České vysoké učení technické v Praze

Fakulta jaderná
a fyzikálně inženýrská

Vlastnosti variant spojené informace

Properties of connected information variants

Bakalářská práce

Autor: Katarína Studeničová

Vedoucí práce: Ing. Mgr. Jaroslav Hlinka, Ph.D.

Konzultant: Ing. Jakub Kořenek

Akademický rok: 2020/2021

Katedra: matematiky Akademický rok: 2020/2021

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Student: Katarína Studeničová

Studijní program: Aplikace přírodních věd

Studijní obor: Matematické inženýrství

Studijní zaměření: Matematické modelování

Název práce (česky): Vlastnosti variant spojené informace

Název práce (anglicky): Properties of connected information variants

Pokyny pro vypracování:

1) Seznamte se s pojmy vektorová náhodná veličina, její rozdělení pravděpodobnosti a marginální rozdělení.
2) Seznamte se s pojmy (Shannonova) entropie, informace, podmíněná vzájemná informace, kopule.
3) Studujte příklady maximálně entropických distribucí za specifických předpokladů (zachování momentů, marginálních rozdělení, nebo entropií do k-tého řádu).
4) Nastudujte pojem spojené informace (connected information) k-tého řádu (Schneidman, 2003). Formalizujte související objekty, založené na maximálně entropických distribucích při zachování momentů či entropií (Martin et al., 2016, 2017), či navrhněte vhodné alternativní konstrukce.
5) Studujte vlastnosti sekvence maximálních entropií a spojených informací k-tého řádu a vztahy mezi jednotlivými konstrukcemi.

Doporučená literatura:

1) E. Schneidman, S. Still, M. J. Berry, W. Bialek, Network Information and Connected Correlations. Physical Review Letters 91, 2003, 238701.
2) E. A. Martin, J. Hlinka, J. Davidsen, Pairwise network information and nonlinear correlations. Physical Review E 94, 2016, 040301.
3) E. A. Martin, J. Hlinka, A. Meinke, F. Děchtěrenko, J. Tintěra, I. Oliver, J. Davidsen, Network inference and maximum entropy estimation on information diagrams. Scientific Reports 7, 2017, 7062.

Jméno a pracoviště vedoucího bakalářské práce:

Ing. Mgr. Jaroslav Hlinka, Ph.D.
Ústav informatiky, Akademie věd České republiky, Pod Vodárenskou věží 271/2, 182 07 Praha 8

Jméno a pracoviště konzultanta:

Ing. Jakub Kořenek
Ústav informatiky, Akademie věd České republiky, Pod Vodárenskou věží 271/2, 182 07 Praha 8, &, Katedra matematiky, Fakulta jaderná a fyzikálně inženýrská, České vysoké učení technické v Praze, Trojanova 13, 120 00 Praha 2

Datum zadání bakalářské práce: 31.10.2020

Datum odevzdání bakalářské práce: 7.7.2021

Doba platnosti zadání je dva roky od data zadání.

V Praze dne 30.10.2020

garant oboru

vedoucí katedry

děkan

Čestné prehlásenie

Prehlasujem, že som túto prácu vypracovala samostatne a uvidela som všetku použitú literatúru.

V Prahe dňa 7.7.2021 Katarína Studeničová

Poďakovanie

Na tomto mieste by som rada poďakovala svojmu školiteľovi za možnosť pracovať na zaujímavej téme, za čas, ktorý mi venoval, ústretový prístup a podnetné pripomienky. Taktiež ďakujem svojmu konzultantovi za odpovede na moje otázky, za pomoc a ochotu. V neposlednom rade ďakujem svojim rodičom, ktorí mi poskytli ideálne zázemie na štúdium a prípravu bakalárskej práce.

Katarína Studeničová

Název práce: **Vlastnosti variant spojené informace**

Autor: Katarína Studeničová

Obor: Matematické inženýrství

Zaměření: Matematické modelování

Druh práce: Bakalářská práce

Vedoucí práce: Ing. Mgr. Jaroslav Hlinka, Ph.D.
Ústav informatiky, AV ČR

Konzultant: Ing. Jakub Kořenek
FJFI ČVUT; Ústav informatiky, AV ČR

Abstrakt: V této práci se věnujeme budování matematického aparátu vhodného pro popis multivariátních interakcí v komplexních systémech pomocí odhadů maximální entropie. Formalizujeme definice pěti různých konstrukcí spojené informace založené na odhadu maximální entropie. Dále zkoumáme jejich základní vlastnosti, jako jsou monotonie a omezenost posloupností hodnot funkcionálů, podmínky nezávislosti veličin a vztahy mezi konstrukcemi. Část práce je věnována otázce přesného vztahu mezi vlastnostmi posloupnosti maximálních entropií a výskytem interakcí mezi složkami tvořícími popisovaný systém. Vyslovujeme a dokazujeme několik vět týkajících se interakcí, které jsou jedním z hlavních výsledků práce. Dalším přínosem je vybudování nové maximalizační konstrukce na informačních diagramech a důkaz některých jejích významných vlastností a vztahů k ostatním konstrukcím.

Klíčová slova: informační diagram, maximální entropie, multivariátní interakce, spojená informace

Title: **Properties of connected information variants**

Author: Katarína Studeničová

Abstract: In this work we focus on building a mathematical apparatus suitable for the description of multivariate interactions in complex systems using maximum entropy estimates. We formalize the definitions of five constructions of connected information based on maximum entropy approximation. We examine their basic properties, such as the monotonicity, boundedness, conditions of independence and relations between constructions. Part of the work is devoted to the question of the precise relationship between the properties of the sequence of maximum entropies and the presence of interactions between the components constituting the described system. We formulate and prove several theorems about interactions, which are one of the main results of the work. Another benefit is the construction of a new maximization structure on information diagrams and proof of some of its important properties and relationships to other structures.

Key words: connected information, information diagram, maximal entropy, multivariate interaction

Obsah

Úvod

Pri skúmaní komplexných systémov, napríklad ľudského mozgu alebo klímy planéty Zem, sme schopní pozorovať niektoré charakteristiky systému a zaznamenávať ich hodnoty. Naším cieľom je na základe týchto údajov vytvoriť model pozorovaného systému, resp. preskúmať niektoré jeho konkrétne vlastnosti. Z matematického hľadiska môžeme popis formalizovať metódami pravdepodobnosti, štatistiky a teórie informácie.

Jedným z dôležitých konceptov, ktorý je používaný v celej skupine pravdepodobnostných modelov, je maximálne entropické rozdelenie. Kľúčovým predpokladom je, že maximálne entropické rozdelenie pri daných podmienkach najpresnejšie popisuje systém. Porovnaním tohto rozdelenia s dátami experimentu potom môžeme určiť, či a do akej miery je systém viazaný podmienkami, ktoré sme dovtedy nebrali do úvahy. V nasledujúcom texte sa budeme s pojmom maximálne entropické rozdelenie často stretávať. Bude nás zaujímať, ako nájsť takéto rozdelenie pri určitých zadaných podmienkach, tiež možnosť interpretácie hodnôt entropie takéhoto rozdelenia.

Pri popise systému reprezentovaného vektorovou náhodnou veličinou je prirodzené položiť otázku, či sú zložky tejto náhodnej veličiny nezávislé alebo nie. V prípade, že zložky nie sú nezávislé, vzniká priestor na ďalšie skúmanie vzťahov medzi nimi. To, že systém reprezentujeme všeobecne vektorovou náhodnou veličinou, nám umožňuje, narozdiel napríklad od popisu systému pomocou grafu, určiť nielen závislosť medzi dvojicami prvkov tvoriacich systém, ale aj interakcie, ktoré sa prejavujú vo väčších skupinách. Práve takéto interakcie môžu v niektorých prípadoch niesť významnú časť celkovej informácie systému. Jedným z našich cieľov preto bude vybudovať matematický aparát tak, aby sme pomocou neho mohli určiť, aké rády interakcií sa v systéme vyskytujú a tiež do akej miery prispievajú k celkovej informácii nesenej systémom. Nami budovaný aparát sa naopak nebude zameriavať na zachytenie vlastností systému meniacich sa v čase, možné rozšírenia, ktoré by zachytávali aj dynamiku prenosu informácie, prenechávame na budúce skúmanie.

V prvej kapitole práce oboznámime čitateľa so základnými pojmami týkajúcimi sa pravdepodobnosti a teórie informácie a so značením, ktoré budeme v celej práci používať. Druhá kapitola práce je venovaná formálnej definícii maximálne entropických funkcionálov, ktorých konštrukcie boli navrhnuté v článkoch [1, 2, 3]. Definujeme tri typy maximálnej entropie a spojenej informácie, tiež preskúmame ich základné vlastnosti, napríklad ohraničenosť, monotóniu postupnosti maximálnych entropií, niektoré vzťahy medzi konštrukciami. Vety a skúmané vlastnosti budeme ilustrovať na viacerých príkladoch. Popíšeme tiež možné použitie konkrétnych konštrukcií, ich výhody a predovšetkým niektoré významné rozdiely medzi nimi.

Po vybudovaní základného matematického aparátu sa otvára priestor zamerať sa na interpretáciu hodnôt definovaných funkcionálov. Ako sme už uviedli skôr, v rámci práce nás bude predovšetkým zaujímať možnosť interpretácie súvisiaca s interakciami v popisovanom systéme. Tomuto skúmaniu budeme venovať celú tretiu kapitolu. Najprv predstavíme spôsob konštrukcie maximálne entropického rozdelenia pri určitých zadaných podmienkach, tiež preskúmame niektoré vlastnosti dát, ktoré sme získali zo systému, v ktorom sa prejavujú interakcie. Tento prieskum nám umožní vysloviť a dokázať vety, v ktorých prepojíme hodnoty maximálne entropických funkcionálov, predovšetkým spojenej informácie, s existenciou interakcií medzi zložkami systému. Tieto tvrdenia patria medzi kľúčové výsledky celej práce.

Posledné dve kapitoly sú venované ďalším dvom možným konštrukciám maximálne entropických funkcionálov. Najprv uvedieme teóriu týkajúcu sa informačných diagramov a predstavíme na nich definovanú mieru, následne uvedieme definíciu štvrtého typu maximálnej entropie spomínaného v článkoch [1, 2]. Opäť preskúmame jeho základné vlastnosti a vzťahy ku ostatným konštrukciám, tiež niektoré výhody tohto typu maximálnej entropie. V piatej kapitole predstavíme ďalší zo zaujímavých prínosov práce - novú nami vybudovanú konštrukciu maximálnej entropie. Preskúmame tiež jej charakteristiky, jej možné výhody a nevýhody. V záverečnej časti práce určíme hodnoty všetkých piatich typov maximálnej entropie pre konkrétny príklad a zhrnieme všetky preskúmané vlastnosti a dokázané vety.

Kapitola 1

Teoretický základ

V úvodnej časti pripomenieme vybrané pojmy z pravdepodobnosti a z teórie informácie, ktoré neskôr budeme potrebovať pri definovaní nových veličín a pri formulácii viet, a to predovšetkým v kapitole 2. V podkapitole 1.1 budeme do veľkej miery využívať značenie použité v [4]. Z uvedenej knihy tiež preberáme spôsob výstavby pravdepodobnostnej teórie, ktorý je pomerne všeobecný. Tento štýl výkladu sme vybrali hlavne preto, že cieľom našej práce je predovšetkým korektne matematicky popísať novodefinované veličiny a ich vlastnosti, nie len skúmať ich využitie v praxi. Podobne v podkapitole 1.2 sa budeme až na menšie úpravy držať značenia a obsahu knihy [5].

1.1 Pravdepodobnosť

1.1.1 Definícia pravdepodobnosti

Značenie 1.1 (Základné pojmy)**.** Elementárny jav budeme značiť ω, základnú množinu, t. j. množinu všetkých elementárnych javov, označíme Ω. Javom nazývame ľubovoľnú podmnožinu Ω.

Aby sme mohli s javmi pohodlne pracovať, definovať na nich pravdepodobnosť a skúmať jej vlastnosti, uvažujeme množinu javov takú, že tvorí σ-algebru, teda je uzavretá na doplnky, uzavretá na spočetné zjednotenia a obsahuje Ω. Na takto zvolenej množine javov môžeme definovať pravdepodobnostnú mieru nasledovne:

Definícia 1.2 (Axiomatická definícia pravdepodobnosti)**.** Majme neprázdnu základnú množinu Ω, na nej σ-algebru javov $\mathcal{A}$. Potom funkciu $P : \mathcal{A} \to \mathbb{R}$ spĺňajúcu

K1. $P(\Omega) = 1$,

K2. $(\forall A \in \mathcal{A})\,(P(A) \geq 0)$,

K3. $\left(\forall\,(A_j)_{j=1}^{+\infty} \subset \mathcal{A},\ A_j \text{ navzájom disjunktné}\right)\left(P\left(\bigcup_{j=1}^{+\infty} A_j\right) = \sum_{j=1}^{+\infty} P(A_j)\right)$,

nazveme pravdepodobnostnou mierou.

Podmienky K1-K3 sa tiež nazývajú Kolmogorove axiómy. Priamo z definície by sme ľahko ukázali aditivitu, monotóniu, ohraničenosť a ďalšie základné vlastnosti pravdepodobnosti. Priestor vybavený σ-algebrou javov a pravdepodobnostnou mierou nazveme pravdepodobnostný priestor, značíme $(\Omega, \mathcal{A}, P)$. Uvedieme ešte jednu z kľúčových viet, ktorá sa často používa pri dokazovaní tvrdení z oblasti pravdepodobnosti, Boolovu nerovnosť.

Veta 1.3 (Boolova nerovnosť)**.** *Majme $(\Omega, \mathcal{A}, P)$. Nech pre všetky $j \in \mathbb{N}$ je $A_j \in \mathcal{A}$, potom platí $P\left(\bigcup_{j=1}^{n,+\infty} A_j\right) \leq \sum_{j=1}^{n,+\infty} P(A_j)$.*

Dôkaz. Postupovalo by sa priamočiaro matematickou indukciou. □

Použili sme zápis $\bigcup_{j=1}^{n,+\infty}$ a $\sum_{j=1}^{n,+\infty}$, čím vyjadrujeme, že veta platí pre konečné, ale aj pre spočítateľné zjednotenia, respektíve sumy. Pripomeňme ešte pojmy podmienená pravdepodobnosť, nezávislosť javov a nezávislosť množín javov.

Definícia 1.4 (Podmienená pravdepodobnosť)**.** Majme $(\Omega, \mathcal{A}, P)$, nech $A, B \in \mathcal{A}$, $P(B) > 0$. Potom definujeme podmienenú pravdepodobnosť javu A, pri predpoklade, že nastal jav B, vzťahom

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)}.$$

Definícia 1.5 (Nezávislosť javov)**.** Majme $(\Omega, \mathcal{A}, P)$. Nech $(A_i)_{i \in I}$, kde I je indexová množina (môže byť aj nekonečná), je súbor javov z $\mathcal{A}$. Ak pre každú J konečnú podmnožinu I platí

$$P\left(\cap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i)\,,$$

tak hovoríme, že javy v súbore $(A_i)_{i \in I}$ sú vzájomne nezávislé.

Definícia 1.6 (Nezávislosť množín javov)**.** Majme konečnú postupnosť množín javov $(\mathcal{C}_j)_{j=1}^{n}$. Potom hovoríme, že množiny javov v postupnosti $(\mathcal{C}_j)_{j=1}^{n}$ sú nezávislé, ak pre každú vybranú n-ticu javov $A_1, \ldots, A_n$ takú, že $A_i \in C_i\ \forall i \in \widehat{n}$ platí, že $A_1, \ldots, A_n$ sú nezávislé javy.

Majme nekonečnú postupnosť množín javov $(\mathcal{C}_j)_{j=1}^{+\infty}$. Potom hovoríme, že množiny javov v postupnosti $(\mathcal{C}_j)_{j=1}^{+\infty}$ sú nezávislé, ak sú množiny javov v ľubovoľnej z nej vybranej konečnej podpostupnosti nezávislé.

1.1.2 Náhodné veličiny

Máme zadefinovaný pravdepodobnostný priestor s mierou, ďalej predstavíme pojem náhodná veličina, jej rozdelenie pravdepodobnosti a marginálne rozdelenie.

Definícia 1.7 (Borelovská σ-algebra)**.** Nech $n \in \mathbb{N}$, $\Omega = \mathbb{R}^n$, označme ďalej množinu $\tau = \{\bigtimes_{i=1}^{n} (a_i, b_i) \mid a_i \in \mathbb{R},\ b_i \in \mathbb{R}\}$. Potom minimálnu σ-algebru nad systémom τ nazývame borelovskou σ-algebrou, budeme označovať $\mathcal{B}_n$.

V definícii sme použili pojem minimálna σ-algebra nad systémom τ, čím myslíme najmenší nadsystém systému τ, ktorý je σ-algebrou.

Definícia 1.8 (Náhodná veličina)**.** Majme $(\Omega, \mathcal{A}, P)$, $n \in \mathbb{N}$. Potom vektorovú funkciu $\boldsymbol{X}\colon (\Omega, \mathcal{A}, P) \to (\mathbb{R}^n, \mathcal{B}_n)$ takú, že platí $(\forall B \in \mathcal{B}_n)\,(\boldsymbol{X}^{-1}(B) \in \mathcal{A})$, nazývame vektorová náhodná veličina.

Častokrát budeme namiesto pojmu vektorová náhodná veličina používať skrátený názov náhodná veličina, budeme zapisovať po zložkách $\boldsymbol{X} = (X_1, ..., X_n)$.

Definícia 1.9 (Rozdelenie pravdepodobnosti)**.** Majme $(\Omega, \mathcal{A}, P)$, náhodnú veličinu $\boldsymbol{X}$. Rozdelenie pravdepodobnosti náhodnej veličiny $\boldsymbol{X}$ budeme označovať $P^{\boldsymbol{X}}$ a definujme ho vzťahom $P^{\boldsymbol{X}} := P \circ \boldsymbol{X}^{-1}$.

Definícia 1.10 (Diskrétna náhodná veličina, diskrétna hustota)**.** Hovoríme, že náhodná veličina je diskrétna, ak je jej obor hodnôt maximálne spočítateľný, môžeme ho zapísať ako $\operatorname{Ran} \boldsymbol{X} = (\boldsymbol{x}_i)_{i=1}^{n,+\infty}$. Ďalej označme $p(\boldsymbol{x}) := P(\boldsymbol{X} = \boldsymbol{x})$. Potom hustotou pravdepodobnosti diskrétnej náhodnej veličiny (frekvenčnú funkciu) definujeme vzťahom

$$f_{\boldsymbol{X}}(\boldsymbol{x}) := \begin{cases} p(\boldsymbol{x}) & \text{pre } \boldsymbol{x} = \boldsymbol{x}_i \text{ z oboru hodnôt } \boldsymbol{X} \\ 0 & \text{inak.} \end{cases}$$

V celom texte sa budeme zameriavať na diskrétne náhodné veličiny, značenie $p(\boldsymbol{x}) = P(\boldsymbol{X} = \boldsymbol{x})$ použité v definícii budeme využívať aj naďalej. Špeciálnym prípadom diskrétnej náhodnej veličiny je binárna náhodná veličina, jej obor hodnôt je $\{0, 1\}$. Keďže nebudeme pracovať so spojitými náhodnými veličinami, nepripomíname ani v teoretickom úvode vety a definície s nimi spojené.

Uveďme teraz definíciu marginálneho rozdelenia náhodnej veličiny. Tento pojem bude jedným z kľúčových v kapitole 2, pretože jednou z podmienok, pri ktorej budeme hľadať maximálne entropické distribúcie, budú zadané marginálne rozdelenia skúmanej náhodnej veličiny.

Lema 1.11. *Majme n-rozmernú diskrétnu vektorovú náhodnú veličinu s hustotou pravdepodobnosti* $f_{\boldsymbol{X}}(\boldsymbol{x})$*. Potom náhodná veličina* $\boldsymbol{X}' = (X_1, ..., X_{i-1}, X_{i+1}, ..., X_n)$ *je diskrétna vektorová náhodná veličina s hustotou pravdepodobnosti*

$$f_{\boldsymbol{X}'}(\boldsymbol{x}') = \sum_{x_i} f_{\boldsymbol{X}}(x_1, ..., x_{i-1}, x_i, x_{i+1}, ..., x_n) \qquad \forall \boldsymbol{x}' \in \mathbb{R}^{n-1},$$

kde sčítame cez všetky možné hodnoty i-tej zložky $\boldsymbol{x}$*. Táto hustota jednoznačne určuje rozdelenie pravdepodobnosti náhodnej veličiny* $\boldsymbol{X}'$*.*

Definícia 1.12 (Marginálna hustota, rozdelenie)**.** Majme n-rozmernú diskrétnu vektorovú náhodnú veličinu s hustotou pravdepodobnosti $f_{\boldsymbol{X}}(\boldsymbol{x})$. Potom hustota a rozdelenie pravdepodobnosti z lemy 1.11 nazývame marginálna hustota a marginálne rozdelenie náhodnej veličiny.

V ďalšom texte budeme používať slovné spojenie marginálne rozdelenie k-teho stupňa, čím myslíme, že vysčítame pôvodné rozdelenie n-rozmernej veličiny cez vybraných $n-k$ indexov, zostane nám teda rozdelenie pravdepodobnosti $k-$rozmernej náhodnej veličiny.

Pripomeňme ešte jeden z dôležitých pojmov, s ktorým sa budeme stretávať predovšetkým v ilustračných príkladoch v kapitole 2.

Definícia 1.13 (Nezávislosť náhodných veličín)**.** Majme postupnosť náhodných veličín $(X_j)_{j=1}^{n,+\infty}$ na $(\Omega, \mathcal{A}, P)$ zobrazujúcich do $(\mathbb{R}, \mathcal{B}_1)$. Ak sú množiny javov v postupnosti $\left(X_j^{-1}(\mathcal{B}_1)\right)_{j=1}^{n,+\infty}$ nezávislé v zmysle definície 1.6, tak hovoríme, že náhodné veličiny $(X_j)_{j=1}^{n,+\infty}$ sú nezávislé.

Na záver sekcie uvedieme dva dôležité príklady rozdelení pravdepodobnosti diskrétnych náhodných veličín, ktoré budú spomínané v nasledujúcich kapitolách:

- Bernoulliho (alternatívne) rozdelenie s parametrom q - rozdelenie binárnej náhodnej veličiny, $p(1) = q$, $p(0) = 1 - q$, $q \in [0, 1]$.

- Uniformné (rovnomerné) rozdelenie - rozdelenie náhodnej veličiny s konečným oborom hodnôt, každá z hodnôt v obore nastáva s rovnakou pravdepodobnosťou.

1.1.3 Charakteristiky náhodných veličín

Podobne ako v predchádzajúcej sekcii sa budeme zameriavať len na definície týkajúce sa diskrétnych náhodných veličín.

Definícia 1.14 (Stredná hodnota)**.** Majme jednorozmernú diskrétnu náhodnú veličinu X na $(\Omega, \mathcal{A}, P)$. Potom strednú hodnotu X definujeme ako

$$\mathrm{E}\, X := \sum_{x_i} x_i\, f_{X_i}(x_i),$$

kde sčítame cez všetky $x_i \in \mathrm{Ran}\, X$. Pre n-rozmernú náhodnú veličinu $\boldsymbol{X} = (X_1, ..., X_n)$ definujeme strednú hodnotu po zložkách $\mathrm{E}\, \boldsymbol{X} = (\mathrm{E}\, X_1, ..., \mathrm{E}\, X_n)$.

Mohlo by nás zaujímať, kedy suma v definícii strednej hodnoty konverguje. Pokiaľ je konečná, je konvergentná, ak však pracujeme s diskrétnou veličinou so spočítateľným oborom hodnôt, jedná sa o nekonečnú sumu, teda konvergenciu nemáme zaručenú. Existujú príklady rozdelení pravdepodobnosti, pri ktorých suma diverguje. V ďalších definíciách preto budeme musieť konvergenciu predpokladať. Pri spracovaní reálnych dát však konvergencia problém nie je, keďže pracujeme len s konečnými súbormi.

Značenie 1.15. Majme pravdepodobnostný priestor $(\Omega, \mathcal{A}, P)$, $p \in \mathbb{N}$. Označme množinu $\mathcal{L}^p = \{X \text{ diskrétna náhodná veličina na } (\Omega, \mathcal{A}, P) \mid \mathrm{E}\, X^p < +\infty\}$. Obvykle sa takéto značenie používa pri popise spojitých náhodných veličín, pretože ich stredná hodnota je definovaná cez integrál a $\mathcal{L}^p$ je typické značenie pre veličiny integrabilné s p-tou mocninou. Stredná hodnota diskrétnej veličiny však tiež môže byť chápaná ako integrál, pričom uvažujeme sčítaciu mieru, preto značenie preberáme.

Jednou z podmienok, pri ktorej budeme v kapitole 2 hľadať maximálne entropické distribúcie, sú zadané momenty rozdelenia, preto ich teraz definujeme spolu s niektorými ďalšími súvisiacimi veličinami.

Definícia 1.16 (Rozptyl, smerodajná odchýlka)**.** Majme $(\Omega, \mathcal{A}, P)$, $X \in \mathcal{L}^2$. Potom rozptylom náhodnej veličiny X rozumieme $\mathrm{D}\, X := \mathrm{E}\, (X - \mathrm{E}\, X)^2$. Smerodajnou odchýlkou náhodnej veličiny X rozumieme $\sqrt{\mathrm{D}\, X}$.

Definícia 1.17 (Momenty rozdelenia)**.** Majme $(\Omega, \mathcal{A}, P)$, $k \in \mathbb{N}$, nech X je diskrétna náhodná veličina. Ak existuje konečná $\mathrm{E}\, X^k$, tak túto hodnotu nazývame k-tym momentom rozdelenia.

V literatúre sa moment rozdelenia z uvedenej definície tiež označuje prívlastkom všeobecný, okrem neho sa zavádza ešte centrálny moment vzťahom $\mathrm{E}\,(X - \mathrm{E}\, X)^k$. Mohlo by nás zaujímať, či záleží na tom, či v našom ďalšom postupe použijeme centrálne alebo všeobecné momenty. Presnejšie formulované, chceme vedieť, koľko informácie o náhodnej veličine nám poskytne znalosť všeobecného momentu daného rádu a či je to ekvivalentné informácii, ktorú by sme získali zo znalosti centrálneho momentu. Skúsme nájsť odpoveď pre jednoduchý prípad.

Majme binárnu náhodnú veličinu. Pokiaľ poznáme prvý všeobecný moment, vieme presne určiť pravdepodobnostné rozdelenie veličiny. Prvý centrálny moment je však nulový pre všetky možné rozdelenia, z jeho hodnoty teda nezískame o rozdelení veličiny žiadnu informáciu. Na tomto príklade sme ilustrovali, že pri hľadaní rozdelení vyhovujúcich danej väzbe je medzi voľbou typu momentov rozdiel. Na druhej strane, pokiaľ poznáme strednú hodnotu, potom druhý, tretí a štvrtý centrálny moment náhodnej veličiny sú v jednoznačnom vzťahu so všeobecnými momentmi daného rádu. Pri voľbe väzbových podmienok vyšších rádov teda máme viacero ekvivalentných možností. My budeme, aj pre jednoduchosť zápisu, používať všeobecné momenty.

Objasnime ešte značenie, ktoré budeme využívať pri práci s vyššími zmiešanými momentmi, ide vlastne o výpočet strednej hodnoty náhodnej veličiny, ktorá vznikne súčinom zložiek vektorovej náhodnej veličiny. Pre jednoduchosť značenie popíšeme na dvojrozmernej diskrétnej náhodnej veličine.

Značenie 1.18. Majme dvojrozmernú diskrétnu náhodnú veličinu $\boldsymbol{X} = (X, Y)$. Potom $\mathrm{E}\,(X \cdot Y)$, budeme značiť aj $\mathrm{E}\, XY$, pri predpoklade existencie vypočítame ako

$$\mathrm{E}\, XY = \sum_{x,y} x\, y\, f_{X,Y}(x, y),$$

kde sčítame cez všetky možné hodnoty x a y.

Definícia 1.19 (Kovariancia)**.** Majme $(\Omega, \mathcal{A}, P)$, nech X, Y sú diskrétne náhodné veličiny. Potom ich kovarianciou nazveme pri predpoklade existencie číslo

$$\mathrm{Cov}\,(X, Y) := \mathrm{E}\,[(X - \mathrm{E}\, X)(Y - \mathrm{E}\, Y)].$$

Ďalej by sme mohli kovarianciu normalizovať a definovať koreláciu a korelačný koeficient, ale platí, že pri použití momentov ako väzbových podmienok je jedno, aké volíme škálovanie. Dokonca ani nepotrebujeme pracovať s kovarianciami, v našom ďalšom postupe ich môžeme ekvivalentne nahradiť momentmi súčinu veličín, ktoré sme popísali v 1.18, pretože kovariancia je jednoznačne určená znalosťou prvých a druhých všeobecných momentov. Ďalší dôležitý poznatok je, že ak sú veličiny nezávislé, tak ich korelácia a kovariancia sú nulové. Nevýhodou však pre naše použitie zostáva, že opačná implikácia neplatí.

V nasledujúcej kapitole budeme používať pojem momenty k-teho rádu n-rozmernej náhodnej veličiny, čím myslíme stredné hodnoty súčinu vybranej k-tice zložiek veličiny, pričom indexy zložiek sa môžu aj opakovať. Napríklad pre veličinu $\boldsymbol{X} = (X, Y, Z)$ je jeden z momentov tretieho rádu pre trojicu indexov $(1, 1, 2)$ rovný $\mathrm{E}\, X^2 Y$.

1.2 Teória informácie

V druhej podkapitole teoretického základu sa zoznámime so základnými pojmami z teórie informácie, ktorými sú napríklad entropia a vzájomná informácia, tiež uvedieme niektoré z ich dôležitých vlastností a vzťahov medzi veličinami. Podobne ako v časti 1.1 sa budeme zameriavať na diskrétne náhodné veličiny. Budeme sa držať konvencie $0 \log_2 0 = 0$.

Definícia 1.20 (Shannonova entropia)**.** Nech $\boldsymbol{X}$ je diskrétna vektorová náhodná veličina na $(\Omega, \mathcal{A}, P)$, označme $P(\boldsymbol{X} = \boldsymbol{x}) = p(\boldsymbol{x})$. Potom entropia náhodnej veličiny $\boldsymbol{X}$ je definovaná ako

$$H(\boldsymbol{X}) := -\sum_{\boldsymbol{x}} p(\boldsymbol{x}) \log_2 p(\boldsymbol{x}), \tag{1.1}$$

kde sčítame cez $\boldsymbol{x} \in \operatorname{supp} p(\boldsymbol{x})$.

Uveďme niekoľko komentárov k definičnému vzťahu:

- V definícii sme vzhľadom na ďalšie použitie v práci zvolili logaritmus so základom 2, jednotkou entropie teda budú bity. Môžeme sa stretnúť aj s inou voľbou základu logaritmu.

- Keďže entropia nie je funkciou hodnôt náhodnej veličiny, ale len jej rozdelenia pravdepodobnosti, často sa namiesto $H(\boldsymbol{X})$ používa značenie $H(p(\boldsymbol{x}))$. To budeme v ďalších kapitolách používať aj my.

- Priamo z definície vidíme nezápornosť entropie. Členy v sume sú totiž všetky nekladné, pretože $p(\boldsymbol{x})$ dosahuje na svojom nosiči hodnoty v intervale $(0, 1]$.

- V definícii sme uvažovali vektorovú náhodnú veličinu, v literatúre sa jej entropia častokrát označuje aj pojmom združená entropia.

- Kvôli skráteniu zápisu zaveďme značenie $\boldsymbol{X} \sim p(\boldsymbol{x})$. Myslíme tým, že pre náhodnú veličinu platí $P(\boldsymbol{X} = \boldsymbol{x}) = p(\boldsymbol{x})$. Tiež by sme mohli pre zjednodušenie vyjadrovania povedať, že $p(\boldsymbol{x})$ je hustotou pravdepodobnosti $\boldsymbol{X}$. Pre úplnú korektnosť by sme mali uvádzať, že $p(\boldsymbol{x})$ určuje hustotu $f_{\boldsymbol{X}}(\boldsymbol{x})$ vzťahom z definície 1.10, no tento možno až príliš presný spôsob vyjadrovania nebudeme ďalej využívať. Často tiež budeme používať výraz, že $\boldsymbol{X}$ má rozdelenie $p(\boldsymbol{x})$, čím samozrejme myslíme, že $\boldsymbol{X}$ má rozdelenie $P^{\boldsymbol{X}}$, pre ktoré platí $P(\boldsymbol{X} = \boldsymbol{x}) = p(\boldsymbol{x})$.

- V nasledujúcej kapitole budeme používať pojem entropia rádu k. Vysvetlime, čo tým myslíme: Majme n-rozmernú náhodnú veličinu. Keď jej rozdelenie vysčítame cez $n - k$ vybraných zložiek, dostaneme marginálne rozdelenie stupňa k, ktoré jednoznačne definuje k-rozmernú náhodnú veličinu. Entropia tejto veličiny je entropia k-teho rádu pre vybranú množinu indexov, pričom budeme uvažovať, že indexy sa neopakujú, to predovšetkým preto, lebo platí $H(X, X) = H(X)$.

Definícia 1.21 (Podmienená entropia)**.** Nech $(X, Y) \sim p\,(x, y)$, potom definujeme podmienenú entropiu $H(X \mid Y)$ vzťahom

$$H(X \mid Y) := -\sum_x \sum_y p\,(x, y) \log_2 p\,(x \mid y), \tag{1.2}$$

kde sčítame cez $x, y \in \operatorname{supp} p(x, y)$ také, že $p(x \mid y)$ je kladné.

Definovali sme pre jednoduchosť podmieňovanie jednorozmernou veličinou, pre vektorovú veličinu $\boldsymbol{Y}$ by sme definovali analogicky, podobne by aj X mohla byť vektorová.

Entropia popisuje mieru neurčitosti náhodnej veličiny. Ďalej predstavíme pojem relatívna entropia, ktorým zachytíme istú formu vzdialenosti dvoch hustôt pravdepodobnosti.

Definícia 1.22 (Relatívna entropia)**.** Majme $p(x)$, $q(x)$ diskrétne hustoty pravdepodobnosti. Potom relatívnu entropiu, nazývanú tiež Kullback-Leiblerova vzdialenosť, definujeme vzťahom

$$D\,(p \,||\, q) := \sum_x p(x)\, \log_2 \frac{p(x)}{q(x)}, \tag{1.3}$$

kde sčítame cez $x \in \operatorname{supp} p(x)$ také, že $\frac{p(x)}{q(x)}$ je kladné konečné.

Špeciálnym prípadom relatívnej entropie dôležitým nielen v našom ďalšom postupe, ale aj v iných aplikáciách, je vzájomná informácia, ktorá predstavuje vzdialenosť združenej hustoty dvojrozmernej vektorovej náhodnej veličiny od súčinu jej marginálnych hustôt. Vieme, že ak by boli zložky nezávislé, tak nastáva rovnosť združenej hustoty a súčinu marginálnych hustôt, teda suma je nulová. Ak však nezávislé nie sú, vzájomná informácia je kladná. Z uvedeného vyplýva, že pomocou vzájomnej informácie sme schopní zachytiť akékoľvek, aj nelineárne, vzťahy medzi veličinami, čo je výhodou oproti kovariancii.

Definícia 1.23 (Vzájomná informácia)**.** Majme X, Y jednorozmerné diskrétne náhodné veličiny so združenou hustotou $p(x, y)$ a marginálnymi hustotami $p(x)$ a $p(y)$. Potom vzájomnú informáciu $I(X;Y)$ definujeme vzťahom

$$I(X;Y) := D\left(p(x,y) \,||\, p(x)\,p(y)\right). \tag{1.4}$$

Predstavíme teraz niekoľko dôležitých vzťahov medzi definovanými veličinami. Zámerne uvedieme vety v znení pre dvojrozmernú náhodnú veličinu, aj keď by sme našli a matematickou indukciou ukázali všeobecnejšie pravidlá aj pre viacrozmerné veličiny. Našim zámerom je totiž prepojiť vety s obrázkom 1.1 a ilustrovať tak pojem informačný diagram, ktorý je jedným z kľúčových v článku [1]. Informačné diagramy budeme podrobne popisovať v kapitole 4.

Dôkazy tvrdení neuvádzame, sú priamočiarym prepisom definičných vzťahov.

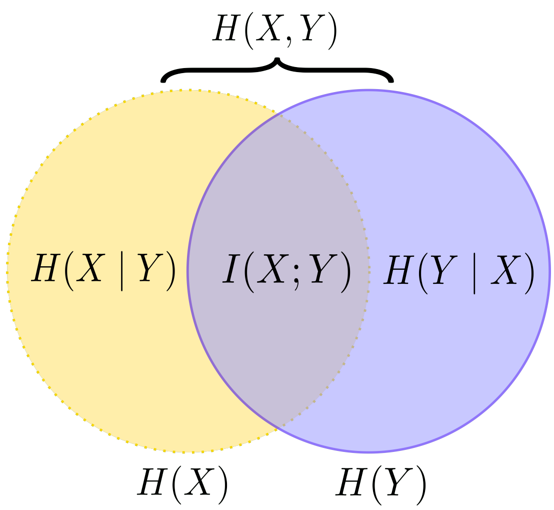
Veta 1.24 (Reťazové pravidlo pre entropiu)**.** *Nech* $(X, Y) \sim p(x, y)$*. Potom platí*

$$H(X, Y) = H(X) + H(Y \mid X). \tag{1.5}$$

Veta 1.25 (Vzťah entropie a informácie)**.** *Nech* $(X, Y) \sim p(x, y)$*. Potom platí*

$$\begin{aligned}
I(X;Y) &= H(X) - H(X \mid Y),\\
I(X;Y) &= H(Y) - H(Y \mid X),\\
I(X;Y) &= H(X) + H(Y) - H(X,Y),\\
I(X;X) &= H(X).
\end{aligned} \tag{1.6}$$

Vzťahy medzi informačnými veličinami by sme mohli vyjadriť Vennovym diagramom, ktorý v takomto prípade nazývame aj informačný diagram. Rozdiel hodnôt veličín chápeme ako rozdiel množín, súčet ako zjednotenie. Pre dvojrozmernú náhodnú veličinu diagram ilustruje nasledujúci obrázok.



Obr. 1.1: Informačný diagram dvojzložkovej náhodnej veličiny

Na obrázku je entropia každej z veličín znázornená kruhom, združená entropia je zjednotením množín, vzájomná informácia je reprezentovaná prienikom. Ako sme už uviedli, konštrukciu diagramov pre tri a viac náhodných veličín popíšeme v samostatnej kapitole 4, v tejto sekcii len definujeme pojmy a popíšeme vzťahy, ktoré budeme pri konštrukcii využívať.

Definícia 1.26 (Podmienená vzájomná informácia)**.** Majme náhodné veličiny X, Y, Z, potom podmienenú vzájomnú informáciu X a Y pri predpoklade Z definujeme vzťahom

$$I(X; Y \mid Z) = H(X \mid Z) - H(X \mid Y, Z). \tag{1.7}$$

Poznamenajme, že rozšírenie definície na vektorové veličiny je priamočiare, keďže podmienená entropia je definovaná aj pre vektorové veličiny.

Intuitívne očakávame, že v diagrame pre tri veličiny budeme musieť nielen popísať rozdiely a prieniky dvoch množín, podobne ako v diagrame dvojrozmernej veličiny, ale budeme tiež chcieť pomenovať časť diagramu reprezentovanú prienikom všetkých troch množín. Chceli by sme, aby táto časť popisovala informáciu medzi troma veličinami. V literatúre však narazíme nielen na problém s názvoslovím, ale taktiež nájdeme rôzne spôsoby, ako kvantifikovať informáciu medzi viacerými náhodnými veličinami. Používajú sa hlavne dve definície, prehľadne ich nájdeme spolu s anglickým názvoslovím uvedené v [6]. V našej práci budeme potrebovať obidva spôsoby definície, preto ich teraz uvedieme.

Definícia 1.27 (Viacrozmerná vzájomná informácia)**.** Majme náhodné veličiny X, Y, Z, potom ich viacrozmernú vzájomnú informáciu definujeme ako

$$I_1(X;Y;Z) := I(X;Y) - I(X;Y \mid Z). \tag{1.8}$$

Oproti článku [6] uvádzame opačnú znamienkovú konvenciu, ktorá je však v súlade s článkom [1]. Takúto voľbu znamienok robíme predovšetkým preto, že takto definovaná veličina popisuje prienik troch množín informačného diagramu. Poznamenajme, že nie je podstatné, ktorou veličinou sa podmieňuje, označenie premenných je zámenné. Vzhľadom na symetriu diagramu je to očakávané.

Definícia 1.28 (Multiinformácia)**.** Majme diskrétnu vektorovú náhodnú veličinu $\boldsymbol{X} = (X_1, \ldots, X_n)$. Potom jej multiinformáciu definujeme ako

$$I_2(X_1; \ldots; X_n) := \sum_{x} p(\boldsymbol{x}) \log_2 \frac{p(\boldsymbol{x})}{\prod_{i=1}^{n} p(x_i)}, \tag{1.9}$$

kde sčítame cez také $\boldsymbol{x} = (x_1, \ldots, x_n) \in \operatorname{supp} \boldsymbol{x}$ také, že $p(x_i)$ je kladné $\forall i \in \widehat{n}$.

Použili sme skrátené značenie $\widehat{n} := \{1, \ldots, n\}$, v texte ho budeme využívať aj naďalej.

Multiinformácia, tiež nazývaná totálna korelácia, reprezentuje celkové množstvo informácie potrebnej na popis daného systému. Vidíme, že definícia je analogická definičnému vzťahu pre vzájomnú informáciu a my ju uvádzame hlavne preto, že v kapitole o interakciách v poznámke 3.54, budeme multiinformáciu rozkladať pomocou novodefinovaných veličín, takzvaných spojených informácií. Na záver kapitoly uvedieme vlastnosti veličín, ktoré budeme potrebovať pri dokazovaní nových tvrdení v kapitole 2. Vety nebudeme dokazovať, v prípade záujmu by sme dôkazy našli napríklad v knihe [5].

Veta 1.29 (Nezápornosť relatívnej entropie)**.** *Nech $p(x)$ a $q(x)$ sú hustoty pravdepodobnosti, potom platí*

$$D\left(p \mid\mid q\right) \geq 0, \tag{1.10}$$

kde rovnosť nastáva práve vtedy, keď $p(x) = q(x)$ pre $\forall x$.

Dôsledok 1.30 (Nezápornosť informácie)**.** *Pre dve náhodné veličiny X, Y platí*

$$I(X;Y) \geq 0, \tag{1.11}$$

kde rovnosť nastáva práve vtedy, keď X a Y sú nezávislé.

Jedným z kľúčových pojmov, s ktorým sa budeme v nasledujúcom texte stretávať, je maximálne entropické rozdelenie pri daných podmienkach. Nasledujúca veta určuje hranicu pre maximálnu hodnotu entropie vektorovej veličiny, tou je súčet entropií jednotlivých zložiek.

Veta 1.31 (Maximálna entropia)**.** *Majme* $\boldsymbol{X} = (X_1, \ldots, X_n) \sim p(x_1, \ldots, x_n)$*, potom*

$$H(X_1, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i), \tag{1.12}$$

kde rovnosť nastáva práve vtedy, keď sú zložky $\boldsymbol{X}$ *navzájom nezávislé.*

Ďalšou z vlastností entropie je, že pridávaním požiadaviek na vlastnosti rozdelenia ju nezvýšime, čo môžeme aj interpretovať - pokiaľ sa dozvieme o veličine nejakú informáciu, naša neznalosť sa zníži. Uveďme preto ešte jedno tvrdenie spojené s touto vlastnosťou.

Veta 1.32 (Maximálne entropické rozdelenie)**.** *Majme* X *náhodnú veličinu, potom*

$$H(X) \leq \log_2 \|\mathcal{X}\|, \tag{1.13}$$

kde $\|\mathcal{X}\|$ *značí počet prvkov v obore hodnôt* X*, rovnosť nastáva práve vtedy, keď má* X *uniformné rozdelenie.*

V ďalšom texte budeme často používať vzťah $H(X_1, \ldots, X_n) \leq \sum_{i=1}^{n} \log_2 \|\mathcal{X}_i\|$, ktorého platnosť je priamočiarym dôsledkom predchádzajúcich viet.

Týmto tvrdením uzatvárame prehľad vybraných definícii a vlastností, ktoré budeme v texte ďalej využívať. V ďalšej kapitole sa bližšie zameriame na maximálne entropické distribúcie, rozoberieme obsah článkov [1, 2, 3], formálne definujeme veličiny používané v textoch a preskúmame niektoré ich vlastnosti.

Kapitola 2

Veličiny maximalizované na rozdeleniach

V úvodnej kapitole sme pripomenuli matematický aparát potrebný pre náš ďalší postup, môžeme teda pristúpiť k formalizácii pojmov z článkov [1, 2, 3] a ku skúmaniu vlastností novodefinovaných objektov. V článkoch boli postupne zavedené viaceré nové veličiny, ktoré sa vzápätí využili pri budovaní modelov popisujúcich komplexný systém. Keďže sú však články do veľkej miery zamerané aplikačne, definície veličín sú formulované len skratkovito. Okrem toho, popisujú sa novovznikajúce koncepty, takže značenie veličín nie je ustálené, podobne nie sú jasné niektoré detaily. Formulácia definícií objektov v článkoch tiež nie je vhodná na skúmanie vlastností veličín matematicky korektným spôsobom. My preto zadefinujeme veličiny presnejšie a ukážeme niektoré ich základné vlastnosti. Budeme sa snažiť zjednotiť značenie, pričom ho do veľkej miery preberieme z článku [1].

2.1 Definície

Definície nových objektov uvedené v tejto podkapitole sa môžu zdať na prvý pohľad komplikované, ujasnime si preto najprv myšlienku, ktorá za nimi stojí. Máme komplexný systém, pozorujeme nejakú jeho vlastnosť. Komplexný systém môžeme chápať ako sieť n prvkov, ktoré spolu môžu interagovať. Pozorovaním danej vlastnosti tohto systému potom myslíme pozorovanie hodnôt určitej n-rozmernej vektorovej náhodnej veličiny. Predpokladáme, že sme schopní na základe pozorovaní zistiť niektoré charakteristiky systému, v našom prípade to budú marginálne rozdelenia k-teho rádu, resp. združené entropie alebo momenty rozdelenia rádu 1 až k pozorovanej náhodnej veličiny. Tieto podmienky použijeme ako väzby, pri ktorých budeme počítať maximálnu prípustnú entropiu, to je veličina, ktorú chceme definovať.

Keďže máme tri druhy väzbových podmienok, budeme definovať tri typy maximálnej entropie. Podľa toho, ktorý maximálne rád momentov, resp. marginálnych rozdelení alebo entropií použijeme ako väzbu, označíme aj maximálnu entropiu ako entropiu daného rádu. Rozdiel maximálnych entropií dvoch po sebe idúcich rádov jedného typu nazveme spojená informácia, jej vlastnostiam a významu v popise systému sa budeme viac venovať v podkapitole 2.3 a v kapitole 3.

Pred samotnou definíciou maximálne entopických funkcionálov zaveďme ešte množinu náhodných veličín, ktoré majú rovnaký obor hodnôt a ujasnime značenie.

Definícia 2.1 (Množina n.v. s daným oborom hodnôt)**.** Majme $n \in \mathbb{N}$, maximálne spočítateľné množiny $R_j \subset \mathbb{R}\ \forall j \in \widehat{n}$. Potom množinu náhodných veličín s daným oborom hodnôt, značíme $\{\boldsymbol{X}\}$, definujeme ako $\{\boldsymbol{X}\} := \{(X_1, \ldots, X_n)\ \mid \operatorname{Ran} X_j \subset R_j\ \forall j \in \widehat{n}\}$.

Poznámka 2.2. V definíciách a vetách budeme množinu náhodných veličín s daným oborom hodnôt niekedy označovať aj symbolom $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$. Naopak, pokiaľ budeme chcieť upozorniť na fakt, že hovoríme o náhodnej veličine, budeme používať zápis $\boldsymbol{X} = (X_1, \ldots, X_n) \sim p(\boldsymbol{x})$, čím zvýrazníme, že sa jedná o jednu náhodnú veličinu s konkrétnym rozdelením pravdepodobnosti.

Značenie 2.3. Vo všetkých troch definíciách maximálnych entropií pri zachovaní podmienok budeme používať podobné značenie pre množiny, ktoré plnia rovnakú úlohu. Množiny sa budú v definíciách vyskytovať s vhodným horným indexom.

Usporiadanú množinu hodnôt väzbových podmienok, ktoré máme zadané, označíme M. V definícii maximálnej entropie pri zachovaní marginálnych rozdelení namiesto množiny M zadávame väzbové podmienky v podobe daných rozdelení veličín $\widetilde{\boldsymbol{X}}_i$.

Ďalej N je indexová množina, ktorá nám umožní správne priradiť hodnoty zadané v M k hodnotám veličín vypočítaným z rozdelení. Prvky N sú vektory indexov, mohli by sme ich tiež chápať ako multiindexy. V prípade maximálnej entropie pri zachovaní momentov sú v indexovej množine všetky možné vybrané 1 až k-tice z $\widehat{n}$ s opakovaním, v prípade zachovaní entropií vyberáme indexy bez opakovania, v prípade zachovania marginálnych rozdelení vyberáme len k-tice bez opakovania.

Množina prípustných rozdelení A bude obsahovať všetky rozdelenia pravdepodobnosti pre náhodné veličiny z $\{\boldsymbol{X}\}$, ktoré spĺňajú zadané väzbové podmienky. V definícii a vo vetách budeme, pokiaľ nebude určené inak, chápať obor hodnôt $\{\boldsymbol{X}\}$ ako zadaný pevne, ale ľubovoľne, nebudeme ho preto explicitne zapisovať.

Pre praktické použitie je pre nás dôležité, aby boli množiny A neprázdne, v definíciách to preto budeme predpokladať. V príkladoch a vetách bude neprázdnosť množiny prípustných rozdelení obvykle zaručená tým, že ako maximalizačné podmienky budeme zadávať hodnoty marginál, resp. entropií alebo momentov vypočítané pre nejakú náhodnú veličinu $\tilde{\boldsymbol{X}}$, ktorej rozdelenie poznáme a má obor hodnôt rovnaký, ako veličina, ktorej entropiu maximalizujeme. Máme teda zaručené, že aspoň rozdelenie veličiny $\tilde{\boldsymbol{X}}$ je v množine A.

Značenie 2.4. Pre $p(\boldsymbol{x})$ rozdelenie pravdepodobnosti veličiny $\boldsymbol{X} = (X_1, \dots, X_n)$ a multiindex $\boldsymbol{n} = (i_1, \dots, i_k)$ budeme označovať $p_{\boldsymbol{n}}(\boldsymbol{x})$ marginálne rozdelenie, ktoré v rozdelení $p(\boldsymbol{x})$ zachová x_j, ak má index v $\boldsymbol{n}$.

Definícia 2.5 (Maximálna entropia pri zachovaní marginálnych rozdelení rádu k)**.** Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $k \leq n$. Nech $\{\boldsymbol{X}\} = \{(X_1, \dots, X_n)\}$ je množina náhodných veličín s daným oborom hodnôt. Nech $\tilde{\boldsymbol{X}}_i = (\tilde{X}_{i_1}, \dots, \tilde{X}_{i_k}) \sim \tilde{p}_i(\boldsymbol{x}) = \tilde{p}_i(x_{i_1}, \dots, x_{i_k})$ sú $\forall i \in \widehat{\binom{n}{k}}$ diskrétne vektorové náhodné veličiny. Majme $N^{(k)} = \{(i_1, \dots, i_k) \mid 1 \leq i_1 < \dots < i_k \leq n\}$ množinu multiindexov, jej prvky označme $\boldsymbol{n}_l$. Nakoniec definujme množinu prípustných rozdelení

$$A^{(k)} = \left\{ p(\boldsymbol{x}) \text{ rozdelenie } \boldsymbol{X} \in \{\boldsymbol{X}\} \;\middle|\; \forall l \in \widehat{\binom{n}{k}} : p_{\boldsymbol{n}_l}(\boldsymbol{x}) = \tilde{p}_l(\boldsymbol{x}) \right\},$$

pričom predpokladáme, že $\tilde{\boldsymbol{X}}_i$ sú volené tak, aby nosiče $p_{\boldsymbol{n}_l}$ a $\tilde{p}_l$ boli rovnaké $\forall l \in \widehat{\binom{n}{k}}$. Potom, pri predpoklade $A^{(k)} \neq \emptyset$, definujeme maximálnu entropiu pri zachovaní marginálnych rozdelení rádu k, značíme $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$, ako

$$H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}})) := \sup_{p(\boldsymbol{x}) \in A^{(k)}} H(p(\boldsymbol{x})). \tag{2.1}$$

Poznámka 2.6. Objasnime, prečo v značení $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$ uvádzame $\tilde{\boldsymbol{X}}$, aj keď tento objekt v definícii nie je spomínaný. V definícii predpokladáme, že množina $A^{(k)}$ je neprázdna, teda existuje nejaká náhodná veličina, ktorá má marginály rovnaké, ako marginály veličín $\tilde{\boldsymbol{X}}_{\boldsymbol{i}}$, ktoré používame ako väzbové podmienky. Túto veličinu môžeme označiť $\tilde{\boldsymbol{X}}$. Podobne, v definíciách 2.11 a 2.15 bude $\tilde{\boldsymbol{X}}$ predstavovať veličinu, ktorá spĺňa podmienky na zadané momenty, resp. entropie. V príkladoch a vetách budeme najčastejšie rovno používať zadávanie väzbových podmienok pomocou veličiny $\tilde{\boldsymbol{X}}$, naše značenie je teda aj intuitívne, vyjadruje totiž funkčnú závislosť funkcionálu maximálnej entropie od $\tilde{\boldsymbol{X}}$.

Poznámka 2.7. Mohlo by sa zdať, že podmienka na maximalizáciu pri marginálnych rozdeleniach sa líši od ostatných podmienok v tom, že maximalizujeme len pri znalosti rozdelení rádu k, zatiaľ čo v nasledujúcich definíciách požadujeme znalosti entropií, resp. momentov od rádu 1 do k vrátane. Stačí si však uvedomiť, že z marginálneho rozdelenia rádu k vieme jednoznačne určiť všetky marginálne rozdelenia nižších rádov. Podmienky pre nižšie rády rozdelení teda nie sú nezávislé od podmienky pre rád k, ich pridaním do väzieb, pri ktorých maximalizujeme, sa nič nezmení.

Poznamenajme ešte, že v prípade diskrétnych veličín s konečným oborom hodnôt sa suprémum na niektorom z rozdelení dosahuje, teda je rovné maximu. Podobne pri ostatných typoch maximálnej entropie. Pre diskrétne vektorové veličiny s konečným oborom hodnôt preto má zmysel uvažovať pojem maximálne entropické rozdelenie, čo je rozdelenie, na ktorom sa dosahuje maximálna entropia pri zadaných väzbách.

Definícia 2.8 (Spojená informácia rádu k pri zachovaní marginálnych rozdelení)**.** Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $2 \leq k \leq n$, ostatné predpoklady ako v definícii 2.5. Nech platí $H(\tilde{P}^{(1)}(\tilde{\boldsymbol{X}})) < +\infty$. Potom definujeme spojenú informáciu rádu k pri zachovaní marginálnych rozdelení, značíme $I_c^{(k)}(\tilde{\boldsymbol{X}})$, ako

$$I_c^{(k)}(\tilde{\boldsymbol{X}}) := H(\tilde{P}^{(k-1)}(\tilde{\boldsymbol{X}})) - H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}})). \tag{2.2}$$

Poznámka 2.9. Konečnosť maximálnej entropie prvého rádu implikuje konečnosť všetkých vyšších rádov, viď veta o monotónii 2.24, preto stačí v definícii požadovať podmienku konečnosti len pre prvý rád. Podobné vysvetlenie platí tiež pre maximalizáciu pri zadaných entropiách alebo momentoch, viď vety 2.29 a 2.34, v definíciách 2.13 a 2.17 preto tiež budeme požadovať konečnosť len pre prvý rád maximálnej entropie.

Značenie 2.10. Pre $p(\boldsymbol{x})$ rozdelenie pravdepodobnosti veličiny $\boldsymbol{X} = (X_1, \ldots, X_n)$ a multiindex $\boldsymbol{n} = (i_1, \ldots, i_k)$ budeme označovať $\mathrm{E}_{\boldsymbol{n}} = \mathrm{E}(X_{i_1} \cdot \ldots \cdot X_{i_k})$.

Definícia 2.11 (Maximálna entropia pri zachovaní momentov rozdelenia do rádu k)**.** Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $k \leq n$. Nech $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$ je množina náhodných veličín s daným oborom hodnôt. Majme usporiadanú množinu hodnôt podmienok $M^{<k>}$, v ktorej je $s = \sum_{j=1}^{k} \binom{n+j-1}{j}$ reálnych čísiel a $N_j = \{(i_1, \ldots, i_j) \mid 1 \leq i_1 \leq \cdots \leq i_j \leq n\}$

množiny multiindexov $\forall j \in \widehat{k}$. Označme $N^{<k>} = \bigcup_{j=1}^{k} N_j$. Ďalej jej prvky označme $\boldsymbol{n}_l$, podobne prvky $M^{<k>}$ označme m_l. Nakoniec definujme množinu prípustných rozdelení

$$A^{<k>} = \{p(\boldsymbol{x}) \text{ rozdelenie } \boldsymbol{X} \in \{\boldsymbol{X}\} \mid \forall l \in \widehat{s} : \mathrm{E}_{\boldsymbol{n}_l} = m_l\}.$$

Potom, pri predpoklade $A^{<k>} \neq \emptyset$, definujeme maximálnu entropiu pri zachovaní momentov rozdelenia do rádu k, značíme $H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$, ako

$$H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}})) := \sup_{p(\boldsymbol{x}) \in A^{<k>}} H(p(\boldsymbol{x})). \tag{2.3}$$

Poznámka 2.12. Množina hodnôt podmienok $M^{<k>}$ predstavuje zadávané hodnoty momentov, pri ktorých maximalizujeme. Vo všeobecnosti za túto množinu môžeme zadať ľubovoľnú usporiadanú množinu reálnych čísiel vhodnej mohutnosti, to nám však nezaručuje existenciu rozdelenia, ktoré by takto zadávané podmienky spĺňalo. V definícii preto musíme predpokladať neprázdnosť množiny $A^{<k>}$. Takýmto spôsobom budeme zadávať podmienky na maximalizáciu pri zadaných momentoch napríklad v príklade 2.31. Častejšie však budeme vo vetách používať iný spôsob zadania množiny $M^{<k>}$. Bude daná vektorová veličina $\tilde{\boldsymbol{X}}$, množina $M^{<k>}$ bude obsahovať hodnoty príslušných momentov veličiny $\tilde{\boldsymbol{X}}$. V tomto prípade môžeme hovoriť, že maximalizujeme entropiu pri momentoch veličiny $\tilde{\boldsymbol{X}}$. Tento spôsob zadávania maximalizačných väzieb využijeme napríklad vo vete o monotónii 2.29. Presnejšie je tento postup zadávania podmienok popísaný v predpokladoch vety 2.48. V tomto prípade máme zaručenú neprázdnosť množiny $A^{<k>}$. Z tohto postupu je tiež jasné, prečo hovoríme o maximalizácii pri zachovaní momentov.

Definícia 2.13 (Spojená informácia rádu k pri zachovaní momentov rozdelenia)**.** Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $2 \leq k \leq n$, ostatné predpoklady ako v definícii 2.11. Nech platí $H(\tilde{P}^{<1>}(\tilde{\boldsymbol{X}})) < +\infty$. Potom definujeme spojenú informáciu rádu k pri zachovaní momentov rozdelenia, značíme $I_c^{<k>}(\tilde{\boldsymbol{X}})$, ako

$$I_c^{<k>}(\tilde{\boldsymbol{X}}) := H(\tilde{P}^{<k-1>}(\tilde{\boldsymbol{X}})) - H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}})). \tag{2.4}$$

Značenie 2.14. Pre $p(\boldsymbol{x})$ rozdelenie pravdepodobnosti veličiny $\boldsymbol{X} = (X_1, \ldots, X_n)$ a multiindex $\boldsymbol{n} = (i_1, \ldots, i_k)$ budeme označovať $H_{\boldsymbol{n}} = H(X_{i_1}, \ldots, X_{i_k})$.

Definícia 2.15 (Maximálna entropia pri zachovaní entropií do rádu k)**.** Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $k \leq n$. Nech $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$ je množina náhodných veličín s daným oborom hodnôt. Majme usporiadanú množinu hodnôt podmienok $M^{[k]}$, v ktorej je $s = \sum_{j=1}^{k} \binom{n}{j}$ nezáporných reálnych čísiel a $N_j = \{(i_1, \ldots, l_k) \mid 1 \leq i_1 < \cdots < i_j \leq n\}$ množiny multiindexov $\forall j \in \widehat{k}$. Označme $N^{[k]} = \bigcup_{j=1}^{k} N_j$. Ďalej označme prvky množiny $M^{[k]}$ ako m_l, podobne prvky $N^{[k]}$ ako $\boldsymbol{n}_l$. Nakoniec definujme množinu prípustných rozdelení

$$A^{[k]} = \{p(\boldsymbol{x}) \text{ rozdelenie } \boldsymbol{X} \in \{\boldsymbol{X}\} \mid \forall l \in \widehat{s} : H_{\boldsymbol{n}_l} = m_l\}.$$

Potom, pri predpoklade $A^{[k]} \neq \emptyset$, definujeme maximálnu entropiu pri zachovaní entropií do rádu k, značíme $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$, ako

$$H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})) := \sup_{p(\boldsymbol{x}) \in A^{[k]}} H(p(\boldsymbol{x})). \tag{2.5}$$

Poznámka 2.16. Podobne ako v definícii 2.11, aj v definícii 2.15 môžeme zadávať množinu podmienok $M^{[k]}$ dvomi spôsobmi. Prvým možným spôsobom je zadanie ľubovoľnej množiny nezáporných reálnych čísiel vhodnej mohutnosti. Nemáme zaručené, že sa tieto hodnoty väzbových podmienok dosahujú na niektorom z rozdelení, preto v definícii predpokladáme neprázdnosť množiny $A^{[k]}$. Druhým spôsobom je vypočítanie príslušných hodnôt entropie z veličiny $\tilde{\boldsymbol{X}}$ so zadaným rozdelením, v tomto prípade máme zaručenú neprázdnosť $A^{[k]}$. Tento spôsob budeme slovne označovať ako maximalizáciu entropie pri entropiách veličiny $\tilde{\boldsymbol{X}}$. Presnejšie bude popísaný v predpokladoch vety 2.45.

Definícia 2.17 (Spojená informácia rádu k pri zachovaní entropií)**.** Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $2 \leq k \leq n$, ostatné predpoklady ako v definícii 2.15. Nech platí $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) < +\infty$. Potom definujeme spojenú informáciu rádu k pri zachovaní entropií, značíme $I_c^{[k]}(\tilde{\boldsymbol{X}})$, ako

$$I_c^{[k]}(\tilde{\boldsymbol{X}}) := H(\tilde{P}^{[k-1]}(\tilde{\boldsymbol{X}})) - H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})). \tag{2.6}$$

Poznámka 2.18. Definovali sme tri druhy maximálnej entropie a tri druhy spojenej informácie. Uveďme ešte, kde sa nami definované veličiny spomínajú v článkoch. S pojmom spojená infomácia sa prvýkrát stretávame v [3], kde sa za maximalizačné podmienky uvažujú marginálne rozdelenia, okrajovo sa spomína možnosť maximalizovať pri daných momentoch. Text tiež poskytuje niekoľko možných aplikácií pri popise systémov a interpretáciu významu spojenej informácie. Poznamenajme, že problém maximalizácie entropie pri zadaných momentoch je známy zo štatistickej fyziky, obvykle sa však uvažujú len prvé a druhé momenty. Rieši sa pomocou Lagrangeových multiplikátorov.

Články [1, 2] problém maximalizácie pri zadaných momentoch spomínajú len krátko, zameriavajú sa predovšetkým na maximalizáciu entropie pri zadaných entropiách do určitého rádu, a to hlavne preto, že autori chcú zachytiť aj nelineárne vzťahy medzi prvkami. Môže sa zdať, že podmienky maximalizácie volia autori iné ako my, napríklad ako maximalizačná podmienka pre dve veličiny je v článku uvedená znalosť vzájomnej informácie medzi dvoma veličinami a znalosť entropií oboch veličín. Táto podmienka je však ekvivalentná znalosti entropií prvého a druhého rádu, ako ľahko vidno zo vzťahov 1.25. Podobne by sme mohli ekvivalentne preformulovať vyššie rády podmienok. Doplňme, ako aj autori v článku navrhujú, že by sme mohli ďalej zavádzať nové druhy maximálnych entropií pri inej voľbe zadaných informačných veličín.

Uvedené tri typy maximálnych entropií boli definované ako suprémum entropie na určitej vybranej množine rozdelení (aj preto sme zvolili názov tejto kapitoly). V článkoch sa stretneme ešte so štvrtým druhom maximálnej entropie a spojenej informácie. Jeho zavedenie vyplýva z problému výpočtu maximálnej entropie pri zadaných čiastočných entropiách, ktorý je ťažko riešiteľný, a tak bol v [2] preformulovaný na optimalizačný problém na informačných diagramoch. Tento typ veličín podrobnejšie popíšeme v kapitolách 4 a 5. V poslednej kapitole tiež navrhneme piaty možný typ maximálnej entropie, ktorý sa pôvodne v článkoch vôbec nevyskytuje.

2.2 Vety o entropii

V tejto podkapitole preskúmame základné vlastnosti sekvencie maximálnych entropií a dokážeme niektoré vzťahy medzi tromi definovanými konštrukciami.

2.2.1 Monotónia a podmienky nezávislosti

Vyslovíme a budeme chcieť dokázať tvrdenia o monotónii troch typov maximálnych entropií. Myšlienky dôkazov by sa opakovali, preto ich pre prehľadnosť uvedieme v sérii pomocných tvrdení, ktoré využijeme aj v dôkazoch v ďalších sekciách. Ako prvé uveďme jednoduché pozorovanie o suprémach:

Lema 2.19. *Pre reálnu funkciu f s definičným oborom B platí pre ľubovoľnú neprázdnu $A \subset B$, že $\sup_A f \leq \sup_B f$.*

Suprémum budeme hľadať na množinách rozdelení pravdepodobnosti vyhovujúcich väzbovým podmienkam, tie budú tvoriť postupnosť usporiadanú inklúziou. Je zrejmé, že platí:

Lema 2.20. *Nech $n \in \mathbb{N}$ a $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$ je množina náhodných veličín s daným oborom hodnôt. Majme dané pre všetky $k \in \widehat{n}$ množiny podmienok M_k, ktorých platnosť pre rozdelenie veličín z $\{\boldsymbol{X}\}$ môžeme požadovať. Nech platí $M_1 \subset M_2 \subset \cdots \subset M_n$. Označme ďalej A_k množinu rozdelení pravdepodobnosti veličín z $\{\boldsymbol{X}\}$, ktoré spĺňajú podmienky v M_k. Potom platí $A_n \subset A_{n-1} \subset \cdots \subset A_1$.*

Aby sme dokázali monotóniu maximálnych entropií, potrebujeme, aby nami zadávané podmienky v tvare zadaných momentov, entropií, resp. marginálnych rozdelení, tvorili postupnosť usporiadanú inklúziou, nie je ťažké vidieť, že je to naozaj tak, teda platí:

Lema 2.21. *Nech $n \in \mathbb{N}$, majme $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim p(\tilde{\boldsymbol{x}})$ diskrétnu vektorovú náhodnú veličinu. Nech má $\tilde{\boldsymbol{X}}$ všetky momenty a entropie k-teho rádu konečné pre všetky $k \in \widehat{n}$. Označme M_k^1 množinu marginálnych rozdelení $\tilde{\boldsymbol{X}}$ do rádu k vrátane, M_k^2 množinu entropií $\tilde{\boldsymbol{X}}$ do rádu k vrátane, M_k^3 množinu momentov $\tilde{\boldsymbol{X}}$ do rádu k vrátane. Potom pre takto označené množiny podmienok platí $M_1^j \subset M_2^j \subset \cdots \subset M_n^j$ pre všetky $j \in \{1, 2, 3\}$.*

Platnosť tvrdenia je zrejmá. Môžeme si však všimnúť, že v tejto leme predpokladáme explicitne zadané podmienky na marginály do rádu k vrátane, čo je potrebné na to, aby platila inklúzia. To je však v praxi zbytočné, pretože marginály rádu k jednoznačne určujú marginály nižších rádov. Môžeme preto zaviesť nové značenie.

Značenie 2.22. Zaveďme označenie $M_i \implies M_j$. Majme danú množinu podmienok M_i, napríklad marginálne rozdelenia určitého stupňa. V tejto množine nie sú explicitne vypísané podmienky napríklad na hodnotu momentov do tohto stupňa, označme množinu týchto hodnôt M_j, ale vieme ich z marginál jednoznačne určiť. Platnosť podmienok v množine M_i teda implikuje platnosť podmienok v množine M_j.

Zjavne platí, že množina väzieb M_k určujúca marginály rádu k síce nie je vo vzťahu inklúzie s množinou M_{k-1}, v ktorej sú určené marginály rádu $k-1$, ale platí $M_n \implies M_{n-1} \implies \cdots \implies M_1$. Vyslovme preto ešte pomocné tvrdenie, ktoré je zovšeobecnením lemy 2.20, budeme ho potrebovať pri dôkaze monotónie entropie pri zadaných marginálach, ale predovšetkým pri dôkaze viet o vzťahoch medzi konštrukciami.

Lema 2.23. *Nech $n \in \mathbb{N}$ a $\{\boldsymbol{X}\} = \{(X_1, \dots, X_n)\}$ je množina náhodných veličín s daným oborom hodnôt. Majme $m \in \widehat{n}$ a majme dané pre všetky $k \in \widehat{m}$ množiny podmienok M_k, ktorých platnosť pre rozdelenie veličiny z $\{\boldsymbol{X}\}$ môžeme požadovať. Nech platí $M_m \Longrightarrow M_{m-1} \Longrightarrow \cdots \Longrightarrow M_1$. Označme ďalej A_k množinu rozdelení pravdepodobnosti veličín z $\{\boldsymbol{X}\}$, ktoré spĺňajú podmienky v M_k. Potom platí $A_m \subset A_{m-1} \subset \cdots \subset A_1$.*

Z uvedených pomocných tvrdení už vyplýva monotónia všetkých troch typov entropie, môžeme teda vysloviť a dokázať tvrdenia 2.24, 2.29, 2.34.

V predpokladoch viet sa v nasledujúcom texte viackrát objaví predpoklad, že náhodná veličina $\tilde{\boldsymbol{X}}$ má rovnaký obor hodnôt, ako je obor hodnôt zadaný pre $\{\boldsymbol{X}\}$ množinu veličín s daným oborom hodnôt. Skrátene budeme hovoriť, že $\tilde{\boldsymbol{X}}$ a $\{\boldsymbol{X}\}$ majú rovnaký obor hodnôt.

Veta 2.24 (Monotónia $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$)**.** *Majme náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \dots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ a množinu $\{\boldsymbol{X}\} = \{(X_1, \dots, X_n)\}$ s rovnakým oborom hodnôt. Maximalizujme entropiu pri marginálnych rozdeleniach veličiny $\tilde{\boldsymbol{X}}$. Potom pre všetky $k \in \widehat{n-1}$ platí*

$$H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}})) \geq H(\tilde{P}^{(k+1)}(\tilde{\boldsymbol{X}})). \tag{2.7}$$

Dôkaz. Vieme, že postupnosť zadaných marginál spĺňa podmienky z lemy 2.23, postupnosť rozdelení, ktoré dané podmienky splňujú je teda usporiadaná inklúziou a klesá. Ak na tejto množine spočítame suprémum hodnôt entropie, z lemy 2.19 dostávame klesajúcu postupnosť maximálnych entropií pri zadaných marginálach. □

Poznámka 2.25. V znení vety pre úplnosť a súvislosť s definíciou maximálnej entropie uvádzame okrem náhodnej veličiny $\tilde{\boldsymbol{X}}$ aj množinu $\{\boldsymbol{X}\}$. V nasledujúcom texte nebudeme vždy túto množinu v znení viet uvádzať, z kontextu je totiž zrejmé, že táto množina z definície maximálnej entropie je jednoznačne určená oborom hodnôt náhodnej veličiny, pri ktorej marginálach, resp. momentoch alebo entropiách maximalizujeme.

Poznámka 2.26. Skôr, než vyslovíme ďalšiu poznámku týkajúcu sa vety, objasnime, ako získame maximálne entropické rozdelenie pri znalosti prvých marginálnych rozdelení, samozrejme pri predpoklade, že takéto rozdelenie existuje, teda suprémum entropie sa dosahuje. Pre jednoduchosť majme $\{\boldsymbol{X}\} = \{(X_1, X_2)\}$ s konečným oborom hodnôt, maximalizujme entropiu pri daných marginálnych rozdeleniach prvého rádu, označme ich, v zhode so značením použitým v definícii 2.5, ako $\tilde{p}(x_1)$, $\tilde{p}(x_2)$. Potom maximálne entropické rozdelenie získame ako $p(x_1, x_2) = \tilde{p}(x_1)\tilde{p}(x_2)$. Intuitívne je to tak preto, že takto vytvorené rozdelenie je rozdelením veličiny pri predpoklade nezávislosti jej zložiek. Takouto voľbou teda do systému nevnášame žiadnu novú informáciu, entropia je maximálna. Rovnosť tiež vyplýva, nie len intuitívne, ale aj matematicky korektne, z vety 1.31. Na danom rozdelení sa totiž zjavne dosahuje horná hranica entropie, teda musí ísť o maximum. Zovšeobecnenie pre viaczložkovú veličinu je priamočiare. Vo všeobecnosti budeme maximálne entropické rozdelenie konzistentné s danými marginálami popisovať v poznámke 3.21.

Poznámka 2.27. Preskúmajme, či je neostrá nerovnosť najlepšou možnou formuláciou vety 2.24. To, že nie vždy nastáva rovnosť, je jasné už z troch uvedených pomocných tvrdení, konkrétne v leme 2.21 vidno, že inklúzie vo všeobecnosti nemôžeme otáčať. Zaujíma nás ešte, či vôbec niekedy môže rovnosť nastať, teda či by sme nemohli namiesto neostrej nerovnosti vo vete použiť ostrú. Preskúmajme nasledujúci príklad. Majme $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ diskrétnu vektorovú náhodnú veličinu s konečným oborom hodnôt a predpokladajme nezávislosť všetkých $\tilde{X}_i$. Vieme určiť marginálne rozdelenia tejto veličiny a použiť ich ako väzbové podmienky, pri ktorých budeme hľadať maximálnu entropiu. Potom vieme, že pri znalosti prvých marginálnych rozdelení určíme maximálnu entropiu ako entropiu rozdelenia $p(x_1, \ldots, x_n) = \tilde{p}(x_1) \cdot \ldots \cdot \tilde{p}(x_n)$. Pri znalosti n-tých marginál však zadáme ako väzbovú podmienku opäť $p(x_1, \ldots, x_n) = \tilde{p}(x_1, \ldots, x_n) = \tilde{p}(x_1) \cdot \ldots \cdot \tilde{p}(x_n)$, pretože zložky $\tilde{\boldsymbol{X}}$ sú nezávislé. Z uvedeného vyplýva, že nastáva rovnosť $H(\tilde{P}^{(1)}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{(n)}(\tilde{\boldsymbol{X}}))$, z monotónie potom dostávame rovnosť maximálnej entropie všetkých rádov. Najlepšou možnou formuláciou vety je teda naozaj neostrá nerovnosť.

Uveďme ešte jedno pozorovanie vyplývajúce z uvedeného príkladu. Zistili sme totiž, že pokiaľ maximalizujeme pri marginálnych rozdeleniach príslušných vektorovej náhodnej veličine s nezávislými zložkami, tak je postupnosť maximálnych entropií konštantná. Platí však aj opačná implikácia, teda môžeme zhrnúť naše pozorovanie do nasledujúcej vety.

Veta 2.28 (Nutná a postačujúca podmienka nezávislosti cez $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$)**.** *Majme náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s konečným oborom hodnôt. Maximalizujme entropiu pri marginálnych rozdeleniach veličiny $\tilde{\boldsymbol{X}}$. Potom zložky $\tilde{\boldsymbol{X}}$ sú nezávislé práve vtedy, keď je postupnosť maximálnych entropií pri zadaných marginálnych rozdeleniach konštantná, teda platí*

$$H(\tilde{P}^{(1)}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{(2)}(\tilde{\boldsymbol{X}})) = \cdots = H(\tilde{P}^{(n-1)}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{(n)}(\tilde{\boldsymbol{X}})).$$

Dôkaz. To, že z nezávislosti vyplýva konštantnosť postupnosti maximálnych entropií, sme ukázali v poznámke 2.27. Odôvodníme teraz opačnú implikáciu, budeme postupovať sporom. Nech je postupnosť maximálnych entropií pri zadaných marginálnych rozdeleniach konštantná a nech $\tilde{\boldsymbol{X}}$ nemá nezávislé zložky. Pracujeme s náhodnými veličinami s konečným oborom hodnôt, teda suprémum entropie sa dosahuje na niektorom z rozdelení. Platia rovnosti $H(\tilde{P}^{(1)}(\tilde{\boldsymbol{X}})) = H(\tilde{p}(x_1) \cdot \ldots \cdot \tilde{p}(x_n)) = \sum_{i=1}^{n} H(\tilde{p}(x_i))$, $H(\tilde{P}^{(n)}(\tilde{\boldsymbol{X}})) = H(\tilde{p}(\boldsymbol{x}))$. Keďže ale zložky $\tilde{\boldsymbol{X}}$ nie sú nezávislé, z vety 1.31 dostávame ostrú nerovnosť $H(\tilde{p}(\boldsymbol{x})) < \sum_{i=1}^{n} H(\tilde{p}(x_i))$, z toho už $H(\tilde{P}^{(n)}(\tilde{\boldsymbol{X}})) < H(\tilde{P}^{(1)}(\tilde{\boldsymbol{X}}))$, čo je spor s konštantnosťou postupnosti. □

Veta 2.29 (Monotónia $H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$). *Majme $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ a množinu $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$ s rovnakým oborom hodnôt. Maximalizujme entropiu pri momentoch veličiny $\tilde{\boldsymbol{X}}$, predpokladajme ich konečnosť. Potom pre všetky $k \in \widehat{n-1}$ platí*

$$H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}})) \geq H(\tilde{P}^{<k+1>}(\tilde{\boldsymbol{X}})). \tag{2.8}$$

Dôkaz. Postupnosť zadaných podmienok je usporiadaná inklúziou, spĺňa podmienky z lemy 2.20, určuje teda postupnosť rozdelení, ktorá je taktiež usporiadaná inklúziou a klesá. Ak na tejto množine spočítame suprémum hodnôt entropie, z lemy 2.19 dostávame klesajúcu postupnosť maximálnych entropií pri zadaných marginálach. □

Poznámka 2.30. Vo všeobecnosti nastáva nerovnosť, podmienka na momenty rádu $k+1$ môže zvýšiť našu znalosť o systéme, znížiť maximálnu možnú entropiu, ale existujú prípady, v ktorých nastáva rovnosť.

Príklad 2.31. Majme $\{\boldsymbol{X}\} = \{(X_1, X_2)\}$, kde X_i môže dosahovať hodnotu 0 alebo 1. Maximalizujme hodnotu pri zadaných prvých momentoch $\mathrm{E}\, X_1 = 0.5$, $\mathrm{E}\, X_2 = 0.5$. Potom maximálna možná entropia pri takto zadaných momentoch prvého rádu je entropia rovnomerného rozdelenia, teda 2 bity. Zadajme druhé momenty $\mathrm{E}\, X_1^2 = \mathrm{E}\, X_2^2 = 0.5$, $\mathrm{E}\, X_1 X_2 = 0.25$. Tieto momenty sú momentmi rovnomerného rozdelenia, teda aj maximálna entropia bude aj v tomto prípade rovná entropii rovnomerného rozdelenia, nastala rovnosť entropii prvého a druhého rádu. Vo vete tak opäť ponechávame neostrú nerovnosť.

Pre entropiu maximalizovanú pri zadaných marginálnych rozdeleniach sme okrem vety o monotónii vyslovili aj nutnú a postačujúcu podmienku nezávislosti veličiny, podľa ktorej marginál maximalizujeme. Chceli by sme vysloviť podobnú vetu aj pre entropiu maximalizovanú pri daných momentoch. Pomocou momentov do určitého rádu však nie sme schopní zachytiť všetky závislosti medzi veličinami, napríklad vieme, že nekorelovanosť veličín neimplikuje ich nezávislosť. Podobná veta teda nemusí platiť.

Príklad 2.32. Uveďme príklad, ktorý vyvráti platnosť implikácie v jednom smere. Ukážeme, že z konštantnosti postupnosti entropií maximalizovaných pri zadaných momentoch nevyplýva nezávislosť zložiek veličiny, podľa ktorej maximalizujeme. Majme dvojzložkovú diskrétnu náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \tilde{X}_2)$, kde $\operatorname{Ran} \tilde{X}_1 = \{0, 1, 2\}$, $\operatorname{Ran} \tilde{X}_1 = \{0, 1\}$, zložky majú rozdelenie pravdepodobnosti uvedené v tabuľke 2.1. Potom ich prvé a druhé momenty sú rovné momentom rovnomerného rozdelenia. Máme dve náhodné veličiny, postupnosť z vety je teda iba v tvare $H(\tilde{P}^{<1>}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{<2>}(\tilde{\boldsymbol{X}}))$. Pokiaľ zadáme ako maximalizačné momenty práve momenty rovnomerného rozdelenia, tak obe maximálne entropie budú rovné entropii rovnomerného rozdelenia, teda $\log_2 6$. Ľahko však nahliadneme, že zložky $\tilde{\boldsymbol{X}}$ nie sú nezávislé, teda neplatí, že by z konštantnosti postupnosti vyplývala nezávislosť.

$\tilde{X}_2 \backslash \tilde{X}_1$	**0**	**1**	**2**
0	1/12	1/3	1/12
1	1/4	0	1/4

Tabuľka 2.1: Združené rozdelenie pravdepodobnosti veličín z príkladu 2.32

Poznámka 2.33. V stručnosti popíšeme, ako by sme postupovali pri konštrukcii protipríkladu na implikáciu v opačnom smere, t. j. chceme ukázať, že ak aj maximalizujeme pri momentoch nezávislej veličiny, nemusí byť postupnosť maximálnych entropií konštantná. Za veličinu, pri ktorej momentoch maximalizujeme, budeme voliť takú, ktorá síce má zložky nezávislé, ale nemá rovnomerné rozdelenie. Ďalej budeme požadovať, aby prvé momenty zložiek boli momentami rovnomerného rozdelenia, vyššie momenty už nemusia byť. Potom pri prvých momentoch bude maximálna entropia entropiou rovnomerného rozdelenia, ale pri zadaní niektorého vyššieho momentu, ktorý už nie je konzistentný s rovnomerným rozdelením maximálna entropia klesne, nebude teda platiť, že postupnosť je konštantná.

Ďalej budeme pokračovať skúmaním monotónie a podmienok nezávislosti pre entropiu maximalizovanú pri zadaných entropiách.

Veta 2.34 (Monotónia $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$). *Majme $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$, množinu $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$ s rovnakým oborom hodnôt. Maximalizujme entropiu pri entro-*

piách veličiny $\tilde{\boldsymbol{X}}$, predpokladajme ich konečnosť. Potom pre všetky $k \in \widehat{n-1}$ platí

$$H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})) \geq H(\tilde{P}^{[k+1]}(\tilde{\boldsymbol{X}})). \tag{2.9}$$

Dôkaz. Argumentovali by sme podobne, ako v dôkaze vety 2.29. □

Poznámka 2.35. Aj v tomto prípade nastáva vo všeobecnosti nerovnosť, rovnosť dostávame v niektorých situáciách, napríklad nech $\{\boldsymbol{X}\} = \{(X_1, X_2)\}$ je, podobne ako v príklade 2.31, množina binárnych veličín. Zadajme ako maximalizačné podmienky hodnoty entropii prvého rádu $H(X_1) = H(X_2) = 1$ bit. Pri týchto podmienkach je maximálna entropia 2 bity. Zadajme maximalizačnú podmienku druhého rádu ako $H(X_1, X_2) = 2$ bity, potom je jasné, že nastáva rovnosť maximálnych entropií prvého a druhého rádu.

Entropie vyšších rádov, podobne ako marginálne rozdelenia, umožňujú zachytiť akúkoľvek závislosť medzi náhodnými veličinami. Preto môžeme aj pre entropiu maximalizovanú pri zadaných k-tych entropiách vysloviť vetu o súvislosti konštantnosti postupnosti maximálnych entropií a nezávislosti veličiny, z ktorej sme k-te entropie určovali. Ešte predtým objasnime, aká je maximálna entropia pri zadaných prvých entropiách, budeme to totiž potrebovať do dôkazu vety.

Poznámka 2.36. Majme $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$ a maximalizujeme hodnotu entropie pri entropiách prvého rádu veličiny $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n)$, predpokladáme ich konečnosť. Požadujeme teda platnosť $H(X_i) = H(\tilde{X}_i)\ \forall i \in \widehat{n}$. Z vety 1.31 vieme, že pre všetky rozdelenia $p(\boldsymbol{x})$, ktoré majú zadané hodnoty prvých entropií, platí $H(p(\boldsymbol{x})) \leq \sum_{i=1}^{n} H(X_i)$, z toho dostávame priamo z definície maximálnej entropie $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) \leq \sum_{i=1}^{n} H(X_i)$. Ukážeme, že vždy medzi rozdeleniami náhodných veličín z množiny $\{\boldsymbol{X}\}$, ktoré spĺňajú podmienky na prvé entropie, nájdeme rozdelenie $q(\boldsymbol{x})$ také, na ktorom nastáva rovnosť $H(q(\boldsymbol{x})) = \sum_{i=1}^{n} H(X_i)$. Zvoľme $q(\boldsymbol{x})$ ako súčin prvých marginálnych rozdelení veličiny $\tilde{\boldsymbol{X}}$. Potom zložky popísané rozdelením $q(\boldsymbol{x})$ sú nezávislé, teda vo vete 1.31 nastáva rovnosť $H(q(\boldsymbol{x})) = \sum_{i=1}^{n} H(q(x_i))$. Z toho, ako sme volili marginály q však vieme, že $H(q(x_i)) = H(\tilde{X}_i) = H(X_i)$. Potom, pretože, že $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}}))$ je suprémom cez všetky možné rozdelenia veličín z $\{\boldsymbol{X}\}$, ktoré spĺňajú podmienky na zadané hodnoty prvých entropií, môžeme odhadovať $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) \geq H(q(\boldsymbol{x})) = \sum_{i=1}^{n} H(q(x_i)) = \sum_{i=1}^{n} H(X_i)$. Celkovo teda musí nastávať rovnosť $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) = \sum_{i=1}^{n} H(X_i) = \sum_{i=1}^{n} H(\tilde{X}_i)$.

Poznámka 2.37. V poznámke 2.36 sme podobne ako v definícii maximálnej entropie pri zadaných entropiách použili značenie $H(X_i)$ pre entropiu náhodnej veličiny z množiny $\{\boldsymbol{X}\}$. Uvedenú rovnosť $H(X_i) = H(\tilde{X}_i)\ \forall i \in \widehat{n}$ teda chápeme tak, že platí pre ľubovoľne vybranú náhodnú veličinu z množiny $\{\boldsymbol{X}\}$. Tento zápis budeme v podobných situáciách používať aj v nasledujúcom texte, predovšetkým v kapitole 3.

Veta 2.38 (Nutná a postačujúca podmienka nezávislosti cez $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$)**.** *Majme náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \dots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s konečným oborom hodnôt. Maximalizujme entropiu pri entropiách veličiny $\tilde{\boldsymbol{X}}$ do určitého rádu. Potom zložky $\tilde{\boldsymbol{X}}$ sú nezávislé práve vtedy, keď je postupnosť maximálnych entropií pri zadaných entropiách konštantná, teda platí*

$$H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{[2]}(\tilde{\boldsymbol{X}})) = \dots = H(\tilde{P}^{[n-1]}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{[n]}(\tilde{\boldsymbol{X}})).$$

Dôkaz. Vďaka monotónii maximálnej entropie stačí, keď overíme, že nezávislosť zložiek $\tilde{\boldsymbol{X}}$ nastáva práve vtedy, keď $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{[n]}(\tilde{\boldsymbol{X}})$. Uvedomíme si, že môžeme prepísať vzťah $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{[n]}(\tilde{\boldsymbol{X}})$ na rovnosť $\sum_{i=1}^{n} H(\tilde{p}(x_i)) = H(\tilde{p}(\boldsymbol{x}))$. Rovnosť $H(\tilde{p}(\boldsymbol{x})) = H(\tilde{P}^{[n]}(\tilde{\boldsymbol{X}}))$ je zrejmá. Rovnosť $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) = \sum_{i=1}^{n} H(\tilde{p}(x_i))$ sme odôvodnili v poznámke 2.36. Dokazované tvrdenie potom vyplýva z vety 1.31, ktorej tvrdením je, že rovnosť sa vo vzťahu $H(\tilde{p}(\boldsymbol{x})) = \sum_{i=1}^{n} H(\tilde{p}(x_i))$ dosahuje práve vtedy, keď má $\tilde{\boldsymbol{X}}$ nezávislé zložky. □

2.2.2 Ohraničenosť

Veta 2.39 (Dolná hranica maximálnych entropií)**.** *Maximalizujme entropiu n-zložkovej náhodnej veličiny. Potom pre všetky $k \in \widehat{n}$ platí*

$$H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}})) \geq 0 \tag{2.10}$$

$$H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})) \geq 0 \tag{2.11}$$

$$H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}})) \geq 0. \tag{2.12}$$

Dôkaz. Vzťahy vyplývajú z nezápornosti entropie ľubovoľného rozdelenia. □

Predpoklady vety chápeme tak, že nezápornosť maximálnych entropií platí pri akokoľvek zadaných väzbových podmienkach, preto bližšie nepopisujeme množiny hodnôt maximalizačných podmienok.

Veta 2.40 (Horná hranica maximálnych entropií)**.** *Nech $n \in \mathbb{N}$, $k \in \widehat{n}$. Majme náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \dots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$. Maximalizujme entropiu pri marginálach $\tilde{\boldsymbol{X}}$ rádu k,* *resp. pri entropiách veličiny $\tilde{\boldsymbol{X}}$ do rádu k alebo pri jej momentoch do rádu k. Potom platí*

$$H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})) \leq \sum_{i=1}^{n} H(\tilde{X}_i) \tag{2.13}$$

$$H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}})) \leq \sum_{i=1}^{n} H(\tilde{X}_i). \tag{2.14}$$

Ďalej platí

$$H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\| \tag{2.15}$$

$$H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}})) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\| \tag{2.16}$$

$$H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}})) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|, \tag{2.17}$$

kde $\|\tilde{\mathcal{X}}_i\|$ značí počet prvkov v obore hodnôt $\tilde{X}_i$.

Dôkaz. Vzťahy (2.15),(2.16) a (2.17) vyplývajú priamo z vety 1.32. Vzťahy (2.13) a (2.14) vyplývajú z toho, že platí $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) = \sum_{i=1}^{n} H(\tilde{X}_i)$ a $H(\tilde{P}^{(1)}(\tilde{\boldsymbol{X}})) = \sum_{i=1}^{n} H(\tilde{X}_i)$, ako sme odôvodnili v poznámkach 2.36 a 2.26, a z monotónie maximálnych entropií, ako sme ju ukázali vo vetách 2.34 a 2.24. □

Poznámka 2.41. Poznamenajme, že pre entropiu maximalizovanú pri zadaných momentoch $\tilde{\boldsymbol{X}}$ nemusí platiť, že by jej hodnota bola zhora ohraničená výrazom $\sum_{i=1}^{n} H(\tilde{X}_i)$. Protipríklad uvidíme v 2.56, tam totiž ukážeme rozdelenie, pre ktoré nastáva nerovnosť $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) < H(\tilde{P}^{<1>}(\tilde{\boldsymbol{X}}))$, čo ako vieme je ekvivalentné nerovnosti $\sum_{i=1}^{n} H(\tilde{X}_i) < H(\tilde{P}^{<1>}(\tilde{\boldsymbol{X}}))$.

Poznámka 2.42. Vety sme vyslovili s pomerne slabými predpokladmi, odhady bez ďalších požiadavok na vlastnosti veličín nevieme zlepšiť. Vieme ľahko nájsť príklady maximálne entropických rozdelení, pri ktorých nastáva rovnosť v dolnom, resp. v hornom odhade.

Príklad 2.43. Majme opäť $\{\boldsymbol{X}\} = \{(X_1, X_2)\}$ s binárnym oborom hodnôt. Rovnosť v dolnom odhade môžeme ilustrovať na voľbe entropií, marginálnych rozdelení a momentov pre rád 1. Keď entropie prvého rádu volíme nulové, aj maximálne entropia pri znalosti entropií prvého stupňa bude nulová. Podobne, keď volíme obidva prvé momenty nulové, jediné rozdelenie, ktoré je v súlade s týmito momentmi, je rozdelenie, kde $p_i(1) = 0$, pre $i \in \{1, 2\}$. Toto rozdelenie má nulovú entropiu. Rozdelenie s nulovou entropiou by sme dostali aj pri voľbe oboch prvých momentov rovných 1. Zostáva nám určiť marginálne rozdelenia, pri ktorých maximálna entropia dosahuje hodnotu 0. Ľahko vidíme, že nulovú hodnotu entropie budú mať práve tie rozdelenia, kde jedna z hodnôt nastáva s pravdepodobnosťou 1, druhá hodnota nenastáva.

Uveďme teraz príklad pre rovnosť v hornom odhade pre maximalizáciu pri zadaných entropiách a marginálach, myslíme tým rovnosť maximálnych entropií a výrazu $\sum_{i=1}^{n} H(\tilde{X}_i)$. Majme $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$, kde X_i majú konečný obor hodnôt, entropiu maximalizujeme pri marginálach a entropiách vypočítaných z $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt, kde $\tilde{X}_i$ sú navzájom nezávislé. Potom vieme, že nastáva rovnosť v hornom odhade, ako sme už videli v dôkaze nutných a postačujúcich podmienok nezávislosti. Takúto voľbu môžeme urobiť pre ľubovoľný rád k. Ak by sme chceli dosiahnuť rovnosť maximálnych entropií s výrazom $\sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|$, stačí voliť maximalizačné podmienky také, ktoré vypočítame z rovnomerného rozdelenia.

Poznámka 2.44. V tejto podkapitole sme ilustrovali rovnosti v horných a v dolných odhadoch maximálnych entropií len jedným príkladom, tiež vyslovené vety o monotónii a ohraničenosti majú síce všeobecnú platnosť, ale ich tvrdenia sú jednoduché. V kapitole 3, kde budeme skúmať interakcie medzi veličinami, nájdeme všeobecnejší popis toho, kedy sa dosahuje horná hranica entropie, uvidíme tiež, že pri určitých predpokladoch budeme vedieť nájsť ďalšie vzťahy medzi entropiami jedného druhu rôznych rádov.

2.2.3 Vzťahy medzi konštrukciami

Preskúmali sme monotóniu a ohraničenosť sekvencií maximálnych entropii každého z troch typov, tiež sme vyslovili kritériá nezávislosti. Mohli by nás zaujímať mnohé iné vlastnosti, napríklad vzťahy medzi jednotlivými druhmi maximálnych entropií. Prieskum niektorých vzťahov bude priamočiary a bude založený na znení pomocných tvrdení 2.19 a 2.20, respektíve na ich zovšeobecnení 2.23. Skôr, než tieto vzťahy vyslovíme korektne vo vetách, popíšme slovne tvrdenie a predpoklady viet.

Chceli by sme preskúmať situáciu, keď máme dáta, z ktorých sme schopní vypočítať napríklad marginálne rozdelenia aj entropie, alebo marginálne rozdelenia aj momenty určitého rádu. Vo vetách budeme dáta reprezentovať náhodnou veličinou $\tilde{\boldsymbol{X}}$. Príslušné momenty, resp. marginálne rozdelenia alebo entropie vypočítané z $\tilde{\boldsymbol{X}}$ použijeme ako väzbové podmienky, nájdeme maximálne entropie. Zaujíma nás, či platí rovnosť alebo nejaký typ nerovnosti medzi dvoma druhmi takto vypočítanej maximálnej entropie. Vety najprv uvedieme v plnom znení, presne popíšeme, ako sa volia množiny väzbových podmienok, v poznámkach potom uvedieme ich skrátené znenie bez technických detailov.

Veta 2.45 (Vzťah $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$ a $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$)**.** *Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $k \leq n$, $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$, ďalej majme $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Majme indexovú množinu $N^{(k)}$ z definície $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$, jej prvky označme $\boldsymbol{n}_i$. Označme $B = \{\tilde{p}_{\boldsymbol{n}_i}(\boldsymbol{x}) \mid \boldsymbol{n}_i \in N^{(k)}\}$. Za vektorové veličiny $\tilde{\boldsymbol{X}}_i$ v definícii* 2.5 *zvoľme vektorové náhodné veličiny určené rozdeleniami v B.*
Majme indexovú množinu $N^{[k]}$ z definície $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$, jej prvky označme $\tilde{\boldsymbol{n}}_l$. Nech existujú konečné $H_{\tilde{\boldsymbol{n}}_l} = H(\tilde{X}_{j_1}, \ldots, \tilde{X}_{j_i})$ pre všetky $(j_1, \ldots, j_i) = \tilde{\boldsymbol{n}}_l$. Za množinu hodnôt podmienok $M^{[k]}$ v definícii 2.15 *položme $M^{[k]} = \{H_{\tilde{\boldsymbol{n}}_l} \mid \tilde{\boldsymbol{n}}_l \in N^{[k]}\}$.*
Potom pre takto definované maximálne entropie platí

$$H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})) \geq H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}})). \tag{2.18}$$

Dôkaz. Máme $M^{[k]}$ množinu podmienok danú entropiami do rádu k vrátane, označme $M^{(k)}$ množinu podmienok danú marginálnymi rozdeleniami rádu k. Stačí overiť, že platí $M^{(k)} \implies M^{[k]}$, zvyšok už plynie z pomocných tvrdení 2.19 a 2.23. Naozaj, vďaka tomu, ako sme vo vete definovali pomocnú veličinu $\tilde{\boldsymbol{X}}$ platí, že entropie, ktoré by sme vypočítali z marginálnych rozdelení v $M^{(k)}$ sú rovnaké, ako tie, pri ktorých maximalizujeme, teda platí $M^{(k)} \implies M^{[k]}$. Opačná implikácia nemusí platiť, znalosť entropií nie je vždy postačujúca na jednoznačné určenie marginálneho rozdelenia daného rádu. □

Poznámka 2.46. Uveďme niekoľko pozorovaní týkajúcich sa vety:

- Ako sme už upozorňovali v komentári za definíciou 1.14, musíme predpokladať konečnosť entropií do rádu k vrátane, pretože entropia je len špeciálnym typom strednej hodnoty a vo všeobecnosti nemáme zaručenú jej konečnosť pre diskrétne náhodné veličiny so spočítateľným oborom hodnôt. Podobne vo vete 2.48 budeme musieť predpokladať existenciu konečných vyšších momentov.

- V dôkaze sme uviedli, že nemusí nastávať ekvivalencia množín podmienok. Uveďme ešte príklad, kedy nastáva rovnosť maximálnych entropií, aby bolo jasné, že v tvrdení vety by nestačila ostrá nerovnosť. Majme $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \tilde{X}_2)$ s nezávislými zložkami, teda rozdelenie druhého stupňa bude v tvare $\tilde{p}(x_1, x_2) = \tilde{p}(x_1)\tilde{p}(x_2)$. Potom maximálna entropia pri danom takomto rozdelení je rovnaká, ako maximálna entropia pri zadaných entropiách, a to pre stupeň 1 aj stupeň 2. Okrem toho rovnosť pre $k = 1$ nastáva pre ľubovoľné rozdelenie veličiny $\tilde{\boldsymbol{X}}$. Ďalší prípad, v ktorom nastáva rovnosť uvidíme v kapitole 3, kde ukážeme, že pre systémy, v ktorých sa prejavujú len určité typy interakcií, bude platiť rovnosť týchto dvoch typov maximálnej entropie.

- To, že v dôkaze nemusí nastávať ekvivalencia podmienok ešte nemusí znamenať, že v skutočnosti neplatí vo vete rovnosť. Na to, aby sme ukázali, že neostrá nerovnosť je najlepšou možnou formuláciou vety by sme potrebovali nájsť príklad, v ktorom nastane ostrá nerovnosť medzi danými dvoma druhmi entropie. Otázku existencie takéhoto príkladu ponechávame otvorenú.

- Poznamenajme, že znalosť entropií nie je dostatočná na určenie marginálneho rozdelenia ani v zmysle takom, že by sme dokázali určiť pravdepodobnosti jednotlivých hodnôt a len ich nevedeli priradiť ku správnym hodnotám z oboru hodnôt. Ilustrujme na príklade, čo presne myslíme. Majme danú entropiu prvého rádu pre jednozložkovú diskrétnu náhodnú veličinu, ktorej obor hodnôt má tri prvky, hľadáme jej hustotu pravdepodobnosti, t.j. p_1, p_2, p_3. Zvoľme hodnotu p_2 fixne z intervalu $[0, 1]$, ďalej nech bez ujmy na všeobecnosti $p_1 \leq p_3$. Potom môžeme voliť p_1 a p_3 tak, aby bola dosiahnutá daná hodnota entropie a aby $\boldsymbol{p}$ bola hustota pravdepodobnosti, táto voľba je pre pevné p_2 práve jedna. Z uvedeného vyplýva, že existuje rozdelenie pravdepodobnosti pre ľubovoľne volené p_2, a teda hodnoty rozdelenia zložiek pravdepodobnosti nie sú dané jednoznačne.

Uveďme ešte skrátené znenie vety bez presného popisu indexovania množín.

Poznámka 2.47 (Vzťah $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$ a $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$, slovne)**.** Majme $k \in \mathbb{N}$, nech $n \in \mathbb{N}$ také, že $k \leq n$, $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$, ďalej majme $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Maximalizujme entropiu pri marginálnych rozdeleniach veličiny $\tilde{\boldsymbol{X}}$ a pri jej entropiách, predpokladajme ich konečnosť. Potom platí

$$H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})) \geq H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}})). \tag{2.19}$$

Preskúmajme teraz vzťah medzi entropiami maximalizovanými pri zadaných momentoch a marginálach. Opäť uvedieme najprv vetu s presnými detailami a potom v poznámke jej skrátené slovné znenie.

Veta 2.48 (Vzťah $H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$ a $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}})))$**.** *Majme $k \in \mathbb{N}$, nech $n \in \mathbb{N}$ také, že $k \leq n$, $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$, ďalej majme $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Majme indexovú množinu $N^{(k)}$ z definície $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$, jej prvky označme $\boldsymbol{n}_i$. Označme $B = \{\tilde{p}_{\boldsymbol{n}_i}(\boldsymbol{x}) \mid \boldsymbol{n}_i \in N^{(k)}\}$. Za vektorové veličiny $\tilde{\boldsymbol{X}}_i$ v definícii 2.5 zvoľme vektorové náhodné veličiny určené rozdeleniami v B.*
Majme indexovú množinu $N^{<k>}$ z definície $H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$, jej prvky označme $\tilde{\boldsymbol{n}}_l$. Nech existujú konečné $\mathrm{E}_{\tilde{\boldsymbol{n}}_l} = \mathrm{E}(\tilde{X}_{j_1}, \ldots, \tilde{X}_{j_i})$ pre všetky $(j_1, \ldots, j_i) = \tilde{\boldsymbol{n}}_l$. Za množinu hodnôt podmienok $M^{<k>}$ v definícii 2.11 položme $M^{<k>} = \{\mathrm{E}_{\tilde{\boldsymbol{n}}_l} \mid \tilde{\boldsymbol{n}}_l \in N\}$.
Potom pre takto definované maximálne entropie platí

$$H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}})) \geq H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}})). \tag{2.20}$$

Dôkaz. Máme $M^{<k>}$ množinu podmienok danú momentmi do rádu k vrátane, označme $M^{(k)}$ množinu podmienok danú marginálnymi rozdeleniami rádu k. Podobne ako v predchádzajúcom dôkaze stačí overiť, že platí $M^{(k)} \implies M^{<k>}$. Je zrejmé, že platí, že momenty, ktoré by sme vypočítali z marginálnych rozdelení v $M^{(k)}$ sú rovnaké, ako tie, pri ktorých maximalizujeme, teda $M^{(k)} \implies M^{<k>}$. Opačná implikácia nemusí platiť, znalosť momentov nie je vždy postačujúca na jednoznačné určenie marginálneho rozdelenia. $\square$

Poznámka 2.49 (Vzťah $H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$ a $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$, slovne)**.** Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $k \leq n$, $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$, ďalej majme $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Maximalizujme entropiu pri marginálnych rozdeleniach $\tilde{\boldsymbol{X}}$ a pri jej momentoch, predpokladajme ich konečnosť. Potom platí

$$H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}})) \geq H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}})). \tag{2.21}$$

Poznámka 2.50. Podobne ako vo vete 2.45, vo všeobecnosti nastáva nerovnosť, podmienky na maximalizáciu v dvoch množinách si nie sú ekvivalentné. Zároveň vieme uviesť príklad, kedy nastáva rovnosť.

Príklad 2.51. Majme binárnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \tilde{X}_2)$. Potom znalosť prvých momentov nám jednoznačne určuje marginálne rozdelenia 1. stupňa, pretože $\mathrm{E}(\tilde{X}_1) = \tilde{p}_1(1)$, $\mathrm{E}(\tilde{X}_2) = \tilde{p}_2(1)$, kde $\tilde{p}_i$ sme označili rozdelenie $\tilde{X}_i$. Podobne zo znalosti prvých a druhých momentov jednoznačne určíme celé rozdelenie $\tilde{p}(\boldsymbol{x})$. Z uvedeného vyplýva, že množiny podmienok, pri ktorých maximalizujeme sú si v tomto prípade ekvivalentné a nastáva rovnosť maximálnych entropií pre rád 1 aj 2.

Okrem toho, že dokážeme nájsť príklad rovnosti, dokážeme nájsť aj príklad, v ktorom nastáva ostrá nerovnosť medzi danými dvoma druhmi entropie. Uveďme najjednoduchší možný takýto príklad a tým je maximalizácia entropie jednozložkovej veličiny.

Príklad 2.52. Majme $\{X\}$ s oborom hodnôt $\{0, 1, 3\}$. Vezmime za $\tilde{X}$ z vety náhodnú veličinu s rovnakým oborom hodnôt a s rozdelením $\tilde{p}(0) = 0$, $\tilde{p}(1) = 1$, $\tilde{p}(3) = 0$. Potom jej prvý moment je $\mathrm{E}\,\tilde{X} = 1$, jej entropia je $H(\tilde{X}) = 0$. Pri týchto hodnotách budeme maximalizovať. Najprv si uvedomíme, že takýto moment má aj rozdelenie $\tilde{p}(0) = 0.5$, $\tilde{p}(1) = 0.25$, $\tilde{p}(3) = 0.25$. Platí teda $H(\tilde{P}^{<1>}(\tilde{X})) \geq 1.5$ bitu, zároveň platí $H(\tilde{P}^{[1]}(\tilde{X})) = H(\tilde{X}) = 0$. Nastáva teda ostrá nerovnosť $H(\tilde{P}^{<1>}(\tilde{X})) > H(\tilde{P}^{(1)}(\tilde{X}))$. Vlastnosti náhodnej veličiny s oborom hodnôt $\{0, 1, 3\}$ sa ľahko skúmajú, preto sa vyskytne ešte aj v príklade 3.55.

Príklad 2.53. Doplňme pozorovanie o binárnych veličinách z príkladu 2.51. Majme množinu $\{\boldsymbol{X}\} = \{(X_1, X_2)\}$, kde obe X_i môžu dosahovať hodnoty 0 alebo 1. Potom platí nezávisle na rozdelení, z ktorého počítame väzbové podmienky rovnosť $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ pre $k \in \{1, 2\}$. Vyplýva to z toho, že zo znalosti momentov príslušného rádu vieme jednoznačne určiť marginálne rozdelenie toho rádu, podobne ho vieme určiť zo znalosti prvých entropií, teda zadané podmienky, pri ktorých maximalizujeme, sú ekvivalentné. Upozornime, že v prípade zadaných entropií je rozdelenie určené až na symetrie - vieme určiť jednotlivé hodnoty pravdepodobnosti, ale máme dve možnosti, ako voliť, ktorá hodnota sa s danou pravdepodobnosťou dosahuje. Pozorovanie o rovnosti entropií všetkých druhov prvého a druhého rádu by sme mohli priamočiaro zovšeobecniť na viacrozmernú vektorovú binárnu veličinu.

Poznámka 2.54. V texte uvádzame viacero príkladov, kde sme schopní explicitne spočítať hodnoty maximálnych entropií. Ľahko sa dajú nájsť napríklad pre binárne veličiny alebo pre jednozložkovú diskrétnu vektorovú náhodnú veličinu. Upozorňujeme však, že úloha na hľadanie maxima entropie vo všeobecnosti nemusí byť ľahko riešiteľná, a to ani pre veličiny s konečným oborom hodnôt. Tvar maximálne entropického rozdelenia pri zadaných momentoch pre konkrétny prípad troch náhodných veličín uvedieme v príklade 2.56, tvar rozdelenia konzistentného s danými marginálnymi rozdeleniami alebo entropiami budeme bližšie skúmať v kapitole 3.

Poznámka 2.55. Poznamenajme, že vzťah medzi poslednými dvoma typmi maximálnej entropie, teda medzi entropiou maximalizovanou pri daných entropiách a pri určených momentoch, nemôžeme odvodiť a ukázať cez lemu 2.23, pretože znalosť momentov neimplikuje znalosť entropií a ani naopak, entropie jednoznačne neurčujú momenty. Na záver sekcie ešte preskúmame príklad, z ktorého bude jasné, že nerovnosť $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})) \geq H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$ určite nie je vždy splnená. Otázku, či vždy platí opačná nerovnosť, ponechávame otvorenú.

Príklad 2.56 (Maximalizácia pri zadaných momentoch)**.** Na základe príkladu 2.53 môžme usúdiť, že ak chceme, aby sa maximálne entropie rôznych druhov nerovnali, potrebujeme pracovať s veličinami, ktoré nie sú binárne. Majme preto $\{X\}$ a $\{Y\}$, nech obor hodnôt $\{X\}$ je $\{0, 1, 2\}$, obor hodnôt $\{Y\}$ je $\{0, 1\}$. Majme zadané hodnoty prvých momentov $\mathrm{E}\, X = A$, $\mathrm{E}\, Y = B$. Potom metódou Lagrangeových multiplikátorov môžeme spočítať hodnotu maximálnej entropie a tiež určiť, na akom rozdelení sa dosahuje. Upozornime na technický detail, ktorý vyplýva z podmienok na diferencovateľnosť funkcie, ktorú maximalizujeme. Budeme požadovať, že hodnoty $p_{X,Y}(x, y)$ sú nenulové pre všetky x, y. V praxi nám to však neprekáža, pretože maximum entropie by sa dosahovalo na rozdelení, ktoré má nejaké $p_{X,Y}(x, y) = 0$ iba vtedy, ak by to bolo jediné možné rozdelenie s danými momentami. Majme teda dané $\mathrm{E}\, X$ a $\mathrm{E}\, Y$ také, že existujú rozdelenia, ktoré majú zadané momenty a $p_{X,Y}(x, y)$ je nenulové pre všetky x, y, potom môžeme hodnoty multiplikátorov môžeme získať vyjadrením zo vzťahov

$$\exp\gamma = \frac{(1-B)\exp(1)}{1+\exp\alpha+\exp 2\alpha} \qquad \exp\beta = \frac{B}{1-B} \qquad A = \frac{\exp\alpha + 2\exp 2\alpha}{1+\exp\alpha+\exp 2\alpha}.$$

Rozdelenie na ktorom sa maximum dosahuje potom nájdeme dosadením do rovností

$$\begin{aligned}
p_{X,Y}(0,0) &= \exp(\gamma - 1) & p_{X,Y}(0,1) &= \exp(\gamma + \beta - 1) \\
p_{X,Y}(1,0) &= \exp(\alpha + \gamma - 1) & p_{X,Y}(1,1) &= \exp(\alpha + \beta + \gamma - 1) \\
p_{X,Y}(2,0) &= \exp(\gamma + 2\alpha - 1) & p_{X,Y}(2,1) &= \exp(2\alpha + \beta + \gamma - 1).
\end{aligned}$$

Chceme nájsť príklad na nerovnosť medzi $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}}))$ a $H(\tilde{P}^{<1>}(\tilde{\boldsymbol{X}}))$. Maximálne entropické rozdelenie pri zadaných prvých momentoch sme našli vyššie. Zostáva nám nájsť maximálnu entropiu pri zadaných entropiách. Vieme, že pre prvý rád zadaných entropií je táto úloha ľahko riešiteľná, maximálna entropia konzistentná s danými entropiami prvého rádu je ich súčet. Môžeme teda vypočítať oba druhy entropie pre prvý rád zadaných podmienok. Zadajme ako väzbové podmienky prvé entropie a prvé momenty spočítané z rozdelenia

$$\begin{aligned} &p_{\tilde{X},\tilde{Y}}(0,0) = 0,1 \qquad p_{\tilde{X},\tilde{Y}}(1,0) = 0,1 \qquad p_{\tilde{X},\tilde{Y}}(2,0) = 0,1 \\ &p_{\tilde{X},\tilde{Y}}(0,1) = 0,1 \qquad p_{\tilde{X},\tilde{Y}}(1,1) = 0,1 \qquad p_{\tilde{X},\tilde{Y}}(2,1) = 0,5. \end{aligned}$$

Potom hodnoty maximálnych entropií sú po zaokrúhlení $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) = 2,252$ a $H(\tilde{P}^{<1>}(\tilde{\boldsymbol{X}})) = 2,286$, nastáva nerovnosť $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) < H(\tilde{P}^{<1>}(\tilde{\boldsymbol{X}}))$. Ak by sme ako rozdelenie, z ktorého počítame väzbové podmienky zadali rovnomerné rozdelenie, nastala by rovnosť. Čitateľovi určite napadne, prečo na tomto príklade neilustrujeme vhodnou voľbou aj opačnú nerovnosť. Náhodne sme generovali 400 000 rozdelení pravdepodobnosti, spočítali ich entropie a momenty prvého stupňa a zadali ich ako maximalizačné podmienky. Ani v jednom z týchto prípadov nenastala opačná nerovnosť. Rozdelenia sme generovali pomocou C++ funkcií *srand* a *rand*, počiatočnú podmienku, takzvaný *seed*, sme znáhodnili pomocou funkcie *time*. Generovali sme každú zložku rozdelenia osobitne a následne normovali súčet na 1. Podobný postup sme realizovali aj pre obory hodnôt {0,1,2}, taktiež sa nám nepodarilo nájsť príklad na opačnú nerovnosť. Ako sme už spomínali, otázku, či táto nerovnosť môže alebo nemôže nastať, ponechávame otvorenú.

Týmto uzatvárame prehľad základných vlastností maximálnych entropií, v nasledujúcej podkapitole budeme skúmať spojené informácie a tiež sa dostaneme k časti zaujímavej aj z pohľadu prípadných aplikácii, konkrétne k skúmaniu vlastností nami definovaných funkcionálov na systéme, v ktorom sa prejavujú interakcie medzi jeho zložkami.

2.3 Vety o spojenej informácii

V ďalšom postupe sa zameriame na vlastnosti spojenej informácie, zväčša priamo vyplývajúce z vlastností sekvencie maximálnych entropií daného typu, ktoré sme preskúmali v predchádzajúcej časti.

2.3.1 Ohraničenosť

Veta 2.57 (Ohraničenosť $I_c^{[k]}(\tilde{\boldsymbol{X}})$). *Nech $n \in \mathbb{N}$, $k \in \widehat{n}$. Majme $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \dots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s konečným oborom hodnôt. Maximalizujme entropiu pri entropiách $\tilde{\boldsymbol{X}}$ do rádu k.*

Potom platí

$$I_c^{[k]}(\tilde{\boldsymbol{X}}) \geq 0 \tag{2.22}$$

$$I_c^{[k]}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} H(\tilde{X}_i) \tag{2.23}$$

$$I_c^{[k]}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|, \tag{2.24}$$

kde $\|\tilde{\mathcal{X}}_i\|$ *značí počet prvkov v obore hodnôt* $\tilde{X}_i$*.*

Dôkaz. Vzťah (2.22) platí vďaka monotónii $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$. Nerovnosti (2.23) a (2.24) získame z odhadu $I_c^{[k]}(\tilde{\boldsymbol{X}}) = H(\tilde{P}^{[k-1]}(\tilde{\boldsymbol{X}})) - H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})) \leq H(\tilde{P}^{[k-1]}(\tilde{\boldsymbol{X}})) \leq H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) = \sum_{i=1}^{n} H(\tilde{X}_i) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|$, kde sme využili vzťahy pre horné a dolné odhady maximálnej entropie pri daných entropiách a vetu 1.32. □

Podobné vety môžeme vysloviť a ukázať aj pre ostatné typy spojenej informácie.

Veta 2.58 (Ohraničenosť $I_c^{<k>}(\tilde{\boldsymbol{X}})$)**.** *Majme* $n \in \mathbb{N}$, $k \in \widehat{n}$ *a* $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ *s konečným oborom hodnôt. Maximalizujme entropiu pri momentoch* $\tilde{\boldsymbol{X}}$ *do rádu* k*. Potom platí*

$$I_c^{<k>}(\tilde{\boldsymbol{X}}) \geq 0 \tag{2.25}$$

$$I_c^{<k>}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|, \tag{2.26}$$

kde $\|\tilde{\mathcal{X}}_i\|$ *značí počet prvkov v obore hodnôt* $\tilde{X}_i$*.*

Dôkaz. Dolný odhad (2.25) vyplýva z monotónie $H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$. Pri maximalizácii pri zadaných momentoch nemusí platiť $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) = \sum_{i=1}^{n} H(\tilde{X}_i)$, preto v znení vety uvádzame len jeden horný odhad (2.26). Jeho platnosť vyplýva z nasledujúceho odhadu $I_c^{<k>}(\tilde{\boldsymbol{X}}) = H(\tilde{P}^{<k-1>}(\tilde{\boldsymbol{X}})) - H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}})) \leq H(\tilde{P}^{<k-1>}(\tilde{\boldsymbol{X}})) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|$, kde sme využili horné a dolné odhady pre maximálnu entropiu. □

Veta 2.59 (Ohraničenosť $I_c^{(k)}(\tilde{\boldsymbol{X}})$)**.** *Nech* $n \in \mathbb{N}$, $k \in \widehat{n}$*. Majme* $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ *s konečným oborom hodnôt. Maximalizujme entropiu pri marginálach* $\tilde{\boldsymbol{X}}$ *rádu* k*.*

Potom platí

$$I_c^{(k)}(\tilde{\boldsymbol{X}}) \geq 0 \tag{2.27}$$

$$I_c^{(k)}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} H(\tilde{X}_i) \tag{2.28}$$

$$I_c^{(k)}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|, \tag{2.29}$$

kde $\|\tilde{\mathcal{X}}_i\|$ *značí počet prvkov v obore hodnôt* $\tilde{X}_i$.

Dôkaz. Ukázali by sme podobne, ako vetu 2.57. □

Vďaka tomu, že spojená informácia je definovaná ako rozdiel maximálnych entropií dvoch po sebe idúcich rádov, okamžitým dôsledkom viet o ohraničenosti maximálnej entropie je aj ohraničenosť sumy spojených informácií.

Veta 2.60 (Ohraničenosť $\sum I_c$). *Nech* $n \in \mathbb{N}$, $k \in \widehat{n}$. *Majme* $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \dots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ *s konečným oborom hodnôt. Potom platí*

$$\sum_{k=2}^{n} I_c^{[k]}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} H(\tilde{X}_i) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\| \tag{2.30}$$

$$\sum_{k=2}^{n} I_c^{<k>}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\| \tag{2.31}$$

$$\sum_{k=2}^{n} I_c^{(k)}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} H(\tilde{X}_i) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|. \tag{2.32}$$

Dôkaz. Ukážeme platnosť odhadu (2.30), odhady (2.31), (2.32) by sa ukázali analogicky. Môžeme odhadovať nasledovne $\sum_{k=2}^{n} I_c^{[k]}(\tilde{\boldsymbol{X}}) = \sum_{k=2}^{n} \Big(H(\tilde{P}^{[k-1]}(\tilde{\boldsymbol{X}})) - H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})) \Big) = \sum_{k=1}^{n-1} H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})) - \sum_{k=2}^{n} H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) - H(\tilde{P}^{[n]}(\tilde{\boldsymbol{X}})) \leq H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) \leq \sum_{i=1}^{n} H(\tilde{X}_i) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|$, kde sme využili odhady z viet 2.39 a 2.40. □

2.3.2 Podmienky nezávislosti

Uveďme teraz dva dôsledky nutných a postačujúcich podmienok nezávislosti formulovaných vo vetách 2.28 a 2.38, ich platnosť vyplýva priamo z uvedených tvrdení.

Veta 2.61 (Nutná a postačujúca podmienka nezávislosti cez $I_c^{(k)}(\tilde{\boldsymbol{X}})$). *Majme náhodnú veličinu* $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \dots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ *s konečným oborom hodnôt. Maximalizujme entropiu* *pri marginálnych rozdeleniach veličiny $\tilde{\boldsymbol{X}}$. Potom zložky $\tilde{\boldsymbol{X}}$ sú nezávislé práve vtedy, keď je postupnosť spojených informácii pri zadaných marginálnych rozdeleniach nulová, teda platí*

$$I_c^{(k)}(\tilde{\boldsymbol{X}}) = 0 \qquad \forall k \in \{2, ..., n\}.$$

Dôkaz. Vyplýva z vety 2.28. □

Veta 2.62 (Nutná a postačujúca podmienka nezávislosti cez $I_c^{[k]}(\tilde{\boldsymbol{X}})$)**.** *Majme náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s konečným oborom hodnôt. Maximalizujme entropiu pri entropiách veličiny $\tilde{\boldsymbol{X}}$ do určitého rádu. Potom zložky $\tilde{\boldsymbol{X}}$ sú nezávislé práve vtedy, keď je postupnosť spojených informácii pri zadaných entropiách nulová, teda platí*

$$I_c^{[k]}(\tilde{\boldsymbol{X}}) = 0 \qquad \forall k \in \{2, ..., n\}.$$

Dôkaz. Vyplýva z vety 2.38. □

Preskúmali sme niekoľko základných vlastností spojenej informácie, ďalej budeme skúmať jej možnú interpretáciu a jej vlastnosti v závislosti na interakciách prítomných medzi jednotlivými zložkami skúmaného systému.

Kapitola 3

Interakcie

Preskúmali sme niektoré jednoduché vlastností spojenej informácie. Mohlo by nás ďalej zaujímať, ako sa dá interpretovať jej význam v rámci popisu komplexného systému. Jednou z možných interpretácii je, že hodnota spojenej informácie odráža množstvo informácie, o ktoré sa naša znalosť o systéme líši, ak máme možnosť skúmať zvolené parametre dvoch po sebe idúcich rádov, t. j. marginálne rozdelenia rádu k a rádu $k+1$, entropie do rádu k a do rádu $k+1$ alebo momenty týchto rádov. V tejto kapitole podrobnejšie preskúmame, či je táto interpretácia vhodná.

V prípade $I_c^{[k]}(\tilde{\boldsymbol{X}})$ a $I_c^{(k)}(\tilde{\boldsymbol{X}})$ sme zistili, že zložky skúmanej vektorovej náhodnej veličiny sú závislé práve vtedy, keď postupnosť týchto dvoch typov spojenej informácie nie je konštantne nulová (pre $I_c^{<k>}(\tilde{\boldsymbol{X}})$ toto tvrdenie neplatí). Týmto sa otvára otázka, či v prípade závislosti zložiek sú niektoré členy postupnosti spojených informácii aj naďalej nulové a ktoré členy naopak určite nulové nie sú. Uvidíme, že túto otázku budeme vedieť sčasti zodpovedať, budeme však na to najprv potrebovať definovať pojem interakcia k-teho rádu a preskúmať maximálne entropické rozdelenia pri určitých väzbových podmienkach. Skôr, než k tomu pristúpime, uveďme dva príklady na veličiny, u ktorých sa prejavuje závislosť vyšších rádov, zatiaľ budeme pojem interakcia využívať voľne, v jeho intuitívnom význame.

Príklad 3.1 (XOR)**.** Príklad, v ktorom máme zadané dve náhodné binárne veličiny a definujeme tretiu náhodnú veličinu ako ich XOR, býva používaný na ilustráciu vzťahov medzi časťami systému v rôznych kontextoch. Podobne aj v našom prípade ho autori diskutovali v článku [1]. Majme vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \tilde{X}_2, \tilde{X}_3)$. Nech $\tilde{X}_1$ a $\tilde{X}_2$ sú nezávislé binárne náhodné veličiny s uniformným rozdelením pravdepodobnosti, ďalej nech $\tilde{X}_3 = \text{XOR}\,(\tilde{X}_1, \tilde{X}_2)$. Rozdelenie pravdepodobnosti týchto veličín je uvedené v tabuľke 3.1.

$\tilde{x}_1, \tilde{x}_2, \tilde{x}_3$	$p(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$	$\tilde{x}_1, \tilde{x}_2, \tilde{x}_3$	$p(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$
0,0,0	1/4	0,1,1	1/4
1,0,0	0	1,0,1	1/4
0,1,0	0	1,1,0	1/4
0,0,1	0	1,1,1	0

Tabuľka 3.1: Rozdelenie pravdepodobnosti veličín z príkladu 3.1, $\tilde{X}_3 = \text{XOR}(\tilde{X}_1, \tilde{X}_2)$

Rozdelenie troch veličín vo vzťahu XOR uvedené v tabuľke 3.1 má nasledovné hodnoty entropie a vzájomnej informácie

$$\begin{aligned} H(\tilde{X}_1) = H(\tilde{X}_2) = H(\tilde{X}_3) = 1, \\ H(\tilde{X}_1, \tilde{X}_2) = H(\tilde{X}_1, \tilde{X}_3) = H(\tilde{X}_2, \tilde{X}_3) = 2, \\ I(\tilde{X}_1; \tilde{X}_2) = I(\tilde{X}_1; \tilde{X}_3) = I(\tilde{X}_2; \tilde{X}_3) = 0, \\ H(\tilde{X}_1, \tilde{X}_2, \tilde{X}_3) = 2. \end{aligned}$$

Vypočítajme všetky typy spojenej informácie druhého a tretieho rádu. Pripomeňme, že v definícii entropie počítame s logaritmom so základom 2, výsledky tak budú v bitoch. $I_c^{[2]}(\tilde{\boldsymbol{X}}) = 0$, pretože $H(\tilde{P}^{[2]}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) = \sum_{i=1}^{3} H(\tilde{X}_i) = 3$. Ak totiž zvolíme za entropie, pri ktorých hľadáme maximálne entropické rozdelenie, entropie 1. a 2. stupňa vypočítané z rozdelenia pre XOR, tak maximálne entropické rozdelenie, ktoré má takéto hodnoty entropií 1. a 2. stupňa, bude uniformné. Toto rozdelenie má entropiu 3 bity. Ďalej $I_c^{[3]}(\tilde{\boldsymbol{X}}) = 1$, lebo $H(\tilde{P}^{[3]}(\tilde{\boldsymbol{X}})) = H(\tilde{X}_1, \tilde{X}_2, \tilde{X}_3) = 2$. Pre spojenú informáciu pri zachovaní marginálnych rozdelení platí $I_c^{(2)}(\tilde{\boldsymbol{X}}) = 0$, $I_c^{(3)}(\tilde{\boldsymbol{X}}) = 1$. Ďalej $I_c^{<2>}(\tilde{\boldsymbol{X}}) = 0$. Rovnosť druhého stupňa spojených informácii všetkých troch typov sme očakávali z rovnosti maximálnych entropií 1. a 2. stupňa pre binárne veličiny, ktorú sme už viedli v príklade 2.53. Zostáva nám určiť hodnotu $I_c^{<3>}(\tilde{\boldsymbol{X}})$. Za zadané momenty teda zvolíme momenty vypočítané z rozdelenia pre XOR. Opäť zistíme, že pri zadaných nosičoch už z hodnôt momentov tretieho rádu vieme jednoznačne určiť, že rozdelenie popisujúce naše veličiny je XOR, je to zároveň rozdelenie s maximálnou entropiou pri zadaných momentoch. Celkovo dostávame, rovnako ako pre spojené informácie tretieho rádu ostatných druhov, $I_c^{<3>}(\tilde{\boldsymbol{X}}) = 3 - 2 = 1$, čo by sme mohli interpretovať ako interakciu len medzi trojicou premenných.

Príklad 3.2 (Interakcie medzi dvojicami)**.** Uveďme príklad na veličiny, u ktorých sa neprejavuje interakcia v trojiciach, ale len v pároch. Majme vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \tilde{X}_2, \tilde{X}_3)$. Nech $\tilde{X}_1$ a $\tilde{X}_2$ sú nezávislé binárne veličiny s uniformným rozdelením pravdepodobnosti. Nech $\tilde{X}_3 = \tilde{X}_1$. Rozdelenie pravdepodobnosti pre takéto veličiny je uvedené v tabuľke 3.2.

$\tilde{x}_1, \tilde{x}_2, \tilde{x}_3$	$p(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$	$\tilde{x}_1, \tilde{x}_2, \tilde{x}_3$	$p(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$
0,0,0	1/4	0,1,1	0
1,0,0	0	1,0,1	1/4
0,1,0	1/4	1,1,0	0
0,0,1	0	1,1,1	1/4

Tabuľka 3.2: Rozdelenie pravdepodobnosti z príkladu 3.2, $\tilde{X}_3 = \tilde{X}_1$

Hodnoty entropií a vzájomnej informácie pre veličiny s rozdelením pravdepodobnosti z tabuľky 3.2 sú nasledovné

$$
\begin{aligned}
H(\tilde{X}_1) = H(\tilde{X}_2) = H(\tilde{X}_3) &= 1,\\
H(\tilde{X}_1, \tilde{X}_2) = H(\tilde{X}_2, \tilde{X}_3) &= 2,\\
H(\tilde{X}_1, \tilde{X}_3) &= 1,\\
I(\tilde{X}_1; \tilde{X}_2) = I(\tilde{X}_2; \tilde{X}_3) &= 0,\\
I(\tilde{X}_1; \tilde{X}_3) &= 1,\\
H(\tilde{X}_1, \tilde{X}_2, \tilde{X}_3) &= 2.
\end{aligned}
$$

Vypočítajme opäť všetky typy spojenej informácie rádu 2 a rádu 3. Pri zadaní entropií, momentov alebo marginálnych rozdelení rádu 1, je maximálna entropia $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{<1>}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{(1)}(\tilde{\boldsymbol{X}})) = \sum_{i=1}^{3} H(\tilde{X}_i) = 3$. Ak však maximalizujeme pri zadaných veličinách prvého aj druhého rádu, dostávame $H(\tilde{P}^{[2]}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{<2>}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{(2)}(\tilde{\boldsymbol{X}})) = 2$, keďže maximálne entropické rozdelenie je v tomto prípade rovné pôvodnému rozdeleniu pravdepodobnosti, a to má entropiu $H(\tilde{X}_1, \tilde{X}_2, \tilde{X}_3) = H(\tilde{X}_1, \tilde{X}_2) = 2$. Znalosťou entropií, resp. momentov alebo marginálnych rozdelení tretieho rádu sa už nič nezmení, platí $H(\tilde{P}^{[3]}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{<3>}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{(3)}(\tilde{\boldsymbol{X}})) = 2$. Z toho dostávame hodnoty spojených informácii druhého rádu $I_c^{[2]}(\tilde{\boldsymbol{X}}) = I_c^{<2>}(\tilde{\boldsymbol{X}}) = I_c^{(2)}(\tilde{\boldsymbol{X}}) = 1$ a hodnoty spojených informácii tretieho rádu $I_c^{[3]}(\tilde{\boldsymbol{X}}) = I_c^{<3>}(\tilde{\boldsymbol{X}}) = I_c^{(3)}(\tilde{\boldsymbol{X}}) = 0$, čo môžeme interpretovať ako interakciu len medzi dvojicami premenných.

Poznámka 3.3. Na príkladoch sme ilustrovali dve zaujímavé vlastnosti. Prvou z nich je už spomínaná možná interpretácia významu spojených informácii. Druhou vlastnosťou, ktorú na príkladoch vidíme je, že spojené informácie ani jedného typu nebudú monotónne. V prípade XOR totiž platí $I_c^{[2]}(\tilde{\boldsymbol{X}}) < I_c^{[3]}(\tilde{\boldsymbol{X}})$, podobne aj pre ostatné typy. V prípade interakcií vo dvojiciach platí opačná nerovnosť.

Pojem interakcia určitého rádu budeme chcieť ďalej využívať pri skúmaní veličín a vyslovovaní viet, definujme ho preto teraz korektne matematicky pomocou vlastností určitých marginálnych rozdelení.

Definícia 3.4 (Interakcia rádu k)**.** Majme $n \in \mathbb{N}$, $k \in \mathbb{N}$, $k \leq n$, diskrétnu vektorovú náhodnú veličinu $\boldsymbol{X} = (X_1, \ldots, X_n) \sim p(\boldsymbol{x}) = p(x_1, \ldots, x_n)$. Potom hovoríme, že medzi zložkami $\boldsymbol{X}$ sa prejavujú interakcie rádu k, ak existuje $(i_1, \ldots, i_k)$ vybraná k-tica indexov z $\widehat{n}$, v ktorej sa indexy neopakujú taká, že marginálne rozdelenie $p(x_{i_1}, \ldots, x_{i_k})$ sa nedá zapísať ako súčin aspoň dvoch marginálnych rozdelení nižších rádov.

Poznámka 3.5. Objasnime motiváciu k definícii. To, že niektoré rozdelenie rádu k sa nedá zapísať ako súčin marginálnych rozdelení nižších rádov znamená, že keď vyberieme ľubovoľných $l \in \widehat{k-1}$ zložiek z danej k-tice, tak ostatné zložky sú od vybraných závislé. Tiež to znamená, že rozdelenie rádu k popisuje časť informácie systému, ktorú nie sme schopní sa dozvedieť, ak máme možnosť pozorovať len rozdelenia rádu $k-1$. Práve túto vlastnosť systému sme doteraz intuitívne chápali ako interakciu daného vyššieho rádu, preto je uvedená definícia v súlade s tým, ako pojem chápeme a používame pri popise systémov.

Poznámka 3.6. Pokiaľ sa rozdelenie nedá zapísať v tvare súčinu marginál nižšieho stupňa, budeme ho označovať ako nedeliteľné alebo nefaktorizovateľné.

Mohlo by nás zaujímať, ako interakcie rôznych rádov navzájom spolu súvisia. Preskúmajme preto teraz niektoré základné vlastnosti nami definovaných interakcií rádu k. Popritom sa nám tiež podarí popísať, ako vo všeobecnosti vyzerá maximálne entropické rozdelenie pri daných marginálnych rozdeleniach určitého stupňa.

Príklad 3.7 (Interakcie rádu 2 a 3)**.** Ako prvé preskúmame, kedy interakcia nižšieho rádu vynucuje existenciu interakcie rádu vyššieho. Popíšme najprv konkrétny príklad. Majme diskrétnu vektorovú náhodnú veličinu $\boldsymbol{X} = (X_1, X_2, X_3, X_4)$, nech sa u nej prejavujú interakcie druhého rádu, to znamená, že existuje minimálne jedno marginálne rozdelenie rádu 2, ktoré sa nedá faktorizovať na dve marginálne rozdelenia prvého rádu. Ľahko nahliadneme, že pokiaľ takéto rozdelenie je len jedno, neimplikuje to existenciu interakcií vyššieho rádu.

Majme teda aspoň dve marginálne rozdelenia rádu 2, ktoré sa nedajú faktorizovať. Ukážeme, že ak tieto dve rozdelenia neobsahujú rovnaké zložky $\boldsymbol{X}$, tak to neimplikuje interakciu vyššieho rádu. V našom prípade je až na permutáciu indexov možná len jedna takáto voľba, preto bez ujmy na všeobecnosti, nech $p(x_1, x_2)$ a $p(x_3, x_4)$ sa nedajú faktorizovať. Potom nájdeme faktorizáciu marginálnych rozdelení tretieho rádu takú, že sú zachované rozdelenia rádu 2. Konkrétne $p(x_1, x_2, x_3) = p(x_3)p(x_1, x_2)$, $p(x_1, x_2, x_4) = p(x_4)p(x_1, x_2)$, $p(x_1, x_3, x_4) = p(x_1)p(x_3, x_4)$, $p(x_2, x_3, x_4) = p(x_2)p(x_3, x_4)$. Faktorizáciu teda stačí voliť tak, že ak trojica indexov obsahuje v sebe dvojicu indexov, pre ktoré rozdelenie druhého stupňa nedá faktorizovať, tak túto dvojicu necháme vo faktorizácii a rozdelenie doplníme súčinom s marginálnym rozdelením poslednej zložky. Tento krok popisujeme detailne hlavne preto, že bude dôležitý pri zovšeobecnení, ktoré budeme skúmať v nasledujúcej poznámke.

Na príklade štvorzložkovej veličiny ešte ilustrujme, že pokiaľ sa indexy zložiek v dvojiciach rozdelení, ktoré nemôžeme faktorizovať, opakujú, tak to vynucuje existenciu interakcie vyššieho rádu. Majme $p(x_1, x_2)$, $p(x_2, x_3)$, ktoré sa nedajú rozložiť na súčin. Potom rozdelenie $p(x_1, x_2, x_3)$ sa nedá rozložiť na súčin rozdelení rádu 1 a 2, neplatí $p(x_1, x_2, x_3) = p(x_1)p(x_2, x_3)$, ani $p(x_1, x_2, x_3) = p(x_2)p(x_1, x_3)$, ani $p(x_1, x_2, x_3) = p(x_3)p(x_1, x_2)$. To, že to neplatí, ľahko overíme vysčítaním oboch strán cez jeden z indexov dvojice, napríklad po vysčítaní $p(x_1, x_2, x_3) = p(x_1)p(x_2, x_3)$ cez x_3 dostávame $p(x_1, x_2) = p(x_1)p(x_2)$, to však nie je pravda, podobne pre ostatné rozklady. Máme teda interakciu tretieho rádu. Ak by sa dokonca žiadne z rozdelení druhého stupňa nedalo faktorizovať, tak by to implikovalo interakciu štvrtého rádu. Stupeň takto implikovanej interakcie bude jednou z vlastností, ktoré preskúmame v nasledujúcej poznámke.

Poznámka 3.8 (Interakcie rôznych rádov všeobecne)**.** Popíšme teraz vo všeobecnosti, ako spolu súvisia interakcie rôznych rádov n-zložkovej diskrétnej náhodnej veličiny. Nech sa medzi zložkami prejavujú interakcie rádu k, t. j. existuje aspoň jedno nefaktorizovateľné rozdelenie rádu k. Ak naviac platí, že každý index zložky sa vyskytuje maximálne v jednom nefaktorizovateľnom rozdelení, pričom uvažujeme všetky nefaktorizovateľné rozdelenia od rádu 2 až po k, potom nie sú implikované interakcie rádu $k + 1$. Pokiaľ sa však indexy opakujú, môže byť implikovaná existencia vyšších rádov interakcií.

Určime konkrétne rád interakcií, ktoré sú takto implikované. V postupe pre $n = 4$ sme videli, že kľúčovou myšlienkou pri tvorbe rozdelení vyššieho rádu konzistentných s danými marginálnymi je, že ak určité marginálne rozdelenie stupňa k nemôžeme faktorizovať, tak aj v rozdeleniach vyššieho rádu sa musí vyskytovať ako celok. Ak však máme nedeliteľné rozdelenia, v ktorých sa indexy zložiek $\boldsymbol{X}$ opakujú a chceme, aby sa všetky vyskytovali v rozdelení vyššieho rádu ako celok, jediná možnosť, ako to urobiť, je zoskupiť dané rozdelenia do marginály vyššieho rádu, ktorá sa nebude dať deliť na súčin. Pokiaľ máme rozdelenie p_1, ktoré sa nedá faktorizovať a popisuje zložku s indexom (alebo viacerými indexmi), ktorý je rovnaký s indexom zložky popisovanej v inom rozdelení p_2, ktoré je tiež nedeliteľné, tak sa všetky zložky popisované týmito rozdeleniami musia vyskytovať v rozdelení vyššieho rádu vždy v spoločnom rozdelení, ako sme diskutovali vyššie. Nefaktorizovateľné rozdelenia p_1 a p_2 teda môžeme chápať ako určitým spôsobom prepojené, t. j. o nefaktorizovateľných rozdeleniach budeme hovoriť, že sú prepojené, ak v nich existuje aspoň jedna zložka s rovnakým indexom. V celkovom rozdelení vektorovej veličiny budú potom prepojené rozdelenia reprezentované spoločnou nedeliteľnou marginálou. Môže existovať ďalšie nedeliteľné rozdelenie p_3, ktoré popisuje rovnakú zložku, aká je popisovaná aj v rozdelení p_1, teda aj p_3 je prepojené s p_1, s jedným rozdelením môže byť prepojených viacero ďalších. Podobne môžu existovať ďalšie rozdelenia, ktoré obsahujú zložku rovnakú s nejakou zložkou popisovanou v p_3, môžme teda ďalej skúmať prepojenia rozdelení, ktoré sú už prepojené s p_1. Celkovo sme schopní nájsť všetky rozdelenia, ktoré sú prepojené s p_1, potom všetky rozdelenia, ktoré sú prepojené s týmito rozdeleniami a tak ďalej. Potom maximálny stupeň interakcie, ktorý rozdelenia p_i, ktoré sú takto postupne prepojené určite implikujú, je rovný rádu ich spoločného marginálneho rozdelenia, t. j. rozdelenia vybraných zložiek $\boldsymbol{X}$, ktoré obsahuje práve všetky indexy veličín v rozdeleniach p_i. Poznamenajme, že ak interakcie rádu k implikujú interakcie rádu $k + l$, tak nemusí platiť, že implikujú aj interakcie rádov $k + 1, \ldots, k + l - 1$.

Zistili sme, že môžu existovať interakcie vyššieho rádu, o ktorých existencii vieme zo znalosti marginál niektorého rádu nižšieho. Zároveň však vieme, že nie všetky interakcie vyšších rádov sa dajú predpovedať len na základe znalostí rozdelení nižších rádov. Zaveďme preto nasledujúce názvoslovie.

Definícia 3.9. (Implikované interakcie) Majme diskrétnu vektorovú náhodnú veličinu $\boldsymbol{X} = (X_1, \ldots, X_n)$. Hovoríme, že interakcie rádu l nie sú implikované zadanými marginálnymi rozdeleniami rádu k pre nejaké $k < l$, ak medzi rozdeleniami veličiny $\boldsymbol{X}$ konzistentnými so zadanými marginálami rádu k sú aj také, pre ktoré platí, že všetky ich marginálne rozdelenia stupňa l sa dajú zapísať v tvare súčinu aspoň dvoch marginál stupňov nižších ako l. Inak hovoríme, že interakcie rádu l sú implikované danými marginálami rádu k.

Poznámka 3.10. Interakcie, ktoré nie sú implikované žiadnymi marginálami nižšieho rádu, budeme nazývať neimplikované. Ďalej ak nebude záležať na tom, ktorým rádom marginál je interakcia implikovaná, budeme skrátene hovoriť o implikovanej interakcii. Myslíme tým teda, že interakcia je implikovaná niektorým z nižších rádov marginál.

Konkrétnym príkladom na neimplikované interakcie tretieho rádu je rozdelenie troch veličín vo vzťahu XOR, ktoré sme skúmali v príklade 3.1. Je tiež jasné všeobecnejšie tvrdenie - ak sa medzi zložkami náhodnej veličiny vyskytuje určitý minimálny rád interakcií, tak interakcie tohto minimálneho rádu sú neimplikované.

Poznámka 3.11. Upozornime, že ak marginály rádu k neimplikujú interakciu žiadneho vyššieho rádu, tak to znamená, že celkové rozdelenie pravdepodobnosti veličiny s týmito marginálami sa dá zapísať v tvare súčinu marginál maximálne rádu k.

Poznámka 3.12. V texte budeme viackrát používať pojem rozdelenie konzistentné so zadanými marginálami, resp. konzistentné s niektorým iným druhom väzbových podmienok. Myslíme tým rozdelenie, na ktorom sú zadané väzbové podmienky splnené.

Poznámka 3.13. V poznámke 3.8 sme preskúmali niektoré súvislosti medzi tvarom marginál a tým, aké interakcie dané marginály implikujú, popísali sme jeden možný dôvod toho, prečo dané marginály vynucujú existenciu interakcií vyšších rádov. Mohlo by nás zaujímať, či všetky implikované interakcie podľa definície 3.9 sú nutne implikované spôsobom, aký sme popísali v poznámke, alebo existuje ešte nejaký iný dôvod, prečo sa interakcie implikujú. Túto otázku prenechávame otvorenú na budúce skúmanie, v poznámke 3.8 sme ju sčasti zodpovedali pre dva po sebe nasledujúce rády interakcií.

3.1 Interakcie a marginálne rozdelenia

V nasledujúcej poznámke zhrnieme, ako na základe doterajších pozorovaní o rozdeleniach interagujúcich veličín môžeme určiť tvar rozdelenia konzistentného s danými marginálnymi rozdeleniami určitého stupňa. Nemyslíme tým však nájdenie presných hodnôt, ktoré toto rozdelenie bude dosahovať na určitých hodnotách náhodných veličín, rozdelení konzistentných s danými marginálami totiž môže byť viac.

Poznámka 3.14 (Rozdelenie pri daných marginálach)**.** Vďaka prieskumu interakcii v poznámke 3.8 môžeme rovno sformulovať, ako všeobecný tvar rozdelenia $\boldsymbol{X} = (X_1, \ldots, X_n)$ konzistentného s marginálnymi rozdeleniami stupňa k nájdeme. Stačí zložky $\boldsymbol{X}$ rozdeliť do skupín podľa toho, ktoré zložky spolu interagujú. Upresnime, čo tým myslíme. Pokiaľ majú dve zložky v zadaných marginálach spoločné rozdelenie, ktoré sa nedá faktorizovať, tak ich budeme brať ako interagujúce. Podobne nielen pre dvojice, ale aj pre m-tice zložiek, ak sa rozdelenie týchto m zložiek nedá faktorizovať, tak tieto zložky spolu interagujú. Toto vnímanie interakcii je opäť v súlade s intuitívnym vnímaním a aj s definíciou 3.4. Takýmto spôsobom preskúmame všetky interakcie v daných marginálach a rozdelíme zložky na disjunktné množiny, medzi ktorými sa interakcie neprejavujú.

Jednou z možností, ako konštruovať rozdelenie $\boldsymbol{X}$, ktoré je konzistentné s danými marginálami, je vytvoriť ho ako súčin rozdelení týchto skupín zložiek. Toto rozdelenie sa už nedá ďalej faktorizovať na nižšie marginály. Označme skupinu takto vytvorených rozdelení $\mathcal{P}_F$. Ďalšie rozdelenia $\boldsymbol{X}$ konzistentné so zadanými marginálami sú tie, v ktorých je niekoľko skupín interagujúcich zložiek reprezentovaných spoločným marginálnym rozdelením, ktoré nie je rovnaké, ako vzniká súčinom marginál daných zložiek. Množinu rozdelení, ktoré takto vznikli označíme $\mathcal{P}_I$. V množine $\mathcal{P}_I$ sa teda už nevyskytujú rozdelenia z množiny $\mathcal{P}_F$. Intuitívne je $\mathcal{P}_I$ množina rozdelení, v ktorých sa síce prejavujú interakcie zadané danými marginálnymi rozdeleniami, ale okrem toho sa v nich prejavujú aj interakcie, ktoré dané marginálne rozdelenia neimplikujú. Stále platí, že zložky veličiny, ktoré spolu interagovali, musia byť v jednom nefaktorizovateľnom marginálnom rozdelení. Postup ilustrujeme na príklade štvorzložkovej náhodnej veličiny v príklade 3.18.

Poznámka 3.15. V poznámke 3.14 predstavuje $\boldsymbol{X}$ nejakú náhodnú veličinu z množiny $\{\boldsymbol{X}\}$ náhodných veličín, ktoré majú obory hodnôt dané zadanými marginálami. V nasledujúcom texte nebudeme na túto skutočnosť upozorňovať, z kontextu bude jasné, čo značkou $\boldsymbol{X}$ myslíme. Toto značenie využijeme napríklad v dôkazoch viet o interakciách, tiež v niektorých príkladoch.

Poznámka 3.16. O $\mathcal{P}_F$ hovoríme ako o skupine rozdelení, pretože nemusí platiť, že každá z marginál, z ktorých súčinu sa $\mathcal{P}_F$ skladá je jednoznačne určená zadanými marginálami stupňa k. Máme teda viac možností, ako môže každý z faktorov vyzerať, jednoznačne je v tomto prípade určené len to, ktoré zložky $\boldsymbol{X}$ sa vyskytujú v spoločnej marginále.

Poznámka 3.17. Podobný popis štruktúry rozdelení pri zadaných marginálach je navrhnutý a krátko popísaný v článku [7]. Autori štruktúru rozdelení popisujú pomocou zväzu a taktiež hľadajú istú formu rozdelenia, ktoré sa už ďalej nedá faktorizovať. V rámci našej práce nebudeme množiny rozdelení formalizovať do zväzu, vystačíme si so zavedeným popisom. V nasledujúcej sekcii tiež uvidíme, že analogický popis sa dá priamočiaro upraviť a použiť aj pre hľadanie rozdelení konzistentných so zadanými entropiami, štruktúra zväzu by nám v tomto ohľade spôsobila skôr zbytočné komplikácie s tým, ako preniesť formalizmus do skúmania entropií.

Príklad 3.18. Uvažujme štvorzložkové vektorové náhodné veličiny, nech sú zadané marginálne rozdelenia druhého stupňa v tvare $p(x_1, x_2) = p(x_1)p(x_2)$, $p(x_1, x_3) = p(x_1)p(x_3)$, $p(x_1, x_4) = p(x_1)p(x_4)$, $p(x_2, x_4) = p(x_2)p(x_4)$, a nedeliteľné $p(x_2, x_3)$, $p(x_3, x_4)$. Potom disjunktné množiny interagujúcich zložiek budú $\{X_1\}$, $\{X_2, X_3, X_4\}$ a tvar rozdelenia $\boldsymbol{X}$ konzistentného s marginálami druhého stupňa, ktorý nemôžeme ďalej faktorizovať, bude $p(x_1, x_2, x_3, x_4) = p(x_1)p(x_2, x_3, x_4)$, kde $p(x_2, x_3, x_4)$ je nedeliteľné. Ďalší prípustný tvar rozdelenia je úplne nefaktorizované, teda $p(x_1, x_2, x_3, x_4)$.

Poznáme všeobecný tvar rozdelení konzistentných s danými marginálami, ďalej nás zaujíma, ako z týchto rozdelení vybrať to, ktoré má maximálnu entropiu. Skôr, než vyriešime túto otázku, vyslovíme lemu o entropiách rozdelenia faktorizovaného na súčin marginál. V leme a v ďalších tvrdeniach budeme pre miesto značenia $p(\boldsymbol{x})$ používať značenie p. Zjednoduší sa nám spôsob zápisu, pretože inak by sme museli vo výraze $p = p_1 \cdot p_2 \cdot \ldots \cdot p_k$ uvádzať aj funkčné závislosti v marginálach na pravej strane rovnosti. Zápis chápeme tak, že každá zo zložiek $\boldsymbol{X}$ sa vyskytuje práve v jednom marginálnom rozdelení na pravej strane, pre nás však nie je podstatné, v ktorom presne.

Lema 3.19. *Majme $n \in \mathbb{N}$, $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$ množinu náhodných veličín s daným oborom hodnôt. Majme p_A a p_B dve rozdelenia pravdepodobnosti nejakých veličín z množiny $\{\boldsymbol{X}\}$. Nech rozdelenie pravdepodobnosti p_A je také, že sa dá zapísať v tvare súčinu ako $p_A = p_1 \cdot p_2 \cdot \ldots \cdot p_k$, kde $k \in \widehat{n-1}$ a p_i sú pre všetky $i \in \widehat{k}$ marginálne rozdelenia niektorého nižšieho stupňa, ktoré sa už ďalej nedajú faktorizovať na súčin marginál. Nech p_1 je rozdelenie stupňa aspoň m pre nejaké $m \geq 2$. Nech ďalej p_B je rozdelenie pravdepodobnosti, ktoré je v tvare $p_B = p_{1_1} \cdot \ldots \cdot p_{1_m} \cdot p_2 \cdot \ldots \cdot p_k$, kde p_{1_i} sú nejaké marginály p_1. Potom pri predpoklade konečnosti entropií oboch rozdelení platí $H(p_A) < H(p_B)$.*

Dôkaz. Rozpíšme obidve skúmané entropie na súčet $H(p_A) = H(p_1) + \sum_{i=2}^{k} H(p_i)$, $H(p_B) = H(p_{1_1}) + \cdots + H(p_{1_m}) + \sum_{i=2}^{k} H(p_i)$. Z tohto zápisu vidíme, že na dokázanie nerovnosti $H(p_A) < H(p_B)$ stačí ukázať, že $H(p_1) < H(p_{1_1}) + \cdots + H(p_{1_m})$. Z predpokladu vieme, že p_1 sa nedá faktorizovať, teda $p_1 \neq p_{1_1} \cdot \ldots \cdot p_{1_m}$, z toho však vyplýva, že zložky náhodnej veličiny popísanej rozdelením p_1 nie sú nezávislé. Z vety 1.31 dostávame ostrú nerovnosť $H(p_1) < H(p_{1_1}) + \cdots + H(p_{1_m})$, pretože rovnosť sa vo vete dosahuje len v prípade nezávislosti. $\square$

Poznámka 3.20. Z dôkazu je zrejmé, že v leme by stačilo reprezentovať súčin nefaktorizovateľných rozdelení $p_2 \cdot \ldots \cdot p_k$ spoločným rozdelením q, o ktorom by nebolo nutné predpokladať nedeliteľnosť, t. j. rozdelenia p_A a p_B by boli v tvare $p_A = p_1 \cdot q$, kde p_1 nedeliteľné a $p_B = p_{1_1} \cdot \ldots \cdot p_{1_m} \cdot q$. Lemu ponechávame vyslovenú s predpokladom na faktorizovaný tvar pre väčšiu názornosť, práve na rozdelenia vo faktorizovanom tvare ju budeme neskôr aplikovať.

Poznamenajme, že lemu by sme mohli použiť opakovane a tým by sme ukázali ostrú nerovnosť medzi entropiami rozdelení, kde je faktorizovaná nielen prvá zložka, ale aj ďalšie zložky rozdelenia.

Poznatky vyplývajúce z lemy nám umožnia nájsť tvar maximálne entropického rozdelenia, čo neskôr využijeme v dôkaze jedného z dôležitých tvrdení 3.27.

Poznámka 3.21 (Maximálne entropické rozdelenie pri marginálach)**.** Ako sme už upozornili, rozdelení konzistentných s danými marginálami určitého rádu môže byť viac, zaujíma nás, ako medzi nimi vybrať to, ktoré má maximálnu entropiu.

Majme $\{\boldsymbol{X}\}$, ďalej nech sú zadané marginálne rozdelenia rádu k, pri ktorých chceme maximalizovať entropiu. Aby bola zaručená konečnosť entropií všetkých rozdelení, s ktorými budeme pracovať, predpokladajme, že $\{\boldsymbol{X}\}$ má konečný obor hodnôt. Prvým krokom je nájdenie všetkých rozdelení konzistentných s danými marginálami. Všetky tieto rozdelenia môžeme napísať v tvare súčinu marginál, ktoré sa už nedajú ďalej faktorizovať a podľa značenia z poznámky 3.14 ich môžeme rozdeliť do množín $\mathcal{P}_F$ a $\mathcal{P}_I$.

Na základe lemy 3.19 vieme, že ak máme k dispozícii dve rozdelenia náhodnej veličiny, kde jedno sa nedá faktorizovať a druhé je rozdelením, v ktorom sú niektoré zložky prvého rozdelenia ďalej faktorizované na súčin nejakých jeho marginál, tak druhé rozdelenie má ostre väčšiu entropiu, než prvé. Z toho vyplýva, že maximálne entropické rozdelenie stačí hľadať v skupine rozdelení $\mathcal{P}_F$. Ku každému rozdeleniu z $\mathcal{P}_I$ totiž v množine $\mathcal{P}_F$ existuje rozdelenie, kde je niektorá z marginál ďalej faktorizovaná, a teda jeho entropia je ostro väčšia.

Preskúmajme, ako vyzerajú rozdelenia v množine $\mathcal{P}_F$. Pokiaľ marginálne rozdelenie rádu k, pri ktorom maximalizujeme, neimplikuje interakcie vyšších rádov, tak v $\mathcal{P}_F$ sa nachádza jediné rozdelenie. Je to rozdelenie, ktoré vzniklo súčinom marginál maximálne rádu k, všetky tieto marginály však poznáme, takže toto rozdelenie je jednoznačne určené.

Nájdime maximálne entropické rozdelenie ešte pre prípad, že sa implikujú interakcie rádu väčšieho, než k, označme ho l. Výsledný ďalej nedeliteľný tvar rozdelenia obsahuje minimálne jedno marginálne rozdelenie rádu l alebo vyššieho, ktoré nepoznáme. Na nájdenie maximálne entropického rozdelenia potrebujeme maximalizovať cez rozdelenia zložiek zahrnutých v marginálach, ktoré nepoznáme. Na týchto zložkách teda riešime opäť úlohu na maximalizáciu entropie pri zadaných marginálach, t. j. úlohu na viazané extrémy, oproti pôvodnej úlohe na maximalizáciu entropie pri zadaných marginálach sa ale znížili rozmery. Poznamenajme, že v tomto prípade množina rozdelení $\mathcal{P}_F$ nie je jednoprvková, my hľadáme maximum entropie na jej prvkoch. Ilustrujme postup hľadania maximálne entropického rozdelenia na príklade.

Príklad 3.22. Majme podobne ako v príklade 3.18 $\{\boldsymbol{X}\} = \{(X_1, X_2, X_3, X_4)\}$ s konečným oborom hodnôt a zadané marginály rádu 2. Preskúmajme najprv prvú situáciu popisovanú v predchádzajúcej poznámke. Nech sú marginálne rozdelenia v nasledujúcom tvare. Nech $p(x_1, x_2)$ je nedeliteľné, ostatné marginály druhého rádu sú v tvare súčinu prvých marginál. Potom maximálne entropické rozdelenie je $p(x_1, x_2)p(x_3)p(x_4)$, všetky zložky súčinu poznáme.

Preskúmajme teraz situáciu, keď zadané marginály implikujú interakcie tretieho rádu. Majme zadané marginály $p(x_1, x_2)$ a $p(x_2, x_3)$ nedeliteľné, ostatné marginály sa dajú faktorizovať na súčin prvých. Potom všeobecný tvar konzistentného rozdelenia je $p(x_1, x_2, x_3)p(x_4)$. Poznáme zložku $p(x_4)$, ale nepoznáme $p(x_1, x_2, x_3)$. Toto rozdelenie nájdeme maximalizáciou pri zadaných marginálach párov zložiek X_1, X_2, X_3, môžeme riešiť napríklad metódou Lagrangeových multiplikátorov. Pri maximalizácii už nemusíme brať do úvahy štvrtú zložku.

Poznámka 3.23. Pre účely dôkazov v tejto kapitole je pre nás dôležité vedieť rozdeliť zložky vektorovej náhodnej veličiny do interagujúcich skupín a určiť tvar maximálne entropického rozdelenia. Pri výpočtoch založených na reálnych dátach však pre nás nie je jednoduché určiť, či sa rozdelenie dá alebo nedá faktorizovať na súčin nižších marginál. V prípade zadaných marginálnych rozdelení obvykle nie je výrazne výhodnejšie rozdeľovať zložky na neinteragujúce skupiny a až potom maximalizovať, často je výpočetne rýchlejšie rovno maximalizovať entropiu všetkých zložiek. Poznamenajme, že v prípade maximalizácii pri zadaných marginálach sú väzbové podmienky na maximalizáciu lineárne, optimalizačná úloha sa dá pomerne priamočiaro riešiť, to je tiež dôvod, prečo nemáme motiváciu znižovať dimenziu riešenej úlohy čo najviac.

Značnú časť podkapitoly sme venovali skúmaniu interakcií medzi veličinami, podarilo sa nám tiež nájsť všeobecný tvar maximálne entropického rozdelenia pri zadaných marginálach. Uvidíme, že získané poznatky využijeme pri dôkaze tvrdení o súvislosti hodnôt postupnosti maximálnych entropií, resp. spojených informácii a interakcií, ktoré sú jedny z kľúčových tvrdení celej práce, a to hlavne preto, že prepájajú nami budovaný matematický aparát s interpretáciou, ktorá bola doteraz možná len na intuitívnej úrovni. Pristúpme teda k vysloveniu a dokázaniu niekoľkých tvrdení o entropiách maximalizovaných pri zadaných marginálach.

Veta 3.24 (O $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$, implikované interakcie)**.** *Majme $n \in \mathbb{N}$, ďalej množinu náhodných veličín s daným konečným oborom hodnôt $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$ a diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Nech sa v rozdelení $\tilde{\boldsymbol{X}}$ prejavujú interakcie rádov $k_1, \ldots, k_m$, kde $k_1 < \cdots < k_m$, $m \leq n$, môžu byť aj implikované. Maximalizujme entropiu pri marginálnych rozdeleniach veličiny $\tilde{\boldsymbol{X}}$. Potom platia neostré nerovnosti*

$$H(\tilde{P}^{(k_i)}(\tilde{\boldsymbol{X}})) \leq H(\tilde{P}^{(k_i-1)}(\tilde{\boldsymbol{X}})) \qquad \forall i \in \widehat{m} \tag{3.1}$$

$$H(\tilde{P}^{(k_i)}(\tilde{\boldsymbol{X}})) \geq H(\tilde{P}^{(k_{i+1})}(\tilde{\boldsymbol{X}})) \qquad \forall i \in \widehat{m-1}. \tag{3.2}$$

Ďalej platia rovnosti

$$H(\tilde{P}^{(k_i)}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{(k_i+1)}(\tilde{\boldsymbol{X}})) = \cdots = H(\tilde{P}^{(k_{i+1}-1)}(\tilde{\boldsymbol{X}})) \qquad \forall i \in \widehat{m} \tag{3.3}$$

$$H(\tilde{P}^{(1)}(\tilde{\boldsymbol{X}})) = \cdots = H(\tilde{P}^{(k_1-1)}(\tilde{\boldsymbol{X}})) = \sum_{j=1}^{n} H(\tilde{X}_j). \tag{3.4}$$

Dôkaz. Nerovnosti (3.1) a (3.2) vyplývajú z monotónie entropie maximalizovanej pri zadaných marginálach, t. j. z vety 2.24. Pracujeme s náhodnými veličinami s konečným oborom hodnôt, na ukázanie ostatných vzťahov nám preto stačí nájsť vzťahy medzi entropiami maximálne entropických rozdelení.

Ukážme rovnosť (3.4). Z toho, že najnižší rád interakcie je k_1 vieme, že maximálne entropické rozdelenie pri zadaných marginálach rádu 1 až $k_1 - 1$ je $\prod_{j=1}^{n} p(x_j) = \prod_{j=1}^{n} p(\tilde{x}_j)$, to už implikuje platnosť vzťahu (3.4).

Dokážme vzťah (3.3). Označme množinu rozdelení $\mathcal{P}_F$ z poznámky 3.14 pri zadaných marginálach rádu k ako $\mathcal{P}_F^k$. Zvoľme ľubovoľne pevne $i \in \widehat{m}$. Kvôli tomu, že sa medzi rádmi k_i a k_{i+1} nevyskytujú žiadne nové interakcie platí, že rozdelenia v množine $\mathcal{P}_F^{k_i}$ ležia aj v množinách $\mathcal{P}_F^{k_i+1}, \dots, \mathcal{P}_F^{k_{i+1}-1}$. Máme teda vzťah inklúzie $\mathcal{P}_F^{k_i} \subset \mathcal{P}_F^j$ pre všetky $j \in \{k_i + 1, \dots, k_{i+1} - 1\}$. Na týchto množinách nájdeme suprémum entropie, potom z lemy 2.19 dostávame $H(\tilde{P}^{(k_i)}(\tilde{\boldsymbol{X}})) \leq H(\tilde{P}^{(j)}(\tilde{\boldsymbol{X}}))$ a z monotónie vyplýva opačná nerovnosť $H(\tilde{P}^{(k_i)}(\tilde{\boldsymbol{X}})) \geq H(\tilde{P}^{(j)}(\tilde{\boldsymbol{X}}))$ pre všetky $j \in \{k_i + 1, \dots, k_{i+1} - 1\}$. Musí teda platiť $H(\tilde{P}^{(k_i)}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{(j)}(\tilde{\boldsymbol{X}}))$ pre všetky $j \in \{k_i + 1, \dots, k_{i+1} - 1\}$. Týmto sme ukázali platnosť rovnosti (3.3). □

Poznámka 3.25. Vzťahy (3.1), (3.2) sú len vyjadrenia monotónie maximálnej entropie pre určité rády, môže sa teda zdať zbytočné ich vo vete uvádzať znova. Novým zistením, ktoré z tejto vety vyplýva sú teda len uvedené rovnosti. Dôvod, pre ktorý nerovnosti vo vete predsa uvádzame je, že v nasledujúcej časti budeme skúmať, či by sme nemohli predpoklady vety upraviť tak, aby bolo možné dokázať dokonca ostrú nerovnosť v týchto výrazoch. V poznámke 3.29 tiež zdôvodníme, že ak zvolíme predpoklady tak, ako sme to urobili vo vete 3.24, tak naozaj nemusí nastávať ostrá nerovnosť. To, prečo je pre nás dôležitá platnosť ostrých nerovností a nie len neostrých, vysvetlíme neskôr, v poznámke 3.48.

Môžme ešte vysloviť priamočiary dôsledok vety 3.24, dokázané vzťahy sa totiž dajú zapísať cez spojenú informáciu príslušného typu.

Dôsledok 3.26 (O $I_c^{(k)}(\tilde{\boldsymbol{X}})$, implikované interakcie)**.** *Majme $n \in \mathbb{N}$, množinu náhodných veličín s daným konečným oborom hodnôt $\{\boldsymbol{X}\} = \{(X_1, \dots, X_n)\}$. Majme ďalej diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \dots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Nech sa v rozdelení $\tilde{\boldsymbol{X}}$ prejavujú interakcie rádov $k_1, \dots, k_m$, kde $k_1 < \cdots < k_m$, $m \leq n$, môžu byť aj implikované. Maximalizujme entropiu pri marginálnych rozdeleniach veličiny $\tilde{\boldsymbol{X}}$. Potom platí neostrá nerovnosť*

$$I_c^{(k_i)}(\tilde{\boldsymbol{X}}) \geq 0 \qquad \forall i \in \widehat{m} \tag{3.5}$$

a spojené informácie všetkých ostatných rádov sú nulové.

Dôkaz. Nerovnosť (3.5) vyplýva zo vzťahu (3.1), ale nezápornosť spojenej informácie je dokonca jej všeobecná vlastnosť, ako sme videli vo vete 2.59. Nulovosť všetkých ostatných spojených informácii vyplýva z (3.3). □

Ako sme už spomínali, v ďalšom postupe bude našou snahou upraviť predpoklady vety 3.24 tak, aby platili ostré nerovnosti. Predstavíme jednu možnú verzie predpokladov, ktorá zaručí ostré nerovnosti. Pre lepšiu prehľadnosť budeme budeme vyslovovať aj tvrdenie o rovnostiach, ktoré je v tomto prípade dôsledkom už dokázaných rovností z vety 3.24.

Veta 3.27 (O $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$, neimplikované interakcie)**.** *Majme $n \in \mathbb{N}$, množinu náhodných veličín s daným konečným oborom hodnôt $\{\boldsymbol{X}\} = \{(X_1, \dots, X_n)\}$. Majme ďalej diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \dots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Nech sa v rozdelení $\tilde{\boldsymbol{X}}$ prejavujú len neimplikované interakcie rádov $k_1, \dots, k_m$, kde $k_1 < \dots < k_m$, $m \leq n$, nech sa neprejavujú žiadne implikované interakcie. Maximalizujme entropiu pri marginálnych rozdeleniach veličiny $\tilde{\boldsymbol{X}}$. Potom platia ostré nerovnosti*

$$H(\tilde{P}^{(k_i)}(\tilde{\boldsymbol{X}})) < H(\tilde{P}^{(k_i-1)}(\tilde{\boldsymbol{X}})) \qquad \forall i \in \widehat{m} \tag{3.6}$$

$$H(\tilde{P}^{(k_i)}(\tilde{\boldsymbol{X}})) > H(\tilde{P}^{(k_{i+1})}(\tilde{\boldsymbol{X}})) \qquad \forall i \in \widehat{m-1}. \tag{3.7}$$

Ďalej platia rovnosti

$$H(\tilde{P}^{(k_i)}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{(k_i+1)}(\tilde{\boldsymbol{X}})) = \dots = H(\tilde{P}^{(k_{i+1}-1)}(\tilde{\boldsymbol{X}})) \qquad \forall i \in \widehat{m} \tag{3.8}$$

$$H(\tilde{P}^{(1)}(\tilde{\boldsymbol{X}})) = \dots = H(\tilde{P}^{(k_1-1)}(\tilde{\boldsymbol{X}})) = \sum_{j=1}^{n} H(\tilde{X}_j). \tag{3.9}$$

Dôkaz. V celom dôkaze budeme využívať to, že maximálna entropia sa v prípade náhodných veličín s konečným oborom hodnôt dosahuje, na určenie maximálnej entropie preto stačí nájsť maximálne entropické rozdelenie a vyčísliť jeho entropiu.

Platnosť rovností (3.8) a (3.9) vyplýva z vety 3.24, jej predpoklady sú splnené aj v tomto prípade. Ukážme nerovnosť (3.6). Zvoľme $i \in \widehat{m}$. Vďaka tomu, že medzi zložkami $\tilde{\boldsymbol{X}}$ sa prejavujú len neimplikované interakcie vieme, že maximálne entropické rozdelenie konzistentné s danými marginálami $\tilde{\boldsymbol{X}}$ rádu k_i, keď sa prejavujú interakcie rádu k_i, sa dá zapísať v tvare $p = p_1 \cdot \ldots \cdot p_l$ pre nejaké $l \in \widehat{n}$, kde p_j sa nedá faktorizovať $\forall j \in \widehat{l}$ a všetky p_j sú maximálne rádu k_i.

Z existencie interakcií rádu k_i ale vieme, že existuje faktor rádu k_i, bez ujmy na všeobecnosti, nech je to p_1. Preskúmajme hodnotu $H(\tilde{P}^{(k_i-1)}(\tilde{\boldsymbol{X}}))$, teda uvažujme rozdelenia, ktoré zachovajú marginálne rozdelenia veličiny $\tilde{\boldsymbol{X}}$ až do rádu $k_i - 1$. Medzi nimi je iste aj vyššie zmienené rozdelenie $p = p_1 \cdot \ldots \cdot p_l$, toto rozdelenie spĺňa podmienky na marginály dokonca do rádu k_i. Naviac ukážeme, že medzi nimi musí byť aj rozdelenie q v tvare $q = q_{1_A} q_{1_B} p_2 \cdot \ldots \cdot p_l$, teda také, ktoré se zhoduje s p vo všetkých faktoroch $p_2 \ldots p_l$, len namiesto p_1 obsahuje faktor $q_1 = q_{1_A} q_{1_B}$. Vyjdeme z toho, že v rozdelení p sa prejavujú interakcie rádu k_i a sú neimplikované. Bez ujmy na všeobecnosti sme naviac vyššie zvolili, že tieto (neimplikované) interakcie sa prejavujú práve v p_1. Preto, že sú neimplikované, musí existovať rozdelenie q_1 konzistentné s marginálami rádu $k_i - 1$, ktoré sa dá faktorizovať ako $q_1 = q_{1_A} q_{1_B}$. Toto rozdelenie q_1 ďalej rozšírime na rozdelenie celej veličiny $\boldsymbol{X}$ prenásobením faktormi $p_2 \ldots p_{l_1}$. Takto vzniknuté rozdelenie je určite konzistentné s marginálami do rádu $k_i - 1$ (podobne tieto marginály má aj p), teda sú splnené požiadavky lemy 3.19, odtiaľ dostávame, že $q = q_{1_A} q_{1_B} p_2 \cdot \ldots \cdot p_l$ má ostre väčšiu entropiu, než rozdelenie p. Nemusí sa samozrejme jednať o maximálne entropické rozdelenie konzistentné s marginálami rádu $k_i - 1$, platí však, že $H(\tilde{P}^{(k_i-1)}(\tilde{\boldsymbol{X}}))$ je väčšia alebo rovná entropii $q = q_{1_A} q_{1_B} p_2 \cdot \ldots \cdot p_l$, a teda ostre väčšia, ako entropia p.

Nerovnosť (3.7) dostávame z už dokázaných vzťahov (3.6) a (3.8). □

Opäť môžme z vety odvodiť dôsledok pre spojené informácie.

Dôsledok 3.28 (O $I_c^{(k)}(\tilde{\boldsymbol{X}})$, neimplikované interakcie)**.** *Majme $n \in \mathbb{N}$, ďalej majme danú $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$ množinu náhodných veličín s daným konečným oborom hodnôt a diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Nech sa v rozdelení $\tilde{\boldsymbol{X}}$ prejavujú len neimplikované interakcie rádov $k_1, \ldots, k_m$, kde $k_1 < \cdots < k_m$, $m \leq n$, nech sa neprejavujú žiadne implikované interakcie. Maximalizujme entropiu pri marginálnych rozdeleniach veličiny $\tilde{\boldsymbol{X}}$. Potom platí ostrá nerovnosť*

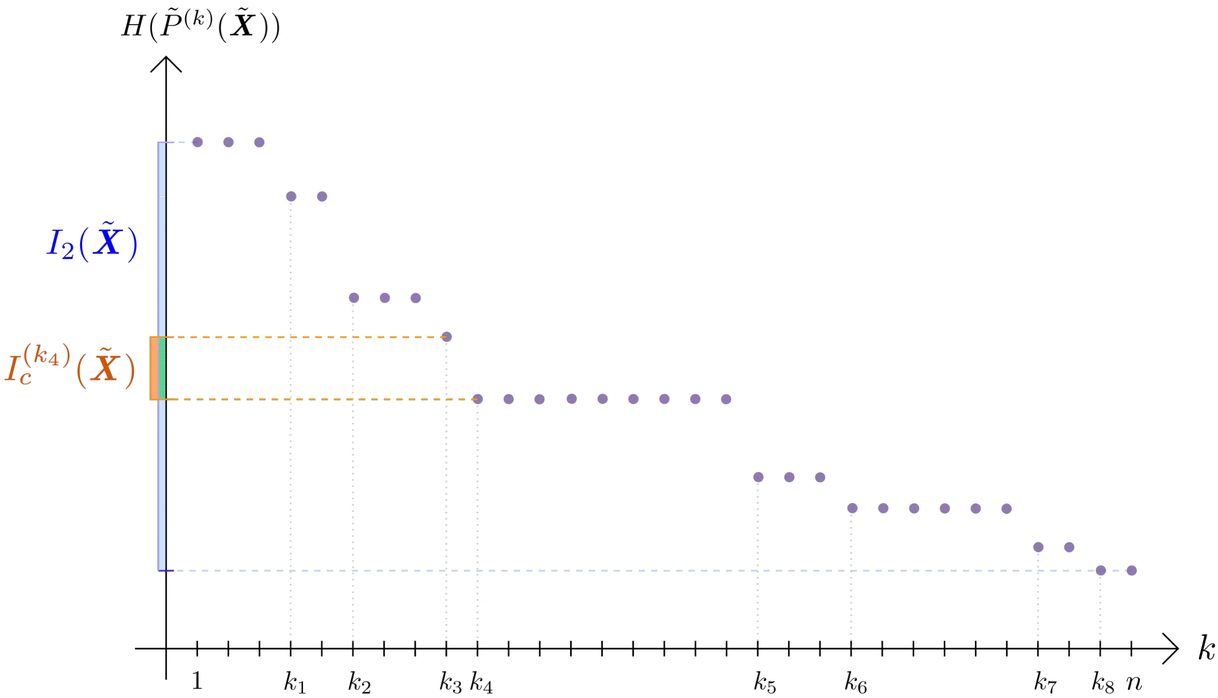
$$I_c^{(k_i)}(\tilde{\boldsymbol{X}}) > 0 \qquad \forall i \in \widehat{m} \tag{3.10}$$

a spojené informácie všetkých ostatných rádov sú nulové.

Dôkaz. Nerovnosť (3.10) vyplýva zo vzťahu (3.6), nulovosť všetkých ostatných spojených informácii vyplýva z (3.8). □

Situáciu popisovanú vetou 3.27 a jej dôsledkom ilustrujeme na grafe 3.1. Hodnoty na osi x predstavujú jednotlivé rády, na os y značíme hodnoty maximálnej entropie daného rádu. Majme vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n)$, nech sa u nej prejavujú len neimplikované interakcie rádov $k_1, \ldots, k_8$. Maximalizujeme entropiu pri marginálach $\tilde{\boldsymbol{X}}$. Potom hodnoty maximálnej entropie tvoria klesajúcu postupnosť a klesajú ostro tam, kde sa vyskytnú interakcie daného rádu. Tam, kde sa interakcie nevyskytujú, zostáva postupnosť konštantná. Dĺžka modrého intervalu reprezentuje celkovú informáciu systému, oranžový interval predstavuje hodnotu spojenej informácie rádu k_4. Podobne by sme mohli intervalmi vyznačiť spojenú informáciu rádu k_i pre $i \in \{1, \dots, 8\}$. Pre ostatné rády je spojená informácia nulová.

Upozornime, že ak by sme nepredpokladali, že interakcie sú len neimplikované, tak by sme nemali zaručené ostré klesanie hodnôt maximálnych entropií pri výskyte interakcií daného rádu.



Obr. 3.1: Postupnosť maximálnych entropií pri daných marginálach pri predpoklade výskytu len neimplikovaných interakcií rádov $k_1, \dots, k_8$

Poznámka 3.29. Ozrejmime, že ostrú nerovnosť naozaj nevieme pri predpokladoch vety 3.24 ukázať. V dôkaze vety 3.27 sme využili, že maximálne entropické rozdelenie pri zadaných marginálach stupňa k_i sa dá faktorizovať na marginály maximálne stupňa k_i, čo sme zaručili predpokladom na výskyt len neimplikovaných interakcií. Ak tento predpoklad úplne vynecháme, tak tvar rozdelenia konzistentného s danými marginálami môže obsahovať marginály vyšších rádov, ktoré sa nedajú faktorizovať. Môže sa teda stať, že maximálne faktorizovaný tvar rozdelenia konzistentného s marginálami rádu k_i a $k_i - 1$ je rovnaký. Nemôžeme tiež zaručiť, že nenastane prípad, že nielen tvar, ale aj konkrétne maximálne entropické rozdelenie týchto dvoch rádov je rovnaké. Postačujúce predpoklady a ďalšie detaily týkajúce sa tejto situácie nebudeme ďalej diskutovať, prenechávame ich na budúce skúmanie.

Poznámka 3.30 (O interpretácii)**.** Z poznámky 3.29 vyplýva, že niektoré implikované interakcie zachytí už spojená informácia rádu nižšieho, než je rád týchto interakcií. Týmto sa otvára otázka, či je takéto správanie funkcionálu vhodné z interpretačného hľadiska. Na túto otázku sa nebudeme pokúšať poskytnúť vyčerpávajúcu odpoveď, uvedieme však dôsledok, ktorý je výrazný pri použití funkcionálov v praxi.

Jedným z hlavných praktických cieľov výstavby celého matematického aparátu popisovaného v práci je určiť, aký rád interakcií nám stačí skúmať na to, aby sme pozorovaný systém popísali dostatočne presne. Inak povedané, aký rád väzbových podmienok musíme poznať, aby bola naša vedomosť o systéme dostatočná. V prípade tohto praktického použitia nám správanie funkcionálu popisované vyššie neprekáža, je pre nás naopak žiadúce, aby boli interakcie v systéme zachytené už pri malom ráde zadaných väzbových podmienok.

Poznámka 3.31. Videli sme, že zachovanie ostrých nerovností vo vete je pre nás síce zaujímavé, ale vyžaduje si dodatočné predpoklady, ktoré nemusia byť splnené pri skúmaní dát popisujúcich systém, ktorý pozorujeme.

Napriek tomu, že sme popísali, ako určiť rád interakcií v systéme priamo zo zadaných marginálnych rozdelení, zistili sme tiež, že v praxi nie je takéto overovanie existencie interakcií jednoduché. Mohlo by nás preto zaujímať, či sa dá rád interakcií určiť z vypočítaných maximálnych entropií. Odpoveď poskytne nasledujúca veta.

Veta 3.32 (Maximálna entropia pri marginálach a interakcie)**.** *Majme $n \in \mathbb{N}$, ďalej majme diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s konečným oborom hodnôt. Maximalizujme entropiu pri marginálnych rozdeleniach veličiny $\tilde{\boldsymbol{X}}$. Ak platí*

$$H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}})) < H(\tilde{P}^{(k-1)}(\tilde{\boldsymbol{X}})),$$

tak sa medzi zložkami $\tilde{\boldsymbol{X}}$ prejavujú interakcie rádu k.

Dôkaz. Ukážeme obmenu implikácie, t. j. ak sa interakcie v danom ráde neprejavujú, nastáva rovnosť maximálnych entropií príslušného rádu. Nech sa neprejavujú interakcie rádu k. Potom platí $\mathcal{P}_F^{k-1} \subset \mathcal{P}_F^{k}$. Z toho už plynie $H(\tilde{P}^{(k-1)}(\tilde{\boldsymbol{X}})) \leq H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$. Z monotónie vyplýva opačná nerovnosť, celkovo teda musí platiť rovnosť. □

Dôsledok 3.33 (Spojená informácia pri marginálach a interakcie)**.** *Majme $n \in \mathbb{N}$, ďalej majme diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s konečným oborom hodnôt, maximalizujme entropiu pri marginálnych rozdeleniach veličiny $\tilde{\boldsymbol{X}}$. Ak platí*

$$I_c^{(k)}(\tilde{\boldsymbol{X}}) > 0,$$

tak sa medzi zložkami $\tilde{\boldsymbol{X}}$ prejavujú interakcie rádu k.

Dôkaz. Tvrdenie vyplýva priamo z vety 3.32. □

Veta 3.32 a jej dôsledok sú zároveň istou formou obrátenej implikácie vo vetách, 3.24, 3.27 a v ich dôsledkoch. Ako sme už upozornili v poznámke 3.31, bez dodatočných predpokladov nemôžeme implikáciu vo vete 3.32 obrátiť.

3.2 Interakcie a entropie

Maximálna entropia pri zadaných entropiách bola v článku [2] pôvodne navrhnutá ako náhrada maximálnej entropie pri zadaných marginálach. Entropie totiž podobne ako marginály umožňujú zachytávať akékoľvek, aj nelineárne vzťahy medzi zložkami. Momenty naopak neumožňujú tak presné zachytávanie nelineárnych interakcií. Preto nás zaujíma, či maximálna entropia pri zadaných entropiách má naozaj podobné vlastnosti, ako maximálna entropia pri zadaných marginálach, resp. v čom sa konštrukcie líšia.

V ďalšom postupe by sme chceli vysloviť podobné tvrdenia, ako sme uviedli pri skúmaní interakcií pre entropie a spojené informácie maximalizované pri zadaných marginálach, aj pre entropie a spojené informácie maximalizované pri zadaných entropiách. Potrebujeme preto preskúmať, aké sú vzťahy medzi maximálnymi entropiami rôznych rádov tohto typu, ak sa v systéme vyskytujú interakcie. Budeme postupovať podobne, ako sme postupovali v prípade zadaných marginálnych rozdelení. Prvou otázkou, na ktorú budeme hľadať odpoveď je, či sme schopní určiť prítomnosť interakcií a to, či sú implikované alebo neimplikované, len na základe nameraných entropií do určitého rádu.

Poznámka 3.34 (Tvar entropií interagujúceho systému)**.** V tejto poznámke určíme, ako vyzerajú namerané entropie, ak vieme, aké interakcie sa v systéme vyskytujú. Majme diskrétnu vektorovú náhodnú veličinu $\boldsymbol{X} = (X_1, \ldots, X_n) \sim p(\boldsymbol{x})$. Nech sa medzi jej zložkami prejavujú interakcie rádov $k_1, \ldots, k_m$, kde $k_1 < \cdots < k_m$, $m \leq n$. Budeme skúmať hodnoty jej entropií. Vyberme i z množiny $\widehat{m}$ a preskúmajme tvar nameraných entropií rádu

k_i. To, že sa v systéme prejavujú interakcie rádu k_i znamená, že existuje marginálne rozdelenie stupňa k_i, ktoré sa nedá faktorizovať na súčin aspoň dvoch marginálnych rozdelení niektorého nižšieho rádu. Nech bez ujmy na všeobecnosti je toto nedeliteľné rozdelenie rozdelením prvých k_i zložiek veličiny $\boldsymbol{X}$, t. j. $p(x_1, \ldots, x_{k_i})$ sa nedá faktorizovať. Vieme, že entropia tohto rozdelenia je ostre menšia, než entropia rozdelenia prvých k_i zložiek, ktoré je v tvare súčinu $p_1 \cdot p_2$, kde p_1 je rozdelenie niektorých l zložiek vybraných z prvých k_i zložiek pre nejaké $l \in \widehat{k_i - 1}$. Platnosť tohto tvrdenia sme ukázali v leme 3.19. Vidíme, že interakcia rádu k_i medzi prvými k_i zložkami náhodnej veličiny sa na nameraných entropiách do rádu k_i prejaví tak, že $H(X_1, \ldots, X_{k_i})$ bude ostre menšia, než súčet $H(p_1) + H(p_2)$ pre všetky možné rozdelenia p_1 a p_2. Zovšeobecnenie na iných vybraných k_i zložiek je priamočiare, ide len o zámenu indexov. Celkovo teda platí, že ak sa v systéme prejavuje interakcia rádu k, tak existuje entropia vybraných k zložiek, ktorá sa nedá zapísať v tvare sumy $H(p_1) + H(p_2)$ pre žiadne p_1 a p_2 také, že $p_1 \cdot p_2$ je rozdelením týchto vybraných zložiek a obidve p_i sú marginály rádu aspoň 1.

Poznámka 3.35 (Určenie interakcií z entropií)**.** Ďalej nás zaujíma, či vieme z toho, aké sú hodnoty nameraných entropií určiť rády interakcií vyskytujúcich sa v systéme. Majme náhodnú veličinu $\boldsymbol{X} = (X_1, \ldots, X_n) \sim p(\boldsymbol{x})$ a majme namerané jej entropie do rádu k. Chceme ukázať že ak pre entropiu niektorých vybraných k zložiek platí, že sa nedá zapísať v tvare $H(p_1) + H(p_2)$ pre žiadne p_1, p_2 také, že $p_1 \cdot p_2$ je rozdelením týchto vybraných zložiek a obe p_i sú marginály rádu aspoň 1, tak v systéme sú interakcie rádu k. Stačí si uvedomiť, že ak sa daná entropia nedá zapísať v tvare sumy, tak to implikuje, že rozdelenie týchto vybraných k zložiek $p(x_{i_1}, \ldots, x_{i_k})$ sa nedá faktorizovať na súčin aspoň dvoch rozdelení nižších rádov, z čoho vyplýva existencia interakcii rádu k. Ak by sa totiž dalo faktorizovať, teda bez ujmy na všeobecnosti, nech je faktorizovaný tvar $p(x_{i_1}, \ldots, x_{i_k}) = p(x_{i_1}, \ldots, x_{i_s}) p(x_{i_{s+1}}, \ldots, x_{i_k}) =: p_1 \cdot p_2$ pre nejaké $s \in \widehat{k - 1}$, potom $H(X_{i_1}, \ldots, X_{i_k}) = H(p(x_{i_1}, \ldots, x_{i_k})) = H(p_1) + H(p_2)$, čo je spor.

Z uvedeného vyplýva, že sme schopní identifikovať existenciu určitého rádu interakcií, ak máme k dispozícii namerané entropie do tohto rádu. Sme tiež schopní určiť, medzi ktorými zložkami táto interakcia nastáva, z toho tiež vieme určiť, či interakcie nami skúmaného rádu implikujú interakcie vyšších rádov, alebo nie. Pozorovania uvedené v poznámkach 3.34 a 3.35 môžeme zhrnúť do nasledujúceho pomocného tvrdenia.

Lema 3.36 (Interakcie a entropie)**.** *Medzi zložkami diskrétnej vektorovej náhodnej veličiny* $\boldsymbol{X}$ *sa vyskytujú interakcie rádu k práve vtedy, keď existuje vybraných k zložiek takých, že* *ich entropia sa nedá zapísať ako súčet $H(p_1) + H(p_2)$ pre žiadne p_1, p_2 také, že $p_1 \cdot p_2$ je rozdelením týchto vybraných zložiek a obe p_i sú marginály rádu aspoň 1.*

Dôkaz. To, že z prítomnosti interakcií v systéme vyplýva existencia entropie, ktorá sa nedá zapísať v tvare uvedenej sumy, sme ukázali v poznámke 3.34. Opačná implikácia bola ukázaná v poznámke 3.35. □

Poznámka 3.37. Upozornime, že toto tvrdenie nám síce poskytuje spôsob, ako interakcie nájsť priamo z entropií, ale overenie toho, či sa naozaj entropia nedá zapísať v tvare uvedenej sumy je pomerne náročné na počet krokov výpočtu, podmienka nie je triviálna, nie je ju ľahko vidieť z nameraných dát.

Ďalším krokom nášho postupu bude nájdenie všeobecného tvaru rozdelenia konzistentného so zadanými entropiami určitého rádu a určenie tvaru maximálne entropického rozdelenia pri daných entropiách.

Poznámka 3.38 (Rozdelenie pri daných entropiách)**.** Majme $\boldsymbol{X} = (X_1, \ldots, X_n)$, majme zadané entropie do rádu k vrátane. Na základe doterajšieho skúmania vieme zo zadaných entropií identifikovať, ktoré zložky $\boldsymbol{X}$ spolu interagujú. Môžeme teda postupovať podobne, ako v prípade hľadania rozdelenia konzistentného so zadanými marginálami. Najprv si rozdelíme zložky do disjunktných skupín podľa toho, ktoré zložky medzi sebou interagujú. Prvý možný tvar rozdelení konzistentných s danými entropiami je súčin marginálnych rozdelení skupín interagujúcich zložiek. Označme množinu rozdelení konzistentných s danými entropiami, ktoré sa dajú zapísať v takomto tvare, ako $\mathcal{P}_{\tilde{F}}$. Druhý možný tvar sú rozdelenia, v ktorých má niekoľko skupín interagujúcich zložiek spoločné marginálne rozdelenie a to nie je rovnaké, ako by vzniklo súčinom týchto skupín zložiek. Označme množinu rozdelení, ktoré sú konzistentné so zadanými entropiami, ale nedajú sa zapísať v tvare súčinu marginál jednotlivých skupín, len v tvare takom, že niektoré skupiny sú predstavované spoločným marginálnym rozdelením, ako $\mathcal{P}_{\tilde{I}}$. Všetky možné rozdelenia konzistentné so zadanými entropiami teda môžeme rozdeliť do dvoch skupín, $\mathcal{P}_{\tilde{F}}$ a $\mathcal{P}_{\tilde{I}}$.

Poznámka 3.39. Poznamenajme, že ak maximalizujeme entropiu pri zadaných marginálach a tieto marginály neimplikujú interakcie žiadneho vyššieho rádu, tak vieme, že množina rozdelení $\mathcal{P}_F$ je jednoprvková. V prípade zadaných entropií to tak obvykle nie je. Rozdelenia v množine $\mathcal{P}_{\tilde{F}}$ sú v tvare súčinu marginál, poznáme len entropiu každého rozdelenia v súčine, ale nepoznáme jeho presný tvar. Z tohto ale vyplýva iná užitočná vlastnosť, že ak sme v situácii, že sa neimplikujú interakcie vyššieho rádu, tak všetky rozdelenia v tejto množine majú rovnakú entropiu, tá je daná súčtom entropií jednotlivých marginál.

Poznámka 3.40 (Maximálne entropické rozdelenie pri entropiách)**.** V poznámke 3.38 sme našli všeobecný tvar rozdelenia konzistentného so zadanými entropiami do rádu k vrátane. Podobne ako pri maximalizácii entropie pri zadaných marginálach nás zaujíma, ako z týchto rozdelení vybrať to, ktoré má maximálnu entropiu. Majme $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$ s daným konečným oborom hodnôt, diskrétnu náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Maximalizujeme entropiu pri entropiách $\tilde{\boldsymbol{X}}$ do rádu k vrátane. Vieme, že maximálna entropia sa dosahuje na niektorom z rozdelení. Všetky rozdelenia pre náhodné veličiny z množiny $\{\boldsymbol{X}\}$ konzistentné so zadanými entropiami rozdelíme do skupín $\mathcal{P}_{\tilde{F}}$ a $\mathcal{P}_{\tilde{I}}$. Podobne ako v prípade maximalizácie pri zadaných marginálach vieme, že ku každému rozdeleniu v množine $\mathcal{P}_{\tilde{I}}$ existuje v množine $\mathcal{P}_{\tilde{F}}$ rozdelenie, ktoré má ostre menšiu entropiu. Maximálne entropické rozdelenie teda stačí hľadať medzi rozdeleniami v množine $\mathcal{P}_{\tilde{F}}$.

Poznámka 3.41. Na množine $\mathcal{P}_{\tilde{F}}$ musíme vo všeobecnosti maximum hľadať optimalizačnou alebo inou metódou na hľadanie maxima, táto úloha už nie je triviálna. Náš postup však umožňuje rozdeliť hľadanie maxima na rozdelení pre n zložiek na viacero úloh na hľadanie maxima na rozdelení menšieho počtu zložiek. Pri zadaných marginálach sme urobili podobné zjednodušenie, avšak vzhľadom na linearitu väzieb zadaných v podobe marginálnych rozdelení nám neprinášalo významnejšiu výpočetnú výhodu. V prípade zadávaných entropií, kde je úloha na hľadanie maxima vo všeobecnosti ťažko riešiteľná kvôli nelinearite väzbových podmienok, nám tento postup môže zjednodušiť hľadanie maxima.

Preskúmali sme, ako súvisia tvary zadávaných entropií s existenciou interakcií, tiež sme našli všeobecný tvar maximálne entropického rozdelenia konzistentného s danými entropiami. Vďaka týmto poznatkom môžeme vysloviť a dokázať tvrdenie o súvislosti hodnôt maximálnej entropie a existencie interakcií v systéme. Najprv dokážeme vetu, v ktorej znení sú neostré nerovnosti medzi maximálnymi entropiami, potom podobne ako pri maximalizácii pri zadaných marginálach skúsime nájsť predpoklady, pri ktorých by sme mohli ukázať dokonca ostré nerovnosti.

Veta 3.42 (O $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$, implikované interakcie)**.** *Majme $n \in \mathbb{N}$, množinu náhodných veličín s daným konečným oborom hodnôt $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$. Ďalej majme diskrétnu* *vektorovú náhodnú veličinu* $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ *s rovnakým oborom hodnôt. Nech sa v rozdelení* $\tilde{\boldsymbol{X}}$ *prejavujú interakcie rádov* $k_1, \ldots, k_m$*, kde* $k_1 < \cdots < k_m$*,* $m \leq n$*, môžu byť aj implikované. Maximalizujme entropiu pri entropiách veličiny* $\tilde{\boldsymbol{X}}$*. Potom platia neostré nerovnosti*

$$H(\tilde{P}^{[k_i]}(\tilde{\boldsymbol{X}})) \leq H(\tilde{P}^{[k_i-1]}(\tilde{\boldsymbol{X}})) \qquad \forall i \in \widehat{m} \tag{3.11}$$

$$H(\tilde{P}^{[k_i]}(\tilde{\boldsymbol{X}})) \geq H(\tilde{P}^{[k_{i+1}]}(\tilde{\boldsymbol{X}})) \qquad \forall i \in \widehat{m-1}. \tag{3.12}$$

Ďalej platia rovnosti

$$H(\tilde{P}^{[k_i]}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{[k_i+1]}(\tilde{\boldsymbol{X}})) = \cdots = H(\tilde{P}^{[k_{i+1}-1]}(\tilde{\boldsymbol{X}})) \qquad \forall i \in \widehat{m} \tag{3.13}$$

$$H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) = \cdots = H(\tilde{P}^{[k_1-1]}(\tilde{\boldsymbol{X}})) = \sum_{j=1}^{n} H(\tilde{X}_j). \tag{3.14}$$

Dôkaz. Nerovnosti (3.11) a (3.12) vyplývajú z vety o monotónii maximálnej entropie pri zadaných entropiách 2.34. Ukážme rovnosť (3.14). Majme $s \in \widehat{k_1 - 1}$. Vďaka tomu, že najmenší rád interakcií prejavujúci sa medzi zložkami $\tilde{\boldsymbol{X}}$ je k_1 vieme, že všeobecný tvar maximálne entropického rozdelenia pre veličiny z množiny $\{\boldsymbol{X}\}$ konzistentného s entropiami $\tilde{\boldsymbol{X}}$ do rádu s je $p(\boldsymbol{x}) = p(x_1) \cdots p(x_n)$. Vieme, že všetky rozdelenia takéhoto tvaru majú entropiu $H(p(\boldsymbol{x})) = H(p(x_1) \cdots p(x_n)) = \sum_{j=1}^{n} H(p(x_j)) = \sum_{j=1}^{n} H(\tilde{X}_j)$. Týmto sme dokázali vzťah (3.14).

Ukážme vzťah (3.13). Zvoľme $i \in \widehat{m}$. V ďalšej argumentácii budeme postupovať podobne, ako v dôkaze vety 3.24. Označme množinu rozdelení $\mathcal{P}_{\tilde{F}}$ z poznámky 3.38 pri zadaných marginálach rádu k ako $\mathcal{P}_{\tilde{F}}^{k}$. Kvôli tomu, že sa medzi rádmi k_i a k_{i+1} nevyskytujú žiadne nové interakcie platí, že rozdelenia v množine $\mathcal{P}_{\tilde{F}}^{k_i}$ ležia aj v množinách $\mathcal{P}_{\tilde{F}}^{k_i+1}, \ldots, \mathcal{P}_{\tilde{F}}^{k_{i+1}-1}$. Máme teda inklúziu $\mathcal{P}_{\tilde{F}}^{k_i} \subset \mathcal{P}_{\tilde{F}}^{j}$ pre všetky $j \in \{k_i + 1, \ldots, k_{i+1} - 1\}$. Na týchto množinách nájdeme suprémum entropie, z lemy 2.19 vyplýva $H(\tilde{P}^{[k_i]}(\tilde{\boldsymbol{X}})) \leq H(\tilde{P}^{[j]}(\tilde{\boldsymbol{X}}))$. Z monotónie maximálnej entropie vyplýva opačná nerovnosť $H(\tilde{P}^{[k_i]}(\tilde{\boldsymbol{X}})) \geq H(\tilde{P}^{[j]}(\tilde{\boldsymbol{X}}))$ pre všetky $j \in \{k_i+1, \ldots, k_{i+1}-1\}$. Musí teda platiť $H(\tilde{P}^{[k_i]}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{[j]}(\tilde{\boldsymbol{X}}))$ pre všetky $j \in \{k_i + 1, \ldots, k_{i+1} - 1\}$, čo dokazuje platnosť rovnosti (3.13). □

Poznámka 3.43. Môžeme si všimnúť, že vzťah (3.14) je zovšeobecnením nutnej podmienky z vety 2.38, podobne vzťah (3.4) je zovšeobecnením nutnej podmienky z vety 2.28.

Dôsledok 3.44 (O $I_c^{[k]}(\tilde{\boldsymbol{X}})$, implikované interakcie)**.** *Majme $n \in \mathbb{N}$, množinu náhodných veličín s daným konečným oborom hodnôt $\{\boldsymbol{X}\} = \{(X_1, \dots, X_n)\}$. Majme ďalej diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \dots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Nech sa v rozdelení $\tilde{\boldsymbol{X}}$ prejavujú interakcie rádov $k_1, \dots, k_m$, kde $k_1 < \dots < k_m$, $m \leq n$, môžu byť aj implikované. Maximalizujme entropiu pri entropiách veličiny $\tilde{\boldsymbol{X}}$. Potom platí neostrá nerovnosť*

$$I_c^{[k_i]}(\tilde{\boldsymbol{X}}) \geq 0 \qquad \forall i \in \widehat{m} \tag{3.15}$$

a spojené informácie všetkých ostatných rádov sú nulové.

Dôkaz. Nerovnosť (3.15) vyplýva zo vzťahu (3.11), resp. aj z vety 2.57. Nulovosť všetkých ostatných spojených informácii vyplýva z (3.13). □

Ďalej vyslovíme a dokážeme tvrdenie, v ktorom predpoklady upravíme tak, aby platili dokonca ostré nerovnosti medzi maximálnymi entropiami. Opäť pre prehľadnosť vo vete ponechávame aj tvrdenie o rovnostiach, aj keď je len priamym dôsledkom vety 3.42.

Veta 3.45 (O $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$, neimplikované interakcie)**.** *Majme $n \in \mathbb{N}$, množinu náhodných veličín s daným konečným oborom hodnôt $\{\boldsymbol{X}\} = \{(X_1, \dots, X_n)\}$. Majme ďalej diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \dots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Nech sa v rozdelení $\tilde{\boldsymbol{X}}$ prejavujú len neimplikované interakcie rádov $k_1, \dots, k_m$, kde $k_1 < \dots < k_m$, $m \leq n$, nech sa neprejavujú žiadne implikované interakcie. Maximalizujme entropiu pri entropiách veličiny $\tilde{\boldsymbol{X}}$. Potom platia ostré nerovnosti*

$$H(\tilde{P}^{[k_i]}(\tilde{\boldsymbol{X}})) < H(\tilde{P}^{[k_i-1]}(\tilde{\boldsymbol{X}})) \qquad \forall i \in \widehat{m} \tag{3.16}$$

$$H(\tilde{P}^{[k_i]}(\tilde{\boldsymbol{X}})) > H(\tilde{P}^{[k_{i+1}]}(\tilde{\boldsymbol{X}})) \qquad \forall i \in \widehat{m-1}. \tag{3.17}$$

Ďalej platia rovnosti

$$H(\tilde{P}^{[k_i]}(\tilde{\boldsymbol{X}})) = H(\tilde{P}^{[k_i+1]}(\tilde{\boldsymbol{X}})) = \dots = H(\tilde{P}^{[k_{i+1}-1]}(\tilde{\boldsymbol{X}})) \qquad \forall i \in \widehat{m} \tag{3.18}$$

$$H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}})) = \dots = H(\tilde{P}^{[k_1-1]}(\tilde{\boldsymbol{X}})) = \sum_{j=1}^{n} H(\tilde{X}_j). \tag{3.19}$$

Dôkaz. Pracujeme s náhodnou veličinou s konečným oborom hodnôt, maximálna entropia sa bude dosahovať na niektorom z rozdelení, v celom dôkaze budeme tento poznatok využívať. Budeme postupovať podobne, ako v dôkaze vety 3.27, niektoré fakty preto už nebudeme tak detailne komentovať. Vzťah (3.18) a (3.19) sme už ukázali vo vete 3.42, jej predpoklady sú aj v tomto prípade splnené. Ukážme platnosť nerovnosti (3.16). Zvoľme $i \in \widehat{m}$. Vďaka tomu, že medzi zložkami $\tilde{\boldsymbol{X}}$ sa prejavujú len neimplikované interakcie vieme, že maximálne entropické rozdelenie konzistentné s danými entropiami $\tilde{\boldsymbol{X}}$ do rádu k_i, keď sa prejavujú interakcie rádu k_i, sa dá zapísať v tvare $p = p_1 \cdot \ldots \cdot p_l$ pre nejaké $l \in \widehat{n}$, kde p_j sa nedá faktorizovať $\forall j \in \widehat{l}$ a všetky p_j sú maximálne rádu k_i.

Z existencie interakcií rádu k_i vieme, že existuje faktor rádu k_i, bez ujmy na všeobecnosti, nech je to p_1. Uvažujme teraz rozdelenia, ktoré zachovajú entropie veličiny $\tilde{\boldsymbol{X}}$ až do rádu $k_i - 1$. Medzi nimi je iste aj vyššie zmienené rozdelenie $p = p_1 \cdot \ldots \cdot p_l$, toto rozdelenie spĺňa podmienky na entropie dokonca do rádu k_i. Naviac ukážeme, že medzi nimi musí byť aj rozdelenie q v tvare $q = q_{1_A} q_{1_B} p_2 \cdot \ldots \cdot p_l$, teda také, ktoré se zhoduje s p vo všetkých faktoroch $p_2 \ldots p_l$, len namiesto p_1 obsahuje faktor $q_1 = q_{1_A} q_{1_B}$, kde q_{1_A} a q_{1_B} sú nejaké marginály p_1. Vyjdeme z toho, že v rozdelení p sa prejavujú interakcie rádu k_i a sú neimplikované. Bez ujmy na všeobecnosti sme naviac vyššie zvolili, že tieto interakcie sa prejavujú v zložke p_1. Interakcie sú neimplikované, teda musí existovať rozdelenie q_1 konzistentné s marginálami p rádu $k_i - 1$, a teda aj s entropiami do rádu $k_i - 1$, ktoré sa dá faktorizovať ako $q_1 = q_{1_A} q_{1_B}$. Toto rozdelenie q_1 ďalej rozšírime na rozdelenie celej veličiny $\boldsymbol{X}$ prenásobením faktormi $p_2 \ldots p_{l_1}$. Takto vzniknuté rozdelenie je určite konzistentné s marginálami p do rádu $k_i - 1$, teda sú splnené požiadavky lemy 3.19, odtiaľ dostávame, že $q = q_{1_A} q_{1_B} p_2 \cdot \ldots \cdot p_l$ má ostre väčšiu entropiu, než rozdelenie p. Platí, že $H(\tilde{P}^{[k_i - 1]}(\tilde{\boldsymbol{X}}))$ je väčšia alebo rovná entropii $q = q_{1_A} q_{1_B} p_2 \cdot \ldots \cdot p_l$, a teda ostre väčšia, ako entropia p.

Nerovnosť (3.17) dostávame z už dokázaných vzťahov (3.16) a (3.18). □

Poznámka 3.46. Podobne ako v prípade vety 3.27, ktorú sme ilustrovali grafom 3.1, aj tu by sme mohli použiť podobnú ilustráciu. Vieme totiž presne určiť, kedy entropia ostro klesá a kedy zostáva konštantná vzhľadom k rádu zadaných entropií, pri ktorých maximalizujeme.

Dôsledok 3.47 (O $I_c^{[k]}(\tilde{\boldsymbol{X}})$, neimplikované interakcie)**.** *Majme $n \in \mathbb{N}$, množinu náhodných veličín s daným konečným oborom hodnôt $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$. Majme ďalej diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Nech sa v rozdelení $\tilde{\boldsymbol{X}}$ prejavujú len neimplikované interakcie rádov $k_1, \ldots, k_m$, kde $k_1 < \cdots < k_m$, $m \leq n$, nech sa neprejavujú žiadne implikované interakcie. Maximalizujme* *entropiu pri entropiách veličiny $\tilde{\boldsymbol{X}}$. Potom platí ostrá nerovnosť*

$$I_c^{[k_i]}(\tilde{\boldsymbol{X}}) > 0 \qquad \forall i \in \widehat{m} \tag{3.20}$$

a spojené informácie všetkých ostatných rádov sú nulové.

Dôkaz. Nerovnosť (3.20) vyplýva zo vzťahu (3.16), nulovosť všetkých ostatných spojených informácii vyplýva z (3.18). □

Poznámka 3.48 (O ostrých nerovnostiach)**.** Môže sa zdať, že dôkazy ostrých nerovností vo vetách o neimplikovaných interakciách pre entropiu maximalizovanú pri zadaných entropiách alebo pri zadaných marginálach, si vyžadujú príliš mnoho opatrnosti a predpoklady viet sú príliš prísne, takže v praxi nám moc nepomôžu. Zaujímavé však je, ako budeme diskutovať v poslednej sekcii tejto podkapitoly, že pre entropiu maximalizovanú pri zadávaných momentoch ostré nerovnosti neplatia ani pri podobne prísnych predpokladoch. Z pohľadu praxe teda vety odrážajú fakt, že sú druhy interakcií, ktoré maximálna entropia pri zadaných marginálach alebo entropiách vždy spoľahlivo identifikuje, ale maximálna entropia pri zadaných momentoch tieto interakcie nijakým spôsobom nezachytí. Pripomeňme, že práve toto pozorovanie je motiváciou článkov [1, 2, 3]. V článkoch autori okrem iného vyhodnocovali na konkrétnych dátach, aká časť informácie skúmaného systému je nesená interakciami určitého rádu. Výsledky ukázali, že spojená informácia pri daných momentoch nezachytila takmer žiadne interakcie, zatiaľ čo pomocou spojenej informácie pri entropiách autori boli interakcie schopní zachytiť. Ostré nerovnosti v nami ukázaných vetách podporujú tieto zistenia.

Poznámka 3.49 (Porovnanie vlastností funkcionálov)**.** V predchádzajúcej poznámke sme diskutovali nedostatočnosť zachytávania interakcií cez entropiu maximalizovanú pri zadaných momentoch. Ako sme už spomenuli v úvode sekcie, ďalšou motiváciou autorov článkov [1, 2] bolo ukázať, že namiesto toho, aby sme pri skúmaní interakcií systému používali maximálnu entropiu pri zadaných marginálnych rozdeleniach, ako je pôvodne navrhnuté v článku [3], môžeme používať maximálnu entropiu pri zadaných entropiách. Nami skúmané vety túto myšlienku podporujú, v prípade neimplikovaných interakcií sme dokonca ukázali, že spojené informácie pri zadaných marginálach a entropiách sú nenulové práve súčasne. Môžeme dokonca vysloviť silnejšie tvrdenie. Z toho, ako konštruujeme maximálne entropické rozdelenie pre veličiny s konečným oborom hodnôt je totiž vidieť, že pokiaľ sa v systéme popísanom veličinou $\tilde{\boldsymbol{X}}$ vyskytujú len neimplikované interakcie, tak maximálna entropia pri zadaných entropiách $\tilde{\boldsymbol{X}}$ má rovnakú hodnotu ako maximálna entropia pri zadaných marginálach $\tilde{\boldsymbol{X}}$, a to pre všetky možné rády maximálnej entropie.

Naším skúmaním sme teda zistili, že $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ a $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$ majú viaceré zaujímavé spoločné vlastnosti, čo podporuje myšlienku v praxi používať $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ namiesto $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$. Na konci kapitoly preskúmame ešte jednu zaujímavú spoločnú vlastnosť, a to možnosť faktorizovať úplne celú informáciu o systéme pomocou spojených informácii týchto dvoch typov.

Pre entropiu maximalizovanú pri zadávaných entropiách do určitého rádu sme formulovali a dokázali vetu 3.45, v ktorej tvrdení sa vyskytuje ostrá nerovnosť a vetu 3.42, v ktorej tvrdení sa vyskytuje neostrá nerovnosť. Prípadné možné zoslabenia predpokladov vety 3.42 také, aby stále ešte platila ostrá nerovnosť ponechávame na budúce skúmanie. Upozornime ešte na to, že v prípade entropií maximalizovaných pri zadaných entropiách môže nastať podobný nesúlad medzi rádom nenulovej spojenej informácie a rádom interakcií v systéme ako nastal v prípade maximalizácie pri zadaných marginálach, konkrétne tým myslíme, že sa môže stať, že niektoré implikované interakcie vyšších rádov zachytí už spojená informácia rádu nižšieho. V prípade entropií maximalizovaných pri zadaných marginálach sme tento problém a jeho praktické dopady diskutovali v poznámke 3.30.

V úvode tejto sekcie sme našli spôsob, ako určiť, či sa v systéme vyskytujú interakcie, len na základe nameraných entropií. Tento spôsob bol podobne ako v prípade zadaných marginálnych rozdelení netriviálny. Vyslovme preto ešte vetu, ktorá nám umožní určiť, či sa v systéme nachádzajú interakcie na základe vypočítanej maximálnej entropie pri zadaných entropiách.

Veta 3.50 (Maximálna entropia pri entropiách a interakcie)**.** *Majme $n \in \mathbb{N}$, ďalej majme diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \dots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s konečným oborom hodnôt, maximalizujme entropiu pri entropiách veličiny $\tilde{\boldsymbol{X}}$. Ak platí*

$$H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})) < H(\tilde{P}^{[k-1]}(\tilde{\boldsymbol{X}})),$$

tak sa medzi zložkami $\tilde{\boldsymbol{X}}$ prejavujú interakcie rádu k.

Dôkaz. Ukázali by sme obmenu implikácie, postupovali by sme analogicky dôkazu vety 3.32. □

Dôsledok 3.51 (Spojená informácia pri entropiách a interakcie)**.** *Majme $n \in \mathbb{N}$, ďalej majme diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \dots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s konečným obo-*

rom hodnôt, maximalizujme entropiu pri entropiách veličiny $\tilde{\boldsymbol{X}}$. Ak platí

$$I_c^{[k]}(\tilde{\boldsymbol{X}}) > 0,$$

tak sa medzi zložkami $\tilde{\boldsymbol{X}}$ prejavujú interakcie rádu k.

Dôkaz. Tvrdenie vyplýva priamo z vety 3.50. □

3.3 Interakcie a momenty

V predchádzajúcej diskusii boli spomenuté niektoré odlišnosti maximálnej entropie pri zadaných momentoch od ostatných typov maximálnej entropie. Uveďme teraz konkrétne, na aké rozdiely by sme prišli, keby sme sa snažili vysloviť a dokázať vety podobné tým v predchádzajúcej kapitole.

Poznámka 3.52 (Rozdelenie pri daných momentoch)**.** Chceli by sme podobne ako v predchádzajúcich sekciách nájsť všeobecný tvar rozdelenia $\boldsymbol{X}$ konzistentného so zadanými momentmi do určitého rádu. Konštrukciu takéhoto rozdelenia pri zadaných marginálach resp. entropiách sme začínali rozdelením zložiek náhodnej veličiny na skupiny podľa prítomných interakcií. Pokiaľ by sme chceli postupovať rovnako, už v tomto kroku sa objaví prvý problém. Momenty nemusia zachytávať všetky interakcie, čo vieme napríklad zo známeho tvrdenia - nekorelovanosť neimplikuje nezávislosť. Nebudeme teda schopní rozdeliť zložky na interagujúce skupiny. Mohli by sme sa pokúsiť o rozdelenie zložiek $\boldsymbol{X}$ do skupín aspoň podľa vzájomných korelácií. Tu opäť narazíme na problém, nevieme totiž, ako z vyšších momentov určiť vlastnosť, ktorá by bola analogická korelácii, ale predstavovala by jej zovšeobecnenie na väčšie skupiny zložiek. Pretože nie je jasné, ako rozdelenie faktorizovať na súčin marginál, budeme vždy maximalizovať entropiu pri zadaných momentoch na celom rozdelení. V dôkazoch viet v predchádzajúcich sekciách sme pritom do veľkej miery využívali všeobecný tvar maximálne entropického rozdelenia a možnosť faktorizovať rozdelenie na súčin, nemôžeme teda postupovať podobným spôsobom.

Poznámka 3.53 (O nepresnom určení rádu interakcii)**.** Ďalším problémom, ktorý môže nastať pri maximalizácii pri zadaných momentoch je, že rád momentu, pri ktorom sa zmení hodnota maximálnej entropie a rád interakcie nemá tak presnú spojitosť, ako sme videli pri entropiách a marginálach. Popíšme, čo tým myslíme na konkrétnom príklade.

Majme n zložkovú diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}}$. Nech sa medzi zložkami $\tilde{\boldsymbol{X}}$ prejavujú interakcie druhého rádu také, že nie sme schopní tieto interakcie zachytiť druhým momentom, teda zložky sú nekorelované, ale nie po dvojiciach nezávislé. Nech sa neprejavujú interakcie vyšších rádov. Maximalizujme entropiu pri momentoch $\tilde{\boldsymbol{X}}$. Môže nastať situácia, keď napríklad maximálna entropia pri zadaných momentoch tretieho rádu je rovnaká, ako maximálna entropia rádu druhého, ale keď zadáme ďalšie väzbové podmienky, napríklad momenty štvrtého rádu, tak už maximálna entropia klesne. V tejto situácii bude $I_c^{<2>}(\tilde{\boldsymbol{X}})$ a $I_c^{<3>}(\tilde{\boldsymbol{X}})$ nulová, ale $I_c^{<4>}(\tilde{\boldsymbol{X}})$ bude nenulová. Rád nenulovej spojenej informácie teda nereflektuje rád interakcií v systéme. Nedochádza k zníženiu maximálnej entropie preto, že by pomocou ďalších podmienok bolo možné zachytiť interakcie medzi viacerými veličinami, ale len preto, že zvyšujeme počet nezávislých väzbových podmienok a tým zužujeme priestor, na ktorom hľadáme maximum. Detailný popis toho, kedy takáto situácia nastáva, prenecháme na budúce skúmanie.

Z uvedených pozorovaní vyplýva, že nemôžeme vysloviť podobné vety, ako v časti o interakciách a entropiách alebo o interakciách a marginálach. Upozornime tiež na to, že spojená informácia, ktorá vznikla maximalizáciou pri entropiách alebo marginálach, je v článkoch interpretovaná ako informácia nesená určitým rádom interakcií, čo do veľkej miery súhlasí s nami vyslovenými vetami. V prípade maximalizácie pri momentoch táto interpretácia nie je vhodná, pretože rád spojenej informácie neodráža rád interakcií, ako sme komentovali v predchádzajúcej poznámke. Ďalší dôvod, prečo nie je vhodné používať entropiu maximalizovanú pri momentoch a jej príslušnú spojenú informáciu na popisovanie interakcií je ten, že niektoré interakcie nezachytí žiadny rád zadaných väzbových podmienok a teda nie sme schopní rozložiť celú informáciu systému na časti príslušné určitým rádom interakcií. Faktorizáciu celkovej informácie systému bližšie popíšeme v nasledujúcej poznámke.

Poznámka 3.54 (Rozklad informácie systému)**.** Možnosť rozdelenia celkovej informácie systému na časti, ktoré sa budú dať vhodne interpretovať, je jednou z myšlienok, ktoré pôvodne viedli ku vzniku celého nami skúmaného matematického aparátu. Prvýkrát sa spomína v článku [3]. Tvrdenie je tam popísané na maximálnej entropii pri znalosti marginálnych rozdelení, v tomto znení ho preto najprv uvedieme aj my. Vezmime celkovú informáciu o systéme, ktorý má n prvkov a teda je popísaný n-rozmernou náhodnou veličinou $\tilde{\boldsymbol{X}} \sim \tilde{p}(\boldsymbol{x})$. Celková informácia je hodnota multiinformácie, ako sme ju definovali v 1.28, pripomeňme definičný vzťah

$$I_2(\tilde{X}_1; \ldots; \tilde{X}_n) := \sum_x \tilde{p}(\boldsymbol{x}) \log_2 \frac{\tilde{p}(\boldsymbol{x})}{\prod_{i=1}^n \tilde{p}_i(x_i)},$$

kde sčitujeme cez všetky $\boldsymbol{x} = (x_1, \ldots, x_n) \in \operatorname{supp} \boldsymbol{x}$ také, že $\tilde{p}_i(x_i)$ je kladné $\forall i \in \hat{n}$. Výraz môžeme prepísať ako

$$I_2(\tilde{X}_1; \ldots; \tilde{X}_n) = H\left(\prod_{i=1}^n \tilde{p}_i(x_i)\right) - H(\tilde{p}(\boldsymbol{x})).$$

Ďalej môžeme upraviť na tvar

$$I_2(\tilde{X}_1; \ldots; \tilde{X}_n) = H(\tilde{P}^{(1)}(\tilde{\boldsymbol{X}})) - H(\tilde{P}^{(n)}(\tilde{\boldsymbol{X}})).$$

Potom už z definície spojenej informácie vidno, že celkovú informáciu môžeme rozložiť na sumu spojených informácii

$$I_2(\tilde{X}_1; \ldots; \tilde{X}_n) = \sum_{k=2}^n I_c^{(k)}(\tilde{\boldsymbol{X}}).$$

Celkovú informáciu o systéme sme teda rozložili na príspevky od jednotlivých spojených informácii, čo opäť pripomína možnú interpretáciu jednotlivých členov ako istej formy množstva informácie prenášanej interakciami daného rádu. Rovnako dobre by sme mohli použiť na faktorizáciu aj maximálnu entropiu pri znalosti entropií nižšieho stupňa, rovnosť $H\left(\prod_{i=1}^n \tilde{p}_i(x_i)\right) = H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}}))$ a aj $H(\tilde{p}(\boldsymbol{x})) = H(\tilde{P}^{[n]}(\tilde{\boldsymbol{X}}))$ sme už diskutovali v 2.27 a vo vete 2.28. Upozornime ešte na to, že v prípade maximalizácie pri daných momentoch nemusí platiť, že $H\left(\prod_{i=1}^n \tilde{p}_i(x_i)\right) = H(\tilde{P}^{<1>}(\tilde{\boldsymbol{X}}))$ a ani $H(\tilde{p}(\boldsymbol{x})) = H(\tilde{P}^{<n>}(\tilde{\boldsymbol{X}}))$. Momenty dobre nezachytávajú všetky nelineárne interakcie medzi veličinami, teda nemôžeme celú informáciu o systéme rozložiť analogickým spôsobom, čo ilustrujeme na nasledujúcom jednoduchom príklade.

Príklad 3.55. Majme X s oborom hodnôt $\{0, 1, 3\}$. Nech sú zadané marginály prvého stupňa $p(0) = 0$, $p(1) = 1$, $p(3) = 0$. Platí $H\left(\prod_{i=1}^n \tilde{p}_i(x_i)\right) = H(\tilde{p}(x)) = 0$, prvá rovnosť vyplýva z toho, že veličina je jednozložková. Prvý moment, pri ktorom budeme maximalizovať má hodnotu $\operatorname{E} X = 1$. Rozdelenie $p(0) = 0.5, p(1) = 0.25, p(3) = 0.25$ má tiež $\operatorname{E} X = 1$, teda maximálna entropia pri zadanom prvom momente je určite väčšia alebo rovná, než entropia tohto rozdelenia, $H(\tilde{P}^{<1>}(X)) \geq 1.5$ bitu, čo sme videli aj v príklade

2.52. Z toho dostávame $H(\tilde{P}^{<1>}(X)) \neq H\left(\prod_{i=1}^{n} \tilde{p}_i(x_i)\right)$, tiež $H(\tilde{p}(x)) \neq H(\tilde{P}^{<n>}(X))$, keďže $n = 1$.

Poznámka 3.56. Videli sme, že prítomnosť určitého rádu interakcií v popisovanom systéme vieme zistiť už na základe zadaných maximalizačných podmienok v podobe entropií alebo marginálnych rozdelení a že tento spôsob určenia rádu interakcií je dokonca presnejší, než určenie rádov interakcií zo spojených informácii typu $I_c^{(k)}(\tilde{\boldsymbol{X}})$ alebo $I_c^{[k]}(\tilde{\boldsymbol{X}})$. Videli sme tiež, že tento spôsob je pomerne zdĺhavý. Spojená informácia nám v tomto smere nepomôže, nepočíta sa jednoducho, umožňuje však zistiť nielen prítomnosť interakcií určitého rádu v systéme, ale tiež určiť, akú časť celkovej informácie o systéme zachytíme pomocou spojenej informácie daného rádu. To nám umožňuje kvantifikovať, do akej miery je náš popis systému presný, ak máme k dispozícii len väzbové podmienky určitého rádu, čo je pre nás dôležité z praktického hľadiska. Je to tiež jednou z hlavných motivácii pre skúmanie spojených informácií, tejto téme je venovaná aj časť článku [2].

Touto poznámkou uzatvárame časť práce venovanú vlastnostiam maximálnej entropie a spojenej informácie prvých troch typov. Preskúmali sme tiež možnú interpretáciu niektorých viet. Tvrdenia o interakciách boli jedny z kľúčových tvrdení celej práce, otvárajú tiež ďalšie otázky a vytvárajú zaujímavé problémy do budúceho skúmania.

Kapitola 4

Informačné diagramy

V prvých dvoch kapitolách práce sme pripomenuli základné pojmy z teórie informácie a pravdepodobnosti, následne sme ich využili pri definovaní nových veličín. Ako sme už spomenuli v poznámke 2.18, v článkoch sa z výpočetných dôvodov používa ešte štvrtý typ maximálnej entropie a spojenej informácie. Na to, aby sme ho mohli korektne definovať a skúmať vlastnosti, potrebujeme poznať pojem informačný diagram a zaviesť na diagramoch informačnú znamienkovú mieru, v tejto kapitole sa preto budeme venovať práve týmto pojmom. Značenie, väčšinu tvrdení v tejto kapitole a spôsob výstavby potrebnej teórie, z veľkej časti preberáme z [8].

4.1 Základné pojmy

Značenie 4.1. V nasledujúcich kapitolách sa budeme stretávať s troma typmi objektov - $\tilde{X}$, $\mathring{X}$, X. Symboly X a $\tilde{X}$ (resp. $\boldsymbol{X}$ a $\tilde{\boldsymbol{X}}$) ponecháme podobne ako v predchádzajúcom texte vyhradené pre náhodné veličiny, symbol $\{\boldsymbol{X}\}$ bude reprezentovať množinu náhodných veličín s daným oborom hodnôt. Upozornime však, že značka $\mathring{X}$ bude predstavovať množinu generujúcu pole.

Definícia 4.2 (Pole $\mathcal{F}_n$)**.** Majme množiny $\mathring{X}_1, \ldots, \mathring{X}_n$ a označme $\Omega := \bigcup_{i=1}^{n} \mathring{X}_i$ univerzálnu množinu. Pole $\mathcal{F}_n$ generované množinami $\mathring{X}_1, \ldots, \mathring{X}_n$ je množina množín, ktoré môžeme získať aplikáciou postupnosti základných množinových operácii na $\mathring{X}_1, \ldots, \mathring{X}_n$. Za základné množinové operácie považujeme prienik, zjednotenie, doplnok a rozdiel množín. Množiny $\mathring{X}_1, \ldots, \mathring{X}_n$ nazývame generátory poľa $\mathcal{F}_n$.

Definícia 4.3 (Atómy poľa $\mathcal{F}_n$)**.** Atómy poľa $\mathcal{F}_n$ sú množiny v tvare $\bigcap_{i=1}^{n} Y_i$, kde Y_i je buď $\mathring{X}_i$ alebo $\mathring{X}_i^C$.

Definícia 4.4 (Znamienková miera)**.** Zobrazenie $\mu : \mathcal{F}_n \to \mathbb{R}$, pre ktoré platí $\mu(\emptyset) = 0$ a je aditívne, nazveme znamienkovou mierou na $\mathcal{F}_n$.

Poznámka 4.5. Narozdiel od miery môže znamienková miera dosahovať aj záporné hodnoty. Poznamenajme, že obvykle sa pri definícii miery, resp. znamienkovej miery požaduje σ-aditivita, v našom kontexte má však pole $\mathcal{F}_n$ len konečný počet podmnožín, preto stačí predpokladať aditivitu.

Poznámka 4.6. Z aditivity znamienkovej miery priamo vyplýva, že znamienková miera na $\mathcal{F}_n$ je jednoznačne určená hodnotami atómov poľa.

Poznámka 4.7. Graficky môžeme pole $\mathcal{F}_n$ generované množinami $\mathring{X}_1, \ldots, \mathring{X}_n$ znázorniť Vennovym diagramom.

4.2 Informačná znamienková miera a diagramy

V ďalšom postupe budeme konštruovať znamienkovú mieru na informačných diagramoch. Najprv objasníme význam pojmu informačný diagram, potom zadefinujeme príslušnú znamienkovú mieru a uvidíme, že definovaná funkcia naozaj nebude mierou klasickou, uvedieme konkrétny príklad, kde bude hodnota informačnej znamienkovej miery μ_I na jednom z atómov diagramu záporná.

Ako sme už naznačili v kapitole 1, vďaka vzťahom 1.25 môžeme stotožniť množiny Vennovho diagramu pre dve veličiny s určitými informačnými veličinami. Chceli by sme toto stotožnenie popísať korektne a zovšeobecniť, ukázať korešpondenciu medzi informačnými veličinami a množinami diagramu a to aj pre viaceré náhodné veličiny.

Majme n náhodných veličín $X_1, \ldots, X_n$. Ďalej majme množiny $\mathring{X}_1, \ldots, \mathring{X}_n$, ktoré generujú pole $\mathcal{F}_n$. Pri takomto značení môžeme $\mathring{X}_i$ chápať ako množinu reprezentujúcu náhodnú veličinu X_i. V definícii 4.2 sme zaviedli univerzálnu množinu $\Omega = \bigcup_{i=1}^{n} \mathring{X}_i$. Potom platí, že $\bigcap_{i=1}^{n} \mathring{X}_i^C = \emptyset$. Ostatné atómy poľa môžeme nazvať neprázdne atómy. Označme množinu všetkých neprázdnych atómov ako $\mathcal{A}$. Potom $|\mathcal{A}| = 2^n - 1$. Ako sme už spomínali v poznámke 4.6, znamienková miera na diagrame generovanom $\mathring{X}_1, \ldots, \mathring{X}_n$ je jednoznačne určená hodnotou atómov. My by sme však chceli definovať znamienkovú mieru cez jej hodnoty na celých množinách $\mathring{X}_i$ a na všetkých ich konečných zjednoteniach, nebudeme definovať hodnotu osobitne pre každý atóm. Vyslovme preto nasledujúcu vetu, v ktorej budeme používať zjednodušené značenie.

Značenie 4.8. Majme náhodné veličiny $X_1, \ldots, X_n$ a množiny $\mathring{X}_1, \ldots, \mathring{X}_n$. Ďalej majme indexovú množinu $\emptyset \neq N \subset \widehat{n}$. Budeme používať označenie $\mathring{\boldsymbol{X}}_N = \bigcup_{i \in N} \mathring{X}_i$ a ďalej $\boldsymbol{X}_N$ označíme vektorovú náhodnú veličinu $(X_{i_1}, \ldots, X_{i_k})$, kde $k = |N|$, $i_j \in N$ pre všetky $j \in \widehat{k}$ a pre indexy platí $i_1 < \cdots < i_k$.

Veta 4.9 (Určenie miery hodnotami)**.** *Majme $\mathcal{F}_n$ pole generované množinami $\mathring{X}_1, \ldots, \mathring{X}_n$. Označme $\mathcal{B} = \{\mathring{\boldsymbol{X}}_N \mid \emptyset \neq N \subset \widehat{n}\}$. Potom znamienková miera μ na $\mathcal{F}_n$ je jednoznačne určená zadaním hodnôt $\mu(B)$ pre všetky $B \in \mathcal{B}$.*

Dôkaz. Vetu ponechávame bez dôkazu, je možné ho nájsť napríklad v [8]. □

Na základe tejto vety teraz môžeme definovať informačnú znamienkovú mieru tým, že určíme jej hodnoty na všetkých $\mathring{\boldsymbol{X}}_N$.

Definícia 4.10 (Informačná znamienková miera)**.** Majme $\mathcal{F}_n$ pole generované množinami $\mathring{X}_1, \ldots, \mathring{X}_n$, ktoré reprezentujú vektorovú náhodnú veličinu $(X_1, \ldots, X_n)$. Informačnú znamienkovú mieru na $\mathcal{F}_n$, skrátene I-mieru, značíme μ_I, definujeme ako $\mu_I(\mathring{\boldsymbol{X}}_N) = H(\boldsymbol{X}_N)$ pre všetky $\emptyset \neq N \subset \widehat{n}$.

Poznámka 4.11. Môžeme si všimnúť, že pri takto definovanej znamienkovej miere by sme hodnotu μ_I mohli interpretovať ako entropiu alebo vzájomnú informáciu príslušnej množiny. Presnejšie povedané, pre dve náhodné veličiny X_1 a X_2 očakávame, že budeme môcť interpretovať I-mieru určitých častí diagramu ako $\mu_I(\mathring{X}_1 \cup \mathring{X}_2) = H(X_1, X_2)$, $\mu_I(\mathring{X}_1 \cap \mathring{X}_2) = I(X;Y)$, $\mu_I(\mathring{X}_1 - \mathring{X}_2) = H(X_1 \mid X_2)$, pretože to je v súlade so vzťahmi medzi informačnými veličinami, ako sme ich uviedli v 1.25, pri správnom stotožnení operácií na množinách s operáciami medzi náhodnými veličinami. Vzťahy medzi týmito dvoma druhmi operácii sú nasledujúce. Symbolicky reprezentujeme podmieňovanie ako rozdiel, čiarku ako zjednotenie, bodkočiarku ako prienik. Podobne by nás pre viacrozmerné vektorové náhodné veličiny mohla zaujímať platnosť analogických vzťahov $H(\boldsymbol{X}_N) = \mu_I(\cup_{i \in N} \mathring{X}_i)$, $I(\boldsymbol{X}_N; \boldsymbol{X}_M) = \mu_I(\mathring{\boldsymbol{X}}_N \cap \mathring{\boldsymbol{X}}_M)$, $H(\boldsymbol{X}_N \mid \boldsymbol{X}_M) = \mu_I(\mathring{\boldsymbol{X}}_N - \mathring{\boldsymbol{X}}_M)$, kde M, N sú neprázdne indexové množiny. Dalo by sa ukázať, že tieto vzťahy naozaj platia a I-miera zachytáva hodnotu určených informačných veličín. Vyslovme toto tvrdenie vo vete.

Veta 4.12 (Konzistencia)**.** *I-miera z definície 4.10 je jednoznačne určená znamienková miera na $\mathcal{F}_n$ a je konzistentná so všetkými informačnými veličinami.*

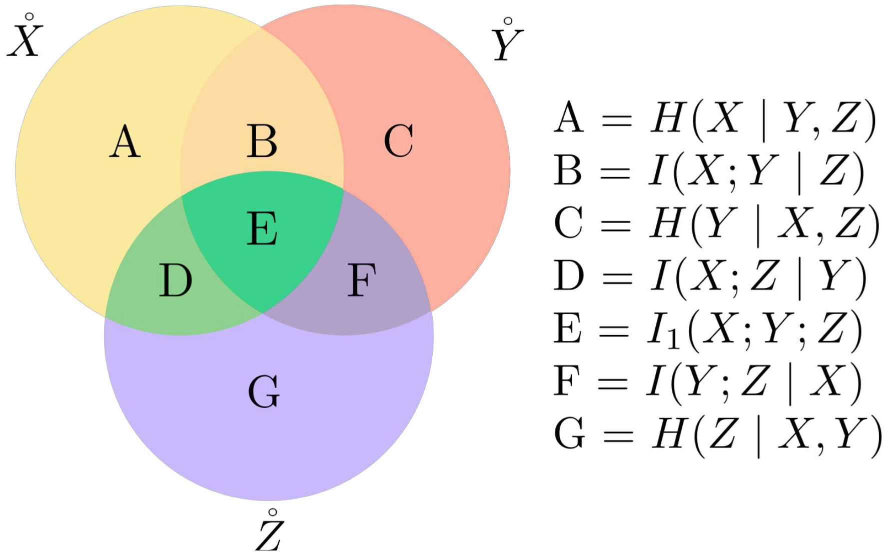
Dôkaz. Tvrdenie ponechávame bez dôkazu, opäť je možné ho nájsť v [8]. □

Poznámka 4.13. Konzistenciou myslíme platnosť vzťahov z poznámky 4.11

$$
\begin{aligned}
H(\boldsymbol{X}_N) &= \mu_I(\cup_{i\in N}\mathring{X}_i),\\
I(\boldsymbol{X}_N;\boldsymbol{X}_M) &= \mu_I(\mathring{\boldsymbol{X}}_N \cap \mathring{\boldsymbol{X}}_M),\\
H(\boldsymbol{X}_N \mid \boldsymbol{X}_M) &= \mu_I(\mathring{\boldsymbol{X}}_N - \mathring{\boldsymbol{X}}_M)
\end{aligned}
$$

pre M, N neprázdne indexové množiny.

Skonštruovali sme informačnú znamienkovú mieru a vieme presne, ako hodnotu I-miery určitého prvku poľa interpretovať v zmysle informačných veličín, korešpondencia je jednoznačná. Z toho vyplýva, že je oprávnené konštruovať Vennove diagramy pre množiny $\mathring{X}_1, \ldots, \mathring{X}_n$ reprezentujúce náhodné veličiny $X_1, \ldots, X_n$, v ktorých budeme jednotlivé časti diagramu označovať práve príslušnými informačnými funkcionálmi. Tieto diagramy budeme nazývať informačné. Hodnota každého z atómov informačného diagramu teda predstavuje hodnotu určitej informačnej veličiny pre $\boldsymbol{X} \sim P^{\boldsymbol{X}}$. Diagram pre dve náhodné veličiny sme ako motiváciu uviedli už v prvej kapitole, viď obrázok 1.1. Uveďme ešte jeden príklad informačného diagramu. Na obrázku 4.1 je diagram pre tri náhodné veličiny, resp. trojzložkovú vektorovú náhodnú veličinu.



Obr. 4.1: Informačný diagram pre tri náhodné veličiny, popísané sú len atómy diagramu, množiny $\mathring{X}, \mathring{Y}, \mathring{Z}$ reprezentujú náhodné veličiny X, Y, Z

Upozornime, že pri značení zavedenom v kapitole 1 je prienik troch množín diagramu reprezentovaný viacrozmernou vzájomnou informáciou, ktorú sme definovali v 1.27 a značíme ju $I_1(X;Y;Z)$. Táto veličina môže byť aj záporná, preto je diagram pre tri veličiny zároveň najjednoduchším príkladom diagramu, v ktorom sa môže vyskytnúť záporný atóm. Tento fakt ešte ilustrujeme na konkrétnom príklade v rámci poznámky 4.14. Diagramy pre štyri a viac náhodných veličín sa už nedajú jednoducho ilustrovať na ploche, obrázky by boli málo prehľadné, preto ich nebudeme uvádzať.

Uveďme ešte vzťahy pre ďalšie informačné veličiny, ktoré sú v diagrame na obrázku 4.1 reprezentované zjednoteniami viacerých atómov:

$$
\begin{aligned}
H(X) &= A + B + D + E & I(X;Y) &= B + E \\
H(Y) &= B + C + E + F & I(X;Z) &= D + E \\
H(Z) &= D + E + F + G & I(Y;Z) &= E + F \\
H(X,Y) &= A + B + C + D + E + F & H(X,Y \mid Z) &= A + B + C \\
H(X,Z) &= A + B + D + E + F + G & H(X,Z \mid Y) &= A + D + G \\
H(Y,Z) &= B + C + D + E + F + G & H(Y,Z \mid X) &= C + F + G \\
H(X,Y,Z) &= A + B + C + D + E + F + G
\end{aligned}
$$

Poznámka 4.14. Mohlo by nás zaujímať, či nami skonštruovaná znamienková miera nie je aj mierou, teda či vôbec môže nastať situácia, že je hodnota informačnej veličiny niektorej časti diagramu záporná. Ako sme prezradili už v úvode podkapitoly, naozaj taký príklad nájsť môžeme, preberáme ho z [8]. Je jasné, že pre dve veličiny budú všetky hodnoty I-miery na $\mathcal{F}_2$ nezáporné, pretože vzájomná informácia dvoch veličín a podmienená entropia sú vždy nezáporné. Majme teda $\mathcal{F}_3$. Nech X_1, X_2 sú binárne náhodné veličiny s rovnomerným rozdelením a nech $X_3 = X_1 + X_2 \bmod 2$, t. j. X_3 je XOR prvých dvoch zložiek. Potom hodnoty entropií sú $H(X_1) = H(X_2) = H(X_3) = 1$, $H(X_1, X_2) = H(X_1, X_3) = H(X_2, X_3) = 2$, $H(X_1, X_2, X_3) = 2$. Vzájomné informácie sú $I(X_1; X_2) = I(X_1; X_3) = I(X_2; X_3) = 0$. Ďalej vypočítame $I(X_1; X_2 \mid X_3) = 1$, potom $I_1(X_1; X_2; X_3) = I(X_1; X_2) - I(X_1; X_2 \mid X_3) = -1$. Vieme, že $I_1(X_1; X_2; X_3)$ je hodnotou atómu v strede informačného diagramu pre tri veličiny, našli sme teda atóm so zápornou hodnotou, z toho usudzujeme, že I-miera nie je miera v klasickom zmysle, ale miera znamienková.

Pre náš ďalší postup je dôležité odpovedať na otázku, aké kombinácie hodnôt môže I-miera na atómoch diagramu dosahovať. Chceme preskúmať nasledujúcu situáciu. Máme Vennov diagram pre n množín, ktorý má ku každej svojej časti priradené číselné hodnoty. Zaujíma nás, či je tento diagram informačný, teda či existujú náhodné veličiny také, že číselné hodnoty v diagrame sú hodnotou príslušného druhu informačných veličín vypočítaných z rozdelení pravdepodobnosti týchto náhodných veličín. Ak sme v situácii, že nemáme zadané podmienky na nosiče hustôt pravdepodobnosti X_i a pokiaľ sú všetky zadané hodnoty na Vennovom diagrame nezáporné, odpoveď nám poskytne nasledujúca veta, ktorú opäť preberáme z [8].

Veta 4.15 (Existencia diagramu pre nezáporné hodnoty)**.** *Majme diskrétnu vektorovú náhodnú veličinu* $(X_1, \ldots, X_n)$ *reprezentovanú množinami* $\mathring{X}_1, \ldots, \mathring{X}_n$*, pole* $\mathcal{F}_n$ *generované týmito množinami. Ak nemáme zadané žiadne obmedzenia na vlastnosti náhodných veličín* $X_1, \ldots, X_n$*, tak* μ_I *môže na* $\mathcal{F}_n$ *dosahovať ľubovoľne zvolené nezáporné hodnoty.*

Dôkaz. Dôkaz neuvádzame, je možné ho nájsť v [8]. □

Pokiaľ obmedzíme obory hodnôt náhodných veličín, tak môže nastať situácia, že diagram už nebude informačný pre dané veličiny. Inak povedané, nemusí existovať rozdelenie pravdepodobnosti, ktoré je rozdelením pre náhodné veličiny s daným oborom hodnôt a zároveň má príslušné hodnoty informačných veličín. Podobne, pokiaľ by boli niektoré zadané hodnoty záporné, diagram nemusí byť informačný. Toto pozorovanie ešte rozvinieme v nasledujúcej kapitole, budeme tiež skúmať niektoré dôsledky tohto faktu.

Na záver kapitoly zavedieme ešte niekoľko užitočných pojmov. Najprv definujeme číselný diagram. Tento pojem bude zahŕňať všetky možné diagramy, teda aj tie, ku ktorým nevieme nájsť príslušné rozdelenie pravdepodobnosti. Medzi diagramami ďalej vyberieme množinu tých, ktoré budú spĺňať takzvané Shannonove nerovnosti.

Definícia 4.16 (Číselný diagram)**.** Majme $n \in \mathbb{N}$, množiny $\mathring{X}_1, \ldots, \mathring{X}_n$ generujúce pole $\mathcal{F}_n$. Vennov diagram poľa $\mathcal{F}_n$ taký, že ku každému jeho atómu je priradené reálne číslo, nazveme číselný diagram poľa $\mathcal{F}_n$.

Poznámka 4.17. Poznamenajme, že každý číselný diagram zároveň predstavuje znamienkovú mieru nad poľom generovaným danými množinami.

Značenie 4.18 (Množiny diagramov)**.** Majme množinu náhodných veličín s daným oborom hodnôt $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$ a množinu množín $\{\mathring{\boldsymbol{X}}\} = \{\mathring{X}_1, \ldots, \mathring{X}_n\}$, pole $\mathcal{F}_n$ generované $\{\mathring{\boldsymbol{X}}\}$. Potom množinu všetkých informačných diagramov náhodných veličín z množiny $\{\boldsymbol{X}\}$ budeme značiť $\mathcal{D}(\boldsymbol{X})$, množinu všetkých číselných diagramov poľa $\mathcal{F}_n$ budeme značiť $\mathcal{D}_N(\mathring{\boldsymbol{X}})$.

Poznámka 4.19. Upozornime, že pre množinu množín $\{\mathring{X}_1, \ldots, \mathring{X}_n\}$ sme zaviedli skrátené označenie $\{\mathring{\boldsymbol{X}}\}$.

Poznámka 4.20. Zhrňme ešte raz vzťahy medzi vyššie popísanými pojmami. Pri našom značení a názvosloví pre každý informačný diagram existuje vektorová náhodná veličina s hodnotami informačných veličín takými, aké zachytáva diagram, a naopak, ku každej vektorovej náhodnej veličine s nejakým konkrétnym rozdelením pravdepodobnosti môžeme skonštruovať informačný diagram. K náhodnej veličine $\boldsymbol{X} = (X_1, \ldots, X_n)$ môžeme priradiť množiny $\mathring{\boldsymbol{X}} = (\mathring{X}_1, \ldots, \mathring{X}_n)$. Pri tomto priradení je potom informačný diagram pre určité rozdelenie pravdepodobnosti veličiny zároveň číselným diagramom poľa $\mathcal{F}_n$ generovaného $\{\mathring{\boldsymbol{X}}\}$, t. j. platí $\mathcal{D}(\boldsymbol{X}) \subset \mathcal{D}_N(\mathring{\boldsymbol{X}})$. Pokiaľ však máme len množiny $\{\mathring{\boldsymbol{X}}\} = \{\mathring{X}_1, \ldots, \mathring{X}_n\}$, vytvoríme nad nimi ľubovoľný číselný diagram, tak nemusí platiť, že k množinám vieme priradiť vektorovú náhodnú veličinu $\boldsymbol{X} \sim P^{\boldsymbol{X}}$ tak, aby čísla v diagrame reprezentovali hodnoty informačných veličín pre túto náhodnú veličinu. Poznamenajme, že v článkoch [1, 2] autori používajú iné názvoslovie. Diagram, ktorý my označujeme ako číselný, sa v článkoch vždy nazýva informačný a ďalej sa označuje prívlastkom dosiahnuteľný resp. nedosiahnuteľný, a to podľa toho, či reprezentuje nejakú náhodnú veličinu alebo nie.

V nasledujúcej kapitole budeme pracovať s diagramami, ktoré naviac spĺňajú Shannonove nerovnosti. Objasnime, ako tieto nerovnosti vyzerajú.

Definícia 4.21 (Elementárne Shannonove nerovnosti)**.** Majme $n \in \mathbb{N}$, diskrétnu vektorovú náhodnú veličinu $\boldsymbol{X} = (X_1, \ldots, X_n)$. Potom elementárne Shannonove nerovnosti sú vzťahy v tvare

$$H(X_i \mid \boldsymbol{X} \backslash X_i) \geq 0 \tag{4.1}$$

$$I(X_i; X_j \mid \boldsymbol{X}_M) \geq 0\,, \tag{4.2}$$

kde $i \neq j$ a $M \subset \{1, \ldots, n\} \backslash \{i, j\}$. Symbol $\boldsymbol{X} \backslash X_i$ reprezentuje náhodnú veličinu $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$.

Poznámka 4.22. Nerovnosti sme zapísali v rovnakom tvare, ako autori článkov [1, 2]. V literatúre sa pojmom Shannonove nerovnosti označujú aj ďalšie vzťahy, ktoré však môžu byť odvodené zo vzťahov uvedených v našej definícii. Poznamenajme, že Shannonove nerovnosti sú vzťahy platné pre každú diskrétnu vektorovú náhodnú veličinu.

Vďaka korešpondencii informačných veličín a atómov informačného diagramu sa dajú Shannonove nerovnosti vyjadriť ako nerovnosti pre hodnoty I-miery atómov diagramu.

Tieto nerovnosti medzi hodnotami atómov majú zmysel aj pre číselné diagramy. Sme teda schopní určiť, či daný číselný diagram spĺňa Shannonove nerovnosti. Zaveďme si označenie pre diagramy, ktoré tieto nerovnosti spĺňajú.

Značenie 4.23. Majme množinu množín $\{\mathring{\boldsymbol{X}}\} = \{\mathring{X}_1, \ldots, \mathring{X}_n\}$. Potom $\mathcal{S}_N(\mathring{\boldsymbol{X}})$ označíme podmnožinu $\mathcal{D}_N(\mathring{\boldsymbol{X}})$ obsahujúcu číselné diagramy, ktoré spĺňajú Shannonove nerovnosti.

Poznámka 4.24. Pre jednoduchosť vyjadrovania budeme diagramy, pre ktoré platia Shannonove nerovnosti, nazývať aj shannonovské diagramy. Tiež poznamenajme, že pre informačné diagramy nie je potrebné zavádzať nové značenie pre množinu tých informačných diagramov, ktoré spĺňajú Shannonove nerovnosti. Vďaka tomu, že informačný diagram je diagram nejakého rozdelenia pravdepodobnosti a I-miera atómov diagramu korešponduje s hodnotami určitých informačných veličín daného rozdelenia, sú Shannonove nerovnosti na informačných diagramoch automaticky splnené.

Týmto ukončujeme prehľad základných pojmov a značenia týkajúceho sa informačných diagramov, ktoré budeme používať v poslednej kapitole.

Kapitola 5

Veličiny maximalizované na diagramoch

V predchádzajúcej kapitole sme objasnili pojmy informačný diagram, číselný diagram a zaviedli sme informačnú znamienkovú mieru μ_I, čo nám umožňuje postupovať ďalej a definovať ďalší druh maximálnej entropie spomínaný v článkoch [1, 2]. Poznamenajme, že autori článkov nerozlišujú medzi maximálnou entropiou pri zadaných entropiách do rádu k ako sme ju my definovali v 2.15 a tou, ktorú budeme definovať v 5.4. V článkoch je navrhnutý postup maximalizácie entropie pomocou informačných diagramov, ktorý my stručne predstavíme v 5.1, autori však maximalizáciu na diagramoch prezentujú len ako výpočetnú schému pre entropiu maximalizovanú na rozdeleniach. Upozorňujú ale na to, že diagram, ktorý vzniká maximalizáciou, nemusí vždy byť informačný, teda nie vždy musí existovať rozdelenie pravdepodobnosti, ktoré by dosahovalo dané hodnoty informačných veličín. Z toho vyplýva, že hodnoty takto maximalizovanej entropie sa nie vždy musia zhodovať s hodnotami $H^{[k]}(\tilde{P}(\tilde{\boldsymbol{X}}))$. Pre účely definície a následného skúmania vlastností maximálnych entropií je preto vhodné oddeliť entropiu maximalizovanú na diagramoch a definovať ju ako osobitný typ maximálnej entropie. My tento typ maximálnej entropie preskúmame v rámci nasledujúcej podkapitoly, túto konštrukciu budeme označovať ako primárna konštrukcia. Okrem nej navrhneme ešte ďalší spôsob maximalizácie entropie na diagramoch, ktorému sa budeme venovať v rámci osobitnej podkapitoly o alternatívnej konštrukcii.

5.1 Primárna konštrukcia

Poznámka 5.1 (Postup maximalizácie)**.** Pri skúmaní vlastností maximálnych entropií sme niekoľkokrát narazili na problém praktického charakteru - ako vyčísliť maximálne entropie. Obzvlášť zložitá sa táto úloha javí pri zadaní entropií do rádu k ako maximalizačných podmienok. Zároveň je pre nás z praktického hľadiska veľmi užitočné vedieť tento problém vyriešiť. V článkoch [1, 2] je preto navrhnutý postup, ktorý umožňuje úlohu riešiť optimalizáciou I-miery atómov diagramu. Zadané optimalizačné väzby sú entropie do rádu k a Shannonove nerovnosti. Vďaka tomu, že môžeme stotožniť hodnoty entropií s I-mierou určitých atómov, väzby sú lineárne. Kľúčové je uvedomiť si, že diagram, ktorý vznikne po maximalizácii, už nemusí byť informačný a odôvodniť, prečo je to tak. Tvarom maximalizačných väzieb môžeme síce zaručiť platnosť Shannonovych nerovností (nezápornosť podmienenej entropie a vzájomnej informácie), nezaručíme však platnosť ďalších nerovností, označovaných ako neshannonovské nerovnosti, ktoré sú nutnou podmienkou toho, aby bol diagram nielen číselný, ale aj informačný. Teoreticky by sme mohli tieto podmienky pridať ako väzbové, avšak tvar všetkých neshannonovských podmienok je stále otvorenou otázkou. Preto my, spolu s autormi článkov, neshannonovské podmienky pri maximalizácii neuvažujeme, dostávame teda po maximalizácii diagram číselný, nie nutne informačný.

5.1.1 Definície

V úvode kapitoly sme vysvetlili motiváciu k maximalizácii entropie na informačných diagramoch. Našim cieľom je teraz definícia štvrtého typu maximálnej entropie a spojenej informácie. Budeme používať značenie množín podobné ako v kapitole 2, teda M bude predstavovať usporiadanú množinu hodnôt väzbových podmienok, N indexovú množinu, A bude množina číselných diagramov, ktoré spĺňajú Shannonove nerovnosti a väzbové podmienky. Množina N v tomto prípade obsahuje všetky možné usporiadané j-tice vybrané bez opakovania z množiny $\{1, ..., k\}$ pre všetky $j \in \widehat{k}$.

Skôr, než vyslovíme definíciu posledného typu maximálnej entropie, je potrebné si uvedomiť, že ešte nemáme definovanú I-mieru ako funkciu na diagramoch, ktoré nie sú informačné, ale sú len číselné. V definícii informačnej znamienkovej miery totiž požadujeme, aby množiny $\mathring{X}_1, \ldots, \mathring{X}_n$ reprezentovali vektorovú náhodnú veličinu $(X_1, \ldots, X_n)$, mieru vybraných častí diagramu potom identifikujeme s určitými entropiami náhodnej veličiny. V prípade, že diagram nie je informačný však neexistujú náhodné veličiny, ktorých reprezentáciou by boli množiny. Preto rozšírime pojem I-miera nasledujúcim spôsobom.

Poznámka 5.2 (Rozšírenie I-miery)**.** Majme $\{\mathring{\boldsymbol{X}}\} = \{\mathring{X}_1, ..., \mathring{X}_n\}$ množinu množín generujúcich pole $\mathcal{F}_n$, nad ktorým máme číselný diagram. Ďalej majme indexovú množinu

$\emptyset \neq N \subset \widehat{n}$, označme obvyklým spôsobom $\mathring{\boldsymbol{X}}_N = \bigcup_{i \in N} \mathring{X}_i$. Potom definujeme $\mu_I(\mathring{\boldsymbol{X}}_N)$ ako súčet hodnôt priradených atómom poľa $\mathcal{F}_n$ obsiahnutým v $\mathring{\boldsymbol{X}}_N$.

Poznamenajme, že takto rozšírená I-miera je konzistentná s definíciou 4.10, čím myslíme, že ak máme informačný diagram vektorovej náhodnej veličiny $(X_1, \ldots, X_n)$ ku ktorej prislúchajú množiny $\{\mathring{X}_1, ..., \mathring{X}_n\}$, tak platí $\mu_I(\mathring{\boldsymbol{X}}_N) = \mu_I(\boldsymbol{X}_N)$ pre všetky $\emptyset \neq N \subset \widehat{n}$.

Kvôli zjednodušeniu zápisu v definícii 5.4 ešte zavedieme pojem I-miera diagramu. Pokiaľ je diagram informačný, chceme týmto pojmom zachytiť celkovú entropiu rozdelenia danej vektorovej náhodnej veličiny. V prípade číselného diagramu I-miera číselného diagramu predstavuje súčet všetkých čísiel priradených jednotlivým atómom diagramu. Vďaka tomu, ako chápeme I-mieru na číselných diagramoch a vďaka inklúzii $\mathcal{D}(\boldsymbol{X}) \subset \mathcal{D}_N(\mathring{\boldsymbol{X}})$, stačí definovať pojem len pre číselné diagramy.

Definícia 5.3 (I-miera diagramu)**.** Majme $\{\mathring{\boldsymbol{X}}\} = \{\mathring{X}_1, ..., \mathring{X}_n\}$ množinu množín, ktorá generuje pole $\mathcal{F}_n$. Pre číselný diagram $D \in \mathcal{D}_N(\mathring{\boldsymbol{X}})$ definujeme I-mieru diagramu ako $\mu_I(D) := \mu_I\{\bigcup_{i \in \widehat{n}} \mathring{X}_i\}$.

Objasnili sme všetky potrebné detaily, pristúpme preto k definícii posledného typu maximálnej entropie a spojenej informácie.

Definícia 5.4 (Maximálna entropia pri zachovaní entropií do rádu k na $\mathcal{S}_N$)**.** Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $k \leq n$. Nech $\{\boldsymbol{X}\} = \{(X_1, ..., X_n)\}$ je množina náhodných veličín s daným oborom hodnôt. Ďalej majme množinu množín $\{\mathring{\boldsymbol{X}}\} = \{\mathring{X}_1, ..., \mathring{X}_n\}$.
Majme usporiadanú množinu hodnôt podmienok M, v ktorej je $s = \sum_{j=1}^{k} \binom{n}{j}$ nezáporných reálnych čísiel a množiny $N_j = \{(i_1, ..., i_j) \mid 1 \leq i_1 < \cdots < i_j \leq n\}$ $\forall j \in \widehat{k}$. Označme $N = \bigcup_{j=1}^{k} N_j$, ďalej označme prvky N ako $\boldsymbol{n}_l$, podobne prvky M ako m_l. Označme $\mathring{\boldsymbol{X}}_{n_l} = \bigcup_{i \in \boldsymbol{n}_l} \mathring{X}_i$. Nakoniec definujme množinu prípustných číselných diagramov

$$A = \{D \in \mathcal{S}_N(\mathring{\boldsymbol{X}}) \mid \forall l \in \widehat{s} : \mu_I(\mathring{\boldsymbol{X}}_{\boldsymbol{n}_l}) = m_l\}.$$

Potom, pri predpoklade $A \neq \emptyset$, definujeme maximálnu entropiu pri zachovaní entropií do rádu k na shannonovských diagramoch, značíme $\tilde{H}^{[k]}(\tilde{\boldsymbol{X}})$, ako

$$\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) := \sup_{D \in A} \mu_I(D). \tag{5.1}$$

Poznámka 5.5. Vyslovme niekoľko pozorovaní súvisiacich s definíciou.

- Vidíme, že maximalizujeme na množine číselných diagramov, používame rozšírenú definíciu I-miery, nepotrebujeme preto požadovať, aby množiny v $\{\mathring{\boldsymbol{X}}\}$ reprezentovali zložky náhodnej veličiny.
- Pre entropiu maximalizovanú na diagramoch sme zaviedli označenie $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$. Na rozdiel od označenia pre entropiu maximalizovanú na rozdeleniach $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$, značenie pre veličinu maximalizovanú na diagramoch neobsahuje rozdelenie P, čím sme chceli vyjadriť, že táto maximálna entropia nie vyjadrením suprémа entropií rozdelenia pravdepodobnosti náhodnej veličiny, ale len suprémom I-mier číselného diagramu. Index S zavádzame preto, lebo maximalizujeme na shannonovských diagramoch.

Definícia 5.6 (Spojená informácia rádu k pri zachovaní entropií na $\mathcal{S}_N$)**.** Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $2 \leq k \leq n$, ostatné predpoklady ako v definícii 5.4. Nech platí $\tilde{H}_S^{[1]}(\tilde{\boldsymbol{X}}) < +\infty$. Potom definujeme spojenú informáciu rádu k pri zachovaní entropií na shannonovských diagramoch, značíme $I_S^{[k]}(\tilde{\boldsymbol{X}})$, ako

$$I_S^{[k]}(\tilde{\boldsymbol{X}}) := \tilde{H}_S^{[k-1]}(\tilde{\boldsymbol{X}}) - \tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}). \tag{5.2}$$

Poznamenajme, že vďaka monotónii maximálnej entropie, ktorú ukážeme v 5.7 stačí na to, aby rozdiel dvoch po sebe idúcich maximálnych entropií bol konečný, predpokladať konečnosť prvých entropií.

5.1.2 Vety

Našim ďalším cieľom bude preskúmať základné vlastnosti entropie maximalizovanej na diagramoch a príslušnej spojenej informácie. V tejto kapitole nebudeme oddeľovať vety o entropii do osobitnej sekcie od viet o spojenej informácii, pretože skúmaných vzťahov je menej. Ukážeme ohraničenosť a monotóniu postupnosti novodefinovaných maximálnych entropií, preskúmame súvislosť konštantnosti postupnosti maximálnych entropií a nezávislosti zložiek veličiny, podľa ktorej entropií maximalizujeme, z týchto viet odvodíme príslušné dôsledky pre vlastnosti spojenej informácie.

Veta 5.7 (Monotónia $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$)**.** *Majme $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ diskrétnu vektorovú náhodnú veličinu. Maximalizujme na shannonovských diagramoch entropiu pri entropiách* *veličiny $\tilde{\boldsymbol{X}}$, predpokladajme ich konečnosť. Potom pre všetky $k \in \widehat{n-1}$ platí*

$$\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) \geq \tilde{H}_S^{[k+1]}(\tilde{\boldsymbol{X}}). \tag{5.3}$$

Dôkaz. Zvoľme $k \in \widehat{n-1}$. Označme $\mathcal{S}_N^k(\mathring{\boldsymbol{X}})$ množinu číselných diagramov, ktoré spĺňajú Shannonove nerovnosti a podmienky určené entropiami do rádu k. Platí $\mathcal{S}_N^{k+1}(\mathring{\boldsymbol{X}}) \subset \mathcal{S}_N^k(\mathring{\boldsymbol{X}})$ a obidve množiny sú neprázdne, určite totiž obsahujú informačný diagram veličiny $\tilde{\boldsymbol{X}}$. Maximálna entropia rádu k je definovaná ako suprémum I-miery na množine $\mathcal{S}_N^k(\mathring{\boldsymbol{X}})$. Z lemy 2.19 dostávame $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) \geq \tilde{H}_S^{[k+1]}(\tilde{\boldsymbol{X}})$. □

Poznámka 5.8. V dôkaze sme využili vzťah $\mathcal{S}_N^{k+1}(\mathring{\boldsymbol{X}}) \subset \mathcal{S}_N^k(\mathring{\boldsymbol{X}})$, opačná inklúzia medzi množinami nemusí platiť a môže nastávať aj ostrá nerovnosť medzi maximálnymi entropiami dvoch po sebe idúcich rádov. Na predstavu najjednoduchší prípad nerovnosti je diagram pre dve množiny $\mathring{X}_1$, $\mathring{X}_2$, kde volíme hodnotu entropie druhého rádu, pri ktorej maximalizujeme tak, aby platilo $H(X_1, X_2) < H(X_1) + H(X_2)$. Na druhej strane, vieme nájsť podmienky také, že nastáva rovnosť, stačí voliť hodnotu entropie druhého rádu tak, aby $H(X_1, X_2) = H(X_1) + H(X_2)$. Neostrá nerovnosť teda vo vete nemôže byť nahradená ostrou nerovnosťou, ale ani rovnosťou.

Poznámka 5.9. Mohlo by nás zaujímať, či je spojená informácia pri maximalizácii na diagramoch monotónna. Pripomeňme, že pre veličiny maximalizované na rozdeleniach sme ukázali, že ani jeden z troch typov spojenej informácie nie je monotónny. Ukázali sme to na príklade XOR a náhodných veličín interagujúcich len po dvojiciach, viď príklady 3.1, 3.2. Na rovnakých príkladoch by sme vyvrátili monotóniu $I_S^{[k]}(\tilde{\boldsymbol{X}})$.

Poznámka 5.10. Je potrebné si uvedomiť, že pri maximalizácii entropie na diagramoch je nutné zahrnúť do maximalizačných väzieb aj Shannonove nerovnosti. Uveďme jednoduchý príklad na to, ako sa maximálna entropia líši podľa toho, či maximalizujeme na všetkých číselných diagramoch alebo len na tých, ktoré spĺňajú Shannonove nerovnosti. Majme diagram pre dve veličiny. Maximalizujme jeho mieru pri zadaných prvých entropiách $H(X)$ a $H(Y)$. Pokiaľ požadujeme platnosť Shannonovych nerovností, tak všetky atómy diagramu musia byť kladné. To implikuje, že maximálna entropia je ohraničená sumou prvých entropií. Ak by sme však platnosť Shannonovych nerovností nepožadovali, vieme nájsť číselný diagram, ktorý má I-mieru rovnú n pre ľubovoľné $n \in \mathbb{N}$. Ľahko odvodíme, že pre $n \in \mathbb{N}$ zvolené ľubovoľne pevne má takýto diagram hodnoty atómov $\mu_I(\mathring{X} \cap \mathring{Y}) = H(X) + H(Y) - n$, $\mu_I(\mathring{X} \cap \mathring{Y}^C) = H(X) - \mu_I(\mathring{X} \cap \mathring{Y})$ a $\mu_I(\mathring{Y} \cap \mathring{X}^C) = H(Y) - \mu_I(\mathring{X} \cap \mathring{Y})$. Maximálna entropia je v tomto prípade nekonečno.

Poznámka 5.11. Skôr, než vyslovíme vetu o ohraničenosti a nutnú a postačujúcu podmienku nezávislosti poznamenajme, že pokiaľ máme zadané ako väzbové podmienky entropie prvého rádu, tak I-miera disjunktného informačného diagramu pre dané veličiny má najväčšiu možnú hodnotu z hodnôt I-mier informačných diagramov pre tieto veličiny. Vyplýva to z vety 1.31 a z korešpondencie I-miery s informačnými veličinami. Pojem disjunktný diagram vysvetlíme v poznámke 5.12. To, že podľa tvrdenia vety 1.31 sa horná hranica hodnoty entropie dosahuje práve pre nezávislé zložky implikuje, že žiadny iný informačný diagram nedosahuje túto hodnotu I-miery. Toto pozorovanie môžeme zovšeobecniť aj na číselné diagramy, pre ktoré platia Shannonove nerovnosti pomocou lemy 5.13.

Poznámka 5.12. Objasnime, čo myslíme pod pojmom disjunktný diagram, aj keď je význam tohto pojmu intuitívne jasný. Za disjunktný diagram množín $\mathring{X}_1, \ldots, \mathring{X}_n$ považujeme taký, v ktorom majú nenulovú I-mieru jedine atómy v tvare $\mathring{X}_i \cap \{\mathring{X}_1 \cup \ldots \cup \mathring{X}_{i-1} \cup \mathring{X}_{i+1} \cup \ldots \cup \mathring{X}_n\}^C$ pre nejaké $i \in \widehat{n}$. Zjednodušene by sme mohli povedať, že disjunktný diagram je taký, v ktorom majú všetky prieniky množín nulovú I-mieru. V disjunktnom diagrame reprezentujúcom vektorovú náhodnú veličinu je vzájomná informácia medzi jednotlivými skupinami zložiek nulová, takýto diagram teda reprezentuje veličinu s nezávislými zložkami.

Lema 5.13 (Odhad miery diagramu)**.** *Majme množinu množín $\{\mathring{\boldsymbol{X}}\} = \{\mathring{X}_1, \ldots, \mathring{X}_n\}$. Potom pre všetky $D \in \mathcal{S}_N(\mathring{\boldsymbol{X}})$ platí $\mu_I(D) \leq \sum_{i=1}^{n} \mu_I(\mathring{X}_i)$.*

Dôkaz. Budeme postupovať matematickou indukciou podľa počtu množín n. Pre $n = 1$ je nerovnosť triviálna, overme preto ešte platnosť pre $n = 2$. Majme číselný diagram D pre dve množiny $\mathring{X}_1$ a $\mathring{X}_2$. Platnosť Shannonovych nerovností implikuje nezápornosť $\mu_I(\mathring{X}_1 \cap \mathring{X}_2)$. Môžeme teda odhadovať $\mu_I(D) = \mu_I(\mathring{X}_1 \cap \mathring{X}_2^C) + \mu_I(\mathring{X}_1 \cap \mathring{X}_2) + \mu_I(\mathring{X}_1^C \cap \mathring{X}_2) \leq \mu_I(\mathring{X}_1 \cap \mathring{X}_2^C) + 2\mu_I(\mathring{X}_1 \cap \mathring{X}_2) + \mu_I(\mathring{X}_1^C \cap \mathring{X}_2) = \mu_I(\mathring{X}_1) + \mu_I(\mathring{X}_2)$. Týmto sme ukázali platnosť pre $n = 2$, môžeme pristúpiť k indukčnému kroku.

Majme $k \in \mathbb{N}$. Nech dokazovaná nerovnosť platí pre shannonovské číselné diagramy pre $k-1$ množín. Nech D je shannonovský diagram pre k množín. Potom jeho I-mieru môžeme vyjadriť ako $\mu_I(D) = \mu_I(\{\mathring{X}_1 \cup \ldots \cup \mathring{X}_{k-1}\} \cap \mathring{X}_k^C) + \mu_I(\{\mathring{X}_1 \cup \ldots \cup \mathring{X}_{k-1}\} \cap \mathring{X}_k) + \mu_I(\{\mathring{X}_1 \cup \ldots \cup \mathring{X}_{k-1}\}^C \cap \mathring{X}_k) \leq \mu_I(\{\mathring{X}_1 \cup \ldots \cup \mathring{X}_{k-1}\} \cap \mathring{X}_k^C) + 2\mu_I(\{\mathring{X}_1 \cup \ldots \cup \mathring{X}_{k-1}\} \cap \mathring{X}_k) + \mu_I(\{\mathring{X}_1 \cup \ldots \cup \mathring{X}_{k-1}\}^C \cap \mathring{X}_k) \leq \sum_{i=1}^{k-1} \mu_I(\mathring{X}_i) + \mu_I(\{\mathring{X}_1 \cup \ldots \cup \mathring{X}_{k-1}\} \cap \mathring{X}_k) + \mu_I(\{\mathring{X}_1 \cup \ldots \cup$

$\mathring{X}_{k-1}\}^C \cap \mathring{X}_k) = \sum_{i=1}^{k} \mu_I(\mathring{X}_i)$ V odhade sme využili, že vďaka platnosti Shannonovych nerovností je hodnota $\mu_I(\{\mathring{X}_1 \cup \ldots \cup \mathring{X}_{k-1}\} \cap \mathring{X}_k)$ nezáporná a podľa indukčného predpokladu platí $\mu_I(\{\mathring{X}_1 \cup \ldots \cup \mathring{X}_{k-1}\} \cap \mathring{X}_k^C) + \mu_I(\{\mathring{X}_1 \cup \ldots \cup \mathring{X}_{k-1}\} \cap \mathring{X}_k) \leq \sum_{i=1}^{k-1} \mu_I(\mathring{X}_i)$. □

Poznámka 5.14. Z lemy vyplýva, že podobne ako sme v poznámke 5.11 komentovali pre informačné diagramy, aj pre shannonovské číselné diagramy bude platiť, že ak maximalizujeme pri prvých entropiách, tak maximálna entropia sa dosahuje na disjunktnom diagrame, jeho miera je totiž práve $\sum_{i=1}^{n} \mu_I(\mathring{X}_i)$. Zároveň bude platiť, že na žiadnom inom, než disjunktnom diagrame sa táto hodnota I-miery nedosahuje. To môžeme vidieť z dôkazu lemy, platí totiž, že na to, aby sa dosahovala rovnosť, musí platiť $\mu_I(\{\mathring{X}_1 \cup \ldots \cup \mathring{X}_{k-1}\} \cap \mathring{X}_k) = 2\mu_I(\{\mathring{X}_1 \cup \ldots \cup \mathring{X}_{k-1}\} \cap \mathring{X}_k)$, čo môžeme upraviť na výraz $\mu_I(\{\mathring{X}_1 \cup \ldots \cup \mathring{X}_{k-1}\} \cap \mathring{X}_k) = 0$, a to nezávisle na voľbe indexov množín a pre každé $k \in \widehat{n}$, prieniky teda majú nulovú I-mieru.

Pomocou vyslovenej lemy teraz dokážeme vetu o ohraničenosti entropie maximalizovanej na diagramoch.

Veta 5.15 (Ohraničenosť $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$)**.** *Nech $n \in \mathbb{N}$, $k \in \widehat{n}$. Majme diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$. Maximalizujme na shannonovských diagramoch entropiu pri entropiách $\tilde{\boldsymbol{X}}$ do rádu k, predpokladajme ich konečnosť. Potom platí*

$$\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) \geq 0 \tag{5.4}$$

$$\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} H(\tilde{X}_i) \tag{5.5}$$

$$\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|, \tag{5.6}$$

kde $\|\tilde{\mathcal{X}}_i\|$ značí počet prvkov v obore hodnôt $\tilde{X}_i$.

Dôkaz. Vzťah (5.4) vyplýva z nerovnosti $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) \geq H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})) \geq 0$. Prvá nerovnosť bude ukázaná vo vete 5.21, nezápornosť $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ sme už odôvodnili v 2.39. Vzťah (5.5) vyplýva z lemy 5.13. Ďalej nerovnosť (5.6) vyplýva z toho, že vďaka vete 1.32 platí $H(\tilde{X}_i) < \log_2 \|\tilde{\mathcal{X}}_i\|$, z toho dostávame $\sum_{i=1}^{n} H(\tilde{X}_i) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|$. V spojení so vzťahom (5.5) dostávame (5.6). □

Príklad 5.16. Uveďme príklad na dosahovanie rovnosti v hornom a dolnom odhade. Maximalizujme entropie na diagrame generovanom n množinami. Ak zadáme entropie prvého rádu nulové, tak $\tilde{H}_S^{[1]}(\tilde{\boldsymbol{X}}) = 0$. Dokonca bude rovnosť platiť aj pre vyššie rády, pretože ak sú zadávané prvé entropie nulové, tak aj zadávané entropie vyšších rádov budú musieť byť nulové, teda platí $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) = 0$ pre všetky $k \in \widehat{n}$. Týmto sme zároveň ilustrovali rovnosť v prvom z horných odhadov, pretože $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) = \sum_{i=1}^{n} H(\tilde{X}_i) = 0$ pre všetky rády. Rovnosť v druhom z horných odhadov sa môže dosahovať, ak sú zadané ako väzbové podmienky entropie rovnomerného rozdelenia.

Z vety o ohraničenosti entropie maximalizovanej na diagramoch vyplýva veta o ohraničenosti príslušnej spojenej informácie.

Veta 5.17 (Ohraničenosť $I_S^{[k]}(\tilde{\boldsymbol{X}})$)**.** *Nech $n \in \mathbb{N}$, $k \in \widehat{n}$. Majme diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$. Maximalizujme na shannonovských diagramoch entropiu pri entropiách $\tilde{\boldsymbol{X}}$ do rádu k, predpokladajme ich konečnosť. Potom platí*

$$I_S^{[k]}(\tilde{\boldsymbol{X}}) \geq 0 \tag{5.7}$$

$$I_S^{[k]}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} H(\tilde{X}_i) \tag{5.8}$$

$$I_S^{[k]}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|, \tag{5.9}$$

kde $\|\tilde{\mathcal{X}}_i\|$ značí počet prvkov v obore hodnôt $\tilde{X}_i$.

Dôkaz. Vzťah (5.7) platí vďaka monotónii $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$. Nerovnosť (5.8) získame odhadom $I_S^{[k]}(\tilde{\boldsymbol{X}}) = \tilde{H}_S^{[k-1]}(\tilde{\boldsymbol{X}}) - \tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) \leq \tilde{H}_S^{[k-1]}(\tilde{\boldsymbol{X}}) \leq \tilde{H}_S^{[1]}(\tilde{\boldsymbol{X}}) = \sum_{i=1}^{n} H(\tilde{X}_i)$, kde sme využili vzťahy pre horné a dolné odhady maximálnej entropie pri daných entropiách, vzťah (5.9) odhadom $I_S^{[k]}(\tilde{\boldsymbol{X}}) = \tilde{H}_S^{[k-1]}(\tilde{\boldsymbol{X}}) - \tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) \leq \tilde{H}_S^{[k-1]}(\tilde{\boldsymbol{X}}) \leq \tilde{H}_S^{[1]}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|$, kde sme opäť využili odhady pre horné a dolné hranice maximálnej entropie. □

Podobne, ako pri skúmaní ohraničenosti spojených informácií v kapitole o veličinách maximalizovaných na rozdeleniach, aj v prípade $I_S^{[k]}(\tilde{\boldsymbol{X}})$ ľahko nájdeme horné hranice sumy spojených informácií.

Veta 5.18 (Ohraničenosť $\sum I_S$)**.** *Nech $n \in \mathbb{N}$, $k \in \widehat{n}$. Majme diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$. Potom platí*

$$\sum_{k=2}^{n} I_S^{[k]}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} H(\tilde{X}_i) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|. \tag{5.10}$$

Dôkaz. Dokázali by sme podobne, ako vetu 2.60. □

Podobne, ako v prípade $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ a $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$, aj pomocou $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$, resp. pomocou jej príslušnej spojenej informácie $I_S^{[k]}(\tilde{\boldsymbol{X}})$, budeme schopní vyjadriť nutnú a postačujúcu podmienku nezávislosti zložiek veličiny, podľa ktorej entropií maximalizujeme.

Veta 5.19 (Nutná a postačujúca podmienka nezávislosti cez $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$)**.** *Majme diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s konečným oborom hodnôt. Maximalizujme entropiu na shannonovských diagramoch pri entropiách veličiny $\tilde{\boldsymbol{X}}$ do určitého rádu. Potom zložky $\tilde{\boldsymbol{X}}$ sú nezávislé práve vtedy, keď je postupnosť maximálnych entropií pri zadaných entropiách konštantná, teda platí*

$$\tilde{H}_S^{[1]}(\tilde{\boldsymbol{X}}) = \tilde{H}_S^{[2]}(\tilde{\boldsymbol{X}}) = \ldots = \tilde{H}_S^{[n-1]}(\tilde{\boldsymbol{X}}) = \tilde{H}_S^{[n]}(\tilde{\boldsymbol{X}}).$$

Dôkaz. Konečný obor hodnôt $\tilde{\boldsymbol{X}}$ zaručuje konečné hodnoty entropií, ktoré použijeme ako väzbové podmienky. Ukážme najprv nutnú podmienku nezávislosti. Majme entropie nezávislej náhodnej veličiny, použime ich ako väzbové podmienky pre maximalizáciu entropie na shannonovských číselných diagramoch. Diagram vytvorený z disjunktných množín tieto podmienky spĺňa, a to platí pre všetky rády zadaných väzieb. Zároveň z lemy 5.13 a poznámky 5.14 vieme, že žiadny shannonovský číselný diagram nemôže mať väčšiu I-mieru, než disjunktný. Maximálna entropia pri zadaní entropií do rádu k pre ľubovoľné $k \in \widehat{n}$ je teda rovná miere disjunktného diagramu, t. j. súčtu entropií jednotlivých zložiek $\tilde{\boldsymbol{X}}$, postupnosť je konštantná. Odôvodnime ešte postačujúcu podmienku. Majme konštantnú postupnosť entropií maximalizovaných na shannonovských diagramoch. Vieme, že $\tilde{H}_S^{[1]}(\tilde{\boldsymbol{X}})$ je I-mierou disjunktného diagramu a ako sme uviedli v poznámke 5.14, túto I-mieru má z možných shannonovských číselných diagramov jedine disjunktný a ten je informačným diagramom $\tilde{\boldsymbol{X}}$, ktorá má nezávislé zložky. □

Veta 5.20 (Nutná a postačujúca podmienka nezávislosti cez $I_S^{[k]}(\tilde{\boldsymbol{X}})$)**.** *Majme diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s konečným oborom hodnôt. Maximalizujme entropiu na shannonovských diagramoch pri entropiách veličiny $\tilde{\boldsymbol{X}}$ do určitého rádu. Potom zložky $\tilde{\boldsymbol{X}}$ sú nezávislé práve vtedy, keď je postupnosť spojených informácii pri zadaných entropiách nulová, teda platí*

$$I_S^{[k]}(\tilde{\boldsymbol{X}}) = 0 \qquad \forall k \in \{2, ..., n\}.$$

Dôkaz. Vyplýva z vety 5.19. □

5.1.3 Vzťahy medzi konštrukciami

Podobne, ako pri konštrukcii troch typov maximálnej entropie, ktoré vznikli maximalizáciou na rozdeleniach pravdepodobnosti, chceli by sme aj pri štvrtom type preskúmať jeho vzťahy ku ostatným typom maximálnych entropií. Zaujíma nás predovšetkým vzťah $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ a $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$, pretože motiváciou k definícii $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$ bola práve náhrada $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$, ktorú obvykle nevieme vyčísliť.

Veta 5.21 (Vzťah $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ a $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$)**.** *Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $k \leq n$ a množinu $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$. Majme ďalej $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Maximalizujme entropiu na rozdeleniach a na shannonovských diagramoch pri entropiách veličiny $\tilde{\boldsymbol{X}}$, predpokladáme ich konečnosť. Potom platí*

$$\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) \geq H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})). \tag{5.11}$$

Dôkaz. $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ vzniká maximalizáciou entropie na rozdeleniach pravdepodobnosti, ktoré spĺňajú hodnoty daných entropií. Ku každému takémuto rozdeleniu môžeme jednoznačne priradiť informačný diagram, označme množinu týchto informačných diagramov B. Potom $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ môžeme chápať ako hodnotu supréma I-miery diagramov z množiny B. Ďalej vieme, že $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$ vzniká maximalizáciou entropie na shannonovských číselných diagramoch, ktoré spĺňajú podmienky na dané entropie. Ich množinu označíme v súlade so značením v definícii $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$ ako A. Vieme, že každý informačný diagram spĺňa Shannonove nerovnosti, môžeme ho teda chápať ako číselný shannonovský diagram. Platí $B \subset A$, medzi všetkými číselnými diagramami splňujúcimi podmienky na entropie sú určite aj všetky vhodné informačné diagramy. Vďaka predpokladom sú množiny A a B neprázdne, určite obsahujú informačný diagram, ktorý prislúcha rozdeleniu veličiny $\tilde{\boldsymbol{X}}$. Z lemy 2.19 potom vyplýva $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) \geq H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$. □

Poznámka 5.22. Maximalizáciou entropie pri entropiách veličiny $\tilde{\boldsymbol{X}}$ myslíme to, že za množinu M v definícii 5.4 a za množinu $M^{[k]}$ v definícii 2.15 volíme množinu hodnôt príslušných entropií vypočítaných z rozdelenia veličiny $\tilde{\boldsymbol{X}}$.

Poznámka 5.23. Vďaka tomu, že inklúziu medzi množinami diagramov nemôžeme vo všeobecnosti obrátiť, neočakávame, že by vždy platila rovnosť. Na druhej strane, existujú príklady, keď rovnosť nastáva. Je to tak napríklad vtedy, keď za väzbové podmienky zadávame entropie diskrétnej vektorovej veličiny s nezávislými zložkami. Potom totiž $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) = H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})) = \sum_{i=1}^n H(\tilde{X}_i)$, túto situáciu sme už preskúmali vo vetách o podmienkach nezávislosti.

Ďalším príkladom na rovnosť daných dvoch typov entropie je situácia, kedy o diagrame vytvorenom pri maximalizácii vieme, že je informačný. Napríklad, ak má výsledný diagram len nezáporné hodnoty atómov a nemáme obmedzený obor hodnôt náhodných veličín, z vety 4.15 vyplýva, že je informačný pre tieto náhodné veličiny, teda existuje rozdelenie pravdepodobnosti, na ktorom sa vypočítaná maximálna entropia dosahuje, nastáva rovnosť maximálnych entropií.

Poznamenajme, že prvý uvedený príklad na rovnosť, t. j. rovnosť nastáva pri zadaní entropií veličiny s nezávislými zložkami, je tiež špeciálnym prípadom situácie, keď o výslednom diagrame vieme, že je informačný.

Vyslovme ešte vetu o vzťahu medzi entropiou maximalizovanou na diagramoch pri zadaných entropiách a entropiou maximalizovanou na rozdeleniach pri zadaných marginálach. Vyslovíme najprv vetu s detailnými predpokladmi, v poznámke potom uvedieme slovné znenie.

Veta 5.24 (Vzťah $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$ a $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$)**.** *Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $k \leq n$, množinu $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$. Majme ďalej diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt a indexovú množinu $N^{(k)}$ z definície $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$, jej prvky označme $\boldsymbol{n}_i$. Označme $B = \{\tilde{p}_{\boldsymbol{n}_i}(\boldsymbol{x}) \mid \boldsymbol{n}_i \in N^{(k)}\}$. Za vektorové veličiny $\tilde{\boldsymbol{X}}_i$ v definícii* 2.5 *zvoľme vektorové náhodné veličiny určené rozdeleniami v B. Majme indexovú množinu N z definície $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$. Jej prvky označme $\tilde{\boldsymbol{n}}_l$. Ďalej označme združenú entropiu $H(\tilde{X}_{i_1}, \ldots, \tilde{X}_{i_j}) = H_{\tilde{\boldsymbol{n}}_l}$ pre všetky $(i_1, \ldots, i_j) = \tilde{\boldsymbol{n}}_l$. Nech sú všetky $H_{\tilde{\boldsymbol{n}}_l}$ konečné. Za množinu hodnôt podmienok M v definícii* 5.4 *položme $M = \{H_{\tilde{\boldsymbol{n}}_l} \mid \tilde{\boldsymbol{n}}_l \in N\}$. Potom pre takto definované maximálne entropie platí*

$$\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) \geq H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}})). \tag{5.12}$$

Dôkaz. Veta je priamočiarym dôsledkom viet 2.45 a 5.21. □

Poznámka 5.25. V oboch vetách spomínaných v dôkaze máme dokázané tvrdenie v tvare neostrej nerovnosti, preto aj v tejto vete ponecháme neostrú nerovnosť.

Poznámka 5.26 (Vzťah $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$ a $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$, slovne)**.** Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $k \leq n$, množinu $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$. Majme ďalej diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Maximalizujme entropiu pri marginálnych rozdeleniach veličiny $\tilde{\boldsymbol{X}}$ a na shannonovských diagramoch pri entropiách veličiny $\tilde{\boldsymbol{X}}$, predpokladáme ich konečnosť. Potom platí

$$\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}) \geq H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}})). \tag{5.13}$$

Touto poznámkou ukončujeme prieskum základných vlastností primárnej konštrukcie entropie maximalizovanej pri zadaných entropiách na diagramoch, v poslednej časti práce ešte predstavíme alternatívnu konštrukciu, ktorá bude taktiež aproximáciou maximálnej entropie pri zadaných entropiách definovanej na rozdeleniach, ktorú sme popísali v druhej kapitole práce, resp. aproximáciou maximálnej entropie z primárnej konštrukcie. Zatiaľ čo primárna konštrukcia bola navrhnutá v článku [2], alternatívna konštrukcia sa v článkoch vôbec nespomína, je to nami novovytvorený spôsob maximalizácie, ktorý je však primárnej schéme blízky.

5.2 Alternatívna konštrukcia

V predchádzajúcej časti sme preskúmali vlastnosti konštrukcie, ktorá bola navrhovaná v článkoch [1, 2]. V tejto konštrukčnej schéme sa maximalizuje entropia pri zadaných entropiách do určitého rádu na diagramoch, ktoré spĺňajú Shannonove nerovnosti. Ako sme ukázali v leme 5.13, platnosť Shannonovych nerovností je postačujúca na to, aby bola miera diagramu konečná. Pokiaľ požadujeme platnosť Shannonovych nerovností na diagrame pre n množín, zadávame n nerovností vyjadrujúcich nezápornosť podmienenej entropie a $\binom{n}{2}$ nerovností vyjadrujúcich nezápornosť podmienenej vzájomnej informácie, nerovnosti sú v tvare danom definíciou 4.21. Mohlo by nás zaujímať, či nemôžeme tieto nerovnosti nahradiť inou sadou nerovností, ktorá by tiež zabezpečila konečnosť entropie. V nasledujúcej časti preto skúsime nahradiť platnosť Shannonovych nerovností požiadavkou na nezápornosť I-miery každého atómu diagramu.

Skôr, než pristúpime k samotným definíciám, uveďme ešte jeden technický detail, na ktorom uvidíme, prečo by naša alternatívna konštrukcia mohla byť vhodná. Pri zadávaní optimalizačného problému na maximalizáciu entropie na diagramoch potrebujeme správne previesť Shannonove nerovnosti na nerovnosti medzi I-mierou atómov diagramu. Tento problém síce nie je zložitý, ale vyžaduje si znalosť základných vzťahov medzi informačnými veličinami. Ak by sme však namiesto týchto nerovností požadovali len nezápornosť I-miery diagramu, zadávame $2^n - 1$ podmienok, všetky majú rovnaký jednoduchý tvar $\mu_I(A) \geq 0$, kde A značí atóm diagramu. Na druhej strane, priamo vidíme aj možné nevýhody konštrukcie. Vieme totiž, že nie všetky informačné diagramy pre tri a viac náhodných veličín musia mať nezáporné hodnoty atómov diagramu, je teda vhodné preskúmať, či sme sa neobmedzili na príliš malú množinu informačných diagramov. Bude nás tiež zaujímať, či táto konštrukcia môže priniesť aj nejaké iné výhody okrem jednoduchšieho zadávania väzbových podmienok a či je naozaj vhodné sa obmedziť len na nezáporné diagramy.

5.2.1 Definície

Definícia 5.27 (Nezáporné diagramy)**.** Majme množinu náhodných veličín s daným oborom hodnôt $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$ a množinu množín $\{\mathring{\boldsymbol{X}}\} = \{\mathring{X}_1, \ldots, \mathring{X}_n\}$. Potom $\mathcal{N}_N(\mathring{\boldsymbol{X}})$ označíme podmnožinu $\mathcal{D}_N(\mathring{\boldsymbol{X}})$ obsahujúcu číselné diagramy, ktoré majú I-mieru všetkých atómov nezápornú.

Poznámka 5.28. Nezáporné diagramy spĺňajú Shannonove nerovnosti, t. j. platí $\mathcal{N}_N(\mathring{\boldsymbol{X}}) \subset \mathcal{S}_N(\mathring{\boldsymbol{X}})$.

Poznámka 5.29 (O nezáporných diagramoch)**.** V článku [1] bol odvodený tvar maximálne entropického diagramu pri zadaných entropiách do rádu dva pre trojzložkovú náhodnú veličinu. Výsledok ukazuje, že tento diagram má všetky hodnoty I-miery na atómoch nezáporné. Autori tohto článku tiež uvádzajú, že vo väčšine nimi testovaných prípadov maximálne entropické diagramy, ktoré vznikli optimalizáciou miery diagramu pri platnosti Shannonovych nerovností, boli práve diagramy s nezápornými atómami. Na druhej strane, pokiaľ by sme hľadali diagramy s minimálnou entropiou, výsledkom by veľmi často boli práve diagramy s negatívnou I-mierou atómov, na túto úlohu teda nie je naše obmedzenie vhodné. Ako však aj autori článku upozorňujú, minimalizáciu entropie pri zadaných entropiách nie je vhodné riešiť ani ich primárnou konštrukciou, pretože pre diagramy so zápornými hodnotami nevieme zaručiť, že existuje rozdelenie, na ktorom sa dosahujú dané hodnoty informačných veličín, pritom práve existencia takéhoto rozdelenia je pre nás dôležitá.

Poznámka 5.30. Ďalšia výhoda alternatívnej konštrukcie súvisí s existenciou rozdelenia k číselnému diagramu. Vďaka vete 4.15 máme zaručené, že pre diagram pre n množín s nezápornými hodnotami atómov, existuje n diskrétnych náhodných veličín, ktoré majú hodnoty informačných veličín také, ako sú určené diagramom. Nevýhodou tejto vety je, že nepoznáme nosiče hustôt pravdepodobnosti týchto veličín. Symbolicky môžeme zapísať, že platí $\mathcal{N}_N(\mathring{\boldsymbol{X}}) \subset \mathcal{D}(\boldsymbol{X})$, kde $\{\mathring{\boldsymbol{X}}\} = \{\mathring{X}_1, \ldots, \mathring{X}_n\}$, $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$ a $\operatorname{Ran} X_i \subset \mathbb{N}_0$ pre všetky $i \in \widehat{n}$, t. j. $\mathcal{D}(\boldsymbol{X})$ predstavuje množinu diagramov, ktoré sú informačné pre nejakú vektorovú náhodnú veličinu, ktorá môže mať nosič hustoty pravdepodobnosti jednotlivých zložiek konečný, ale aj spočítateľný, teda neohraničený.

Zaviedli sme množinu diagramov, na ktorej budeme maximalizovať hodnotu I-miery a preskúmali jej vlastnosti a vzťahy ku ostatným množinám diagramov, pristúpme teraz už k samotnej definícii posledného typu maximálnej entropie a spojenej informácie.

Definícia 5.31 (Maximálna entropia pri zachovaní entropií do rádu k na $\mathcal{N}_N$)**.** Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $k \leq n$. Ďalej majme množinu množín $\{\mathring{\boldsymbol{X}}\} = \{\mathring{X}_1, ..., \mathring{X}_n\}$. Majme usporiadanú množinu hodnôt podmienok M_+, v ktorej je $s = \sum_{j=1}^{k} \binom{n}{j}$ nezáporných reálnych čísiel a množiny $N_j = \{(i_1, ..., i_j) \mid 1 \leq i_1 < \cdots < i_j \leq n\}$ $\forall j \in \widehat{k}$. Označme $N_+ = \bigcup_{j=1}^{k} N_j$ indexovú množinu, ďalej označme prvky N_+ ako $\boldsymbol{n}_l$, podobne prvky M_+ ako m_l. Označme $\mathring{\boldsymbol{X}}_{n_l} = \bigcup_{i \in \boldsymbol{n}_l} \mathring{X}_i$. Nakoniec definujme množinu prípustných číselných diagramov

$$A_+ = \{D \in \mathcal{N}_N(\mathring{\boldsymbol{X}}) \mid \forall l \in \widehat{s} : \mu_I(\mathring{\boldsymbol{X}}_{\boldsymbol{n}_l}) = m_l\}.$$

Potom, pri predpoklade $A_+ \neq \emptyset$, definujeme maximálnu entropiu pri zachovaní entropií do rádu k na nezáporných diagramoch, značíme $\tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}})$, ako

$$\tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}}) := \sup_{D \in A_+} \mu_I(D). \tag{5.14}$$

Definícia 5.32 (Spojená informácia pri zachovaní entropií do rádu k na $\mathcal{N}_N$)**.** Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $2 \leq k \leq n$, ostatné predpoklady ako v definícii 5.4. Nech platí $\tilde{H}_+^{[1]}(\tilde{\boldsymbol{X}}) < +\infty$. Potom definujeme spojenú informáciu rádu k pri zachovaní entropií na nezáporných diagramoch, značíme $I_+^{[k]}(\tilde{\boldsymbol{X}})$, ako

$$I_+^{[k]}(\tilde{\boldsymbol{X}}) := \tilde{H}_+^{[k-1]}(\tilde{\boldsymbol{X}}) - \tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}}). \tag{5.15}$$

5.2.2 Vety

Podobne, ako pri všetkých ostatných konštrukciách, vyslovíme a dokážeme najprv vety o základných vlastnostiach maximálnej entropie a spojenej informácie na nezáporných diagramoch, potom preskúmame vzťahy tejto konštrukcie a niektorých ostatných.

Predovšetkým nás zaujímajú vzťahy ku ostatným entropiám maximalizovaným pri zadaných entropiách, t .j. vzťah k $\tilde{H}^{[k]}(\tilde{\boldsymbol{X}})$ a $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$.

Veta 5.33 (Monotónia $\tilde{H}^{[k]}_{+}(\tilde{\boldsymbol{X}})$). *Majme $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$. Maximalizujme na nezáporných diagramoch entropiu pri entropiách veličiny $\tilde{\boldsymbol{X}}$, predpokladajme ich konečnosť. Predpokladajme ďalej, že existuje nezáporný diagram spĺňajúci dané väzbové podmienky pre všetky $k \in \widehat{n}$. Potom pre všetky $k \in \widehat{n-1}$ platí*

$$\tilde{H}^{[k]}_{+}(\tilde{\boldsymbol{X}}) \geq \tilde{H}^{[k+1]}_{+}(\tilde{\boldsymbol{X}}). \tag{5.16}$$

Dôkaz. Zvoľme $k \in \widehat{n-1}$. Označme $\mathcal{N}^k_N(\mathring{\boldsymbol{X}})$ množinu nezáporných diagramov, ktoré spĺňajú podmienky určené entropiami do rádu k. Platí $\mathcal{N}^{k+1}_N(\mathring{\boldsymbol{X}}) \subset \mathcal{N}^k_N(\mathring{\boldsymbol{X}})$ a obidve množiny sú podľa predpokladu neprázdne. Maximálna entropia rádu k je definovaná ako suprémum I-miery na množine $\mathcal{N}^k_N(\mathring{\boldsymbol{X}})$. Z lemy 2.19 dostávame $\tilde{H}^{[k]}_{+}(\tilde{\boldsymbol{X}}) \geq \tilde{H}^{[k+1]}_{+}(\tilde{\boldsymbol{X}})$. □

Veta 5.34 (Ohraničenosť $\tilde{H}^{[k]}_{+}(\tilde{\boldsymbol{X}})$). *Nech $n \in \mathbb{N}$, $k \in \widehat{n}$. Majme diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$. Maximalizujme na nezáporných diagramoch entropiu pri entropiách $\tilde{\boldsymbol{X}}$ do rádu k, predpokladajme existenciu nezáporného diagramu spĺňajúceho väzbové podmienky. Potom platí*

$$\tilde{H}^{[k]}_{+}(\tilde{\boldsymbol{X}}) \geq 0 \tag{5.17}$$

$$\tilde{H}^{[k]}_{+}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} H(\tilde{X}_i) \tag{5.18}$$

$$\tilde{H}^{[k]}_{+}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|, \tag{5.19}$$

kde $\|\tilde{\mathcal{X}}_i\|$ značí počet prvkov v obore hodnôt $\tilde{X}_i$.

Dôkaz. Vzťah (5.17) vyplýva z toho, že I-miera diagramu je súčtom I-mier jednotlivých atómov diagramu a tie sú nezáporné. Vzťah (5.18) vyplýva z lemy 5.13. Lemu sme totiž ukázali pre všetky diagramy z množiny $\mathcal{S}_N(\mathring{\boldsymbol{X}})$ a vieme, že $\mathcal{N}_N(\mathring{\boldsymbol{X}}) \subset \mathcal{S}_N(\mathring{\boldsymbol{X}})$. Nerovnosť (5.19) vyplýva z vety 1.32, detaily nebudeme popisovať, postupovali by sme podobne, ako v dôkaze vety 5.15. □

Veta 5.35 (Ohraničenosť $I^{[k]}_{+}(\tilde{\boldsymbol{X}})$). *Nech $n \in \mathbb{N}$, $k \in \widehat{n}$. Majme diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$. Maximalizujme na nezáporných diagramoch entropiu pri entropiách $\tilde{\boldsymbol{X}}$ do rádu k, predpokladajme existenciu nezáporného diagramu* *splňajúceho väzbové podmienky. Potom platí*

$$I_+^{[k]}(\tilde{\boldsymbol{X}}) \geq 0 \tag{5.20}$$

$$I_+^{[k]}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} H(\tilde{X}_i) \tag{5.21}$$

$$I_+^{[k]}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|, \tag{5.22}$$

kde $\|\tilde{\mathcal{X}}_i\|$ *značí počet prvkov v obore hodnôt* $\tilde{X}_i$.

Dôkaz. Postupovali by sme podobne, ako v dôkaze vety 5.17. □

Veta 5.36 (Ohraničenosť $\sum I_+$)**.** *Nech* $n \in \mathbb{N}$, $k \in \widehat{n}$. *Majme* $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$. *Maximalizujme na nezáporných diagramoch entropiu pri entropiách* $\tilde{\boldsymbol{X}}$, *predpokladajme existenciu nezáporného diagramu splňajúceho väzbové podmienky. Potom platí*

$$\sum_{k=2}^{n} I_+^{[k]}(\tilde{\boldsymbol{X}}) \leq \sum_{i=1}^{n} H(\tilde{X}_i) \leq \sum_{i=1}^{n} \log_2 \|\tilde{\mathcal{X}}_i\|. \tag{5.23}$$

Dôkaz. Dokázali by sme podobne, ako vetu 2.60. □

Veta 5.37 (Nutná a postačujúca podmienka nezávislosti cez $\tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}})$)**.** *Majme diskrétnu vektorovú náhodnú veličinu* $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ *s konečným oborom hodnôt. Maximalizujme na nezáporných diagramoch entropiu pri entropiách veličiny* $\tilde{\boldsymbol{X}}$ *do určitého rádu. Potom zložky* $\tilde{\boldsymbol{X}}$ *sú nezávislé práve vtedy, keď je postupnosť maximálnych entropií pri zadaných entropiách konštantná, teda platí*

$$\tilde{H}_+^{[1]}(\tilde{\boldsymbol{X}}) = \tilde{H}_+^{[2]}(\tilde{\boldsymbol{X}}) = \ldots = \tilde{H}_+^{[n-1]}(\tilde{\boldsymbol{X}}) = \tilde{H}_+^{[n]}(\tilde{\boldsymbol{X}}).$$

Dôkaz. Budeme postupovať podobne, ako v dôkaze vety 5.19. Konečnosť oboru hodnôt $\tilde{\boldsymbol{X}}$ zaručuje konečné hodnoty entropií, ktoré použijeme ako väzbové podmienky. Ukážeme najprv nutnú podmienku nezávislosti. Majme entropie nezávislej náhodnej veličiny, použime ich ako väzbové podmienky pre maximalizáciu entropie na nezáporných diagramoch. Diagram vytvorený z disjunktných množín tieto podmienky splňa, a to platí pre všetky rády zadaných väzieb. Vďaka tomu, že entropia diskrétnej náhodnej veličiny je vždy nezáporná vieme, že disjunktný diagram je nezáporný. Ďalej z lemy 5.13 a poznámky 5.14 vieme, že žiadny shannonovský číselný diagram nemôže mať väčšiu I-mieru, než disjunktný, a keďže nezáporné diagramy sú shannonovské, platí to aj pre ne. Z uvedeného vyplýva, že maximálna entropia na nezáporných diagramoch pri zadaní entropií do rádu k pre ľubovoľné $k \in \widehat{n}$ je rovná I-miere disjunktného diagramu, t. j. súčtu entropií jednotlivých zložiek $\tilde{\boldsymbol{X}}$, a teda postupnosť je konštantná. Odôvodnime teraz postačujúcu podmienku. Majme konštantnú postupnosť entropií maximalizovaných na diagramoch. Vieme, že $\tilde{H}_+^{[1]}(\tilde{\boldsymbol{X}})$ je súčtom zadaných prvých entropií, a ako vieme z poznámky 5.14, túto I-mieru má z možných shannonovských a teda aj z možných nezáporných diagramov jedine disjunktný a ten je informačným diagramom $\tilde{\boldsymbol{X}}$, ktorá má nezávislé zložky. □

Veta 5.38 (Nutná a postačujúca podmienka nezávislosti cez $I_+^{[k]}(\tilde{\boldsymbol{X}})$)**.** *Majme diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s konečným oborom hodnôt. Maximalizujme na nezáporných diagramoch entropiu pri entropiách veličiny $\tilde{\boldsymbol{X}}$ do určitého rádu. Potom zložky $\tilde{\boldsymbol{X}}$ sú nezávislé práve vtedy, keď je postupnosť spojených informácii pri zadaných entropiách nulová, teda platí*

$$I_+^{[k]}(\tilde{\boldsymbol{X}}) = 0 \qquad \forall k \in \{2, ..., n\}.$$

Dôkaz. Vyplýva z vety 5.37. □

5.2.3 Vzťahy medzi konštrukciami

Ukázali sme niektoré základné vlastnosti alternatívnej konštrukcie maximálnej entropie pri zadaných entropiách počítanej na diagramoch, teraz preskúmame vzťahy ku ostatným konštrukciám. Budeme sa tiež zameriavať na to, aké výhody by mohla naša konštrukcia priniesť.

Veta 5.39 (Vzťah $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ a $\tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}})$)**.** *Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $k \leq n$, množinu $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$ takú, že $\operatorname{Ran} X_i \subset \mathbb{N}_0$ pre každé $i \in \widehat{n}$. Majme ďalej diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Nech informačný diagram pre veličinu $\tilde{\boldsymbol{X}}$ je nezáporný. Maximalizujme entropiu pri entropiách veličiny $\tilde{\boldsymbol{X}}$, predpokladáme ich konečnosť, maximalizujme na rozdeleniach a na nezáporných diagramoch. Potom platí*

$$\tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}}) \leq H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}})). \tag{5.24}$$

Dôkaz. $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ vzniká maximalizáciou entropie na rozdeleniach pravdepodobnosti, ktoré spĺňajú hodnoty daných entropií. Ku každému takémuto rozdeleniu môžeme jednoznačne priradiť informačný diagram, označme množinu týchto informačných diagramov B. Potom $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ môžeme chápať ako suprémum I-miery diagramov z množiny B. Ďalej $\tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}})$ vzniká maximalizáciou entropie na nezáporných diagramoch, ktoré spĺňajú podmienky na dané entropie. Ich množinu označíme ako A. Z vety 4.15 vieme, že ku každému nezápornému diagramu existuje rozdelenie pravdepodobnosti diskrétnej vektorovej náhodnej veličiny z $\{\boldsymbol{X}\}$, pokiaľ obor hodnôt X_i nie je ohraničený, čo je v našom prípade podľa predpokladu splnené. Platí teda, že všetky diagramy v množine A sú informačné pre nejaké $\boldsymbol{X} \in \{\boldsymbol{X}\}$, a teda $A \subset B$. Platí, že množiny A a B neprázdne, z lemy 2.19 potom vyplýva $\tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}}) \leq H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$. $\square$

Poznámka 5.40. Podobne ako v prípade vety 5.21, aj tu myslíme maximalizáciou entropie pri entropiách $\tilde{\boldsymbol{X}}$ to, že za množiny hodnôt maximalizačných podmienok z definícií $\tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}})$ a $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ volíme hodnoty príslušných entropií vypočítaných z rozdelenia veličiny $\tilde{\boldsymbol{X}}$.

Poznámka 5.41. Veta je zaujímavá tým, že novovytvorená alternatívna konštrukcia pri určitých predpokladoch poskytuje dolný odhad $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$, zatiaľ čo primárna konštrukcia poskytovala odhad zhora. Vieme, že dolný odhad $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ nám poskytuje aj $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$, na výpočet tejto veličiny však potrebujeme poznať marginály príslušného rádu. Na druhej strane, problémom na praktické použitie dolného odhadu pomocou alternatívnej konštrukcie je predpoklad neohraničeného oboru hodnôt skúmanej náhodnej veličiny.

Poznámka 5.42. Poznamenajme, že pre dvojzložkovú náhodnú veličinu majú všetky atómy diagramu, ktorý spĺňa Shannonove nerovnosti nezápornú I-mieru, pre tieto veličiny teda nastáva rovnosť entropie maximalizovanej na shannonovských a na nezáporných diagramoch.

Ďalej vyslovíme vzťah medzi novou alternatívnou a pôvodnou primárnou konštrukciou maximálnej entropie počítanej na diagramoch.

Veta 5.43 (Vzťah $\tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}})$ a $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$)**.** *Majme $k \in \mathbb{N}$, $n \in \mathbb{N}$ také, že $k \leq n$, množinu $\{\boldsymbol{X}\} = \{(X_1, \ldots, X_n)\}$ takú, že $\operatorname{Ran} X_i \subset \mathbb{N}_0$ pre každé $i \in \widehat{n}$. Majme ďalej diskrétnu vektorovú náhodnú veličinu $\tilde{\boldsymbol{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n) \sim \tilde{p}(\boldsymbol{x})$ s rovnakým oborom hodnôt. Nech informačný diagram pre veličinu $\tilde{\boldsymbol{X}}$ je nezáporný. Maximalizujme entropiu pri entropiách veličiny $\tilde{\boldsymbol{X}}$, predpokladáme ich konečnosť, maximalizujme na shannonovských a na nezá-*

porných diagramoch. Potom platí

$$\tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}}) \leq \tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}}). \tag{5.25}$$

Dôkaz. Veta je priamočiarym dôsledkom tvrdení 5.39 a 5.21. $\square$

Vzťahy medzi alternatívnou konštrukciou a maximálnou entropiou pri ostatných typoch väzbových podmienok prenechávame na budúce skúmanie. Na záver uvedieme príklad, v ktorom predstavíme numericky získané hodnoty viacerých typov maximálnej entropie a spojenej informácie pre konkrétny prípad rozdelenia pravdepodobnosti trojzložkovej náhodnej veličiny.

Príklad 5.44 (Numerický výpočet všetkých typov entropie)**.** Majme $\{\boldsymbol{X}\} = \{(X, Y, Z)\}$, s obormi hodnôt $\operatorname{Ran} X = \operatorname{Ran} Y = \operatorname{Ran} Z = \{1, 2, 3\}$. Vypočítame maximálnu entropiu typov $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$, $H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$, $\tilde{H}^{[k]}(\tilde{\boldsymbol{X}})$, $\tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}})$ pre $k \in \{1, 2, 3\}$, $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ pre $k \in \{1, 3\}$ a spojenú informáciu typov $I_c^{(k)}(\tilde{\boldsymbol{X}})$, $I_c^{<k>}(\tilde{\boldsymbol{X}})$, $I_S^{[k]}(\tilde{\boldsymbol{X}})$, $I_+^{[k]}(\tilde{\boldsymbol{X}})$ pre $k \in \{2, 3\}$ pri marginálach, resp. momentoch alebo entropiách získaných z rozdelenia veličiny $(\tilde{X}, \tilde{Y}, \tilde{Z}) \sim p$ s rovnakým oborom hodnôt, ako má (X, Y, Z), ktoré uvádzame v tabuľke 5.1.

$\tilde{x},\tilde{y},\tilde{z}$	$p(\tilde{x},\tilde{y},\tilde{z})$	$\tilde{x},\tilde{y},\tilde{z}$	$p(\tilde{x},\tilde{y},\tilde{z})$	$\tilde{x},\tilde{y},\tilde{z}$	$p(\tilde{x},\tilde{y},\tilde{z})$
1,1,1	1/32	2,2,1	1/32	2,2,3	1/32
2,2,2	1/32	2,1,2	1/32	2,3,2	1/32
3,3,3	1/32	1,2,2	1/32	3,2,2	1/32
2,1,1	1/32	3,3,1	1/16	1,2,3	1/32
1,2,1	1/32	3,1,3	1/16	2,3,1	1/16
1,1,2	1/32	1,3,3	1/32	3,1,2	1/16
3,1,1	1/32	2,3,3	1/32	1,3,2	1/16
1,3,1	1/32	3,2,3	1/32	3,2,1	1/32
1,1,3	1/32	3,3,2	1/32	2,1,3	1/32

Tabuľka 5.1: Rozdelenie pravdepodobnosti $(\tilde{X}, \tilde{Y}, \tilde{Z})$ z príkladu 5.44

Explicitne nebudeme uvádzať hodnoty momentov a entropií pri ktorých maximalizujeme, je možné ich ľahko dopočítať z hodnôt v tabuľke. Riešenie sme hľadali numericky pomocou lineárnej, resp. nelineárnej optimalizácie. Využili sme programovací jazyk Julia, balíček JuMP.jl a nelineárny riešič Ipopt. Požadovali sme splnenie väzbových podmienok s presnosťou 10^{-14}. Pomocou tohto riešiča sme neboli schopní nájsť hodnotu $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$, tento problém sme však očakávali, spomínajú ho už autori článkov [1, 2]. Pre $k \in \{1, 3\}$ vieme hodnotu tohto typu entropie určiť aj bez riešenia maximalizačného problému, pre $k = 1$ je totiž maximálna entropia rovná súčtu zadaných hodnôt prvej entropie, pre $k = 3$ je maximálna entropia rovná hodnote združenej entropie, ktorá je zadaná ako väzba.

V tabuľke 5.2 uvádzame vypočítané hodnoty maximálnej entropie, v tabuľke 5.3 hodnoty spojenej informácie v bitoch zaokrúhlené na 4 desatinné miesta.

k	$H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$	$H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$	$\tilde{H}^{[k]}(\tilde{\boldsymbol{X}})$	$\tilde{H}^{[k]}_{+}(\tilde{\boldsymbol{X}})$	$H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$
1	4,7379	4,7485	4,7379	4,7379	4,7379
2	4,7132	4,7183	4,7136	4,7136	-
3	4,6875	4,7034	4,6875	4,6875	4,6875

Tabuľka 5.2: Hodnoty maximálnej entropie z príkladu 5.44

k	$I_c^{(k)}(\tilde{\boldsymbol{X}})$	$I_c^{<k>}(\tilde{\boldsymbol{X}})$	$I_S^{[k]}(\tilde{\boldsymbol{X}})$	$I_{+}^{[k]}(\tilde{\boldsymbol{X}})$
2	0,0247	0,0275	0,0243	0,0243
3	0,0257	0,0149	0,0261	0,0261

Tabuľka 5.3: Hodnoty spojenej informácie z príkladu 5.44

Jedna z horných hraníc z vety o ohraničenosti maximálnych entropií je v tvare $H(X)+H(Y)+H(Z)$, v našom prípade má táto hranica hodnotu po zaokrúhlení $4,7379$. Vieme, že túto hodnotu dosiahneme v prípade $H(\tilde{P}^{(1)}(\tilde{\boldsymbol{X}}))$, $H(\tilde{P}^{[1]}(\tilde{\boldsymbol{X}}))$, $\tilde{H}^{[1]}(\tilde{\boldsymbol{X}})$, a keďže v prípade maximalizácie pri prvých entropiách je maximálne entropický diagram nezáporný, tak hranicu dosiahneme aj pre $\tilde{H}^{[1]}_{+}(\tilde{\boldsymbol{X}})$. Z tabuľky 5.2 tiež vidíme, že $H(\tilde{P}^{<1>}(\tilde{\boldsymbol{X}}))$ hodnotu tejto hranice prekračuje. Okrem toho, hodnota $H(\tilde{P}^{<3>}(\tilde{\boldsymbol{X}}))$ je väčšia, než hodnota skutočnej združenej entropie. Ďalej môžeme vidieť monotóniu maximálnych entropií jednotlivých typov.

V tabuľke 5.2 ďalej môžeme vidieť, že $\tilde{H}^{[2]}(\tilde{\boldsymbol{X}}) = \tilde{H}^{[2]}_{+}(\tilde{\boldsymbol{X}})$. Na základe článku [1] je táto rovnosť očakávaná, autori totiž našli tvar maximálne entropického diagramu pre zadané entropie do druhého rádu, a ako sme už spomínali v poznámke 5.29, tento diagram je nezáporný. Pokiaľ by sme vypočítali I-mieru maximálne entropického diagramu z článku pre naše rozdelenie, označme tento diagram ako D, dostali by sme po zaokrúhlení výsledok $\mu_I(D) = H(X)+H(Y)+H(Z)-I(X;Y)-I(Y;Z) \doteq 4,7136$, čo je hodnota rovnaká, ako sme získali optimalizáciou.

Maximálnu entropiu pri zadaných entropiách do druhého rádu vypočítanú na rozdeleniach, t. j. $H(\tilde{P}^{[2]}(\tilde{\boldsymbol{X}}))$, sa nám nepodarilo nájsť numericky a ani analyticky, máme však k dispozícii horný odhad, ktorý nám udáva hodnota $\tilde{H}^{[2]}(\tilde{\boldsymbol{X}})$. Zároveň vieme, že

$H(\tilde{P}^{[2]}(\tilde{\boldsymbol{X}})) \geq H(\tilde{P}^{(2)}(\tilde{\boldsymbol{X}}))$, tento vzťah udáva dolný odhad maximálnej entropie, celkovo teda $H(\tilde{P}^{[2]}(\tilde{\boldsymbol{X}})) \in (4,7132; 4,7136)$. Pripomeňme, že nemôžeme tvrdiť, že by náhodná veličina s nami zadaným oborom hodnôt mala maximálnu entropiu pri zadaných entropiách druhého rádu počítanú na rozdeleniach rovnú $\tilde{H}^{[2]}(\tilde{\boldsymbol{X}})$. Vieme síce, že pre náš prípad existuje rozdelenie trojzložkovej náhodnej veličiny, ktoré má informačný diagram rovnaký, ako maximálne entropický diagram z článku, nemáme však zaistené, že táto trojzložková veličina má rovnaký obor hodnôt, ako $\{\boldsymbol{X}\}$.

Uveďme ešte niekoľko komentárov k hodnotám spojenej informácie v tabuľke 5.3. Vidíme, že všetky druhy spojenej informácie všetkých rádov sú nenulové, môžeme tvrdiť, že medzi zložkami popisovanej náhodnej veličiny sa vyskytujú interakcie rádov 2 a 3. Hodnotu spojenej informácie typu $I_c^{[k]}(\tilde{\boldsymbol{X}})$ sme v tabuľke neuvideli, pretože nepoznáme jej presnú hodnotu. Na základe odhadu hodnoty $H(\tilde{P}^{[2]}(\tilde{\boldsymbol{X}}))$ však môžeme určiť aspoň interval, v ktorom hodnoty príslušnej spojenej informácie určite ležia. Platí $I_c^{[2]}(\tilde{\boldsymbol{X}}) \in (0,0243; 0,0247)$, $I_c^{[3]}(\tilde{\boldsymbol{X}}) \in (0,0257; 0,0261)$.

Týmto príkladom uzatvárame skúmanie poslednej konštrukcie maximálnej entropie a tým pádom aj celý prieskum vlastností maximálne entropických funkcionálov.

Záver

V práci sme preskúmali jednu z možností popisu interakcií v komplexných systémoch pomocou maximálne entropických funkcionálov. V prvej časti textu sme predstavili definície troch typov maximálnej entropie, ktoré boli pôvodne popísané v článkoch [1, 2, 3], kde však objekty neboli definované rigorózne, ale len skratkovito. Formálna definícia nám následne umožnila preskúmať vlastnosti funkcionálov a formulovať ich vo vetách. Ukázali sme monotóniu postupnosti maximálnych entropií, našli sme horné a dolné hranice hodnôt funkcionálov, ukázali sme nutnú a postačujúcu podmienku nezávislosti zložiek vektorovej náhodnej veličiny, podľa ktorej entropií, resp. marginál maximalizujeme. Keďže tvrdenia mali obvykle tvar neostrých nerovností, snažili sme sa dopĺňať vety príkladmi na situácie, v ktorých nastáva ostrá nerovnosť resp. rovnosť.

vlastnosť	$H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$	$H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$	$H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$	$\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$	$\tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}})$
prvok množiny, kde sa maximalizuje H	rozdelenie	rozdelenie	rozdelenie	diagram	diagram
optimalizovaná funkcia	nelineárna	nelineárna	nelineárna	lineárna	lineárna
optimalizačné podmienky	lineárne	lineárne	nelineárne	lineárne	lineárne
možnosť riešiť pomocou JuMP.jl	✓	✓	✗	✓	✓
monotónia	✓	✓	✓	✓	✓
dolný odhad 0	✓	✓	✓	✓	✓
horný odhad $\sum_{i=1}^n H(\tilde{X}_i)$	✓	✗	✓	✓	✓
horný odhad $\sum_{i=1}^n \log_2 \|\tilde{\mathcal{X}}_i\|$	✓	✓	✓	✓	✓
vyjadrenie N a P podmienky nezávislosti	✓	✗	✓	✓	✓

Tabuľka 5.4: Zhrnutie základných vlastností maximálnej entropie

Už v tejto časti práce sme prvýkrát objavili výrazné odlišnosti medzi vlastnosťami rôznych konštrukcií, zistili sme totiž, že pre $H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$, teda pre maximálnu entropiu pri zachovaní momentov, neplatí jeden z horných odhadov jej hodnôt a tiež že nie sme schopní vyjadriť pomocou $H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$ nutnú a ani postačujúcu podmienku nezávislosti. Stretli sme sa aj so špecifickým problémom, na ktorý upozorňovali už autori článkov [1, 2], a tým je komplikované hľadanie maxima entropie, pokiaľ sú zadanými podmienkami entropie do určitého rádu a chceme maximalizovať na rozdeleniach. Celkový prehľad preskúmaných vlastností uvádzame v tabuľke 5.4.

Prvé tri typy maximálnej entropie nám umožnili definovať aj príslušné tri druhy spojenej informácie. Ukázali sme, že postupnosť spojených informácií ani jedného typu nie je monotónna, našli sme hornú a dolnú hranicu ich hodnôt a tiež hornú hranicu sumy spojených informácií, vyjadrili sme podmienky nezávislosti. Podobne, ako v prípade $H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$, aj príslušná spojená informácia $I_c^{<k>}(\tilde{\boldsymbol{X}})$ prejavuje odlišnosti od ostatných konštrukcií, opäť v jednom z horných odhadov a pri snahe o formuláciu podmienok nezávislosti. Prehľad vlastností spojenej informácie uvádzame v tabuľke 5.5.

vlastnosť	$I_c^{(k)}(\tilde{\boldsymbol{X}})$	$I_c^{<k>}(\tilde{\boldsymbol{X}})$	$I_c^{[k]}(\tilde{\boldsymbol{X}})$	$I_S^{[k]}(\tilde{\boldsymbol{X}})$	$I_+^{[k]}(\tilde{\boldsymbol{X}})$
monotónia	✗	✗	✗	✗	✗
dolný odhad 0	✓	✓	✓	✓	✓
horný odhad $\sum_{i=1}^n H(X_i)$	✓	✗	✓	✓	✓
horný odhad $\sum_{i=1}^n \log_2 \lVert\mathcal{X}_i\rVert$	✓	✓	✓	✓	✓
odhad $\sum_{k=2}^n I^k$ výrazom $\sum_{i=1}^n H(\tilde{X}_i)$	✓	✗	✓	✓	✓
odhad $\sum_{k=2}^n I^k$ výrazom $\sum_{i=1}^n \log_2 \lVert\tilde{\mathcal{X}}_i\rVert$	✓	✓	✓	✓	✓
vyjadrenie N a P podmienky nezávislosti	✓	✗	✓	✓	✓

Tabuľka 5.5: Zhrnutie základných vlastností spojenej informácie

Po prieskume základných vlastností prvých troch konštrukcií sme sa zamerali na otázku súvislosti vlastností postupností maximálnych entropií, resp. spojených informácií a prítomnosti interakcií v pozorovanom systéme. Definovali sme pojem interakcia a jeho dva druhy - implikovaná a neimplikovaná interakcia. V prvej časti tretej kapitoly sme sa sústredili na entropiu maximalizovanú pri zadaných marginálach. V rámci prípravy na vyslovenie a dôkaz viet o interakciách, ktoré sú jedny z kľúčových tvrdení celej práce, sme popísali možnosť faktorizácie rozdelenia vektorovej náhodnej veličiny s danými marginálami na súčin marginál určitých skupín zložiek. Následne sme formulovali vetu o implikovaných interakciách, ktorej hlavným tvrdením bola konštantnosť hodnôt postupnosti $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$ v neprítomnosti interakcií. Ďalej sme dokázali jeden možný varianty vety, v ktorom sme volili predpoklady tak, aby bolo zaručené ostré klesanie maximálnej entropie v prítomnosti interakcií. Následne sme dokázali, že ostrá nerovnosť medzi dvoma po sebe idúcimi rádmi maximálnej entropie implikuje prítomnosť interakcií. Celkovo sme teda formulovali podmienky na to, aby nenulovosť spojenej informácie bola ekvivalentná prítomnosti interakcií daného rádu, pre tento prípad sa teda veľkosť spojenej informácie stáva možným kandidátom na popis sily tejto interakcie.

V ďalšom postupe sme podobnú konštrukciu vybudovali pre maximálnu entropiu pri zadaných entropiách. Vyslovili sme vetu o implikovaných interakciách a jej možnú variantu so silnejšími predpokladmi. Podobne, ako v prípade maximalizácie pri marginálach sme ukázali aj tvrdenie o prítomnosti interakcií tam, kde postupnosť $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ ostro klesá. Opäť sme narazili na odlišnosť konštrukcie $H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$. Pre tento typ maximálnej entropie nebolo možné realizovať podobnú konštrukciu a neplatia analogické vzťahy medzi interakciami v systéme a vlastnosťami postupnosti maximálnych entropií. Tvrdenia o postupnosti maximálnych entropií súvisiace s interakciami v systéme sú zhrnuté v tabuľke 5.6.

vlastnosť	$H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$	$H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$	$H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$
zachytenie všetkých nelineárnych vzťahov	✓	✗	✓
konštrukcia maximálne entropického rozdelenia v tvare súčinu nižších marginál	✓	✗	✓
ostré klesanie entropie pri silne neimplikovaných interakciách	✓	✗	✓
rovnosť H rádu k a $k-1$ pri neprítomnosti interakcií rádu k	✓	✗	✓
nerovnosť H rádov k a $k-1$ implikuje prítomnosť interakcií rádu k	✓	✗	✓

Tabuľka 5.6: Zhrnutie vlastností maximálnej entropie súvisiacich s interakciami

Priamočiarym dôsledkom viet o interakciách pre maximálne entropie sú tieto vety vyslovené pre spojené informácie. Prehľad týchto tvrdení sa nachádza v tabuľke 5.7. Okrem nich posledný riadok tabuľky odkazuje na príklad, v ktorom sme ukázali, ako môžeme rozložiť celkovú informáciu systému na sumu spojených informácii, avšak opäť len pre dve možné konštrukcie $I_c^{(k)}(\tilde{\boldsymbol{X}})$ a $I_c^{[k]}(\tilde{\boldsymbol{X}})$.

vlastnosť	$I_c^{(k)}(\tilde{\boldsymbol{X}})$	$I_c^{<k>}(\tilde{\boldsymbol{X}})$	$I_c^{[k]}(\tilde{\boldsymbol{X}})$
kladná hodnota pri silne neimplikovaných interakciách	✓	✗	✓
nulovosť rádu k pri neprítomnosti interakcií rádu k	✓	✗	✓
nenulovosť I_c rádu k implikuje prítomnosť interakcií rádu k	✓	✗	✓
možnosť rozložiť celkovú informáciu systému	✓	✗	✓

Tabuľka 5.7: Zhrnutie vlastností spojenej informácie súvisiacich s interakciami

V štvrtej kapitole sme sa venovali teórii potrebnej na vybudovanie ďalších dvoch konštrukcií maximálnej entropie. Ich hlavným významom je aproximovať hodnoty $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$. Konštrukcia $\tilde{H}_S^{[k]}(\tilde{\boldsymbol{X}})$ je prvýkrát spomínaná už v článku [2], avšak $\tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}})$ je nový nami navrhnutý funkcionál. Časť piatej kapitoly je venovaná základným vlastnostiam týchto konštrukcií, opäť sme ukázali monotóniu, ohraničenosť, podmienky nezávislosti, tieto vlastnosti sú zhrnuté v tabuľke 5.4. Podobne, ako pre entropie maximalizované na rozdeleniach, aj pre tieto dva typy maximálnej entropie sme zaviedli príslušné spojené informácie a ukázali ich základné vlastnosti. Zhrnuté sú v tabuľke 5.5.

V priebehu práce bolo našou snahou všímať si spoločné vlastnosti a rozdiely medzi jednotlivými konštrukciami. Niektoré zo vzťahov sme boli schopní vyjadriť v podobe nerovností, iné však zostávajú otvorené na budúce skúmanie. Sú to predovšetkým nerovnosti medzi $H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$ a ostatnými konštrukciami, teda aj v tomto prípade sa prejavili špecifické vlastnosti konštrukcie maximálnej entropie pri zadaných momentoch. Ďalšími vlastnosťami, ktoré sú pre nás zaujímavé, sú vzťahy pôvodných štyroch konštrukcií a nami novodefinovanej $\tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}})$, t. j. entropie maximalizovanej na nezáporných diagramoch. Okrem iného sme ukázali, že $\tilde{H}_+^{[k]}(\tilde{\boldsymbol{X}})$ je dolným odhadom $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$, čo je dôležité zistenie. Toto tvrdenie sme však dokázali len pre spočítateľné nosiče hustôt skúmaných náhodných veličín. Otvorenou otázkou teda zostáva prípadné zovšeobecnenie vety aj pre konečné nosiče.

Ďalšia perspektívna oblasť budúceho skúmania sa otvára v rámci prieskumu súvislosti postupnosti maximálnych entropií a interakcií. Vytvorili sme spôsob zápisu rozdelení konzistentných so zadanými marginálami, resp. entropiami, čo nám umožnilo dokázať niekoľko viet o interakciách. Zaujímavou otázkou na budúce skúmanie sú ďalšie možné varianty už dokázaných viet, ktoré by upresnili popis toho, kedy postupnosť maximálnych entropií musí ostro klesať, kedy naopak vieme ukázať len neostrú nerovnosť, pričom môže byť nápomocný práve nami vybudovaný dôkazový aparát. Skúmanie týchto otázok je dôležité aj z hľadiska interpretácie. V práci sme už upozornili na to, že rád interakcií a rád, v ktorom klesá postupnosť maximálnych entropií, nemusí mať tak jasnú spojitosť, ako je pôvodne navrhnuté v článkoch [1, 2]. Bolo by prínosné túto informáciu ďalej spresniť. Ďalšou otázkou súvisiacou s touto oblasťou je prieskum toho, či $H(\tilde{P}^{[k]}(\tilde{\boldsymbol{X}}))$ a $H(\tilde{P}^{(k)}(\tilde{\boldsymbol{X}}))$ klesajú v prítomnosti interakcií vždy práve súčasne, alebo sú rozdiely medzi týmito dvoma konštrukciami. V práci sme tento problém vyriešili pre prípad neimplikovaných interakcií.

Otázky spojené s postupnosťou maximálnych entropií samozrejme vytvárajú aj priestor na skúmanie príslušných spojených informácii. Zaujímavým problémom je napríklad explicitné vytvorenie postupnosti spojených informácii niektorého typu s danými hodnotami a následné skúmanie, aké sú spojené informácie ostatných druhov pri takýchto podmienkach. Táto otázka úzko súvisí s prieskumom interakcií, jedným z perspektívnych spôsobov dôkazu by mohlo byť práve vytváranie konštrukcie postupným pridávaním interakcií do systému, na čo sme už v práci poskytli viaceré užitočné informácie.

Okrem otvorených otázok v oblasti dôkazov nových vlastností vzniká aj priestor na hľadanie príkladov alebo protipríkladov na konkrétne vlastnosti, nerovnosti a situácie. V práci sme ich uviedli značné množstvo, hľadali sme numerické a aj analytické riešenia problémov, ďalších možností je ale stále mnoho. Napokon, zaujímavou oblasťou na premýšľanie je aj spôsob naprogramovania optimalizácie pri rôznych podmienkach.

Poslednou skupinou otázok, na ktorú upozorníme, sú vlastnosti konštrukcie $H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$, t. j. entropie maximalizovanej pri zadaných momentoch. V priebehu práce sme objavili jej viaceré odlišnosti od ostatných typov maximálnej entropie. V príkladoch sme videli možnosť špecifikovať nosiče hustoty náhodnej veličiny, pri ktorej momentoch maximalizujeme, skúmali sme napríklad binárne veličiny. Otázkou, ktorú sme v práci neskúmali je, či konkrétne určenie nosičov, napríklad fixovanie na množinu $\{1, \ldots, m\}$ pre nejaké $m \in \mathbb{N}$, môže byť zaujímavé aj v kontexte ďalších tvrdení. Zaujímali by nás hlavne predpoklady, pri ktorých by platili vlastnosti, v ktorých sa táto konštrukcia odlišuje od ostatných, čo sú napríklad možnosť ohraničenia hodnoty maximálnej entropie zhora sumou prvých entropií alebo vyjadrenie nutnej a postačujúcej podmienky nezávislosti zložiek popisovaného systému pomocou $H(\tilde{P}^{<k>}(\tilde{\boldsymbol{X}}))$.

Vzhľadom na to, že sme budovali všetku potrebnú teóriu od základu, nie je množstvo otázok, ktoré naša práca otvára prekvapivé, je skôr reflexiou faktu, že vybudované štruktúry majú perspektívne vlastnosti a možnosti využitia a budúceho skúmania. K veľkej časti otázok sme tiež popísali potrebné súvislosti a vytvorili sme nové koncepty niekoľkých dôkazových schém, ktoré môžu byť užitočné. Za hlavný prínos práce považujeme okrem formalizácie a prieskumu základných vlastností konštrukcií hlavne prepojenie výskytu interakcií s hodnotami funkcionálov a tiež vytvorenie novej konštrukcie odhadu maximálnej entropie.

Literatúra

[1] E. A. Martin, J. Hlinka, A. Meinke, F. Děchtěrenko, J. Tintěra, I. Oliver, and J. Davidsen, "Network inference and maximum entropy estimation on information diagrams," *Scientific Reports*, vol. 7, p. 7062, 2017.

[2] E. A. Martin, J. Hlinka, and J. Davidsen, "Pairwise network information and nonlinear correlations," *Physical Review E*, vol. 94, p. 040301, 2016.

[3] E. Schneidman, S. Still, M. J. Berry, and W. Bialek, "Network information and connected correlations.," *Physical review letters*, vol. 91, p. 238701, 2003.

[4] J. Jacod and P. Protter, *Probability Essentials.* Springer, 2012.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley-Interscience, 2006.

[6] T. Van de Cruys, "Two multivariate generalizations of pointwise mutual information," in *Proceedings of the Workshop on Distributional Semantics and Compositionality*, p. 16, Association for Computational Linguistics, 2011.

[7] N. Ay, D. Polani, and N. Virgo, "Information decomposition based on cooperative game theory," *CoRR*, vol. 1910, p. 05979, 2019.

[8] R. Yeung, *Information Theory and Network Coding.* Springer, 2008.