

O POUŽITÍ ŘETĚZCOVÝCH GRAFŮ PRO POPIS STRUKTUR PODMÍNĚNÉ NEZÁVISLOSTI

MILAN STUDENÝ

ABSTRAKT. První část příspěvku obsahuje motivační příklady, které ilustrují možné použití grafických metod popisu struktur podmíněné nezávislosti při zpracování kontingenčních tabulek, v mnohorozměrné statistické analýze a v pravděpodobnostním rozhodování. Poté je formálně zaveden pojem struktury podmíněné nezávislosti a pojem řetězového grafu. Následuje přehled podmínek, které svazují strukturu řetězového grafu s pravděpodobnostní strukturou. Jednoduchý příklad ilustruje tři typy Markovských podmínek a faktorizační podmínku.

Abstract. The first part of the contribution contains motivation examples which illustrate the possible use of graphical methods of description of conditional independence structures in contingency tables processing, in multivariate statistical analysis and in probabilistic decision-making. Then formal definitions of the concept of conditional independence structure and of the concept of chain graph are given. After that an overview of conditions which relate structure of a chain graph and probabilistic structure follows. A simple example illustrates three types of Markovian conditions and a factorization condition.

Резюме. Первая часть статьи содержит мотивирующие примеры, которые показывают возможное применение графических методов описания структур относительной независимости при обработке таблиц сопряженности признаков, в многомерном статистическом анализе и в вероятностном решении. Затем формально вводится понятие структуры относительной независимости и понятие цепного графа. После этого следует обзор условий, которые связывают структуру цепного графа и вероятностную структуру. Простой пример иллюстрирует три типа условий Маркова и условие факторизации.

1. PŘEDMLUVA

Tento příspěvek není klasický matematický text. Je to spíše přehledový článek, jehož účelem je informovat čtenáře o jedné oblasti výzkumu na pomezí matematické statistiky a umělé inteligence. Téma příspěvku bylo tématem přednášky na setkání ROBUST 2000 konaném v září roku 2000 v Nečtinách. Tato přednáška byla pod vlivem (rad) některých tradičních účastníků ROBUSTu koncipována spíše jako didaktická promluva. Proto také velká část tohoto příspěvku je věnována motivačním příkladům, jejichž účelem je osvětlit, kde se ve statistice a umělé inteligenci mohou využít struktury podmíněné nezávislosti a jak se pro jejich popis (a interpretaci) mohou použít (diskrétní) grafy, jejichž uzly odpovídají náhodným veličinám.

2000 *Mathematics Subject Classification.* 62H02 68T30 05C90.

Klíčová slova. Podmíněná nezávislost, kontingenční tabulky, mnohorozměrná statistická analýza, pravděpodobnostní rozhodování, řetězový graf.

Autorův výzkum je podporován z grantů GAČR č. 201/01/1482 a GA AVČR č. A107104 a K101901.

Další zvláštností příspěvku (za kterou se tímto způsobem ještě jednou omlouvám organizátorům ROBUSTu) je více než roční zpoždění způsobené nejruznějšími důvody, zejména mým extrémním časovým zaneprázdněním koncem roku 2000. Proto byl tento příspěvek nabídnut do sborníku ROBUST 2002.

2. MOTIVAČNÍ PŘÍKLADY

V tomto oddíle jsou uvedeny některé uměle vytvořené příklady, které mají demonstrovat, jak se pojem podmíněné nezávislosti může použít pro modelování kvalitativních vztahů mezi veličinami. Grafické metody popisu struktur podmíněné nezávislosti se vyskytují především v následujících třech oblastech:

- při *analýze kontingenčních tabulek*, což spadá do oblasti diskrétní statistiky pracující s kategoriálními daty,
- v *mnohorozměrné statistické analýze*, čímž se rozumí oblast statistiky zabývající se (zejména funkcionálními) vztahy mezi spojitými náhodnými veličinami,
- v *pravděpodobnostním rozhodování*, kteréžto se uplatňuje zejména v oblasti počítačové vědy, přesněji řečeno v oboru umělé inteligence, totiž tam kde je rozhodování za nejistoty založeno na pravděpodobnostním modelu.

Pro snadnější porozumění jsou zde zopakovány některé elementární pojmy z uvedených oblastí.

2.1. Kontingenční tabulky. Více podrobností o grafických modelech používaných při analýze kontingenčních tabulek lze nalézt ve čtvrté kapitole Lauritzenovy knihy [20]. Zde je uveden pouze letmý nástin základních myšlenek.

V této oblasti jde o to klasifikovat určité objekty do nějakých kategorií. Je tedy dána určitá neprázdná konečná množina *klasifikačních kritérií* či *faktorů* N , přičemž každé kritérium $i \in N$ má přiřazenu určitou neprázdnou konečnou množinu *úrovní klasifikace* X_i .

PŘÍKLAD 1. Představme si, že milovník čaje klasifikuje jednotlivé dodávky čaje podle tří kritérií: typ čaje, velikost lístků a chuť. Uvažujme tedy $N = \{a, b, c\}$, přičemž význam faktorů a, b, c je popsán v tabulce 1.

FAKTOR	KRITERIUM	MNOŽINA ÚROVNÍ KLASIFIKACE
a	typ čaje	$X_a = \{\text{Zelený, Černý}\}$
b	velikost lístků	$X_b = \{\text{Malé, Střední, Velké}\}$
c	chuť	$X_c = \{\text{Přijatelná, Nepřijatelná}\}$

Tab. 1 Kritéria klasifikace čaje.

Buňky příslušné kontingenční tabulky pak jsou prvky kartézského součinu $X_N = \prod_{i \in N} X_i$. Je-li $\emptyset \neq A \subseteq N$ pak příslušná *marginální tabulka* má za buňky prvky kartézského součinu $X_A \equiv \prod_{i \in A} X_i$ (níže budu používat symbol X_A coby zkrácené označení $\prod_{i \in A} X_i$). Data jsou předpokládána ve formě posloupnosti objektů, které jsou již klasifikovány do příslušných buněk. V užším matematickém smyslu je to tedy posloupnost ohodnocení těchto objektů, tedy prvků X_N . Z těchto dat se pak vytvoří *kontingenční tabulka* tak, že se do každé buňky napíše příslušný počet výsledků v dané posloupnosti.

PŘÍKLAD 2. V situaci popsané v příkladě 1 uvažujeme například posloupnost šesti ohodnocení dodávek čaje (faktory jsou řazeny v pořadí a, b, c):

$$d_1 = (Z, V, P), \quad d_2 = (\check{C}, S, N), \quad d_3 = (\check{C}, S, P), \\ d_4 = (Z, V, P), \quad d_5 = (\check{C}, M, N), \quad d_6 = (\check{C}, M, P),$$

kde Z zkracuje Zelený a podobně tomu je s ostatními klasifikačními úrovněmi. Protože klasifikujeme podle tří kritérií, odpovídající kontingenční tabulka je třírozměrná; v tabulce 2 je však zapsána ve formě dvou tabulek.

$a = \check{C}$	$b = M$	$b = S$	$b = V$
$c = P$	1	1	0
$c = N$	1	1	0

$a = Z$	$b = M$	$b = S$	$b = V$
$c = P$	0	0	2
$c = N$	0	0	0

Tab. 2 Třírozměrná kontingenční tabulka.

Statistická interpretace těchto dat je následující. Považujeme je za realizaci náhodného výběru z nějakého rozdělení na X_N . Přesněji řečeno, považujeme je za realizace náhodných veličin $\zeta_1(\omega), \dots, \zeta_n(\omega)$, $n \geq 1$ s hodnotami v konečné množině X_N a o těchto náhodných veličinách předpokládáme, že jsou nezávislé a stejně rozdělené a mají nějaké (společné) rozdělení P . Toto diskrétní rozdělení je určeno svojí hustotou p vůči aritmetické míře, to jest souborem hodnot $p(x) \in [0, 1]$, $x \in X_N$ splňujícím $\sum_{x \in X_N} p(x) = 1$.

POZNÁMKA 1. Pro doplnění dodávám známý fakt, že za výše uvedených předpokladů se dá snadno spočítat pravděpodobnost výskytu každé jednotlivé tabulky $\{n(x); x \in X_N\}$. Jedná se o multinomické rozdělení, jehož parametry jsou (teoretické) pravděpodobnosti jednotlivých buněk, přesněji

$$\text{Pravděpodobnost } \{\omega; \forall x \in X_N \ n(x, \omega) = n(x)\} = \frac{n!}{\prod_{x \in X_N} n(x)!} \cdot \prod_{x \in X_N} p(x)^{n(x)},$$

kde $n(x, \omega)$ je počet $i \in \{1, \dots, n\}$ splňujících $\zeta_i(\omega) = x$.

Nicméně, o rozdělení P se činí (a později ověřují) další předpoklady. Tím se vlastně přijímá nějaký *statistický model* (čímž obecně rozumím nějak vymezenou třídu pravděpodobnostních rozdělení). Velmi populární třída modelů používaných při analýze kontingenčních tabulek jsou *log-lineární modely*. Takový model je specifikován následovně. Předpokládá se, že P je 'striktně kladná', to jest $p(x) > 0$ pro každé $x \in X_N$. Potom lze aplikovat logaritmus a psát

$$(1) \quad \ln p(x) = \sum_{A \subseteq N} \lambda_A(x_A),$$

kde $x_A = [x_i]_{i \in A}$ označuje projekci vektoru $x = [x_i]_{i \in N} \in X_N$ na X_A a reálné funkce $\lambda_A : X_A \rightarrow \mathbb{R}$ se nazývají *interakce*. Po přijetí určitých standardisačních podmínek (t.j. požadavků na funkce λ_A , $A \subseteq N$) je rozklad (1) jednoznačný. Konkrétní log-lineární model je pak určen požadavkem, že některé interakce se nulují. Příkladem standardisačních podmínek je požadavek

$$(2) \quad \forall B \subset A \quad \forall y \in X_B \quad \sum \{\lambda_A(z); z \in X_A, z_B = y\} = 0,$$

kde $B \subset A$ značí $B \subseteq A$, $B \neq A$; tedy požadavek 'nulování vlastních marginál' pokud interakce interpretujeme jakožto znaménkové míry. Dodejme, že jednotlivé

interakce se poté klasifikují: rozlišují se hlavní interakce (když $|A| = 2$), interakce druhého řádu (když $|A| = 3$) a tak dále. Toto názvosloví pochází asi z fyziky neboť tyto statistické modely byly inspirovány právě (statistickou) fyzikou. Nicméně, z matematického hlediska není rozklad (1) nic jiného než specifická parametrisace třídy (striktně kladných) rozdělení na X_N .

Hierarchický log-lineární model je pak určen takzvanou generující třídou $\mathcal{G} \subseteq \mathcal{P}(N)$ (kde $\mathcal{P}(N)$ označuje potenční množinu množiny N), čímž se rozumí systém neporovnatelných podmnožin N (t.j. pro různé $A, B \in \mathcal{G}$ nenastane ani $A \subset B$ ani $B \subset A$). Pak lze zavést příslušný generovaný ‘dolů usměrněný’ soubor podmnožin

$$\mathcal{G}^\downarrow = \{B \subseteq N; \exists A \in \mathcal{G} \ B \subseteq A\},$$

a příslušný (hierarchický log-lineární) model se určí požadavkem, že interakce se nulují pro množiny mimo \mathcal{G}^\downarrow , to jest

$$(3) \quad \{P; \text{v rozkladu (1) } \lambda_C \equiv 0 \text{ kdykoliv } C \subseteq N, C \notin \mathcal{G}^\downarrow\}.$$

PŘÍKLAD 3. V situaci popsané v příkladě 1, t.j. $N = \{a, b, c\}$, můžeme například mít $\mathcal{G} = \{\{a, c\}, \{b, c\}\}$. To vede ke konkrétnímu log-lineárnímu modelu

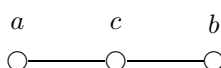
$$(4) \quad \{P > 0; \text{v rozkladu (1) } \lambda_{\{a,b,c\}} \equiv 0 \text{ a } \lambda_{\{a,b\}} \equiv 0\}.$$

Speciální třídou hierarchických log-lineárních modelů jsou *grafické modely*, které byly zavedeny počátkem 70-tých let dvacátého století Goodmanem a Habermanovou [9, 10] (pro podrobnější přehled viz [32]). U těchto modelů je generující třída souborem klik nějakého neorientovaného grafu.

DEFINICE 1. Neorientovaný graf G je určen neprázdnou konečnou množinou *uzlů* N a množinou \mathcal{E} dvouprvkových podmnožin N zvaných *hran* (jinak řečeno, množinou dvojic $\{u, v\}$ kde $u \neq v$). Množina uzlů $A \subseteq N$ je *úplná* v G jestliže $\{u, v\} \in \mathcal{E}$ kdykoliv $u, v \in A$, $u \neq v$. *Klikou* grafu G se rozumí maximální (vzhledem k inklusi) úplná množina uzlů.

Takže, je-li \mathcal{G} soubor klik nějakého neorientovaného grafu, pak jím generovaný dolů usměrněný systém \mathcal{G}^\downarrow není nic jiného než soubor úplných množin onoho neorientovaného grafu. Log-lineární model (4) z příkladu 3 je také příklad grafického modelu:

PŘÍKLAD 4. Třída $\mathcal{G} = \{\{a, c\}, \{b, c\}\}$ je souborem klik neorientovaného grafu z obrázku 1.



Obr. 1 Jednoduchý neorientovaný graf.

Klíčovým sdělením tohoto oddílu je, že výše zmíněné

grafické (log-lineární) modely je možno ekvivalentně zavést coby modely určené (zadanou) strukturou podmíněné nezávislosti.

To znamená, že třídu rozdělení na X_N , které tvoří příslušný statistický model, to jest třídu (3), lze ekvivalentně zavést jako třídu těch (striktně kladných) rozdělení, která splňují jistá nezávislostní tvrzení. *Nezávislostním tvrzením* (či údajem) se rozumí požadavek či fakt, že platí nějaký konkrétní vztah podmíněné nezávislosti (vzhledem k uvažovanému rozdělení). Každý takový nezávislostní údaj je popsán trojicí

$\langle A, B|C \rangle$ navzájem disjunktních podmnožin N . Zápis, který se používá pro označení takového nezávislostního tvrzení je $A \perp\!\!\!\perp B | C [P]$. Znamená že A je podmíněně nezávislé na B dáno (= podmíněno) C vzhledem k rozdělení P . Definice pojmu podmíněná nezávislost je v § 3.1. Cílem tohoto oddílu je pouze naznačit, jakou má pojem podmíněné nezávislosti souvislost s teorií kontingenčních tabulek; proto vynechávám i důkaz výše zmíněné teze.

Další důležité sdělení (opět bez důkazu) je že soubor nezávislostních tvrzení, která definují grafický (log-lineární) model indukovaný neorientovaným grafem G , lze určit přímo z tohoto grafu pomocí určitého grafického kritéria. Rozdělení splňující tato nezávislostní tvrzení určená grafem G se pak nazývají *Markovská rozdělení* vzhledem k G .

DEFINICE 2. *Cestou* v neorientovaném grafu $G = (N, \mathcal{E})$ se rozumí posloupnost u_1, \dots, u_n , $n \geq 1$ různých uzlů grafu taková, že pro každé $i = 1, \dots, n-1$ platí $\{u_i, u_{i+1}\} \in \mathcal{E}$. Nechť A, B, C jsou navzájem disjunktí podmnožiny N . Řekneme, že C separuje A a B v G jestliže pro každou cestu u_1, \dots, u_n , $n \geq 1$ v G takovou, že $u_1 \in A$ a $u_n \in B$ existuje u_i , $1 < i < n$ takové, že $u_i \in C$. V tomto případě říkáme, že $\langle A, B|C \rangle$ je representována v G podle *separačního kritéria* (a píšeme $A \perp\!\!\!\perp B | C [G]$). Rozdělení P na X_N je (*globálně*) *Markovské* vzhledem k G jestliže splňuje všechna nezávislostní tvrzení $A \perp\!\!\!\perp B | C [P]$ pro trojice $\langle A, B|C \rangle$ representované v G .

PŘÍKLAD 5. Podle kritéria z předchozí definice jsou v grafu G na obrázku 1 representovány trojice $\langle \{a\}, \{b\} | \{c\} \rangle$ a $\langle \{b\}, \{a\} | \{c\} \rangle$ a pak tzv. triviální trojice $\langle A, B|C \rangle$ v nichž je buď $A = \emptyset$ či $B = \emptyset$. Protože $\{b\} \perp\!\!\!\perp \{a\} | \{c\} [P]$ je ekvivalentní $\{a\} \perp\!\!\!\perp \{b\} | \{c\} [P]$ pro každé rozdělení P a triviální trojice lze vynechat (viz vysvětlení v pozdější poznámce 6) má třída Markovských rozdělení vzhledem k G tvar

$$(5) \quad \{ P > 0; \{a\} \perp\!\!\!\perp \{b\} | \{c\} [P] \}.$$

Klíčové sdělení tohoto oddílu tedy v tomto případě říká, že třídy (4) a (5) se rovnají.

2.2. Mnohorozměrná statistická analýza. Grafické modely pro popis vztahů mezi veličinami jsou také používány v mnohorozměrné statistické analýze, přesněji řečeno v té části tohoto oboru, kde se popisují vztahy mezi reálnými Gaussovskými náhodnými veličinami pomocí lineárních rovnic. V těchto modelech jsou (některé) náhodné veličiny vlastně perturbovanými lineárními funkcemi jiných. Anglicky se modely tohoto typu nazývají *linear structural equation models* a typickým příkladem jsou tzv. LISREL modely [13]. Více podrobností o tom, kterak lze modely tohoto typu interpretovat coby grafické modely struktur podmíněné nezávislosti lze nalézt v článcích [2, 15].

POZNÁMKA 2. V tomto příspěvku záměrně neopakuji komplikovanou definici LISREL modelu. Tyto složité hierarchicky strukturované modely zahrnují mnoho náhodných veličin, z nichž pouze některé jsou ‘pozorovatelné’ a ostatní hrají roli skrytých veličin v příslušném ‘generujícím’ mechanismu (například chybové veličiny). Navíc, pozorovatelné veličiny se ještě dělí na *exogenní* (či vysvětlující, anglicky *explanatory*) a *endogenní* (či reagující, anglicky *response*). Zájemce o podrobnosti odkazují na článek [15], který obsahuje jednu z matematicky důsledných definic LISREL modelu včetně přesné specifikace předpokladů. V citovaném článku je navíc ukázáno, že každému modelu tohoto typu lze přiřadit jistý (poměrně komplikovaný) *reciprokový graf*, který vlastně popisuje určitou strukturu podmíněné nezávislosti.

Následující motivační příklad je velmi speciálním případem výše zmíněných obecných modelů. Ve statistické analýze jsou tyto speciální modely někdy označovány termínem *rekursivní (lineární) modely*. Rekursivní modely mohou být popsány pomocí acyklických orientovaných grafů. Dodejme, že terminologie není jednotná a že někteří autoři [2] označují tyto speciální modely termínem ‘regresní model’ snad kvůli formální podobnosti matematické formulace s klasickou úlohou nalezení odhadu koeficientů lineární regrese (viz [1], str. 33).

PŘÍKLAD 6. Uvažujme pět pozorovatelných veličin X_a, X_b, X_c, X_d, X_e a chybové veličiny odpovídající těmto veličinám. Předpokládejme následující vztahy:

$$\begin{aligned} X_c &= X_a + 3 \cdot X_b + \varepsilon_c, \\ X_d &= 2 \cdot X_b + \varepsilon_d, \\ X_e &= X_c + X_d + \varepsilon_e, \end{aligned}$$

kde ε_c popisuje ‘chybu měření’ veličiny X_c atd. Tyto veličiny lze uspořádat do posloupnosti tak, že každá z nich je lineární funkcí předchozích perturbovaná svojí ‘chybovou’ veličinou:

$$(6) \quad \begin{aligned} X_a &= \varepsilon_a, \\ X_b &= \varepsilon_b, \\ X_c &= \alpha \cdot X_a + \beta \cdot X_b + \varepsilon_c, \\ X_d &= \gamma \cdot X_b + \varepsilon_d, \\ X_e &= \delta \cdot X_c + \varrho \cdot X_d + \varepsilon_e, \end{aligned}$$

kde $\alpha = 1, \beta = 3, \gamma = 2, \delta = \varrho = 1$.

Obecně je rekursivní model určen systémem lineárních rovnic pro pozorovatelné veličiny $X_1, \dots, X_n, n \geq 1$, které jsou uspořádány do posloupnosti tak, že každá veličina je lineární funkcí předchozích veličin a své chybové veličiny:

$$(7) \quad \begin{aligned} X_1(\omega) &= \varepsilon_1(\omega), \\ X_2(\omega) &= \alpha_1^2 \cdot X_1(\omega) + \varepsilon_2(\omega), \\ &\dots \\ X_i(\omega) &= \alpha_1^i \cdot X_1(\omega) + \dots + \alpha_{i-1}^i \cdot X_{i-1}(\omega) + \varepsilon_i(\omega) \\ &\dots \\ X_n(\omega) &= \alpha_1^n \cdot X_1(\omega) + \dots + \alpha_{n-1}^n \cdot X_{n-1}(\omega) + \varepsilon_n(\omega) \end{aligned}$$

přičemž některé z koeficientů $\alpha_j^i, j < i$ nemusí být uvedeny (neboli implicitně se předpokládá, že se neuvedené koeficienty identicky nulují stejně jako koeficienty α_j^i pro $j \geq i$). O *uvedených* koeficientech se naopak předpokládá, že mohou být nenulové, ale jsou neznámé. O chybových veličinách se předpokládá, že jsou nezávislé Gaussovské s nenulovým rozptylem a řekněme nulovou střední hodnotou (střední hodnoty vlastně nehrají roli z hlediska zde studovaných kvalitativních vztahů náhodných veličin), to jest

$$\mathbf{E}(\omega) \equiv [\varepsilon_1(\omega), \dots, \varepsilon_n(\omega)] \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$$

kde $\mathbf{0}$ je nulový vektor dimenze n a $\mathbf{\Gamma}$ je $n \times n$ -matice se striktně kladnými čísly na diagonále a nulami mimo diagonálu. Ty pozorovatelné veličiny, které jsou (díky identickému nulování některých koeficientů) funkcemi pouze svých chybových veličin se potom nazývají *exogenní* zatímco ty zbývající *endogenní*.

POZNÁMKA 3. Ovšem i výše uvedené předpoklady jsou někdy modifikovány (viz [3]) obecnějším předpokladem že náhodný vektor $\mathbf{E}(\omega)$ má nedegenerované Gaussovské rozdělení s neznámou kovarianční maticí o níž se případně ví (jinými slovy, předpokládá se), že některé její koeficienty jsou nulové. Odpovídající struktury podmíněné nezávislosti jsou potom popisovány pomocí tzv. řetězcových grafů s alternativní interpretací.

Výše zmíněné předpoklady na systém (7) umožňují vypočítat ‘rekursivně’ pozorovatelné veličiny jakožto lineární funkce chybových veličin (tento fakt možná motivoval název ‘rekursivní model’).

PŘÍKLAD 7. Za situace popsané v příkladě 6 můžeme snadno postupným dosazováním upravit (6) následovně:

$$(8) \quad \begin{aligned} X_a &= \varepsilon_a, \\ X_b &= \varepsilon_b, \\ X_c &= \alpha \cdot \varepsilon_a + \beta \cdot \varepsilon_b + \varepsilon_c, \\ X_d &= \gamma \cdot \varepsilon_b + \varepsilon_d, \\ X_e &= \alpha \cdot \delta \cdot \varepsilon_a + [\beta \cdot \delta + \gamma \cdot \varrho] \cdot \varepsilon_b + \delta \cdot \varepsilon_c + \varrho \cdot \varepsilon_d + \varepsilon_e. \end{aligned}$$

Jinými slovy, lze napsat $\mathbf{X}(\omega) = \mathbf{A} \cdot \mathbf{E}(\omega)$ kde \mathbf{A} je (nenáhodná) dolní trojúhelníková matice s jednotkami na diagonále, tedy nutně regulární matice. Odtud plyne, že náhodný vektor \mathbf{X} má nedegenerované Gaussovské rozdělení, přesněji

$$\mathbf{X}(\omega) \equiv [X_1(\omega), \dots, X_n(\omega)] \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}) \quad \text{kde} \quad \mathbf{\Sigma} = \mathbf{A} \cdot \mathbf{\Gamma} \cdot \mathbf{A}^\top,$$

a to pro každou možnou volbu parametrů, t.j. koeficientů α_j^i , $j < i$ uvedených v původním systému perturbovaných lineárních rovnic (7) a každou volbu (striktně kladných) diagonálních prvků matice $\mathbf{\Gamma}$. Další myšlenkový krok je, že rekursivní model lze interpretovat jako statistický model, to jest třídu pravděpodobnostních rozdělení určenou následovně:

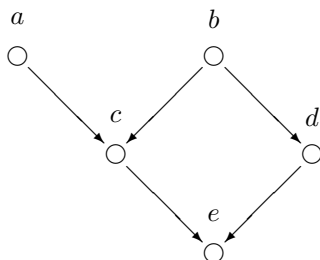
$$(9) \quad \{ P; \quad P \text{ je rozdělení vektoru } \mathbf{X}(\omega) \text{ splňující (7) pro nějaké} \\ \text{hodnoty uvedených koeficientů } \alpha_j^i, j < i \text{ a prvků matice } \mathbf{\Gamma} \}.$$

Je samozřejmé, že (9) nemusí zahrnovat všechny nedegenerované Gaussovské rozdělení. Příslušná omezení lze formulovat v termínech kovarianční matice $\mathbf{\Sigma}$ jakožto nějaké algebraické vztahy mezi jejími prvky. Standardním tvarem takového vztahu může být požadavek, že se nuluje některý prvek matice $\mathbf{\Sigma}$ anebo prvek tzv. *podmíněné kovarianční matice*, to jest Schurova komplementu $\mathbf{\Sigma}_{A|C} \equiv \mathbf{\Sigma}_{A \cdot A} - \mathbf{\Sigma}_{A \cdot C} \cdot (\mathbf{\Sigma}_{C \cdot C})^{-1} \cdot \mathbf{\Sigma}_{C \cdot A}$ pro nějaké disjunktní množiny $A, C \subseteq \{1, \dots, n\}$ (zde $\mathbf{\Sigma}_{A \cdot C}$ označuje příslušnou podmatici $\mathbf{\Sigma}$). Zásadní pozorování je, že tyto algebraické vztahy (mezi prvky kovarianční matice $\mathbf{\Sigma}$) odpovídají právě nezávislostním tvrzením a tedy

rekursivní modely lze ekvivalentně popsat coby statistické modely se zadanou strukturou podmíněné nezávislosti.

Navíc, tuto strukturu podmíněné nezávislosti lze charakterizovat s pomocí jistého acyklického orientovaného grafu. Tento graf se vytvoří snadno na základě původního systému lineárních rovnic následovně: množina jeho uzlů je $\{1, \dots, n\}$ a graf má šipku $j \rightarrow i$ právě když v systému (7) je uveden koeficient α_j^i . S pomocí tohoto grafu lze podle odpovídajících grafických kritérií určit soubor nezávislostních údajů popisujících tuto strukturu podmíněné nezávislosti. Tímto způsobem lze každému acyklickému orientovanému grafu přiřadit odpovídající třídu Markovských rozdělení (v uvažovaném distribučním rámci nedegenerovaných Gaussovských rozdělení).

Podrobný důkaz tvrzení, že třída rozdělení vektorů $[X_1, \dots, X_n]$ splňujících (7) se shoduje s třídou nedegenerovaných Gaussovských rozdělení, která jsou Markovská vůči takto zkonstruovanému grafu, je relativně dlouhý a neúměrně by prodloužil tento příspěvek. Z toho důvodu a též kvůli snazší srozumitelnosti textu důkaz vynechávám.



Obr. 2 Acyklický orientovaný graf.

PŘÍKLAD 8. Systém rovnic (6) popsany v příkladě 6 indukuje acyklický orientovaný graf G na obrázku 2. Lokální Markovská podmínka pro G při uspořádání uzlů a, b, c, d, e (viz definice 11 v § 4.2) identifikuje tři netriviální nezávislostní údaje. Odpovídající třídu Markovských rozdělení vůči G lze pak popsat takto:

$$(10) \quad \{P \text{ nedegenerované Gaussovské ; splňující } \{b\} \perp\!\!\!\perp \{a\} \mid \emptyset [P], \\ \{d\} \perp\!\!\!\perp \{a, c\} \mid \{b\} [P] \text{ a } \{e\} \perp\!\!\!\perp \{a, b\} \mid \{c, d\} [P]\}.$$

Dodejme, že Markovská rozdělení vůči G splňují i další nezávislostní tvrzení, která lze identifikovat pomocí tzv. globální Markovské podmínky (viz § 4.3). Účelem tohoto příkladu je pouze ilustrovat hlavní tezi tohoto oddílu, podle které je třída rozdělení (9) totožná s třídou (10).

2.3. Pravděpodobnostní rozhodování. Více podrobností o pravděpodobnostních expertních systémech lze nalézt v knize [5]; širší otázce rozhodování za nejistoty je věnována kniha [11]. Rozhodování za nejistoty spadá do oblasti umělé inteligence a je teoretickým základem pro vývoj (počítačových) systémů pro podporu rozhodování. Tyto systémy mohou být používány například v medicíně.

Rovněž v této oblasti je zpravidla určena nějaká neprázdná konečná množina *příznaků* (symptomů) či *faktorů* N a každý faktor $i \in N$ má určenu *množinu možných hodnot* X_i , což je neprázdná konečná množina. Často bývá množina X_i dvouprvková a hodnoty mají význam přítomnosti či absence určitého příznaku anebo kladné či záporné odpovědi na nějakou otázku. Nicméně, v medicíně může být počet možných hodnot nějakého symptomu i poměrně vysoký - například pět až deset stupňů nějaké škály. To ovšem není případ následujícího uměle zkonstruovaného příkladu (snad aspoň trochu aktuálního v době konání ROBUSTu 2000), jehož účelem je pouze volně ilustrovat pojmy z této oblasti.

PŘÍKLAD 9. Představme si, že účastník setkání ROBUST 2000 má rozhodovací problém jaký program zvolit pro středeční odpoledne po přednášce o řetězcových grafech v souvislosti s plánovaným výletem na zámek Manětín. Předpokládejme, že své rozhodnutí založí na ranním pohledu na oblohu a případně poledním zjištěním, jaký terén je v okolí ubytovacího zařízení v Nečtinách. Uvažujme tedy $N = \{a, b, c, d\}$ přičemž význam faktorů je popsán v tabulce 3.

FAKTOR	VÝZNAM	MNOŽINA MOŽNÝCH HODNOT
a	stav oblohy ráno	$X_a = \{\text{zataženo, oblačno, jasno}\}$
b	vlhkost terénu v poledne	$X_b = \{\text{mokro, sucho}\}$
c	rozhodnutí stran výletu	$X_c = \{\text{jít, nejít}\}$
d	uvidí kvečeru Manětín?	$X_d = \{\text{ANO, NE}\}$

Tab. 3 Dilema účastníka setkání ROBUST 2000.

POZNÁMKA 4. Expertní systémy se obecně dělí na systémy založené na logickém odvozování (t.j. bez nejistoty) a systémy pracující s nejistotou. Zatímco v první skupině expertních systémů se uplatňuje zejména matematická logika, v druhé skupině záleží na tom, jakým způsobem je nejistota matematicky popsána a formulována. Teorie pravděpodobnosti sice asi nebyla historicky první ‘kalkul’ pro popis nejistoty v této oblasti umělé inteligence, ale v průběhu 80-tých a 90-tých let dvacátého století postupně převzala dominantní úlohu mezi kalkulami nejistoty (alespoň podle subjektivního autorova názoru s kterýmžto jiní výzkumníci pracující v oblasti rozhodování za nejistoty nemusí nutně souhlasit). Proto diskrétní teorie pravděpodobnosti dosud hraje klíčovou roli v této oblasti. Nicméně, v poslední době se v oblasti umělé inteligence využívají i jiné nástroje pro popis nejistoty, například tzv. teorie možností (anglicky *possibility theory*) či Dempster-Shaferova teorie evidence. Také v rámci těchto alternativních kalkulů nejistoty byl zaveden a studován pojem podmíněné nezávislosti.

Klasický (logický) expertní systém zpracovává znalosti ve formě *inferenčních pravidel*, které mají tvar implikací. V systémech pracujících s nejistotou je však bezpodmínečná platnost takových implikací zpochybněna a platnost každého pravidla je ohodnocena *mírou věrohodnosti*. Tato míra věrohodnosti je obvykle normalizována, většinou tak aby to bylo číslo mezi 0 a 1 (ale objevily se i jiné způsoby normalizace). V pravděpodobnostním přístupu k popisu nejistoty je pak míra věrohodnosti interpretována jakožto *podmíněná pravděpodobnost*, že nastane nějaký jev za předpokladu, že nastal jiný jev.

PŘÍKLAD 10. Příkladem inferenčního pravidla v případě popsaném v příkladě 9 může být implikace

$$\text{POKUD } a = \text{zataženo} \quad \text{PAK } b = \text{mokro}.$$

V systému pracujícím s nejistotou se platnost takové implikace ještě může ohodnotit mírou věrohodnosti, například

$$\text{POKUD } a = \text{zataženo} \quad \text{PAK } b = \text{mokro} \quad \text{S VAHOU } 0.9,$$

což lze zapsat ve formě $a = \text{zataženo} \longrightarrow b = \text{mokro} [0.9]$. Toto inferenční pravidlo můžeme interpretovat jakožto podmíněnou pravděpodobnost

$$\text{Pravděpodobnost } (b = \text{mokro} \mid a = \text{zataženo}) = 0.9.$$

Pokud má mít výše zmíněná interpretace smysl je nutné, aby se předpokládala existence takového pravděpodobnostního (simultánního) rozdělení P na $X_N = \prod_{i \in N} X_i$, že jednotlivé zadané podmíněné pravděpodobnosti (= míry věrohodnosti) se vypočítají z P .

POZNÁMKA 5. Výše zmíněná otázka je vlastně otázka *konsistence* souboru vstupních (expertních) znalostí z hlediska pravděpodobnostního přístupu. Mnohorozměrné rozdělení P na X_N hraje roli *globální znalostní base*, zatímco dílčí znalosti ve formě

‘zadaných’ podmíněných pravděpodobností hrají roli *dílčích kvantitativních expertních znalostí*. Tyto dílčí znalosti mohou být přetransformovány na omezení na méněrozměrné marginály P . Proto jedna ze základních úloh souvisejících s tzv. *intensionálním přístupem* k pravděpodobnostním expertním systémům [11] je otázka existence simultánního pravděpodobnostního rozdělení s předepsanými méněrozměrnými marginály. Problém konsistence intensionálního pravděpodobnostního expertního systému tudíž souvisí s diskrétní verzí tzv. *marginálního problému* [12].

Dodejme, že prosazení teorie pravděpodobnosti coby nástroje pro popis nejistoty v umělé inteligenci koncem 80-tých let dvacátého století nebylo zdaleka snadné. Výzkumníci v oblasti umělé inteligence měli velkou nedůvěru v plně pravděpodobnostní přístup. Pokud pomineme konservatismus, pak hlavní námitka byla, že takový přístup není výpočetně zvládnutelný. Tuto nedůvěru se podařilo překonat právě s pomocí grafických modelů a metody lokálních výpočtů zmíněné níže.

Nicméně, výše zmíněná ohodnocená inferenční pravidla představují pouze jeden typ vstupní informace od expertů. Druhým typem informace, který může tvůrce takového rozhodovacího systému získat od expertů je *kvalitativní informace* o tom, jak se faktory navzájem ovlivňují. Tyto informace mohou být ve formě *irelevančních tvrzení* typu „při znalosti hodnoty prvního symptomu hodnota druhého symptomu není ovlivněna hodnotou třetího symptomu“. V podstatě bez předpokladů tohoto typu není možno z teoretického hlediska oprávněně omezit třídu simultánních rozdělení na taková rozdělení, která jsou reprezentovatelná v paměti počítače v redukované formě (totiž ve formě souboru svých méněrozměrných marginál). Podstatné pozorování je, že v rámci pravděpodobnostního přístupu jsou irelevanční tvrzení výše zmíněného typu interpretovatelná jakožto určité nezávislostní údaje.

PŘÍKLAD 11. V situaci popsané v příkladě 9 můžeme na základě významu jednotlivých faktorů provést například následující úvahu. Ví-li onen účastník ROBUSTu zda je v poledne sucho či mokro (t.j. zná-li hodnotu faktoru b), pak na jeho/její rozhodnutí, zda půjde na výlet (t.j. na hodnotu faktoru c) již nemá bezprostřední vliv stav oblohy ráno (t.j. hodnota faktoru a). Samozřejmě, stav oblohy ráno (t.j. faktor a) ovlivňuje rozhodování onoho hypotetického účastníka (t.j. faktor c) ale pouze zprostředkovaně přes hodnotu faktoru b , který je pro onoho účastníka rozhodující. Toto irelevanční tvrzení pak odpovídá v příslušné interpretaci nezávislostnímu tvrzení $\{c\} \perp\!\!\!\perp \{a\} \mid \{b\} [P]$, kde P je příslušné simultánní rozdělení na X_N .

Avšak získání kvalitativní informace od expertů ve formě seznamu irelevančních tvrzení je dosti těžkopádné. Proto v praxi tzv. znalostní inženýři, kteří vytvářejí pravděpodobnostní expertní systémy, chtějí po expertech spíše, aby vyjádřili strukturální informaci ve formě nějakého grafu, který vystihuje představu experta o tom, jak se uvažované faktory ovlivňují. Takto získaná kvalitativní informace se pak může interpretovat jako omezení třídy uvažovaných simultánních rozdělení P na třídu rozdělení, se zadanými nezávislostními údaji. To lze shrnout takto:

soubor irelevančních tvrzení využívaný při konstrukci pravděpodobnostních expertních systémů lze interpretovat jakožto přijetí nějakého modelu se zadanou strukturou podmíněné nezávislosti.

PŘÍKLAD 12. V situaci popsané v příkladě 9 může expert vyjádřit strukturální informaci ve formě acyklického orientovaného grafu na obrázku 3. Obrázek lze volně interpretovat třeba následovně: ranní stav oblohy (t.j. faktor a) přináší relevantní informaci o tom zda bude pršet a tedy ovlivňuje s určitou nejistotou, zda bude terén v poledne mokrá či suchý (t.j. faktor b). Rozhodnutí o výletu (t.j. faktor c) pak

závisí přímo na tomto faktoru a to, zda účastník k večeru uvidí zámek Manětín (t.j. faktor d) zase závisí víceméně na tomto rozhodnutí. Lokální Markovská podmínka vzhledem k tomuto grafu (viz definice 11 v § 4.2) určuje dva netriviální nezávislostní údaje. Odpovídající třída Markovských rozdělání je tedy

$$(11) \quad \{P \text{ na } X_N; \{c\} \perp\!\!\!\perp \{a\} \mid \{b\} [P], \{d\} \perp\!\!\!\perp \{a, b\} \mid \{c\} [P]\}.$$



Obr. 3 Jednoduchý acyklický orientovaný graf.

Význam pojmu podmíněná nezávislost v pravděpodobnostním rozhodování si uvědomil Pearl v polovině 80-tých let dvacátého století a posléze jej zdůraznil ve své knize [24]. Kromě výše zmíněné otázky interpretace tohoto pojmu se ukázala důležitou i otázka *representace znalostní base v paměti počítače*. V této souvislosti vyšel najevo kardinální význam jedné zvláštní třídy grafických modelů, totiž tzv. *rozložitelných modelů*, které odpovídají chordálním neorientovaným grafům.

DEFINICE 3. *Cyklem* v neorientovaném grafu $G = (N, \mathcal{E})$ se rozumí posloupnost $u_1, \dots, u_n, u_{n+1} = u_1$, $n \geq 3$ uzlů grafu kde u_1, \dots, u_n jsou různé a pro každé $i = 1, \dots, n$ platí $\{u_i, u_{i+1}\} \in \mathcal{E}$. Číslo n je pak *délka cyklu*. *Tětivou* cyklu délky $n \geq 4$ se rozumí hrana $\{u_i, u_j\} \in \mathcal{E}$ kde $1 < j - i < n - 1$ a $1 \leq i, j \leq n$. Graf G je *chordální* jestliže v něm každý cyklus délky alespoň 4 má tětivu.

Důležitý fakt je, že kliky chordálního (někdy se říká *triangulovaného*) grafu lze uspořádat do posloupnosti C_1, \dots, C_m , $m \geq 1$ splňující podmínku, která se anglicky nazývá *running intersection property*:

$$\forall i > 1 \quad \exists j < i \quad C_i \cap \left(\bigcup_{k < i} C_k \right) \subseteq C_j,$$

kteřá má klíčový význam pro použití *metody lokálních výpočtů* (v angličtině *local computation method*) [17, 5]. Podstatou této metody je, že simultánní mnohorozměrné rozdělání na X_N je v počítači representováno úsporně ve formě svých méně-rozměrných marginál (či potenciálů), které odpovídají klikám nějakého chordálního neorientovaného grafu, což podstatně snižuje paměťové nároky. Navíc, veškeré výpočty se provádějí manipulací s těmito méněrozměrnými marginálami či potenciály. Tento přístup je dotažen až k metodě, která používá jisté speciální ‘nadgrafy’, nazývané anglicky *junction trees* kteréžto vážený recenzent navrhuje nazývat v češtině půvabným termínem *stromy spojení*. Stromy spojení mají za uzly kliky chordálního grafu a výpočty se realizují ve formě ‘posílání zpráv’ mezi uzly tohoto stromu spojení – viz [5].

PŘÍKLAD 13. Model struktury podmíněné nezávislosti zmíněný v příkladě 12 lze ekvivalentně popsat pomocí neorientovaného grafu s uzly a, b, c, d , který vznikne z orientovaného grafu na obrázku 3 vypuštěním orientace hran. Tento graf je chordální. Markovské rozdělání vůči chordálnímu neorientovanému grafu pak lze ekvivalentně popsat pomocí jisté součinné formule [30], která vyjadřuje toto rozdělání

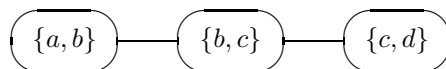
s pomocí jeho marginál, které odpovídají klikám onoho grafu. V uvažovaném příkladě to konkrétně znamená, že třídu (11) lze popsat následovně:

$$(12) \quad \left\{ P \text{ pravděp. rozdělení na } X_N; p = \frac{P_{\{a,b\}} \cdot P_{\{b,c\}} \cdot P_{\{c,d\}}}{P_{\{b\}} \cdot P_{\{c\}}} \right\},$$

kde p_S pro $S \subseteq \{a, b, c, d\}$ označuje marginální hustotu P pro S určenou vzorcem

$$p_S(x) = \sum_{y \in X_{N \setminus S}} p(x, y) \quad \text{pro } x \in X_S$$

na základě hustoty p rozdělení P vzhledem k aritmetické míře na X_N . Odpovídající strom spojený je na obrázku 4. V uvedeném příkladě je paměťový nárok na reprezentaci obecného pravděpodobnostního rozdělení na X_N $3 \times 2 \times 2 \times 2 - 1 = 23$ čísel, zatímco paměťový nárok na reprezentaci marginál P pro $\{a, b\}$, $\{b, c\}$ a $\{c, d\}$ pouze $(3 \times 2 - 1) + (2 \times 2 - 1) + (2 \times 2 - 1) = 11$ čísel (viz tabulka 3). Dodejme, že maximálně úsporná reprezentace ve formě tabulek podmíněných pravděpodobností určených grafem na obrázku 3 vyžaduje pouze 9 čísel.



Obr. 4 Reprezentace ve formě stromu spojení.

3. POJEM PODMÍNĚNĚ NEZÁVISLOSTI

V předchozím motivačním přehledu byl zmiňován pojem podmíněné nezávislosti aniž byl zaveden. Je na čase uvést nějakou rozumnou definici. Zcela obecná definice podmíněné nezávislosti v termínech σ -algeber [22, 7] vyžaduje příliš mnoho technického aparátu. Proto je výhodnější zde uvést jednodušší a relativně názornější definici ve speciálním případě marginálně spojitých rozdělení. Navíc tento speciální případ zahrnuje dva konkrétní podpřípady zmíněné dříve, totiž diskrétní a nedegenerované Gaussovské rozdělení. Nejprve popíšeme situaci, ve které zavádíme pojem podmíněné nezávislosti.

3.1. Několik základních definic. V následujícím již poněkud více matematickém textu budeme identifikovat dříve zmíněné faktory či příznaky (viz § 2.1 a § 2.3) přímo s prvky nějaké abstraktní neprázdné konečné indexové množiny N .

DEFINICE 4. Buď (X_i, \mathcal{X}_i) , $i \in N$ neprázdný konečný soubor měřitelných prostorů. Nechť (X_A, \mathcal{X}_A) pro $\emptyset \neq A \subseteq N$ označuje zkratkovitě kartézský součin těchto měřitelných prostorů pro $i \in A$ a P je pravděpodobnostní rozdělení s výběrovým prostorem (X_N, \mathcal{X}_N) . Pak jeho marginálu na (X_A, \mathcal{X}_A) pro $\emptyset \neq A \subseteq N$ budeme označovat symbolem P^A . Je-li $x = [x_i]_{i \in N}$ prvek X_N a $\emptyset \neq A \subseteq N$, pak symbol x_A bude označovat projekci x na X_A , tedy $x_A = [x_i]_{i \in A}$. Pravděpodobnostní rozdělení P na (X_N, \mathcal{X}_N) se nazývá *marginálně spojitě* jestliže je absolutně spojitě vůči součinu svých jednorozměrných marginál: $P \ll \prod_{i \in N} P^{\{i\}}$.

Dodejme, že ekvivalentní definice (viz Lemma 2.3 [30]) říká, že existuje soubor σ -konečných měr μ_i na (X_i, \mathcal{X}_i) , $i \in N$ takových, že platí $P \ll \prod_{i \in N} \mu_i$. Míra $\mu = \prod_{i \in N} \mu_i$ se pak nazývá *dominující míra* pro P . Lze snadno nahlédnout, že $P^A \ll \mu_A \equiv \prod_{i \in A} \mu_i$ pro $\emptyset \neq A \subseteq N$.

DEFINICE 5. Je-li P marginálně spojitě rozdělení a $\mu = \prod_{i \in N} \mu_i$ pevně zvolená dominující míra pro P , pak pro každou $\emptyset \neq A \subseteq N$ se Radon-Nikodymova derivace P^A vzhledem k μ_A nazývá *marginální hustotou* P pro A (automaticky se omezuje na nezáporné verze Radon-Nikodymových derivací). Budeme ji označovat symbolem f_A a můžeme ji chápat jako funkci na X_N , která každému $x \in X_N$ přiřadí hodnotu $f_A(x_A)$. Zavedeme konvenci, že f_\emptyset je konstantní funkce na X_N nabývající hodnoty 1 a tedy symbol $f_\emptyset(x_\emptyset)$ ve formulích níže je defintoricky jednička. Důležité pozorování je, že kdykoliv $A \subseteq B \subseteq N$, pak

$$f_A(x_A) = 0 \Rightarrow f_B(x_B) = 0 \quad \text{pro } \mu\text{-s.v. } x \in X_N,$$

a tedy lze volit takové verze marginálních hustot, pro něž tato implikace platí všude. V tom případě lze pro každou dvojici $A, C \subseteq N$, $A \cap C = \emptyset$ zavést *podmíněnou hustotu* P pro A dáno C označovanou symbolem $f_{A|C}$ jakožto podíl marginálních hustot

$$f_{A|C}(x_A|x_C) = \frac{f_{A \cup C}(x_{A \cup C})}{f_C(x_C)} \quad \text{pro } x \in X_N \text{ splňující } f_C(x_C) > 0.$$

V případě $f_C(x_C) = 0$ lze přijmout konvenci $f_{A|C}(x_A|x_C) = 0$ a podmíněná hustota je pak rovněž definována všude.

Marginální hustoty jsou tedy určeny jednoznačně jen ve smyslu skoro všude vůči zvolené dominující míře μ . Nyní již lze korektně zavést pojem podmíněné nezávislosti.

DEFINICE 6. Označme symbolem $\mathcal{T}(N)$ systém všech uspořádaných trojic $\langle A, B|C \rangle$ disjunktních podmnožin N . Za situace popsané v definici 4 buď P marginálně spojitě rozdělení na (X_N, \mathcal{X}_N) , μ dominující míra pro P a $\langle A, B|C \rangle \in \mathcal{T}(N)$. Pak píšeme $A \perp\!\!\!\perp B|C [P]$ a říkáme že A a B jsou *podmíněně nezávislé dáno C vzhledem k P* jestliže

$$(13) \quad f_{ABC}(x) \cdot f_C(x_C) = f_{AC}(x) \cdot f_{BC}(x) \quad \text{pro } \mu\text{-s.v. } x \in X_N.$$

Dodejme, že v (13) a ve formulích níže je použita konvence podle níž symbol AB označuje sjednocení (disjunktních) množin $A \cup B$. Další konvence bude že při zápisu jednoprvkové podmnožiny N budeme vynechávat závorky a tedy budeme psát a namísto $\{a\}$.

Lze dokázat, že výše definovaný pojem podmíněné nezávislosti nezávisí na volbě dominující míry (plyne to z Lemmatu 2.4 v [30]). V tomto příspěvku zdůrazníme dva základní speciální případy.

Diskrétním případem se rozumí situace, kdy X_i je neprázdná konečná množina a \mathcal{X}_i je potenční množina $\mathcal{P}(X_i)$ pro každé $i \in N$. Pak lze na místě dominující míry vždy uvažovat aritmetickou míru na X_N , což vede k následujícím (marginálním) hustotám:

$$p_N(x) = P(\{x\}), \quad p_A(x_A) = P(\{x_A\} \times X_{N \setminus A}) \quad \text{pro } \emptyset \neq A \subset N, \quad x \in X_N.$$

Navíc, rovnost skoro všude vzhledem k této dominující míře je rovnost všude a tedy $A \perp\!\!\!\perp B|C [P]$ právě když

$$\forall x \in X_N \quad p_{ABC}(x) \cdot p_C(x) = p_{AC}(x) \cdot p_{BC}(x).$$

Lze snadno ověřit, že tato podmínka je ekvivalentní podmínce

$$\forall x \in X_N, p_{BC}(x_{BC}) > 0 \quad \text{platí} \quad p_{A|BC}(x_A|x_{BC}) = p_{A|C}(x_A|x_C),$$

kde $p_{A|C}(x_A|x_C)$ označuje podmíněnou hustotu definovanou podílem $\frac{p_{AC}(x_{AC})}{p_C(x_C)}$ v případě že $p_C(x_C) > 0$. Čtenář může snadno nahlédnout, že tato ekvivalentní definice odpovídá interpretaci ‘podmíněné irrelevance’ běžné v oblasti pravděpodobnostního rozhodování: „při znalosti hodnoty x_C pravděpodobnost výskytu x_A není ovlivněna tím, zda nastane x_B “.

V *nedegenerovaném Gaussovském případě* je X_i množina reálných čísel a \mathcal{X}_i třída borelovských množin pro každé $i \in N$. Za dominující míru pro nedegenerované Gaussovské rozdělení $P = \mathcal{N}(e, \Sigma)$ kde $e \in \mathbb{R}^N$ a Σ je pozitivně definitní $N \times N$ -matice lze volit Lebesgueovu míru λ_N na $X_N = \mathbb{R}^N$. Samozřejmě, hustota je určena jen v rámci rovnosti skoro všude vůči této míře, ale ve statistice je běžná konvence, že se automaticky vybírá jediná spojitá verze této hustoty určená vzorcem

$$f(x) = \frac{1}{\sqrt{(2\pi)^{|N|} \cdot \det(\Sigma)}} \cdot \exp^{-\frac{(x-e)^\top \cdot \Sigma^{-1} \cdot (x-e)}{2}} \quad \text{for } x \in \mathbb{R}^N.$$

Při stejné konvenci dostaneme jednoznačné vyjádření pro marginální hustotu f_A , $\emptyset \neq A \subseteq N$ tak, že ve formuli výše se nahradí symbol N symbolem A a Σ svou hlavní podmaticí $\Sigma_{A,A}$. Protože rovnost skoro všude vůči Lebesgueově míře pro spojitě verze hustot je ekvivalentní rovnosti všude je podmínka $A \perp\!\!\!\perp B | C [P]$ ekvivalentní požadavku (viz § 4.3.2 v [30])

$$\forall x \in X_N \quad f_{ABC}(x_{ABC}) \cdot f_C(x_C) = f_{AC}(x_{AC}) \cdot f_{BC}(x_{BC}).$$

Nicméně, v Gaussovském případě lze charakterisovat nezávislostní tvrzení ještě lépe - přímo v termínech kovarianční matice. Pro disjunktní $A, C \subseteq N$ lze zavést Schurův doplněk podmatice $\Sigma_{A,A}$ v matici $\Sigma_{AC,AC}$, ve statistice též nazývaný *podmíněnou kovarianční maticí*, vztahem

$$\Sigma_{A|C} = \Sigma_{A,A} - \Sigma_{A,C} \cdot (\Sigma_{C,C})^{-1} \cdot \Sigma_{C,A}.$$

Potom platí (viz Lemma 2.8 v [30]):

$$A \perp\!\!\!\perp B | C [\mathcal{N}(e, \Sigma)] \Leftrightarrow \text{všechny prvky matice } (\Sigma_{AB|C})_{A,B} \text{ jsou nuly.}$$

3.2. Pojem struktury podmíněné nezávislosti. Nyní už je možné zavést pojem struktury podmíněné nezávislosti.

DEFINICE 7. Za situace popsané v definici 4 buď P marginálně spojitě rozdělení na (X_N, \mathcal{X}_N) . *Strukturou podmíněné nezávislosti* indukovanou P se rozumí množina trojic $\langle A, B | C \rangle \in \mathcal{T}(N)$ pro něž platí $A \perp\!\!\!\perp B | C [P]$. Často je užitečné rozlišit typ nezávislostních tvrzení. Nezávislostní údaj $A \perp\!\!\!\perp B | C$ se nazývá *triviální* jestliže $A = \emptyset$ či $B = \emptyset$, jinak je *netriviální*. *Symetrickým údajem* k údaji $A \perp\!\!\!\perp B | C$ se rozumí údaj $B \perp\!\!\!\perp A | C$. *Elementárním* nezávislostním údajem se rozumí údaj $A \perp\!\!\!\perp B | C$ kde A i B jsou jednoprvkové množiny. Systém všech odpovídajících ‘elementárních’ uspořádaných trojic $\langle a, b | C \rangle$ kde $a, b \in N, a \neq b, C \subseteq N \setminus \{a, b\}$ bude označován symbolem $\mathcal{E}(N)$.

Struktury podmíněné nezávislosti vykazují jisté formální vlastnosti, které někteří autoři [24] nazývali poněkud nepřesně axiomy (podmíněné nezávislosti). Je známo, že například platí následující *semi-grafoidové vlastnosti*, které uvádím pod jejich anglickými názvy.

- | | |
|--|----------------|
| 1. $A \perp\!\!\!\perp \emptyset D [P]$ | triviality, |
| 2. $A \perp\!\!\!\perp B D [P] \Rightarrow B \perp\!\!\!\perp A D [P]$ | symmetry, |
| 3. $A \perp\!\!\!\perp BC D [P] \Rightarrow A \perp\!\!\!\perp C D [P]$ | decomposition, |
| 4. $A \perp\!\!\!\perp BC D [P] \Rightarrow A \perp\!\!\!\perp B CD [P]$ | weak union, |

5. $\{A \perp\!\!\!\perp B \mid CD [P] \ \& \ A \perp\!\!\!\perp C \mid D [P]\} \Rightarrow A \perp\!\!\!\perp BC \mid D [P]$ contraction.

Navíc, pokud je indukující rozdělení speciální, mohou platit další formální vlastnosti (viz § 2.3 v [30]). Například, pokud existuje hustota vůči nějaké dominující míře, která je všude kladná, pak indukovaná struktura podmíněné nezávislosti splňuje

6. $A \perp\!\!\!\perp B \mid CD [P] \ \& \ A \perp\!\!\!\perp C \mid BD [P] \Rightarrow A \perp\!\!\!\perp BC \mid D [P]$ intersection.

Nicméně, struktury podmíněné nezávislosti indukované diskretními pravděpodobnostními rozděleními obecně nelze charakterisovat s pomocí konečného počtu formálních vlastností tohoto typu – důkaz viz [29].

Pokud je pevně určen nějaký distribuční rámec, to jest třída rozdělení Φ nad N , například třída všech diskretních rozdělení (na X_N) anebo třída nedegenerovaných Gaussovských rozdělení (na \mathbb{R}^N), pak každou strukturu podmíněné nezávislosti \mathcal{S} lze chápat jako *statistický model*. Definuje totiž třídu pravděpodobnostních rozdělení

$$\{P \in \Phi; A \perp\!\!\!\perp B \mid C [P] \text{ kdykoliv } \langle A, B \mid C \rangle \in \mathcal{S}\}.$$

Hlavním účelem § 2 bylo vlastně ukázat, že některé modely používané ve statistice lze chápat jako statistické modely určené nějakou strukturou podmíněné nezávislosti. Navíc, tyto struktury podmíněné nezávislosti je často možné popsat s pomocí nějakého grafu, který má N za množinu uzlů.

POZNÁMKA 6. Výše zmíněné statistické modely však není třeba popisovat úplným výčtem všech nezávislostních údajů. Například lze vynechat vždy platné triviální nezávislostní údaje neboť pro každou strukturu podmíněné nezávislosti \mathcal{S} indukovanou rozdělením P platí výše zmíněné semi-grafoidové vlastnosti, speciálně vlastnost *triviality*. Analogické pozorování je, že strukturu podmíněné nezávislosti \mathcal{S} lze v zápise statistického modelu nahradit pouze seznamem elementárních nezávislostních údajů, tedy lze psát

$$\{P \in \Phi; a \perp\!\!\!\perp b \mid C [P] \text{ kdykoliv } \langle a, b \mid C \rangle \in \mathcal{S} \cap \mathcal{E}(N)\}.$$

Konečně, díky vlastnosti *symmetry* lze vždy z dvojice navzájem symetrických údajů uvést pouze jeden. Tyto konvence umožňují úspornější zápis a jsou v oblasti grafických modelů běžné. Podstata této poznámky je, že ve formálním zápise statistického modelu lze seznam \mathcal{S} nahradit jakýmkoliv kratším seznamem nezávislostních údajů \mathcal{L} jehož semi-grafoidový uzávěr je \mathcal{S} .

4. ŘETĚZCOVÉ GRAFY

Statistické modely struktur podmíněné nezávislosti popisované neorientovanými grafy a modely popisované acyklickými orientovanými grafy lze shrnout do jednotícího rámce popsaných *řetězcovými grafy*, anglicky nazývanými *chain graphs*. Řetězcové grafy zavedli Lauritzen a Wermuthová v půli 80-tých let dvacátého století [16], přičemž v téže době byly navrženy i tzv. rekursivní kauzální grafy [14], které lze považovat za speciální případ řetězcových grafů. V obou případech se jedná o hybridní grafy, které mají jak neorientované tak orientované hrany.

4.1. Formální definice.

DEFINICE 8. *Hybridní graf* H je určen neprázdnou konečnou množinou uzlů N , množinou *neorientovaných hran* \mathcal{E} (t.j. systémem dvouprvkových podmnožin N) a množinou *orientovaných hran* či *šipek* \mathcal{A} , což je systém uspořádaných dvojic $(u, v) \in N \times N$ kde $u \neq v$. Jestliže $(u, v) \in \mathcal{A}$, pak píšeme $u \rightarrow v$ (v H) a v případě $\{u, v\} \in \mathcal{E}$ píšeme $u - v$ (v H); v obou těchto případech říkáme, že (neuspořádaná

dvojice) $[u, v]$ je *hrana* v H . Jinými slovy $[u, v]$ je hrana H jestliže buď $u - v$ v H (t.j. $v - u$ v H) nebo $u \rightarrow v$ v H či $v \rightarrow u$ v H . Výše uvedené značení odráží způsob jakým jsou hrany hybridního grafu znázorňovány v obrázcích. Dále se implicitně předpokládá, že v hybridním grafu mezi dvěma uzly existuje maximálně jedna hrana, jinými slovy že platí

$$\forall u, v \in N \quad \text{pokud } (u, v) \in \mathcal{A}, \text{ pak } (v, u) \notin \mathcal{A} \text{ a } \{u, v\} \notin \mathcal{E}.$$

Je-li H hybridní graf s množinou uzlů N a $\emptyset \neq A \subseteq N$ pak jeho *podgraf indukovaný* A označený H_A je hybridní graf s množinou uzlů A takový, že $u - v$ v H_A právě když $u, v \in A$, $u - v$ v H a $u \rightarrow v$ v H_A právě když $u, v \in A$, $u \rightarrow v$ v H . Je-li $u \in N$, pak příslušná *neorientovaná komponenta* grafu H je tvořena všemi uzly v , které jsou spojeny s u neorientovanou cestou $u = w_1 - \dots - w_n = v$, $n \geq 1$. Množina uzlů N se tedy rozpadá na disjunkttní neorientované komponenty grafu H .

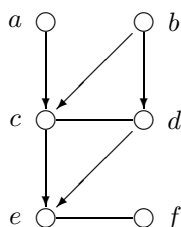
Řetězcové grafy jsou hybridní grafy, které vzniknou z neorientovaných grafů tak že se zorientují některé hrany podle následujícího scénáře.

DEFINICE 9. Nechť G je neorientovaný graf s množinou uzlů N a množinou hran $\mathcal{E}(G)$. Bud' B_1, \dots, B_n , $n \geq 1$ řetězec pro N , to jest uspořádaná posloupnost neprázdných disjunkttních podmnožin N jejichž sjednocením je N . Vytvořme z G hybridní graf H mající stejnou množinu uzlů následovně:

- (i) $\{u, v\} \in \mathcal{E}(H)$ pokud $\{u, v\} \in \mathcal{E}(G)$ a $\exists 1 \leq i \leq n$ takové, že $u, v \in B_i$,
- (ii) $(u, v) \in \mathcal{A}(H)$ pokud $\{u, v\} \in \mathcal{E}(G)$ a $\exists 1 \leq i < j \leq n$ že $u \in B_i$ a $v \in B_j$.

Každý hybridní graf, který lze vytvořit takovouto konstrukcí se nazývá *řetězcový graf*.

Lze snadno dokázat, že pro každý řetězcový graf H existuje řetězec tvořený neorientovanými komponentami H . Řetězcové grafy lze také ekvivalentně zavést jako hybridní grafy, ve kterých neexistují jisté polo-orientované cykly; jsou to tedy určité acyklické hybridní grafy. Jednoduchý příklad řetězcového grafu je na obrázku 5.



Obr. 5 Jednoduchý řetězcový graf.

POZNÁMKA 7. Dříve než přejdeme k pravděpodobnostní interpretaci řetězcových grafů bych rád poznamenal (a tím zároveň varoval čtenáře), že dokonce v některých klasických textech o řetězcových grafech [33, 18] se vyskytují poněkud vágní slovní formulace, které mohou svést čtenáře k nesprávnému pochopení významu hran v řetězcovém grafu. Tyto formulace mohou totiž vzbudit dojem, že jednotlivé šipky lze interpretovat coby 'kauzální vztahy' zatímco jednotlivé neorientované hrany lze interpretovat coby 'symetrické asociace'. Ve skutečnosti interpretace jednotlivé hrany jakožto objektu samého o sobě je nesprávná; hrany získávají správný smysl pouze v kontextu s ostatními hranami. Rozumnou interpretaci má tedy pouze řetězcový

graf jako celek a vskutku původní interpretace byla v termínech podmíněné nezávislosti. Zbytek § 4 je věnován klasické interpretaci řetězových grafů. Nicméně v zájmu vyváženosti je třeba dodat, že zejména v poslední době se objevily další možné (i odlišné) interpretace řetězových grafů; například kauzální interpretace [21] či alternativní interpretace coby struktury podmíněné nezávislosti [3].

4.2. Párová a lokální Markovská podmínka. Je několik způsobů, jak formálně interpretovat řetězový graf coby pravděpodobnostní strukturu. Požadavky, které svazují grafickou strukturu a tvar pravděpodobnostního rozdělení jsou známy jako Markovské respektive faktorisační podmínky. Níže uvedené Markovské a faktorisační podmínky pro řetězové grafy zobecňují analogické podmínky pro neorientované a acyklické orientované grafy, které byly historicky zavedeny dříve. Podmínky rozlišujeme podle síly. Nejslabší je párová Markovská podmínka.

DEFINICE 10. Za situace popsané v definici 4 buď H řetězový graf s množinou uzlů N a řetězcem $B_1, \dots, B_n, n \geq 1$. Řekneme, že (marginálně spojitě) rozdělení P na (X_N, \mathcal{X}_N) splňuje *párovou Markovskou podmínku (PM)* vzhledem k H jestliže

$$(14) \quad \forall u, v \in N, u \neq v, [u, v] \text{ není hrana v } H, \\ \text{platí } u \perp\!\!\!\perp v \mid D(u, v) \setminus \{u, v\} [P],$$

kde $D(u, v)$ označuje sjednocení množin $B_1, \dots, B_k, k \leq n$ přičemž B_k je poslední množina řetězce mající neprázdný průnik s $\{u, v\}$.

V případě, že H je neorientovaný graf, dostaneme pro triviální řetězec $B_1 = N$ klasickou párovou Markovskou podmínkou pro neorientované grafy; v případě, že H je acyklický orientovaný graf a řetězec $B_1, \dots, B_n, n \geq 1$ je tvořen jednoprvkovými množinami pak dostaneme podmínku, která by se anglicky mohla nazývat *pairwise well-numbering Markov property* [19]. Následující příklad ilustruje zavedený pojem.

PŘÍKLAD 14. Uvažujme řetězový graf na obrázku 5 a řetězec $B_1 = \{a, b\}$, $B_2 = \{c, d\}$, $B_3 = \{e, f\}$. Párová Markovská podmínka na rozdělení P pak implikuje následující elementární nezávislostní tvrzení (z dvojice symetrických tvrzení uvádím pouze jedno - viz poznámka 6):

- $a \perp\!\!\!\perp b \mid \emptyset [P]$,
- $a \perp\!\!\!\perp d \mid \{b, c\} [P]$,
- $a \perp\!\!\!\perp e \mid \{b, c, d, f\} [P]$,
- $a \perp\!\!\!\perp f \mid \{b, c, d, e\} [P]$,
- $b \perp\!\!\!\perp e \mid \{a, c, d, f\} [P]$,
- $b \perp\!\!\!\perp f \mid \{a, c, d, e\} [P]$,
- $c \perp\!\!\!\perp f \mid \{a, b, d, e\} [P]$,
- $d \perp\!\!\!\perp f \mid \{a, b, c, e\} [P]$.

Pravděpodobně nejekonomičtější zápis statistického modelu struktury podmíněné nezávislosti indukovaného řetězovým grafem umožňuje lokální Markovská podmínka.

DEFINICE 11. Za situace popsané v definici 4 buď H řetězový graf a $B_1, \dots, B_n, n \geq 1$ příslušný řetězec. Řekneme, že (marginálně spojitě) rozdělení P na (X_N, \mathcal{X}_N) splňuje *lokální Markovskou podmínku (LM)* vzhledem k H jestliže

$$(15) \quad \forall u \in N \quad u \perp\!\!\!\perp D(u) \setminus (\{u\} \cup \text{bd}(u)) \mid \text{bd}(u) [P]$$

kde symbol $D(u)$ označuje sjednocení množin $B_1, \dots, B_k, k \leq n$ přičemž B_k obsahuje u a $\text{bd}(u) = \{v \in N; v \text{ --- } u \text{ či } v \rightarrow u \text{ v } H\}$ označuje *hranici* uzlu u .

PŘÍKLAD 15. Za situace popsané v příkladě 14 požaduje lokální Markovská podmínka následující netriviální nezávislostní tvrzení (opět vynechávám symetrická nezávislostní tvrzení):

- $a \perp\!\!\!\perp b \mid \emptyset [P]$,
- $d \perp\!\!\!\perp a \mid \{b, c\} [P]$,
- $e \perp\!\!\!\perp \{a, b\} \mid \{c, d, f\} [P]$,
- $f \perp\!\!\!\perp \{a, b, c, d\} \mid e [P]$.

4.3. Globální Markovská podmínka. Pro formulaci nejsilnější globální Markovské podmínky je třeba zavést grafické kritérium, podle kterého je možné rozhodnout, zda nějaká trojice $\langle A, B|C \rangle \in \mathcal{T}(N)$ (a tedy odpovídající nezávislostní údaj) je reprezentována v řetězcovém grafu. V případě neorientovaného grafu je takové kritérium jediné, totiž separační kritérium uvedené v definici 2. Avšak v případě acyklického orientovaného grafu si již lze vybrat mezi dvěma různými, ale ekvivalentními grafickými kritérii. Lauritzen a jeho spolupracovníci [19] navrhli *moralisační kritérium* zatímco Pearl a jeho následovníci [24] dávali přednost *d-separačnímu kritériu*. Moralisační kritérium je nepřímé v tom smyslu, že původní orientovaný graf se transformuje na jistý neorientovaný graf (nazývaný z jistého důvodu morální graf) a poté se aplikuje separační kritérium pro neorientované grafy. Naopak, *d-separační kritérium* (*d* značí *directional*, tedy usměrněné) je přímé v tom smyslu, že se testují cesty (resp. sledy) v původním grafu zda jsou blokovány nějakou množinou uzlů, přičemž pojem ‘blokování’ závisí na orientaci hran.

Také v případě obecných řetězcových grafů můžeme odlišit dvě různá ekvivalentní kritéria, která zobecňují kritéria pro acyklické orientované grafy. Původní moralisační kritérium pro řetězcové grafy zavedl Lauritzen [20].

DEFINICE 12. *Morálním grafem* řetězcového grafu H s množinou uzlů N se rozumí neorientovaný graf H^{mor} s toutéž množinou uzlů takový, že $\{u, v\}$ je (neorientovanou) hranou v H^{mor} právě tehdy, když $[u, v]$ je hrana v H anebo v H existuje posloupnost uzlů $u \rightarrow w_1 \leftarrow \dots \leftarrow w_k \leftarrow v$, $k \geq 1$ (zápis značí, že w_1 a w_k jsou v případě $k > 1$ spojeny cestou z neorientovaných hran, jinými slovy že uzly náleží do stejné neorientované komponenty grafu H).

Množina uzlů $D \subseteq N$ se nazývá *ancestrální množinou* v H jestliže $bd(u) \subseteq D$ pro každé $u \in D$ (viz definice 11). Je-li $A \subseteq N$ pak nejmenší ancestrální množinu obsahující A (to jest průnik všech ancestrálních množin obsahujících A) označujeme $an(A)$.

Jestliže $\langle A, B|C \rangle \in \mathcal{T}(N)$, pak test zda je tato trojice reprezentována v řetězcovém grafu H (s množinou uzlů N) má tři následující kroky.

- (i) uvažujeme $H_{an(ABC)}$, to jest podgraf H indukovaný ancestrální množinou $an(ABC)$ (viz definice 8),
- (ii) zkonstruujeme morální graf $(H_{an(ABC)})^{mor}$ zmíněného indukovaného podgrafu $H_{an(ABC)}$,
- (iii) testujeme, zda je $\langle A, B|C \rangle$ reprezentována v tomto morálním grafu $(H_{an(ABC)})^{mor}$ podle separačního kritéria pro neorientované grafy (viz definice 2).

Je-li $\langle A, B|C \rangle$ reprezentována v onom morálním grafu, pak řekneme že je reprezentována v řetězcovém grafu H podle *moralisačního kritéria*.

V případě neorientovaného grafu H je výše definované moralisační kritérium ekvivalentní separačnímu kritériu pro neorientované grafy zatímco v případě acyklického orientovaného grafu H se shoduje s moralisačním kritériem pro tyto grafy. Původní verze *c-separačního kritéria* pro řetězcové grafy (*c* značí *chain*, tedy řetězec) byla

navržena v [4]. Později toto kritérium bylo zjednodušeno; nejjednodušší forma je popsána v následující definici.

DEFINICE 13. Buď H řetězcový graf s množinou uzlů N . *Sledem* v H se rozumí posloupnost uzlů $\rho : v_1, \dots, v_n, n \geq 1$ (ne nutně různých!) taková, že $[v_i, v_{i+1}]$ je hrana v H pro $i = 1, \dots, n - 1$. *Úsekem* sledu se rozumí jeho maximální neorientovaný podsled $v_i - \dots - v_j$, kde $1 \leq i \leq j \leq n$ (to jest pro žádné $i' \leq i \leq j \leq j'$ takové, že $j' - i' > j - i$ není $v_{i'} - \dots - v_{j'}$). Dodejme, že úsek sledu ρ může být tvořen i jediným uzlem, což nastane v případě $i = j$. Úsek se nazývá *kolisní* jestliže $1 < i \leq j < n$ a platí $v_{i-1} \rightarrow v_i - \dots - v_j \leftarrow v_{j+1}$; v opačném případě je *nekolisní*.

Buď $C \subseteq N$ množina uzlů; sled ρ se nazývá *superaktivní* vzhledem k C jestliže

- každý kolisní úsek ρ obsahuje uzel z množiny C ,
- každý nekolisní úsek ρ je tvořen uzly mimo množinu C .

Pokud ρ není superaktivní vzhledem k C , pak je *blokován* množinou C .

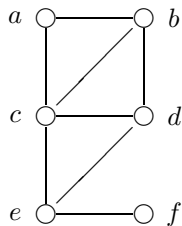
Trojice $\langle A, B | C \rangle \in \mathcal{T}(N)$ je representována v řetězcovém grafu H podle *c-separačního kritéria* jestliže každý sled $v_1, \dots, v_n, n \geq 1$ v H takový, že $v_1 \in A$ a $v_n \in B$ je blokován množinou C .

Laskavý čtenář může sám nahlédnout, že v případě neorientovaného grafu H je *c-separační kritérium* ekvivalentní *separačnímu kritériu* z definice 2. V případě acyklického orientovaného grafu H jsou úseky jakéhokoliv sledu tvořeny vždy jedním uzlem a klasifikace tedy rozlišuje kolisní a nekolisní uzly na uvažovaném sledu. Čtenář obeznámený s *d-separačním kritériem* [24] tedy může po mírném intelektuálním úsilí nahlédnout, že *c-separační kritérium* se vlastně shoduje s příslušnou verzí *d-separačního kritéria*.

Dodejme, že v případě řetězcových grafů je vskutku podstatné uvažovat třídu všech sledů v grafu (kde se uzly mohou opakovat) a nelze se omezit na třídu cest (v nichž se uzly neopakují). Nicméně, přestože třída všech sledů v grafu je často nekonečná, *c-separační kritérium* je rozhodnutelné (díky konečnosti grafu H) a lze snadno implementovat na počítači s pomocí efektivního algoritmu 'lokální propagace' [27]. Ekvivalence *moralisačního a c-separačního kritéria* pro řetězcové grafy byla dokázána v [28].

PŘÍKLAD 16. Uvažujme opět graf H z obrázku 5. Při testování trojice $\langle a, f | b \rangle$ pomocí *c-separačního kritéria* zjistíme, že sled $a \rightarrow c - d \rightarrow e - f$ je aktivní vzhledem k $\{b\}$, a tedy tato trojice není v grafu representována. Při testování pomocí *moralisačního kritéria* zjistíme, že $\text{an}(\{a, b, f\}) = \{a, b, c, d, e, f\}$ a tedy příslušný indukovaný podgraf je samotný graf H . Příslušný *morální graf* je na obrázku 6. Vidíme, že odpovídající cesta $a - c - d - e - f$ je cesta z a do f mimo $\{b\}$ a tedy $\langle a, f | b \rangle$ není v H representována podle *moralisačního kritéria*.

Naopak, při testování $\langle a, f | \{b, c\} \rangle$ podle *c-separačního kritéria* pozorujeme, že sled z a do f v H může buď obsahovat nekolisní úsek obsahující uzel c a pak je tomto úseku blokován množinou $\{b, c\}$ anebo tento sled nutně obsahuje nekolisní úsek $\leftarrow b \rightarrow$ a je rovněž blokován touto množinou. Pročež $\langle a, f | \{b, c\} \rangle$ je representována v řetězcovém grafu H . Tento fakt lze lépe nahlédnout s pomocí *moralisačního kritéria*: každá cesta z a do f v příslušném *morálním grafu* na obrázku 6 totiž obsahuje uzel množiny $\{b, c\}$.



Obr. 6 Morální graf řetězcového grafu z obrázku 5.

Nejsilnější Markovská podmínka je globální Markovská podmínka.

DEFINICE 14. Za situace popsané v definici 4 buď H řetězcový graf s množinou uzlů N . Řekněme, že (marginálně spojitě) rozdělení P na (X_N, \mathcal{X}_N) splňuje *globální Markovskou podmínku (GM)* vzhledem k H jestliže

$$(16) \quad A \perp\!\!\!\perp B \mid C [P] \quad \text{je-li } \langle A, B \mid C \rangle \in \mathcal{T}(N) \text{ representována v } H,$$

přičemž representace se rozumí buď ve smyslu definice 12 nebo definice 13.

4.4. Faktorisační podmínka. Další možnost jak svázat strukturu řetězcového grafu a pravděpodobnostního rozdělení je faktorisační podmínka.

DEFINICE 15. Buď H řetězcový graf s množinou uzlů N , nechť \mathcal{C} je soubor všech neorientovaných komponent H . Pro každou komponentu $C \in \mathcal{C}$ zavedme množinu jejích *rodičů* $\text{pa}(C) = \{v \in N \setminus C; v \rightarrow u \text{ pro } u \in C\}$ a příslušný *uzávěrový graf* jakožto morální graf podgrafu H indukovaného množinou $C \cup \text{pa}(C)$ (viz definice 8 a 12). Nechť $\mathcal{H}(C)$ označuje soubor klik tohoto uzávěrového grafu $(H_{C \cup \text{pa}(C)})^{\text{mor}}$.

Za situace v definici 4 řekneme, že (marginálně spojitě) rozdělení P na (X_N, \mathcal{X}_N) se *faktorizuje vzhledem k H* (t.j. splňuje podmínku *(FP)* vůči H) jestliže platí následující dvě podmínky. Podmínka *globální faktorisace* požaduje, aby se hustota dala vyjádřit jako součin podmíněných hustot:

$$f_N(x) = \prod_{C \in \mathcal{C}} f_{C \mid \text{pa}(C)}(x_C \mid x_{\text{pa}(C)}) \quad \text{pro } \mu\text{-s.v. } x \in X_N.$$

Podmínka *lokální faktorisace* požaduje, aby se každá z podmíněných hustot dále faktorisovala podle příslušného uzávěrového grafu:

$$\forall C \in \mathcal{C} \quad f_{C \mid \text{pa}(C)}(x_C \mid x_{\text{pa}(C)}) = \prod_{S \in \mathcal{H}(C)} \psi_S(x_S) \quad \text{pro } \mu\text{-s.v. } x \in X_N,$$

kde ψ_S , $S \subseteq N$ jsou nějaké (obecné) nezáporné \mathcal{X}_S -měřitelné funkce na X_S (taková funkce se často nazývá *potenciál* na S).

Jak čtenář může očekávat, v případě neorientovaného grafu je faktorisační podmínka ekvivalentní běžné faktorisaci rozdělení podle (klik) tohoto neorientovaného grafu [24]. Podobně v případě acyklického orientovaného grafu dostaneme obvyklou podmínku rekursivní faktorisace [20].

PŘÍKLAD 17. Opět uvažujeme graf na obrázku 5. V tomto případě podmínka globální faktorisace říká

$$f_{abcdef} = f_a \cdot f_b \cdot f_{cd \mid ab} \cdot f_{ef \mid cd},$$

a podmínka lokální faktorisace dále říká

$$f_{cd \mid ab} = \psi_{abc} \cdot \psi_{bcd}, \quad f_{ef \mid cd} = \psi_{cde} \cdot \psi_{ef}.$$

Vztahy výše uvedených podmínek svazujících grafickou a pravděpodobnostní strukturu jsou následující. Faktorisační podmínka implikuje globální Markovskost, ta lokální Markovskost a tato pak párovou Markovskost (viz [20]). Schematicky:

$$(FP) \Rightarrow (GM) \Rightarrow (LM) \Rightarrow (PM).$$

Pro ta marginálně spojitá rozdělení, která mají všude kladnou hustotu (vzhledem k nějaké dominující míře) platí i obrácená implikace a tedy všechny výše uvedené podmínky jsou ekvivalentní [8]. V případě distribučního rámce Φ tvořeného rozděleními tohoto typu je tedy přiřazen řetězcovému grafu všemi čtyřmi způsoby jediný statistický model.

Navíc, v případě rámce diskrétních rozdělení globální Markovská podmínka popisuje právě jen nezávislostní tvrzení platná pro všechna rozdělení příslušná tomuto statistickému modelu. V [28] bylo totiž dokázáno, že pro každý řetězcový graf H existuje *perfektní Markovské diskrétní rozdělení* (dokonce diskrétní rozdělení s všude kladnou hustotou), to jest rozdělení, které splňuje právě ta nezávislostní tvrzení, která jsou charakterisována moralisačním resp. c -separačním kritériem. Pro důkaz tohoto základního tvrzení o korektnosti řetězcových grafů z hlediska popisu diskrétních pravděpodobnostních struktur podmíněné nezávislosti bylo podstatné c -separační kritérium.

5. ZÁVĚR

Výše uvedený text lze chápat jako úvod do problematiky řetězcových grafů a jimi indukovaných struktur podmíněné nezávislosti. Samozřejmě, že byly opomenuty mnohé další teoretické výsledky. Například nebyly zmíněny následující výsledky: grafická charakterisace Markovsky ekvivalentních řetězcových grafů (t.j. grafů indukujících tentýž statistický model) [8], tvrzení o existenci tzv. největšího řetězcového grafu v rámci každé třídy této ekvivalence [8] a grafické charakterisace těchto největších řetězcových grafů včetně příslušných algoritmů na jejich získávání [26, 31].

Opomenuta byla též otázka možné interpretace řetězcových grafů z hlediska výstavby expertních systémů. Faktorisační podmínka z definice 15 by například mohla být interpretována tak že by mohla být využita při určení řetězcových grafů na základě informace od expertů přičemž každý z expertů určí pouze příslušný uzávěrový graf (viz příklad z [27]).

Dodejme, že oblast grafických modelů (pro popis struktur podmíněné nezávislosti) je dosti rychle se rozvíjející obor. V nedávné době byly zavedeny a studovány další, obecnější třídy grafů pro popis pravděpodobnostních struktur, například obecné orientované grafy [25], reciproké grafy [15], alternativně interpretované řetězcové grafy [3], tzv. *joint-response* řetězcové grafy [6], anotační grafy [23] a grafy s latentními (skrytými) veličinami. Letmý přehled těchto přístupů je v třetí kapitole práce [30].

Poděkování. Chtěl bych poděkovat Dr. Jaromíru Antochovi, hlavnímu organizátorovi ROBUSTů, za jeho trpělivost a ochotu tolerovat mou předem avizovanou neschopnost dodat příspěvek do sborníku včas. Dále bych chtěl poděkovat Dr. Daniele Jaruškové za její rady a instrukce stran mé přednášky v Nečtinách a Dr. Ivanu Saxlovi též za rady, ale rovněž i za vytrvalé povzbuzování v mém úsilí sepsat příspěvek do sborníku. Též děkuji Dr. Markovi Malému za poskytnutí jistých textů o zpracování medicínských dat, v nichž jsou také použity grafy pro popis vztahů mezi veličinami. Nakonec jsem se po úvaze rozhodl tyto texty v příspěvku nezmiňovat, neboť zmíněné grafy byly získány specifickou metodou a jejich interpretace

tedy nemusí být zcela v souladu se zde uvažovanou interpretací řetězcových grafů (to jest coby nástrojů na popis struktury podmíněné nezávislosti).

Dále bych rád vyjádřil své díky recenzentovi Prof. Radimu Jirouškovi, který opravdu pečlivě příspěvek přečetl a poskytl celou řadu cenných připomínek, na které jsem se snažil zareagovat. Abych jej potěšil a odměnil se za jeho úsilí rozhodl jsem se nakonec přijmout jeho češtinářský názor a v rámci příspěvku systematicky používat frázi ‘acyklický orientovaný graf’ namísto původně používané fráze ‘orientovaný acyklický graf’.

Nemělo by zde jistě scházet ani poděkování paní Marii Kolářové za přípravu první verze tohoto příspěvku (tedy přepsání rukopisu do \LaTeX), paní Ivě Marešové za poskytnutí elektronické předlohy pro konečnou verzi příspěvku jakož i další užitečné rady technického charakteru a Dr. Sergeji Čelikovskému za kontrolu gramatické správnosti ruského abstraktu.

LITERATURA

- [1] Anděl J.: *Matematická Statistika*, SNTL (Praha) 1985.
- [2] Andersson S. A., Madigan D., Perlman M. D., Richardson T.S.: *Graphical Markov models in multivariate statistics, a chapter in Multivariate Analysis, Design of Experiments, and Survey Sampling*, Marcel Dekker 2001, pp. 187-229.
- [3] Andersson S. A., Madigan D., Perlman M. D.: *Alternative Markov properties for chain graphs*, *Scandinavian Journal of Statistics* 28 (2001), n. 1, pp. 33-85.
- [4] Bouckaert R. R., Studený M.: *Chain graphs: semantics and expressiveness*, in *Symbolic and Quantitative Approaches to Reasoning and Uncertainty* (Ch. Froidevaux, J. Kohlas eds.), *Lecture Notes in Artificial Intelligence* 946, Springer-Verlag 1995, pp. 67-76.
- [5] Cowell R. G., Dawid A. P., Lauritzen S. L., Spiegelhalter D. J.: *Probabilistic Networks and Expert Systems*, Springer-Verlag 1999.
- [6] Cox D. R., Wermuth N.: *Multivariate Dependencies – Models, Analysis and Interpretation*, Chapman and Hall 1996.
- [7] Florens J. -P., Mouchart M., Rolin J. -M.: *Elements of Bayesian Statistics*, Marcel Dekker 1990.
- [8] Frydenberg M.: *The chain graph Markov property*, *Scandinavian Journal of Statistics* 17 (1990), n. 4, pp. 333-353.
- [9] Goodman L. A.: *The multivariate analysis of qualitative data, interactions among multiple classifications*, *Journal of American Statistical Association* 65 (1970), pp. 226-256.
- [10] Haberman S.: *The Analysis of Frequency Data*, University of Chicago Press 1974.
- [11] Hájek P., Havránek T., Jiroušek R.: *Uncertain Information Processing in Expert Systems*, CRC Press 1992.
- [12] Jiroušek R.: *Solution of the marginal problem and decomposable distributions*, *Kybernetika* 27 (1991), pp. 403-412.
- [13] Jöreskog K. G., Sörbom D.: *LISREL 7 – A Guide to the Program and Application*, SPSS Inc. 1989.
- [14] Kiiveri H., Speed T. P., Carlin J. B.: *Recursive causal models*, *Journal of Australian Mathematical Society series A* 36 (1984), pp. 30-52.
- [15] Koster J. T. A.: *Markov properties of nonrecursive causal models*, *Annals of Statistics* 24 (1996), n. 5, pp. 2148-2177.
- [16] Lauritzen S. L., Wermuth N.: *Mixed interaction models*, research report R-84-8, Inst. Elec. Sys., University of Aalborg 1984.
- [17] Lauritzen S. L., Spiegelhalter D. J.: *Local computation with probabilities on graphical structures and their application to expert systems*, *Journal of the Royal Statistical Society series B* 50 (1988), n. 2, pp. 157-224.
- [18] Lauritzen S. L., Wermuth N.: *Graphical models for associations between variables, some of which are qualitative and some quantitative*, *Annals of Statistics* 17 (1989), n. 1, pp. 31-57.
- [19] Lauritzen S. L., Dawid A. P., Larsen B. N., Leimer H. -G.: *Independence properties of directed Markov fields*, *Networks* 20 (1990), n. 5, pp. 491-505.
- [20] Lauritzen S. L.: *Graphical Models*, Clarendon Press 1996.
- [21] Lauritzen S. L., Richardson T. S.: *Chain graph models and their causal interpretations, to appear in Journal of the Royal Statistical Society series B* 64 (2002).

- [22] Loève M.: Probability Theory, Foundations, Random Processes, D. van Nostrand 1955.
- [23] Paz A., Geva R. Y., Studený M.: Representation of irrelevance relations by annotated graphs, *Fundamenta Informaticae* 42 (2000), pp. 149-199.
- [24] Pearl J.: Probabilistic Reasoning in Intelligent Systems, Networks of Plausible Inference, Morgan Kaufmann 1988.
- [25] Spirtes P.: Directed cyclic graphical representations of feedback models, in *Uncertainty in Artificial Intelligence 11* (P. Besnard, S. Hanks eds.), Morgan Kaufmann 1995, pp. 491-498.
- [26] Studený M.: A recovery algorithm for chain graphs, *International Journal of Approximate Reasoning* 17 (1997), n. 2-3, pp. 265-293.
- [27] Studený M.: Bayesian networks from the point of view of chain graphs, in *Uncertainty in Artificial Intelligence 14* (G. F. Cooper, S. Moral eds.), Morgan Kaufmann 1998, pp. 496-503.
- [28] Studený M., Bouckaert R. R.: On chain graph models for description of conditional independence structures, *Annals of Statistics* 26 (1998), n. 4, pp. 1434-1495.
- [29] Studený M., Vejnarová J.: The multiinformation function as a tool for measuring stochastic dependence, in *Learning in Graphical Models* (M. I. Jordan ed.), Kluwer 1998, pp. 261-298.
- [30] Studený M.: On mathematical description of probabilistic conditional independence structures, thesis for DrSc degree, Institute of Information Theory and Automation, Prague May 2001.
- [31] Volf M., Studený M.: A graphical characterization of the largest chain graphs, *International Journal of Approximate Reasoning* 20 (1999), n. 3, pp. 209-236.
- [32] Vomlel J.: Methods of probabilistic knowledge integration, thesis for PhD degree, Czech Technical University, Faculty of Electrical Engineering, Prague July 1999.
- [33] Wermuth N., Lauritzen S. L.: On substantive research hypotheses, conditional independence graphs and graphical chain models, *Journal of the Royal Statistical Society series B* 52 (1990), n. 1, pp. 21-50.

INSTITUTE OF INFORMATION THEORY AND AUTOMATION, ACADEMY OF SCIENCES OF THE CZECH REPUBLIC, POD VODÁRENSKOU VĚŽÍ 4, 182 08 PRAGUE, CZECH REPUBLIC
E-MAIL: studenym@utia.cas.cz