál funkce f přes míru μ
$\int_{\mathbb{R}^n} f dx$	Lebesgueův integrál funkce f přes Lebesgueovu míru dx na $\mathbb{R}^n$
$f^+ f^-$	kladná, záporná část funkce f
$E(f)$	střední hodnota funkce f
$Var(f)$	rozptyl funkce f
G	grupa
$\alpha_g x$	akce α prvku g grupyv bodě x
V	Vitaliho množina
χ_A	charakteristická funkce množiny A
$D(x)$	obecné zobrazení diskrétního dynamického systému

$T(x)$	tent map
$L_\alpha(x)$	logistické zobrazení
$\mathbf{XOR}(x,y)$	exkluzivní disjunkce z logických proměnných x, y
f^{-1}	inverzní funkce
$f^{-1}(Y)$	vzor množiny Y při zobrazení f
$F(x)$	distribuční funkce
f^*	zobrazení
$H(x)$	Heavisideova funkce
M_{ij}	stochastická matice konečného Markovova řetězce
$x \to j$	j je dosažitelné z i
$x \leftrightarrow j$	i komunikuje s j
T_i	doba prvního návratu
$q(z)$	rozdělení pravděpodobnosti pro krok z náhodné procházky
$\pi(x)$	stacionární stav Markovova řetězce
$\mathcal{P}$	prostor uvažovaných rozdělení pravděpodobnosti
Θ	prostor parametrů rozdělení pravděpodobností uvažovaného tvaru
θ	parametr rozdělení pravděpodobností uvažovaného tvaru
$v \in G$	vrchol v orientovaného acyklického grafu G
$Parents(v)$	rodiče vrcholu v
$Children(v)$	děti vrcholu v
$Descendants(v)$	potomci vrcholu v
$Non - Descendants(v)$	vrcholy nemající vztah přímého potomstva s vrcholem v
$O_{j_k I_k}$	stochastická matice určující pozorování skrytého Markovova řetězce v čase k
$\mathcal{X}$	prostor vzorků (pozorování)
$\mathcal{D}$	prostor rozhodnutí
d	rozhodnutí z prostoru rozhodnutí
$\delta(x)$	odhad (rozhodnutí z prostoru rozhodnutí v případě $\mathcal{D} = \Theta$) na základě vzorku x
$\delta^\pi(x)$	Bayesovský odhad pro apriorní rozdělení pravděpodobnosti $\pi(\theta)$
$L(\theta, d)$	ztrátová funkce
$U(\theta, d)$	utilita
$\rho(\pi, d \mid x)$	apriorní očekávaná ztráta
$R(\theta, \delta)$	průměrná ztráta (frekventistické riziko)
$r(\pi, \theta)$	integrované riziko
$\bar{f}_m$	průměr z m hodnot veličiny f
$\Gamma(x)$	Gamma funkce, $\Gamma(n) = n!$ pro $n \in \mathbb{N}$
Δ_A, $\Delta(s)$	rovnostranný trojúhelník A resp. se stranou s

Literatura

1. Wikipedia, Directed acyclic graph (DAG). 62
2. Shun-ichi Amari and Hiroshi Nagaoka, *Methods of information geometry*, AMS Translations of Mathematical Monographs, Oxford University Press, AMS, Providence, Rhode Island, 2007. 54
3. Thomas Bayes, *An Essay towards Solving a Problem in the Doctrine of Chances*, Phil. Trans. **53** (1763), 370–418, Royal Society journal archive. v
4. Vladimir I. Bogachev, *Measure theory*, Springer, Berlin, Heidelberg, 2006. 15, 20
5. Andrew M. Bruckner, Judith B. Bruckner, and Brian S. Thomson, *Real analysis*, Prentice Hall, Upper Saddle River, New Jersey, 1997. 11, 15, 18, 21
6. A. Darwiche, *Modeling and reasoning with Bayesian networks*, Cambridge University Press, New York, 2009. 64
7. Gert de Cooman, *Possibility theory I: the measure- and integral-theoretic groundwork*, International Journal of General Systems **25** (1997), 291–323. 2
8. J.-P. Eckmann, S. Oliffson Kamphorst, and D. Ruelle, *Recurrence plots of dynamical systems*, EPL (Europhysics Letters) **4** (1987), 973. 28
9. Richard P. Feynman, *To nemyslíte vážně, pane Feynmane!*, Aurora, 2001, Z anglického originálu „Surely You're Joking, Mr. Feynman!“ přeložil Jan Klíma. 72
10. C. A. Fuchs and R. Schack, *Quantum Bayesian coherence*, ArXiV, 2007, 0906.2187v1 [quant-ph]. 1
11. Charles M. Grinstead and J. Laurie Snell, *Introduction to probability*, American Mathematical Society, 1997. 39
12. R. Jeffrey, *Subjective probability. the real thing*, Cambridge University Press, Cambridge, 2004. 1
13. O. Kallenberg, *Foundations of modern probability*, Springer-Verlag, New York, 1997. 76
14. Andrey Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, Berlin, 1933. 9
15. M. Matsumoto and T. Nishimura, *Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator*, ACM Transactions on Modeling and Computer Simulation **8** (1998), 3–30. 26
16. K. L. Mengersen and R. L. Tweedie, *Rates of convergence of the Hastings and Metropolis algorithms*, Ann. Statist. **24** (1996), 101–121. 41
17. S.P. Meyn and R. I. Tweedie, *Markov chains and stochastic stability*, Springer, London, 1993. 37, 40, 41
18. Christian P. Robert, *The Bayesian choice: From decision-theoretic foundations to computational implementation*, 2 ed., Springer, New York, 2007. 1, 72
19. Christian P. Robert and George Casella, *Monte Carlo statistical methods*, 2 ed., Springer, New York, 2004. 26, 37, 40, 41, 43, 74
20. Erwin Schrödinger, *Die gegenwärtige Situation in der Quantenmechanik*, Naturwissenschaften **23** (1935), 807–812,823–828,844–849. v
21. B. Skyrms, *Coherence*, Scientific Inquiry in Philosophical Perspective (N. Rescher, ed.), pp. 225–242. 1
22. M. Švec, T. Šalát, and T. Neubrunn, *Matematická analýza funkcií reálnej premennej*, Alfa, Bratislava, 1987. 11, 15, 18

23. John von Neumann and Oskar Morgenstern, *Theory of games and economic behavior*, Princeton University Press, Princeton, New Jersey, 1944. 72
24. Lotfi A. Zadeh, *Fuzzy sets as the basis for a theory of possibility*, Fuzzy Sets and Systems **1** (1978), 3–28. 2

Rejstřík