

Poznámky k předmětu Statistika 3

Jan Kracík

11. dubna 2014

Značení

- *i.i.d.* - nezávislé a stejně rozdělené (náhodné veličiny)

Není-li řečeno jinak, pak:

- Nerozlišujeme náhodnou veličinu a její hodnotu. Význam bude vždy zřejmý z kontextu.
- Hustoty pravděpodobnosti jsou zapisovány bez explicitně uvedených náhodných veličin. Ty jsou jednoznačně určeny argumenty.

1 Úvod

Kdykoliv stojíme před volbou jedné z alespoň dvou různých možností, řešíme jistou rozhodovací úlohu. V případě, že nejsme z nějakého důvodu schopni přesně určit důsledky jednotlivých rozhodnutí, mluvíme o úloze rozhodování za neurčitosti. Téměř každá skutečná rozhodovací úloha je přitom do jisté míry neurčitostí zatížena. Je přirozené požadovat, aby rozhodnutí byla určitým způsobem racionální, nebo dokonce v jistém smyslu co nejlepší. Abychom byli schopni tomuto požadavku vyhovět, je obecně potřeba v rámci možností minimalizovat míru neurčitosti v dané úloze, to znamená získat a využít informace, které umožní co nejlépe předpovídat důsledky jednotlivých rozhodnutí.

Existuje řada možností, jak v konkrétních případech reprezentovat neurčitost, například pomocí nástrojů teorie pravděpodobnosti, fuzzy teorie, nepřesné pravděpodobnosti, nebo possibilistické teorie. Dominantní roli mezi těmito přístupy nepochybně hrají pravděpodobnostní modely. Nástrojem, který v takovém případě umožňuje získávat a zpracovávat informace vedoucí ke snížení neurčitosti je pak matematická statistika.

Patrně nejpropracovanějším přístupem k rozhodování za neurčitosti je tzv. Bayesovská teorie. Tato teorie bývá označována jako normativní, což znamená, že rozhodovací strategie na ní založené jsou navrhovány tak, aby splňovaly určité předem dané podmínky, díky kterým lze tyto strategie považovat za v jistém smyslu racionální a konzistentní. Nejde tedy o deskriptivní teorii, která by si kladla za cíl modelovat skutečné rozhodování nějaké skupiny v konkrétních podmínkách. V rámci bayesovské teorie je neurčitost v rozhodovací úloze reprezentována pomocí pravděpodobnosti a veličiny jejichž hodnoty nelze přesně určit jsou považovány za náhodné. Cíle rozhodování jsou popsány pomocí tzv. ztrátové funkce, která závisí na zvoleném rozhodnutí a dalších (náhodných) veličinách. Optimální rozhodnutí jsou pak navrhována tak, aby minimalizovala střední hodnotu ztrátové funkce.

Nedílnou součástí návrhu optimální strategie je získávání informací z dat, díky čemuž je snižována míra neurčitosti. Protože všechny neznámé veličiny jsou považovány za náhodné, redukuje se tento proces na využití Bayesova vzorce.

2 Bayesovská rozhodovací úloha

Bayesovské metody jsou založeny na jednoduché ale přitom geniální myšlence, se kterou přišel v 18. století Thomas Bayes: *Všechny neznámé parametry vyskytující se v úloze jsou považovány za náhodné veličiny.*

Uvažujeme určitý systém popsaný (vektorovou) náhodnou veličinou x , jejíž rozdělení závisí parametru θ , jehož hodnota není známa. Z Bayesovy myšlenky okamžitě plynou následující důsledky:

- *Statistický model* ve smyslu parametrizovaného systému pravděpodobnostních rozdělení¹ náhodné veličiny x je z bayesovského pohledu reprezentován podmíněnou pravděpodobností²

$$f(x|\theta). \quad (1)$$

- Pro náhodnou veličinu θ reprezentující neznámý parametr v modelu (1) lze zvolit hustotu pravděpodobnosti

$$f(\theta). \quad (2)$$

Tato hustota se nazývá *apriorní hustota* a v bayesovských metodách má zásadní význam. Volbou apriorní hustoty lze do úlohy vnést tzv. apriorní informaci, což je informace o parametru, která je dostupná před tím, než jsou pozorována jakákoli data. Zdrojem takovéto informace může být např. fyzikální model nebo dřívější zkušenost s podobnými problémy. Prostřednictvím apriorní hustoty lze také vyjádřit předem známá omezení pro hodnotu neznámého parametru. Apriorní rozdělení ale může být rovněž zdrojem problémů, neboť v některých případech nemusí být vhodná apriorní informace dostupná. Bez apriorního rozdělení přitom bayesovskou rozhodovací úlohu nelze zformulovat. V takových případech lze ale využít tzv. neinformativní apriorní rozdělení.

- Z apriorního rozdělení (2) a modelu (1) lze pomocí Bayesova vzorce odvodit tzv. *aposteriorní hustotu* $f(\theta|x)$,

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)}, \quad (3)$$

kde marginální hustotu $f(x)$ získáme integrací sdružené hustoty $f(x, \theta)$,

$$f(x) = \int f(x, \theta)d\theta = \int f(x|\theta)f(\theta)d\theta.^3$$

Podmíněná hustota (3) reprezentuje celkovou informaci o parametru θ , která v sobě zahrnuje jednak apriorní informaci a dále informaci získanou pozorováním dat x .

- Pokud v úloze uvažujeme další náhodnou veličinu y , která na parametru θ a veličině x závisí prostřednictvím podmíněné hustoty

$$f(y|x, \theta),$$

lze z ní s využitím aposteriorní hustoty (3) odvodit podmíněnou hustotu

$$f(y|x) = \int f(y, \theta|x)d\theta = \int f(y|x, \theta)f(\theta|x)d\theta. \quad (4)$$

Podmíněná hustota (4) se nazývá *prediktivní hustota*. Tato hustota v jistém smyslu předpovídá “chování” náhodné veličiny y na základě pozorování veličiny x ale také apriorní informace $f(\theta)$.

¹Všude v textu pro zjednodušení předpokládáme, že uvažovaná pravděpodobnostní rozdělení mají hustoty vzhledem k Lebesgueově míře (pro spojité náhodné veličiny) nebo k čítcí míře (pro diskrétní náhodné veličiny).

²V celém textu nebudeme značením rozlišovat náhodnou veličinu a její hodnotu, bude-li význam zřejmý z kontextu. Dále nebudeme u hustot pravděpodobnosti apod. explicitně zapisovat, kterých náhodných veličin se týkají, neboť toto bude jednoznačně určeno argumenty. Např. zápis $f(x|y)$ je zjednodušené vyjádření podmíněné hustoty $f_{X|Y}(x|y)$ náhodné veličiny X za podmínky $Y = y$.

³Nebude-li řečeno jinak, chápeme všechny intergály v textu jako určité integrály přes celý obor hodnot příslušné náhodné veličiny.

Poznámky:

1. Neznámý parametr θ nemusí mít jen význam parametru pravděpodobnostního rozdělení, ale může jít obecně o jakoukoliv veličinu, na níž rozdělení veličiny x závisí prostřednictvím modelu (1). Kromě parametrů rozdělení v běžném smyslu může jít například o fyzikální veličiny, které nelze pozorovat přímo. Příkladem takové veličiny může být poloha a rychlost letadla v určitém čase, přičemž pozorovanou veličinou x by v tomto případě mohly být údaje naměřené radarem.
2. Náhodná veličina x stejně jako parametr θ mohou být obecně vektorové. Náhodný vektor x může reprezentovat posloupnost nezávislých stejně rozdělených pozorování, nebo také nějakým způsobem závislá data. Například v případě statistického zpracování obrazové informace může x reprezentovat náhodný vektor dimenze v řádu milionů obsahující pozorované intenzity jasu pro jednotlivé obrazové body, parametr θ pak může představovat vektor skutečných hodnot jasu. Význam modelu (1) i apriorní hustoty (2) se tím nijak nemění, stejně jako zůstávají v platnosti vztahy (3) a (4). Z praktického hlediska ale může jít o velmi složité objekty a například integrály v (3) a (4) nejen, že často nelze spočítat analyticky, ale kvůli obrovské dimenzi nelze využít ani běžných numerických metod. Díky vzorkovacím metodám (MCMC) nemusí ani takto složité úlohy představovat neřešitelný problém.
3. Bayesův vzorec představuje mechanismus učení, kdy je k apriorní informaci reprezentované apriorní hustotou $f(\theta)$ přidávána informace získaná z dat a celková dostupná informace o parametru θ je pak reprezentována aposteriorní hustotou $f(\theta|x)$. Tento proces učení je názorně vidět v případě, kdy vektorová veličina $x = (x_1, \dots, x_t)$ představuje posloupnost náhodných veličin, jejichž hodnoty jsou získávány postupně a index tak lze interpretovat jako čas. Pro zjednodušení předpokládáme, že pozorování jsou nezávislá a pro model tedy platí

$$f(x_1, x_2, \dots, x_t|\theta) = \prod_{\tau=1}^t f(x_\tau|\theta),$$

z čehož plyne pro všechna $\tau \in \{1, \dots, t\}$

$$f(x_\tau|x_1, \dots, x_{\tau-1}, \theta) = f(x_\tau|\theta). \quad (5)$$

Aposteriorní hustotu $f(\theta|x)$ lze vyjádřit rekurentně pomocí vztahů

$$f(\theta|x_1) = \frac{f(x_1|\theta)f(\theta)}{\int f(x_1|\theta)f(\theta)d\theta}$$

a pro $\tau = 2, \dots, t$

$$f(\theta|x_1, \dots, x_\tau) = \frac{f(x_\tau|x_1, \dots, x_{\tau-1}, \theta)f(\theta|x_1, \dots, x_{\tau-1})}{\int f(x_\tau|x_1, \dots, x_{\tau-1}, \theta)f(\theta|x_1, \dots, x_{\tau-1})d\theta} \quad (6)$$

$$= \frac{f(x_\tau|\theta)f(\theta|x_1, \dots, x_{\tau-1})}{\int f(x_\tau|\theta)f(\theta|x_1, \dots, x_{\tau-1})d\theta}, \quad (7)$$

kde (7) plyne z (6) díky (5). Podmíněnou hustotu $f(\theta|x_1, \dots, x_\tau)$ můžeme interpretovat jako aposteriorní hustotu v čase τ . Vztah (7) pak není nic jiného než Bayesův vzorec, v němž je jako apriorní hustota použita aposteriorní hustota $f(\theta|x_1, \dots, x_{\tau-1})$ z času $\tau - 1$ a jako data zde vystupuje pouze nové pozorování x_τ .

4. Vztah (3) pro výpočet aposteriorní hustoty se často zapisuje ve tvaru

$$f(\theta|x) \propto f(x|\theta)f(\theta), \quad (8)$$

kde symbol \propto značí, že levá strana je úměrná pravé až na normalizační člen nezávislý na θ . Tento člen je jednoznačně určen podmínkou $\int f(\theta|x)d\theta = 1$ a je roven

$$\frac{1}{f(x)} = \frac{1}{\int f(x|\theta)f(\theta)d\theta}.$$

Použití tohoto zápisu v praxi zjednodušuje postup odvození aposteriorní hustoty.

Aposteriorní hustota hraje v bayesovském přístupu k rozhodování za neurčitosti klíčovou roli, neboť reprezentuje celkovou informaci o neznámém parametru. Přesto je většinou jen meziproductem v rozhodovací procesu, který směřuje k volbě určitého rozhodnutí. Tímto rozhodnutím může být určité rozhodnutí v běžném smyslu (prodat nebo neprodat akcie), řídicí zásah v nějakém systému (nastavitelná velikost proudu procházející motorem), stanovení předpovědi (množství srážek během následujícího dne), ale může například i běžné statistické odhady. Stanovení hodnoty bodového odhadu, nebo třeba test hypotézy jsou z bayesovského pohledu rozhodovací úlohy.

Každé racionální rozhodování sleduje určitý cíl. V rámci bayesovského přístupu jsou cíle rozhodování specifikovány pomocí tzv. ztrátové funkce, která každému rozhodnutí přiřazuje určitou hodnotu ve smyslu ztráty, kterou toto rozhodnutí způsobí, a to v závislosti na hodnotě parametru θ . Označíme-li \mathcal{A} množinu všech rozhodnutí, pak *ztrátová funkce* je libovolná funkce

$$L : \mathcal{A} \times \Theta \rightarrow \mathbb{R}^+, \quad (9)$$

kde Θ je obor hodnot náhodné veličiny θ . Za optimální rozhodnutí a^{opt} pak považujeme rozhodnutí, které minimalizuje střední hodnotu ztrátové funkce vzhledem k aposteriornímu rozdělení, tj.

$$a^{opt} \in \underset{a \in \mathcal{A}}{\operatorname{Argmin}} \int L(a, \theta) f(\theta|x) d\theta. \quad (10)$$

Optimální rozhodnutí obecně nemusí existovat a pokud existuje, nemusí být jednoznačné.

2.1 Shrnutí

Formulaci bayesovské úlohy můžeme shrnout do následujících kroků:

1. Pro daný problém specifikujeme
 - statistický model: $f(x|\theta)$,
 - apriorní hustotu pravděpodobnosti: $f(\theta)$,
 - množinu rozhodnutí: \mathcal{A} ,
 - ztrátovou funkci: $L : \mathcal{A} \times \Theta \rightarrow \mathbb{R}^+$.
2. Optimální rozhodnutí $a^{opt} \in \mathcal{A}$ hledáme tak, aby splňovalo podmínku

$$a^{opt} \in \underset{a \in \mathcal{A}}{\operatorname{Argmin}} \int L(a, \theta) f(\theta|x) d\theta,$$

kde

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta}.$$

Poznámky:

⁴Argmin je použito pro označení množiny všech bodů, pro něž daná funkce nabývá svého minima.

1. Výše uvedená struktura rozhodovací úlohy je přímo aplikovatelná na relativně jednoduché úlohy, kdy na základě apriorní informace a dat hledáme optimální rozhodnutí. Takovéto úlohy bývají označovány jako statické. Stejný princip však můžeme uplatnit např. i při řízení dynamického systému, kdy jsou rozhodnutí generována sekvenčně v návaznosti na postupně získávaná data, přičemž rozhodnutí zároveň ovlivňují budoucí chování celého systému. V takovém případě mluvíme o dynamickém rozhodování. Příkladem dynamického rozhodování může být např. řízení dopravy v určité oblasti, řízení mobilního robota v neznámém prostředí apod.
2. Z praktického hlediska mají bayesovské metody dvě hlavní výhody:
 - (a) Úlohy rozhodování za neurčitosti včetně dynamických úloh lze formulovat relativně jednoduše, přičemž je zaručena určitá racionalita rozhodovacího procesu (minimalizace očekávané ztráty).
 - (b) Prostřednictvím apriorní hustoty je explicitně reprezentována apriorní informace. V úlohách, kde se potýkáme z nedostatkem dat, je přitom důsledné využití apriorní informace nezbytné.
3. I když je struktura bayesovské rozhodovací úlohy v principu jednoduchá, jednotlivé kroky samy o sobě mohou být obtížné a často se neobejdou se bez vhodných aproximačních nástrojů. Až na jednodušší případy bývá zdrojem obtíží aposteriorní hustota, kterou často nelze vhodně vyjádřit. V případě dynamických úloh pak návrh optimální rozhodovací strategie vede k natolik výpočetně náročným úlohám, že je prakticky vždy nutné omezit se na hledání sub-optimálního řešení (zjednodušené úlohy).

3 Ilustrační příklady

Následující příklady jsou ukázkou možných aplikací bayesovské teorie.

Uvažujme nejprve běžnou úlohu odhadu parametru statistického modelu. Pro srovnání tuto úlohu budeme nejprve řešit klasicky. Sestrojíme maximálně věrohodný odhad a poté ještě najdeme intervalový odhad.

Příklad 1 Odhad střední hodnoty normálního rozdělení

Nechť x_1, x_2, \dots, x_t jsou i.i.d. náhodné veličiny s rozdělením s hustotou pravděpodobnosti

$$f_\mu(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2}\right) \quad (11)$$

závisející na neznámém parametru $\mu \in \mathbb{R}$. Díky předpokládané nezávislosti pro sdruženou hustotu pravděpodobnosti platí

$$f_\mu(x_1, x_2, \dots, x_t) = \prod_{\tau=1}^t f_\mu(x_\tau) = (2\pi)^{-\frac{t}{2}} \exp\left(-\frac{1}{2} \sum_{\tau=1}^t (x_\tau - \mu)^2\right). \quad (12)$$

Jestliže při daných hodnotách x_1, x_2, \dots, x_t chápeme sdruženou hustotu (12) jako funkci parametru μ , nazýváme ji věrohodnostní funkcí parametru μ . Často používaným typem bodového odhadu je tzv. maximálně věrohodný odhad, $\hat{\mu}_{ML} : \mathbb{R}^t \rightarrow \mathbb{R}$, který každé posloupnosti x_1, \dots, x_t přiřazuje hodnotu parametru, v níž věrohodnostní funkce nabývá svého maxima:

$$\hat{\mu}_{ML}(x_1, x_2, \dots, x_t) = \arg \max_{\mu \in \mathbb{R}} f_\mu(x_1, x_2, \dots, x_t).$$

Maximalizací věrohodnostní funkce (12) dostaneme maximálně věrohodný odhad modelu (11) ve tvaru

$$\hat{\mu}_{ML}(x_1, \dots, x_t) = \frac{1}{t} \sum_{\tau=1}^t x_\tau.$$

Při hledání intervalového odhadu vyjdeme ze skutečnosti, že při známém rozptylu σ^2 má statistika

$$\frac{\frac{1}{t} \sum_{\tau=1}^t x_{\tau} - \mu}{\sigma} \sqrt{t}$$

rozdělení $\mathcal{N}(0, 1)$. Odtud dojdeme k intervalovému odhadu

$$\left\langle \frac{1}{t} \sum_{\tau=1}^t x_{\tau} - \frac{1}{\sqrt{t}} z_{1-\frac{\alpha}{2}}, \frac{1}{t} \sum_{\tau=1}^t x_{\tau} + \frac{t}{\sqrt{t}} z_{1-\frac{\alpha}{2}} \right\rangle,$$

kde $z_{\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}$ jsou kvantily standardizovaného normálního rozdělení. Platí tedy, že

$$P \left(\frac{1}{t} \sum_{\tau=1}^t x_{\tau} - \frac{1}{\sqrt{t}} z_{1-\frac{\alpha}{2}} \leq \mu \leq \frac{1}{t} \sum_{\tau=1}^t x_{\tau} + \frac{t}{\sqrt{t}} z_{1-\frac{\alpha}{2}} \right) = 1 - \alpha.$$

Příklad 2 Bayesovský odhad střední hodnoty normálního rozdělení

Uvažujme nyní stejnou úlohu formulovanou a řešenou pomocí bayesovského přístupu. Statistický model je reprezentován podmíněnou hustotou pravděpodobnosti

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{(x-\mu)^2}{2} \right) \quad (13)$$

kde neznámý parametr μ je chápán jako náhodná veličina. Hustota pravděpodobnosti náhodného vektoru x_1, x_2, \dots, x_t má tvar

$$f(x_1, x_2, \dots, x_t|\mu) = \prod_{\tau=1}^t f(x_{\tau}|\mu) = (2\pi)^{-\frac{t}{2}} \exp \left(-\frac{1}{2} \sum_{\tau=1}^t (x_{\tau} - \mu)^2 \right). \quad (14)$$

Dále je potřeba zvolit apriorní hustotu pro parametr μ . Z čistě praktických důvodů zvolíme normální apriorní rozdělení. Jelikož kromě dat nemáme žádnou další informaci o skutečné hodnotě parametru μ , zvolíme rozdělení s velkým rozptylem a střední hodnotou 0, např.

$$f(\mu) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} \frac{\mu^2}{100} \right).^5 \quad (15)$$

S využitím zjednodušeného zápisu (8) můžeme vyjádřit aposteriorní hustotu

$$\begin{aligned} f(\mu|x_1, \dots, x_t) &\propto f(x_1, \dots, x_t|\mu) f(\mu) \\ &\propto \exp \left(-\frac{1}{2} \sum_{\tau=1}^t (x_{\tau} - \mu)^2 \right) \exp \left(-\frac{1}{2} \frac{\mu^2}{100} \right) \\ &\propto \exp \left(-\frac{1}{2} \left(\left(t + \frac{1}{100} \right) \mu^2 - 2 \left(\sum_{\tau=1}^t x_{\tau} \right) \mu \right) \right) \\ &\propto \exp \left(-\frac{1}{2} \frac{\left(\mu - \frac{1}{t + \frac{1}{100}} \sum_{\tau=1}^t x_{\tau} \right)^2}{\frac{1}{t + \frac{1}{100}}} \right). \end{aligned} \quad (16)$$

Odtud vidíme, že aposteriorní rozdělení je rovněž normální, a to konkrétně

$$\mathcal{N} \left(\frac{1}{t + \frac{1}{100}} \sum_{\tau=1}^t x_{\tau}, \frac{1}{t + \frac{1}{100}} \right). \quad (17)$$

⁵Pro případy, kdy nemáme prakticky žádnou apriorní informaci, existují vhodnější způsoby volby apriorního rozdělení - tzv. neinformativní apriorní rozdělení.

Střední hodnota aposteriorního rozdělení se tedy přibližně rovná výběrovému průměru z hodnot x_1, \dots, x_t a směrodatná odchylka je přibližně nepřímo úměrná \sqrt{t} .

Abychom nyní mohli stanovit bodový odhad parametru μ , je potřeba zvolit ztrátovou funkci (9). Množina rozhodnutí je zřejmě množina všech přípustných hodnot pro parametr μ , tedy \mathbb{R} . Použijeme v praxi často využívanou kvadratickou ztrátovou funkci, která rozhodnutí přiřazuje druhou mocninu eukleidovské vzdálenosti od skutečné hodnoty parametru, tedy

$$L(a, \mu) = (a - \mu)^2. \quad (18)$$

Optimálním rozhodnutím je v našem případě hodnota bodového odhadu, budeme jej proto značit $\hat{\mu}$. Podle (10) hledáme $\hat{\mu}$ tak, aby platilo

$$\hat{\mu} \in \underset{a \in \mathbb{R}}{\text{Argmin}} \int (a - \mu)^2 f(\mu|x_1, \dots, x_t) d\mu. \quad (19)$$

Integrál v (19) je kvadratická funkce proměnné a , kterou lze vyjádřit ve tvaru

$$a^2 - 2a \int \mu f(\mu|x_1, \dots, x_t) d\mu + \int \mu^2 f(\mu|x_1, \dots, x_t) d\mu.$$

Tato funkce má globální minimum v bodě

$$\int \mu f(\mu|x_1, \dots, x_t) d\mu,$$

ktej je roven střední hodnotě parametru vzhledem k aposteriornímu rozdělení. Pro hodnotu bodového odhadu $\hat{\mu}$ tedy platí

$$\hat{\mu} = \frac{1}{t + \frac{1}{100}} \sum_{\tau=1}^t x_{\tau}.$$

Jinou možností stanovení bodového odhadu parametru na základě aposteriorního rozdělení je bayesovská obdoba maximálně věrohodného odhadu (18), kdy je hodnota odhadu stanovena tak, aby pro každé x_1, x_2, \dots, x_t maximalizovala hodnotu aposteriorní hustoty,

$$\hat{\mu}_{MAP}(x_1, x_2, \dots, x_t) = \arg \max_{a \in \mathcal{A}} f(\mu|x_1, x_2, \dots, x_t) \quad (20)$$

Tento odhad je označován jako maximální aposteriorní odhad. Vzhledem k tomu, že aposteriorní rozdělení (16) je normální, dostaneme v tomto případě $\hat{\mu}_{MAP} = \hat{\mu}$, kde $\hat{\mu}$ je již nalezený odhad minimalizující střední hodnotu kvadratické ztrátové funkce. Obecně takováto rovnost neplatí. Poznamenejme, že i když je tento odhad v praxi velmi často využíván (např. při zpracování obrazu), nejde o typicky bayesovské řešení, neboť jej nelze vyjádřit jako argument minima střední hodnoty konkrétní ztrátové funkce.

Bayesovskou obdobou intervalového odhadu (credible interval) bude množina C , pro niž platí

$$P(\mu \in C|x_1, \dots, x_t) = 1 - \alpha$$

pro malé α , např. $\alpha = 0.05$. Je přitom přirozené požadovat, aby tato množina byla co nejmenší. Množinu C tedy budeme hledat ve tvaru

$$C = \{\mu \in \mathbb{R} | f(\mu|x_1, \dots, x_t) \geq k\}$$

pro nějaké $k > 0$. Vzhledem k tomu, že aposteriorní rozdělení (17) je normální, bude množina C interval

$$C = \left\langle \frac{1}{t + \frac{1}{100}} \sum_{\tau=1}^t x_{\tau} - \frac{1}{\sqrt{t + \frac{1}{100}}} z_{1-\frac{\alpha}{2}}, \frac{1}{t + \frac{1}{100}} \sum_{\tau=1}^t x_{\tau} + \frac{1}{\sqrt{t + \frac{1}{100}}} z_{1-\frac{\alpha}{2}} \right\rangle \quad (21)$$

Povšimněme si několika detailů příkladech 1 a (2):

- Neznámý parametr μ v bayesovském modelu je považován za náhodnou veličinu s hodnotami v množině \mathbb{R} , zatímco v klasickém modelu jde o bod z této množiny.
- Statistický model v klasickém přístupu chápeme jako rozdělení pravděpodobnosti závislé na parametru μ , kdežto model v bayesovském pojetí je reprezentován podmíněnou hustotou pravděpodobnosti. Pro libovolnou konkrétní hodnotu parametru $\mu_0 \in \mathbb{R}$ ale oba modely $f_{\mu_0}(x_\tau)$ i $f(x_\tau|\mu = \mu_0)$ představují stejná rozdělení pravděpodobnosti veličiny x_τ .
- Klasický intervalový odhad parametru (v příkladu 1) je dvojice náhodných veličin $T_l(x), T_u(x)$, pro které platí

$$\forall \mu \in \mathbb{R} : P(T_l(x) \leq \mu \leq T_u(x)) = 1 - \alpha,$$

přičemž rozdělení statistik $T_l(x), T_u(x)$ je určeno parametrem μ . Naproti tomu bayesovský interval spolehlivosti je množina C , pro kterou platí $P(\mu \in C|x_1, \dots, x_t) = 1 - \alpha$. Díky tomu, že neznámý parametr je považován za náhodnou veličinu, je interpretace bayesovského intervalového odhadu přímočará v porovnání s klasickým intervalovým odhadem.

- Zatímco veličiny x_1, x_2, \dots, x_t byly v rámci klasického přístupu podle předpokladu nezávislé (viz. (12)), z bayesovského pohledu už tomu tak není. Sdruženou hustotu veličin x_1, x_2, \dots, x_t lze vyjádřit jako

$$f(x_1, x_2, \dots, x_t) = \int f(\mu) \prod_{\tau=1}^t f(x_\tau|\mu) d\mu, \quad (22)$$

kdežto pro součin marginálních hustot veličin x_1, x_2, \dots, x_t platí

$$\prod_{\tau=1}^t f(x_\tau) = \prod_{\tau=1}^t \int f(\mu) f(x_\tau|\mu) d\mu. \quad (23)$$

Sdružené hustoty (22) a (23) se přitom obecně nerovnjají, až na extrémní případy, kdy veličiny x_τ na μ ve skutečnosti nezávisí, nebo kdy je hodnota parametru μ předem známa, což lze ne zcela korektně vyjádřit apriorní hustotou ve tvaru $f(\mu) = \delta(\mu - \mu_0)$ pro nějaké $\mu_0 \in \mathbb{R}$, kde $\delta(\cdot)$ představuje Diracovu δ funkci.

Z bayesovského pohledu tedy veličiny x_1, x_2, \dots, x_t obecně nejsou nezávislé, ale jsou podmíněně nezávislé při dané hodnotě μ , viz. (14). Tento pohled je přitom ve shodě s běžnou představou: V případě, že opakovaně pozorujeme realizace náhodných veličin se stejným rozdělením, které ale není přesně známo, pak s přibývajícím počtem pozorování zpravidla umíme stále přesněji předpovídat hodnoty budoucích pozorování.

- Uvažujeme-li všechna pravděpodobnostní rozdělení libovolné z náhodných veličin x_τ , pak tato rozdělení tvoří obecně jakýsi nekonečně rozměrný prostor. Předpoklad v příkladu 1, že data pochází z rozdělení $f_\mu(x)$ pro nějaké $\mu \in \mathbb{R}$ přináší informaci, že skutečné rozdělení veličin x_τ leží v konkrétním konečně rozměrném podprostoru. Předpoklad konkrétního statistického modelu $f_\mu(x)$ je tedy silnou apriorní informaci o neznámém rozdělení veličin x_τ .
- V rámci klasického přístupu je veškerá apriorní informace, která do úlohy vstupuje, vyjádřena pouze statistickým modelem. Známe “tvar” skutečného rozdělení, ale nemáme možnost do úlohy jednoduše zanezt další informaci týkající se hodnoty neznámého parametru. Naproti tomu bayesovský přístup kromě informace o tvaru rozdělení reprezentované modelem (13) vnáší do úlohy detailnější informaci o hodnotě neznámého parametru μ , která je explicitně reprezentována apriorní hustotou $f(\mu)$.

Příklad 3 *Odhad parametru a predikce posloupnosti Bernoulliových pokusů* Předpokládejme, že x_1, x_2, \dots, x_t tvoří posloupnost Bernoulliových pokusů. Jde tedy o nezávislé stejně rozdělené náhodné veličiny s hodno-

tami v množině $\{0, 1\}$. Odpovídající statistický model má tvar

$$\begin{aligned} f(x_\tau|p) &= p^{\delta(x_\tau,1)}(1-p)^{\delta(x_\tau,0)}, \\ f(x_1, x_2, \dots, x_t|p) &= \prod_{\tau=1}^t f(x_\tau|p), \end{aligned}$$

kde $p \in (0, 1)$ je neznámý parametr.

Zvolme apriorní rozdělení v podobě beta rozdělení (opět hlavně z praktických důvodů) s parametry $\nu_0, \nu_1 \in \mathbb{R}^+$:

$$f(p) = \frac{(1-p)^{\nu_0-1} p^{\nu_1-1}}{B(\nu_0, \nu_1)},$$

kde $B: \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$ je tzv. beta funkce definovaná vztahem

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt.$$

Pro aposteriorní hustotu potom platí

$$\begin{aligned} f(p|x_1, \dots, x_t) &\propto f(p)f(x_1, \dots, x_t|p) \\ &\propto f(p) \prod_{\tau=1}^t f(x_\tau|p) \\ &\propto (1-p)^{\nu_0-1} p^{\nu_1-1} \prod_{\tau=1}^t p^{\delta(x_\tau,1)} (1-p)^{\delta(x_\tau,0)}. \end{aligned} \quad (24)$$

Označíme-li $V = \sum_{\tau=1}^t \delta(x_\tau, 1)$, můžeme aposteriorní hustotu vyjádřit ve tvaru

$$\begin{aligned} f(p|x_1, \dots, x_t) &\propto (1-p)^{\nu_0-1} p^{\nu_1-1} p^V (1-p)^{t-V} \\ &\propto (1-p)^{\nu_0+t-V-1} p^{\nu_1+V-1}. \end{aligned}$$

Odtud plyne, že aposteriorní rozdělení parametru p je také beta rozdělení, a to s parametry $\nu_0 + t - V$ a $\nu_1 + V$,

$$f(p|x_1, \dots, x_t) = \frac{(1-p)^{\nu_0+t-V-1} p^{\nu_1+V-1}}{B(\nu_0+t-V, \nu_1+V)}. \quad (25)$$

Pokud bychom na základě pozorování x_1, \dots, x_t chtěli předpovědět hodnotu (ve smyslu rozdělení pravděpodobnosti) veličiny x_{t+1} , potřebujeme určit podmíněnou hustotu $f(x_{t+1}|x_1, \dots, x_t)$. K tomu využijeme aposteriorní hustotu (25).

$$\begin{aligned} f(x_{t+1}|x_1, \dots, x_t) &= \int f(x_{t+1}, p|x_1, \dots, x_t) dp \\ &= \int f(x_{t+1}|p, x_1, \dots, x_t) f(p|x_1, \dots, x_t) dp \\ &= \int f(x_{t+1}|p) f(p|x_1, \dots, x_t) dp \\ &= \int \frac{(1-p)^{\delta(x_{t+1},0)+\nu_0+t-V-1} p^{\delta(x_{t+1},1)+\nu_1+V-1}}{B(\nu_0+t-V, \nu_1+V)} \\ &= \frac{B(\delta(x_{t+1},0)+\nu_0+t-V, \delta(x_{t+1},1)+\nu_1+V)}{B(\nu_0+t-V, \nu_1+V)} \end{aligned}$$

Pro $x_{t+1} = 0$ pak dostaneme

$$\begin{aligned} f(x_{t+1} = 0 | x_1, \dots, x_t) &= \frac{B(\nu_0 + t - V + 1, \nu_1 + V)}{B(\nu_0 + t - V, \nu_1 + V)} \\ &= \frac{\Gamma(\nu_0 + t - V + 1)\Gamma(\nu_1 + V)}{\Gamma(\nu_0 + \nu_1 + t + 1)} \frac{\Gamma(\nu_0 + \nu_1 + t)}{\Gamma(\nu_0 + t - V)\Gamma(\nu_1 + V)} \\ &= \frac{\nu_0 + t - V}{\nu_0 + \nu_1 + t} \end{aligned}$$

Obdobně pro $x_{t+1} = 1$ bychom dostali

$$f(x_{t+1} = 1 | x_1, \dots, x_t) = \frac{\nu_1 + V}{\nu_0 + \nu_1 + t}.$$

Bayesovský přístup je často používán také při zpracování obrazu. Statistické modely používané v této oblasti ale mají mnohem širší využití - lze je použít v případě, že náhodné veličiny mají určité uspořádání (např. prostorové), na jehož základě lze předpokládat nějakou (ne)závislostní strukturu. Následující příklad ukazuje, jak důležitou roli v takových případech hraje apriorní informace.

Příklad 4 (Segmentace obrazu)

Uvažujme digitální obraz o rozměrech $n \times n$ bodů. Pro jednoduchost budeme předpokládat, že každý bod obrazu je popsán jedním reálným číslem, např. hodnotou jasu. Dále předpokládejme, že obraz zachycuje scénu, v níž se vyskytují větší jednoduché objekty vyplněné náhodnou texturou (všechny stejnou), která je odlišná od textury, již je vyplněno pozadí scény, přičemž známe pravděpodobnostní modely těchto dvou textur. Takovýto obraz je tedy pokryt dvěma druhy ploch s odlišnou náhodnou vyplní. Pokusíme se o segmentaci takového obrazu, tj. pro každý bod obrazu určíme (odhadneme), zda je zaplněn objektem nebo pozadím.

Obrazová data v tomto případě tvoří náhodný vektor y s hodnotami v $\mathbb{R}^{(n^2)}$. Označíme-li

$$S = \{1, \dots, n\} \times \{1, \dots, n\},$$

kde prvky množiny S chápeme jako souřadnice jednotlivých bodů, můžeme náhodný vektor y reprezentovat po složkách jako

$$y = (y_{ij})_{(i,j) \in S}.$$

Podobně můžeme reprezentovat druhy výplně jako vektor

$$x = (x_{ij})_{(i,j) \in S},$$

s hodnotami v $\{0, 1\}^{(n^2)}$. Skutečná hodnota tohoto vektoru není známa, proto vektor x považujeme za náhodný a v úloze vystupuje v roli parametru.

Předpokládejme pro jednoduchost, že pozorované veličiny y_{ij} v libovolném bodě (i, j) závisí pouze na druhu textury v tomto bodě (hodnotě x_{ij}) a že mají normální rozdělení se známými parametry, tedy

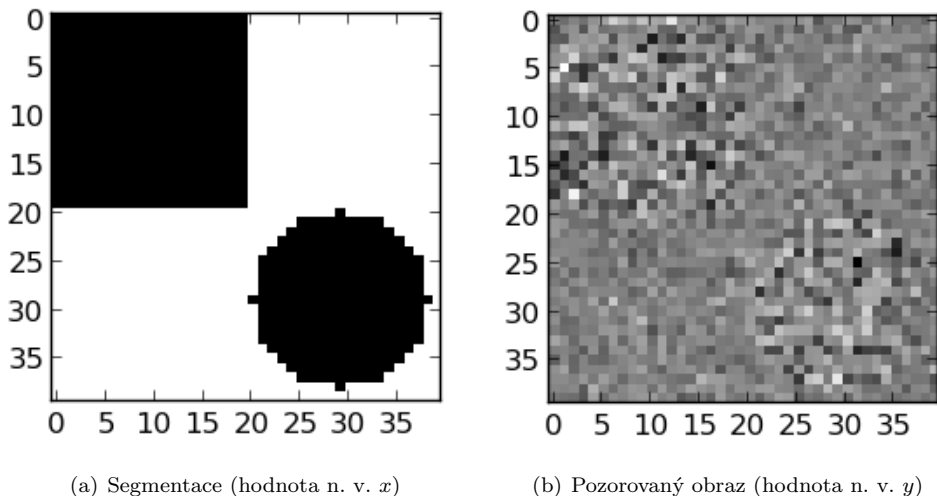
$$f(y|x) = \prod_{(i,j) \in S} f(y_{ij}|x_{ij}), \quad (26)$$

kde pro všechny $(i, j) \in S, c \in \{0, 1\}$

$$f(y_{ij}|x_{ij} = c) = \frac{1}{\sqrt{2\pi}\sigma_c} \exp\left(-\frac{(y_{ij} - \mu_c)^2}{2\sigma_c^2}\right) \quad (27)$$

a $(\mu_0, \sigma_0^2), (\mu_1, \sigma_1^2) \in \mathbb{R} \times \mathbb{R}^+$ jsou známé parametry. Hustotu (27) lze také vyjádřit ve tvaru

$$f(y_{ij}|x_{ij}) = \prod_{c \in \{0,1\}} \left(\frac{1}{\sqrt{2\pi}\sigma_c} \exp\left(-\frac{(y_{ij} - \mu_c)^2}{2\sigma_c^2}\right) \right)^{\delta(x_{ij}, c)}, \quad (28)$$



Obrázek 1: Příklad segmentace a pozorovaného obrazu

kde funkce $\delta(\cdot, \cdot)$ je Kroneckerovo δ , definované

$$\delta(a, b) = \begin{cases} 0 & \text{pro } a \neq b \\ 1 & \text{pro } a = b. \end{cases}$$

Tvar (28) je vhodnější pro další výpočty, protože podmíněná hustota je vyjádřena jako funkce proměnných x_{ij} a y_{ij} a přitom jsme se vyhnuli hodnotám x_{ij} vystupujícím v indexu.

Příklad segmentace a obrazu o rozměru 40×40 , tedy konkrétních realizací náhodných vektorů x a y je na obrázku 1. Body (i, j) , s hodnotou $x_{ij} = 0$ jsou zobrazeny černě, s hodnotou $x_{ij} = 1$ bíle. Obraz závisí na segmentaci prostřednictvím modelu (27), kde $\mu_0 = \mu_1$ a $\sigma_0 > \sigma_1$. Textury se tedy liší pouze rozptylem.

Nalezení odhadu segmentace odpovídá nalezení odhadu parametru x , k čemuž je potřeba vyjádřit a posteriorní hustotu $f(x|y)$. Napřed je tedy nutno zvolit vhodnou apriorní hustotu $f(x)$. Zkusme nejprve uvažovat takto: “Nemáme žádnou informaci o tom, jak jsou jednotlivé typy výplně zastoupeny ani informaci o tom, kde se jednotlivé objekty v obraze nachází a proto žádné hodnoty segmentace nemůžeme preferovat před jinými.” To by znamenalo, apriorní hustotu pravděpodobnosti volíme jako rovnoměrnou, tedy

$$f(x) = \frac{1}{2^{(n^2)}}, \quad (29)$$

protože vektor x může nabývat $2^{(n^2)}$ různých hodnot. Z (4) je zřejmé, pro každé $(i, j) \in S$ je $f(x_{ij}) = \frac{1}{2}$ a veličiny x_{ij} jsou navzájem nezávislé, tj.

$$f(x) = \prod_{(i,j) \in S} f(x_{ij}). \quad (30)$$

Pro model (26) a apriorní hustotu dostaneme aposteriorní hustotu ve tvaru

$$\begin{aligned}
f(x|y) &= \frac{f(y|x)f(x)}{f(y)} \\
&= \frac{\left(\prod_{(i,j) \in S} f(y_{ij}|x_{ij})\right) \left(\prod_{(i,j) \in S} f(x_{ij})\right)}{\sum_{\tilde{x} \in \{0,1\}^{(n^2)}} \left(\prod_{(i,j) \in S} f(y_{ij}|x_{ij} = \tilde{x}_{ij})\right) \left(\prod_{(i,j) \in S} f(x_{ij} = \tilde{x}_{ij})\right)} \\
&= \frac{\prod_{(i,j) \in S} f(y_{ij}|x_{ij})f(x_{ij})}{\sum_{\tilde{x} \in \{0,1\}^{(n^2)}} \prod_{(i,j) \in S} f(y_{ij}|x_{ij} = \tilde{x}_{ij})f(x_{ij} = \tilde{x}_{ij})} \\
&= \frac{\prod_{(i,j) \in S} f(y_{ij}|x_{ij})f(x_{ij})}{\prod_{(i,j) \in S} \sum_{\tilde{x}_{ij} \in \{0,1\}} f(y_{ij}|x_{ij} = \tilde{x}_{ij})f(x_{ij} = \tilde{x}_{ij})} \\
&= \prod_{(i,j) \in S} \frac{f(y_{ij}|x_{ij})f(x_{ij})}{\sum_{\tilde{x}_{ij} \in \{0,1\}} f(y_{ij}|x_{ij} = \tilde{x}_{ij})f(x_{ij} = \tilde{x}_{ij})} \\
&= \prod_{(i,j) \in S} f(x_{ij}|y_{ij}). \tag{31}
\end{aligned}$$

Z tvaru aposteriorní hustoty plyne, že $f(x|y) = \prod_{(i,j) \in S} f(x_{ij}|y)$, což znamená, že x_{ij} jsou podmíněně nezávislé při dané hodnotě y a dále $f(x_{ij}|y) = f(x_{ij}|y_{ij})$, tedy že x_{ij} jsou podmíněně nezávislé na $y_{\tilde{i}\tilde{j}}$ pro $(\tilde{i}, \tilde{j}) \in S, (\tilde{i}, \tilde{j}) \neq (i, j)$ při dané hodnotě y_{ij} . Jinými slovy, aposteriorní rozdělení x_{ij} závisí pouze na hodnotě pixelu y_{ij} a hodnoty ostatních pixelů už žádnou informaci o typu textury v bodě (i, j) nepřináší.

Pro model (27) a apriorní hustotu (4) pak dostaneme $f(x_{ij}|y_{ij})$ ve tvaru

$$f(x_{ij} = c|y_{ij}) = \frac{\frac{1}{\sigma_c} \exp\left(-\frac{(y_{ij} - \mu_c)^2}{2\sigma_c^2}\right)}{\sum_{\tilde{c} \in \{0,1\}} \frac{1}{\sigma_{\tilde{c}}} \exp\left(-\frac{(y_{ij} - \mu_{\tilde{c}})^2}{2\sigma_{\tilde{c}}^2}\right)},$$

pro $c \in \{0, 1\}$.

Abychom mohli určit konkrétní odhad segmentace \hat{x} , musíme nyní zvolit vhodnou ztrátovou funkci. V oblasti zpracování obrazu se často volí tzv. 0 – 1 ztrátová funkce

$$L(a, x) = 1 - \delta(a, x),$$

která nabývá hodnoty 0 pouze pro $x = a$, jinak je rovna 1. Pro střední hodnotu ztrátové funkce vzhledem k aposteriornímu rozdělení dostaneme

$$E[L(a, x)|y] = 1 - f(x = a|y)$$

a tedy platí

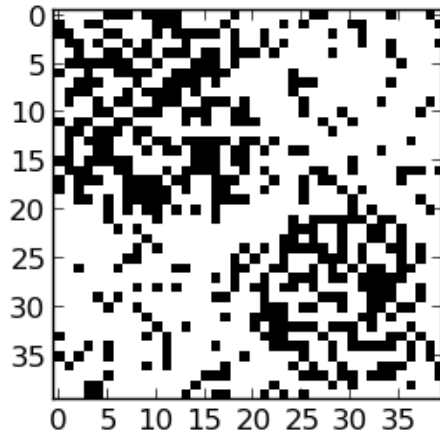
$$\hat{x} \in \underset{a \in \{0,1\}^{(n^2)}}{\text{Argmax}} f(a|y).$$

Protože má aposteriorní hustota díky volbě uniformní apriorní hustoty tvar (31), dostáváme pro očekávanou ztrátu

$$E[L(a, x)|y] = 1 - \prod_{(i,j) \in S} f(x_{ij} = a_{ij}|y_{ij}). \tag{32}$$

Vzhledem k tomu, že jednotlivé členy v součinu (32) můžeme maximalizovat nezávisle, dostaneme pro

$$\begin{aligned}
\hat{x} &\in \underset{a \in \{0,1\}^{(n^2)}}{\text{Argmin}} E[L(a, x)|y] \\
\hat{x}_{ij} &\in \underset{x_{ij} \in \{0,1\}}{\text{Argmax}} f(x_{ij}|y_{ij}). \tag{33}
\end{aligned}$$



Obrázek 2: Odhad segmentace pro uniformní apriorní hustotu

Výsledný odhad segmentace založený na datech z obrázku 1(b) je zobrazen na obrázku 2. I když je v odhadu \hat{x} patrná podobnost se skutečnou segmentací (viz obrázek 1(a)), je výsledek mírně řečeno neuspokojivý. Z (33) je přitom zřejmé, že stejný odhad bychom v tomto případě dostali odhadem metodou maximální věrohodnosti.

Bayesovský odhad segmentace lze ale udělat mnohem lépe, pokud důsledně využijeme dostupnou apriorní informaci. Při formulaci úlohy bylo řečeno, že scéna zachycuje “větší jednoduché objekty”, což znamená, že oblasti se stejnou hodnotou segmentace budou s větší pravděpodobností tvořit větší souvislé plochy, což však uniformní apriorní hustota (4) nijak nereflektuje. Podaří-li se nám sestavit apriorní hustotu, která vyjadřuje takovouto apriorní informaci, můžeme očekávat zlepšení výsledného odhadu. K tomuto účelu lze využít takzvané Gibbsovy distribuce, což jsou pravděpodobnostní rozdělení s hustotami ve tvaru

$$f(x) = \frac{1}{Z(\beta)} \exp(-\beta H(x)), \quad (34)$$

kde

$$Z(\beta) = \int \exp(-\beta H(x)) dx. \quad (35)$$

Tato rozdělení ve statistické mechanice popisují rozdělení pravděpodobnosti stavů velkého systému částic v rovnovážném stavu. Funkce $H(x)$ má význam energie a koeficient $\beta > 0$ odpovídá převrácené hodnotě teploty. Z tvaru hustoty (34) je vidět, že větší pravděpodobnost je přiřazována stavům s nižší hodnotou energie a naopak. S rostoucí hodnotou parametru β (tj. s klesající teplotou) roste rozdíl hustoty pravděpodobnosti pro stavy s vysokou a nízkou hodnotou energie. Zajímavý případ nastane, pokud energii $H(x)$ můžeme vyjádřit jako součet příspěvků, které jsou funkcí stavů malých skupin částic. Zpravidla jde o částice, které spolu v nějakém smyslu sousedí. Tyto modely našly uplatnění mimo jiné v oblasti zpracování obrazu. Původní fyzikální terminologie se přitom přenesla i do těchto oblastí.

V našem případě můžeme energii $H(x)$ volit úměrnou počtu sousedních pixelů s odlišnými hodnotami segmentace. Označíme-li $N(i, j) \subset S$ množinu všech bodů, které přímo sousedí s bodem $(i, j) \in S$, tj.

$$N(i, j) = \{(k, l) \in S : |i - k| + |j - l| = 1\},$$

lze energii $H(x)$ vyjádřit ve tvaru

$$H(x) = \sum_{(i,j) \in S} \left(\sum_{(k,l) \in N(i,j)} \delta(x_{ij}, x_{kl}) \right). \quad (36)$$

Lze snadno nahlédnout, že energie (36) je přímo úměrná celkové délce hranice mezi jednotlivými oblastmi.

Pro apriorní rozdělení (34) s energií (36) a model daný vztahy (26) a (28) dostaneme aposteriorní hustotu ve tvaru

$$f(x|y) \propto \left(\prod_{(i,j) \in S} \prod_{c \in \{0,1\}} \left(\frac{1}{\sigma_c} \exp \left(-\frac{(y_{ij} - \mu_c)^2}{2\sigma_c^2} \right) \right)^{\delta(x_{ij},c)} \right) \exp(-\beta H(x)). \quad (37)$$

Vzhledem k tomu, že x_{ij} nejsou vzhledem k apriorní hustotě (34) nezávislé, nelze aposteriorní hustotu (37) faktorizovat, jak tomu bylo v případě aposteriorní hustoty (31). Další postup při minimalizaci střední hodnoty ztrátové funkce se nyní neobejde bez numerického řešení. Povšimněme si alespoň dvou detailů:

- Vztah (37) určuje aposteriorní hustotu až na normalizační konstantu, která je rovna

$$\left(\sum_{x \in \{0,1\}^{(n^2)}} \left(\prod_{(i,j) \in S} \prod_{c \in \{0,1\}} \left(\frac{1}{\sigma_c} \exp \left(-\frac{(y_{ij} - \mu_c)^2}{2\sigma_c^2} \right) \right)^{\delta(x_{ij},c)} \right) \right)^{-1},$$

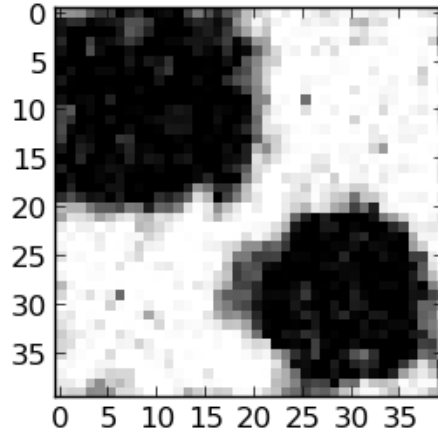
k čemuž je potřeba řádově $2^{(n^2)}$ operací, což je i pro relativně malá n prakticky nemožné. Pro obrázek 40×40 pixelů jde řádově o 10^{481} operací.

- Pokud by veličina x byla spojitá, spočítal by výpočet např. normalizační konstanty, nebo stř. hodnoty parametru x ve výpočtu integrálu přes množinu $\mathbb{R}^{(n^2)}$. Toto opět lze pomocí běžných numerických metod pro přibližný výpočet integrálu realizovat jen pro velmi malá n .

Pro výpočet integrálů, které lze vyjádřit jako střední hodnotu určité funkce náhodné veličiny vzhledem k nějakému rozdělení (v naše případě aposteriornímu), se v praxi používají numerické metody založené na generování velkého množství vzorků z dané hustoty, přičemž hledaná střední hodnota je pak aproximována výběrovým průměrem. Vzhledem k dimenzi úlohy a obtížnému určení normalizační konstanty, je i samotné generování vzorků netriviální. Tento problém lze řešit pomocí tzv. Markov Chain Monte Carlo metod (MCMC), kterými lze přibližně generovat vzorky i z mnohorozměrných rozdělení, určených až na normalizační konstantu.

Aproximace aposteriorní hustoty pro data z obrázku 1(b) a apriorní hustotu ve tvaru Gibbsovy distribuce s parametrem $\beta = 0.4$ vzorkována pomocí tzv. Gibbsova sampleru je na obrázku 3. Přesněji jde o aproximace aposteriorních marginálních hustot parametrů x_{ij} .

Výsledek ilustruje skutečnost, že důsledné využití dostupné apriorní informace může být pro úspěšné řešení úlohy rozhodující, a to zvláště v situacích, kdy je k dispozici jen malé množství dat (vzhledem k počtu parametrů). Dále je z příkladu zřejmé, že s využitím bayesovského přístupu a MCMC metod lze pracovat i s modely, které mají tisíce parametrů.



Obrázek 3: Vizualizace marginálních aposteriorních hustot veličin x_{ij} . Bod na souřadnici (i, j) je zobrazen s odstínem šedi úměrným $f(x_{ij} = 1|y)$ (pro $f(x_{ij} = 1|y) = 0$ černě, $f(x_{ij} = 1|y) = 1$ bíle). Skutečná hodnota vektoru x je na obrázku 1(a).

4 Dodatek: Základní pojmy a vztahy z teorie pravděpodobnosti

Bayesovská statistika se opírá o několik základních vztahů z teorie pravděpodobnosti. Připomeňme nejprve základní pojmy a důležité vztahy týkající se náhodných jevů a jejich pravděpodobností.

Nechť A, B, C jsou náhodné jevy ze společného pravděpodobnostního prostoru s pravděpodobností P . Předpokládejme pro jednoduchost, že $P(A) > 0, P(B) > 0, P(C) > 0$.

- Podmíněná pravděpodobnost: Podmíněná pravděpodobnost jev A za podmínky, že nastal jev B je definována takto:

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

Odtud plyne $P(A \cap B) = P(A|B)P(B)$.

- Řetězové pravidlo: Z definice podmíněné pravděpodobnosti plyne následující “řetězové pravidlo”:

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

- Nezávislost: Náhodné jevy A a B jsou nezávislé, právě když $P(A \cap B) = P(A)P(B)$. Odtud pro nezávislé jevy plyne $P(A|B) = P(A)$ a $P(B|A) = P(B)$.
- Podmíněná nezávislost: Náhodné jevy A a B jsou podmíněně nezávislé za podmínky, že nastal jev C , právě když platí

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

Odtud plyne

$$P(A|B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{P(A|C)P(B|C)P(C)}{P(B|C)P(C)} = P(A|C).$$

- Věta o úplné pravděpodobnosti: Nechť B_1, B_2, \dots, B_n tvoří úplný systém vzájemně disjunktčních náhodných jevů. Pak platí

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i) = \sum_{i=1}^n P(A \cap B_i).$$

- Bayesova věta: Nechť B_1, B_2, \dots, B_n tvoří úplný systém vzájemně disjunktčních náhodných jevů. Pak platí

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^n P(A|B_i)P(B_i)}.$$

Častěji než se samotnými náhodnými jevy a jejich pravděpodobnostmi pracujeme s náhodnými veličinami a jejich hustotami pravděpodobnosti. Výše uvedené pojmy a vztahy budou platit obdobně pro náhodné veličiny.

Nechť X, Y, Z jsou náhodné veličiny se sdruženou hustotou pravděpodobnosti $f_{X,Y,Z}(x, y, z)$. Předpokládejme pro jednoduchost, že $f_{X,Y,Z}(x, y, z) > 0$ pro všechna x, y, z .

- Marginalizace: Pro marginální hustoty pravděpodobnosti veličin X a Y platí

$$f_X(x) = \int f_{X,Y}(x, y)dy, f_Y(y) = \int f_{X,Y}(x, y)dx.$$

- Podmíněná hustota: Pro podmíněnou hustotu pravděpodobnosti náhodné veličiny X za podmínky $Y = y$ platí:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

Odtud plyne $f_{X,Y}(x, y) = f_{X|Y}(x|y)f_Y(y)$.

- Řetězové pravidlo:

$$f_{X,Y,Z}(x, y, z) = f_{X|Y,Z}(x|y, z)f_{Y|Z}(y|z)f_Z(z)$$

- Nezávislost: Náhodné veličiny X a Y jsou nezávislé, právě když

$$f_{X,Y}(x, y) = f_X(x)f_Y(y).$$

Odtud pro nezávislé veličiny plyne

$$f_{X|Y}(x|y) = f_X(x).$$

- Podmíněná nezávislost: Náhodné veličiny X a Y jsou podmíněně nezávislé za podmínky $Z = z$, právě když platí

$$f_{X,Y|Z}(x, y|z) = f_{X|Z}(x|z)f_{Y|Z}(y|z).$$

Odtud plyne

$$f_{X|Y,Z}(x|y, z) = f_{X|Z}(x|z)$$

- Bayesův vzorec:

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x)f_X(x)}{\int f_{Y|X}(y|x)f_X(x)dx}.$$