

Aplikace bayesovského přístupu při zpracování geodetických měření

Doktorská disertační práce

obor : Geodézie a kartografie

Ing. Lubomír Soukup

Stavební fakulta ČVUT

Praha, červen 1998

Obsah

1	Úvod	5
1.1	Cíl disertace	5
1.2	Měřický a zpracovatelský proces	5
1.2.1	Měření	5
1.2.2	Zpracování měření	7
1.2.3	Interpretace výsledků zpracování	7
1.3	Základní pojmy teorie pravděpodobnosti	8
1.3.1	Náhodná veličina	8
1.3.2	Rozdělení pravděpodobnosti	9
1.3.3	Hustota pravděpodobnosti	9
1.4	Přesnost a pravděpodobnost	10
1.5	Statistická indukce	11
2	Bayesovský přístup	12
2.1	Historie bayesovského přístupu	12
2.2	Současnost bayesovského přístupu, dostupná literatura	12
2.3	Podstata bayesovského přístupu — Bayesův vzorec	13
2.4	Vyrovnaní přímých měření v diskrétním případě	14
3	Vyrovnaní zprostředkujících měření	20
3.1	Formulace problému	20
3.2	Obecné řešení problému	22
3.3	Bayesovský přístup a teorie odhadu	23
4	Lineární model s normálním rozdělením chyb měření	25
4.1	Předpoklady	25
4.2	Odvození aposteriorní hustoty	26
4.2.1	Volba apriorní hustoty pravděpodobnosti	29
4.2.2	Neinformativní apriorní rozdělení	33
4.2.3	Výsledné aposteriorní rozdělení	33
4.3	Poznámka o nevlastních hustotách a teorii distribucí	34
5	Normální rozdělení pravděpodobnosti	36
5.1	Způsoby odvození normálního rozdělení	36
5.2	Gaussovo odvození normálního rozdělení	37
5.2.1	První krok Gaussova postupu	37
5.2.2	Druhý krok Gaussova postupu	38
5.3	Rozbor Gaussova odvození	40
5.3.1	První krok Gaussova postupu	41
5.3.2	Druhý krok Gaussova postupu	42
5.4	Přípustnost neomezeně velkých chyb měření	42

6	Modifikace Gaussova odvození normálního rozdělení	44
6.1	Usekuté normální rozdělení	44
6.1.1	První krok Gaussova postupu	44
6.1.2	Druhý krok Gaussova postupu	44
6.2	Von Misesovo rozdělení	45
6.2.1	První krok Gaussova postupu	46
6.2.2	Druhý krok Gaussova postupu	46
6.3	Pearsonovo rozdělení	50
6.3.1	První krok Gaussova postupu	50
6.3.2	Druhý krok Gaussova postupu	50
6.3.3	Vlastnosti Pearsonova rozdělení	53
6.3.4	Integrace hustoty Pearsonova rozdělení	55
6.3.5	Závěrečné poznámky k Pearsonovu rozdělení	56
6.4	Význam modifikace normálního rozdělení pro bayesovské výpočty	57
7	Nelineární model	59
7.1	Přehled existujících metod	59
7.2	Přesný výpočet posteriorní hustoty	60
7.3	Návrh spolehlivého odhadu posteriorní hustoty	63
7.4	Modifikace rozdělení pravděpodobnosti chyb měření	64
7.4.1	Jednorozměrný případ	65
7.4.2	Závěrečné poznámky k navrženému postupu modifikace . .	71
7.4.3	Příklad — trigonometrické určení výšky	72
8	Shrnutí	77

Seznam symbolů

v, θ, v_j, θ_i	veličiny
$\tilde{\theta}, \hat{\theta}, \theta'$	hodnoty veličiny θ
V, Θ, V_j, Θ_i	náhodné veličiny
$\mathbf{v}, \boldsymbol{\theta}$	sloupcové vektory
\mathbf{v}^T	vektor transponovaný k vektoru \mathbf{v}
$\mathbf{A}, \mathbf{P}, \mathbf{Q}$	matice
\mathbf{A}^T	transponovaná matice k matici \mathbf{A}
$\mathbf{V}, \boldsymbol{\Theta}$	náhodné vektory
\mathcal{X}, \mathcal{Y}	množiny
\emptyset	prázdná množina
\mathbb{N}	množina všech přirozených čísel
\mathbb{R}	množina všech reálných čísel
\mathbb{C}	množina všech komplexních čísel
$[\theta_1, \theta_2, \dots, \theta_m]$	uspořádaná množina
$v_j \triangleq \hat{\theta} - \tilde{l}_j$	definice, výraz $\hat{\theta} - \tilde{l}_j$ je označen symbolem v_j
$f : \mathcal{X} \rightarrow \mathcal{Y} : x \mapsto f(x)$	zobrazení z množiny \mathcal{X} do množiny \mathcal{Y} přiřazující prvku $x \in \mathcal{X}$ hodnotu $f(x) \in \mathcal{Y}$
$\frac{da}{d\theta}(\hat{\theta})$	derivace funkce a podle proměnné θ v bodě $\hat{\theta}$
$\frac{\partial a}{\partial \theta_i}(\hat{\boldsymbol{\theta}})$	parciální derivace funkce a podle proměnné θ_i v bodě $\hat{\boldsymbol{\theta}}$
$x \rightsquigarrow X$	k fyzikální veličině x je přiřazena náhodná veličina X
f_X	hustota pravděpodobnosti náhodné veličiny X
$X \sim f_X$	náhodná veličina X má rozdělení pravděpodobnosti s hustotou f_X
$f_{\boldsymbol{\Theta}}(\cdot \tilde{\mathbf{I}})$	podmíněná hustota pravděpodobnosti náhodného vektoru $\boldsymbol{\Theta}$

$\mathcal{N}(\mu, \sigma^2)$	normální rozdělení se střední hodnotou μ a variancí σ^2
θ, θ_i	určovaná veličina
l, l_j	měřená veličina
\tilde{l}, \tilde{l}_j	měřená hodnota, měření
ϵ, ϵ_j	skutečná chyba měření
v, v_j	oprava vůči nějakému odhadu
$\sigma, \sigma_j, \hat{\sigma}$	základní střední chyba, směrodatná odchylka

1 Úvod

V dnešní době je stále více žádoucí mít spolehlivé a přesné informace o okolním světě. Tyto informace umožňují optimalizovat činnost člověka z mnoha různých hledisek: ekologického, ekonomického, sociálního a dalších odvozených či speciálních kritérií úspěšnosti a účelnosti lidského konání. O tomto trendu svědčí především neutuchající zájem o různé typy informačních systémů, které jsou dnes budovány pro nejrůznější účely, např. geografické informační systémy (GIS), systémy pro řízení technologických procesů, bankovní informační systémy apod.

Takovéto informace jsou získávány pomocí měření nejrůznějších fyzikálních veličin, určování rozmanitých parametrů, ukazatelů, koeficientů apod. Na současné úrovni technického rozvoje většiny oborů již většinou není problémem vlastní měření nebo sběr dat, ale zpracování dat a interpretace výsledků zpracování. Z toho vychází i námět této práce. Budeme předpokládat, že měření již byla provedena a důraz bude položen na zpracování naměřených dat. Bude navržen takový způsob zpracování dat, který umožní snadnou a objektivní interpretaci výsledků zpracování.

1.1 Cíl disertace

Cílem této práce je využití výhod, které poskytuje bayesovský přístup ke zpracování měření oproti standardně používané metodě nejmenších čtverců. Mezi tyto výhody patří především jednotná formulace problému zpracování experimentálních dat pro velmi obecně stanovené podmínky experimentu a možnost zahrnout do zpracovatelského procesu apriorní informace o těchto podmínkách.

Tyto výhody budou demonstrovány na příkladu vyrovnání zprostředkujících měření (kapitola 3). Budou též použity ke stanovení vhodného rozdělení pravděpodobnosti měřených veličin (kapitola 6) a k návrhu spolehlivého odhadu rozdělení pravděpodobnosti určených veličin v nelineárních případech (kapitola 7).

1.2 Měřický a zpracovatelský proces

Měření, zpracování a interpretace představují základní fáze procesu poznávání objektivní reality - *poznávacího procesu*. Tyto jednotlivé fáze poznávacího procesu budou nyní stručně charakterizovány, aby pak bylo možno přesněji vymezit námět této práce.

1.2.1 Měření

Pojem měření lze chápat mnoha různými způsoby. Bylo učiněno několik pokusů o sjednocení různých přístupů a vytvoření jednotné teorie měření. Analýzou základních složek obecné teorie měření se zabývá kniha [Ber77] v kap. 6 (str. 122).

Tato kniha pojednává o základních metodologických a filosofických problémech měření a obsahuje rozsáhlý seznam literatury o teorii měření od jejich historických počátků (2. pol. 19. stol.) až po 70. léta tohoto stol. Pro účely této práce však většinou vystačíme s intuitivním názorem na měření a teoretické úvahy o základních principech měření využijeme jen zřídka.

Měření je proces, při němž se získávají hodnoty nějaké dané veličiny. Přitom se předpokládá, že daná veličina, nazývejme ji *měřená veličina*, má konstantní velikost. Kromě této konstantní hodnoty měřené veličiny má na výsledek měření vliv ještě mnoho dalších faktorů, které se souhrnně nazývají podmínky měření. Každé metodě měření určité veličiny odpovídají nějaké charakteristické podmínky měření, jež jsou vyjádřitelné pomocí dalších fyzikálních veličin. Např. při měření úhlů a délek v geodézii mají na měřenou hodnotu vliv fyzikální veličiny, na kterých závisí optické vlastnosti vzduchu, jímž prochází světelný paprsek od dalekohledu k cíli (teplota, tlak, vlhkost, index lomu apod.), okamžitá poloha měřicího přístroje (teodolitu či dálkoměru) nad stanoviskem a poloha cílové značky nad cílovým bodem (chyby centrace), přístrojové vady, osobní chyby měřiče apod.

Jednotlivé faktory ovlivňující výsledek měření lze rozdělit na dvě základní skupiny: stálé a proměnlivé.

Do skupiny stálých faktorů by měla patřit přinejmenším měřená veličina a ty faktory, jejichž vliv je po celý průběh měření stejný. V uvedeném příkladě geodetického měření úhlů nebo délek mohou být mezi stálé faktory zahrnuty chyby centrace stanoviště a cíle, pokud je bezvětří a stativ přístroje i cílové značky je chráněn před sluncem nebo deštěm. Některé ze stálých faktorů lze změřit a jejich vliv eliminovat formou korekcí (např. některé přístrojové chyby, teplotní korekce při měření délek pásmem, určení refrakčního úhlu při měření zenitových úhlů apod.). Ostatní stálé faktory, které nelze eliminovat nebo by jejich určování bylo z ekonomického či časového hlediska pro daný účel zbytečné, zatěžují výsledek měření tzv. systematickou chybou. O systematických chybách je podrobně pojednáno v [BRH90] na str. 103–104.

Proměnlivé faktory se se v průběhu měření náhodně mění a způsobují tak neurčitost výsledku měření. Považují se proto za *náhodné jevy*. Každý náhodný jev lze rozložit na *elementární jevy*, tj. vyjádřit je jako sjednocení několika (i nekonečně mnoha) elementárních jevů. Náhodný jev lze také chápat jako podmnožinu množiny všech elementárních jevů. Tento dvojí způsob znázornění náhodného jevu je možno přirovnat k vyjádření intervalu buď jako množiny reálných čísel nebo jako sjednocení jednoprvkových číselných množin. Tato analogie bude užitečná při definování náhodné veličiny v odstavci 1.3.1.

Pokud neurčitost výsledku měření způsobená proměnlivými faktory vykazuje *statistickou stabilitu* (relativní četnosti výskytu měřených hodnot ve stanovených intervalech jsou při dostatečně velkém počtu opakování měření téměř stejné — podrobněji viz [Tut78], str. 13, 14), je možné přistoupit ke statistickému zpracování měření. V opačném případě je nutné změnit podmínky měření tak, aby

statistické stability bylo dosaženo, nebo zvolit jiný, alternativní přístup ke zpracování neurčité informace. Některé takové přístupy jsou uvedeny v následujícím oddíle.

1.2.2 Zpracování měření

Účelem zpracování naměřených dat je odhad hodnot jistých veličin, které nemusí být přístupné přímému měření, na základě hodnot měřených veličin. Ony veličiny, jejichž hodnoty jsou odhadovány, budeme nazývat *určované veličiny*. Vzájemný vztah určených a měřených veličin je v geodetických problémech obvykle popsán soustavou algebraických rovnic, v nichž jako neznámé vystupují určované veličiny.

Pokud je brána v úvahu nepřesnost měření hodnot měřených veličin, vyvstává závažná otázka, jak se neurčitost hodnot měřených veličin projeví na hodnotách určených veličin. Tuto otázku lze formulovat několika odlišnými způsoby založenými na různých pohledech na neurčitost. Mezi nejznámější způsoby zacházení s neurčitou informací patří vedle teorie pravděpodobnosti např. teorie fuzzy množin (viz např. [Nov90], [Vie96]), neuronové sítě (viz např. [Nov92]), teorie deterministického chaosu (nelineární systémy, fraktální geometrie — viz např. [Gle96]), gnostická teorie neurčitých dat (viz např. [Kov90]) a další. Výsledkem zpracování měření jsou tedy hodnoty určených veličin spolu s údaji o jejich přesnosti.

Otázky přesnosti zpracování neurčitých dat lze téměř vždy formulovat jazykem teorie pravděpodobnosti a ptát se po pravděpodobnosti výskytu hodnoty určované veličiny v daném intervalu. Hledáním odpovědi na takovéto otázky se zabývá odvětví matematické statistiky zvané *teorie odhadu* (viz např. [Mac80], [Kub83]) (estimation theory). V geodézii je proces zpracování měření tradičně nazýván *vyrovnání měření*. Označení příslušné teoretické disciplíny, tzv. *vyrovnávací počet*, pak lze považovat za alternativní název pro teorii odhadu, neboť většina metod vyrovnání odpovídá příslušným metodám odhadu. Rozdílný je pouze způsob podání, terminologie a symbolika — viz [BRH90]. O zpracování geodetických měření z hlediska teorie odhadu pojednává kniha [KKK82].

1.2.3 Interpretace výsledků zpracování

Interpretace představuje závěrečnou fázi poznávacího procesu, kdy jsou výsledky zpracování měření využity k předpovědi nějakého jevu, který je z určitého důvodu významný. Výsledná informace pak bývá ve formě odpovědi na otázku, zda nastane či nenastane onen zkoumaný jev. Takovou otázkou může být např.: “Bude dané místo během příštích sta let zaplaveno vodou?”, “Ovlivní znečištění půdy ropnými produkty v jisté lokalitě kvalitu určitého zdroje pitné vody v sousední oblasti?”, “Odpovídá daná oblast na družicovém snímku jehličnatému lesu?” apod. Takovéto otázky mohou být zodpovězeny na základě hodnot příslušných veličin, které o zkoumaném jevu nějakým způsobem vypovídají a které byly

za tím účelem určovány. Určované veličiny však nemusí zkoumaný jev popisovat jednoznačně a jejich hodnoty většinou nemohou být odhadnuty absolutně přesně. Tvrzení o zkoumaném jevu proto nemusí být naprosto jisté a platí jen s určitou pravděpodobností. Předpokládejme pro jednoduchost, že určované veličiny byly vybrány tak, že uspokojivě reprezentují zkoumaný jev, a nejistota předpovědi je tedy způsobena pouze nepřesností určených veličin. Aby bylo možno určit pravděpodobnost zkoumaného jevu, je třeba, aby nepřesnost odhadovaných hodnot určených veličin byla vyjádřena v pojmech teorie pravděpodobnosti.

Univerzálním prostředkem pro popis přesnosti nějaké veličiny je její *rozdělení pravděpodobnosti*. Příslušnou veličinu, jejíž přesnost je vyšetřována, je však třeba nejprve definovat jako *náhodnou veličinu* (viz např. [And78], str. 11). Protože tyto pojmy budeme v dalším textu velmi často používat, připomeneme si je spolu s dalšími základními pojmy teorie pravděpodobnosti v následující části.

1.3 Základní pojmy teorie pravděpodobnosti

1.3.1 Náhodná veličina

Náhodná veličina je zobrazení, které elementárním jevům přiřazuje reálná čísla. Přitom se předpokládá existence systému náhodných jevů a systému podmnožin reálných čísel tak, aby bylo možno na těchto množinových systémech zavést pravděpodobnostní míry. Podle úrovně abstrakce, s jakou je specifikován systém náhodných jevů a systém podmnožin reálných čísel lze náhodnou veličinu formálně definovat na různých úrovních obecnosti.

Nejobecněji (např. v [KKK82], def. 1.4.4.2) je náhodná veličina definována jako měřitelné zobrazení X

$$X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}) : \omega \mapsto X(\omega) , \quad (1)$$

kde (Ω, \mathcal{A}) a $(\mathbb{R}, \mathcal{B})$ představují měřitelné prostory. Měřitelný prostor je množina, na níž je definován systém jejích podmnožin, jenž má jisté výhodné vlastnosti (s každou podmnožinou zahrnuje i její doplněk, sjednocení několika podmnožin, jež jsou prvky systému, patří rovněž do systému, celá původní množina je též prvkem systému), tzv. množinová σ -algebra (viz např. [Nag76], def. 30.1). Jednotlivé složky měřitelných prostorů (Ω, \mathcal{A}) , $(\mathbb{R}, \mathcal{B})$ jsou:

Ω ... množina všech elementárních jevů,

\mathcal{A} ... σ -algebra náhodných jevů,

\mathbb{R} ... množina všech reálných čísel,

\mathcal{B} ... σ -algebra otevřených intervalů reálných čísel (systém všech borelovských množin na \mathbb{R} — viz např. [Nag76], odst. 30.4 nebo [KKK82], def. 1.3.3.1).

Měřitelné zobrazení je takové zobrazení X , pro které platí:

$$\forall B \in \mathcal{B} : X^{-1}(B) \in \mathcal{A} .$$

1.3.2 Rozdělení pravděpodobnosti

Rozdělení pravděpodobnosti náhodné veličiny umožňuje určit pravděpodobnost, s jakou lze očekávat výskyt hodnoty náhodné veličiny v předem daném intervalu. K tomu je třeba nejprve definovat *pravděpodobnostní prostor* (Ω, \mathcal{A}, P) , kde P je *pravděpodobnostní míra*, tj. množinová funkce (viz např. [LM81], str. 32).

$$P : \mathcal{A} \rightarrow \langle 0, 1 \rangle : A \mapsto P(A) .$$

Rozdělení pravděpodobnosti pak lze formálně definovat jako míru Q (viz např. [Nag76], def. 30.14) indukovanou pravděpodobnostní mírou P , tj.

$$\forall B \in \mathcal{B} : Q(B) = P(X^{-1}(B)) .$$

Náhodná veličina X tedy vlastně převádí pravděpodobnost nějakého náhodného jevu $A \in \mathcal{A}$ na pravděpodobnost toho, že naměřená hodnota $X(\omega)$, $\omega \in A$ padne do určité množiny reálných čísel (borelovské množiny) $B = X(A) \in \mathcal{B}$. Tento převod je zprostředkován rozdělením pravděpodobnosti Q .

Rozdělení pravděpodobnosti náhodné veličiny lze zadat různými způsoby, např. prostřednictvím distribuční funkce nebo hustoty pravděpodobnosti, popř. charakteristické funkce (viz např. [And78], str. 12, 16). V této práci budeme používat především hustoty pravděpodobnosti.

1.3.3 Hustota pravděpodobnosti

Jak již bylo řečeno, při vyšetřování přesnosti náhodné veličiny má význam se ptát, s jakou pravděpodobností padne náhodně zvolená hodnota této veličiny do určitého, předem stanoveného intervalu. Hustota pravděpodobnosti je funkce, jež umožňuje vyčíslit tuto pravděpodobnost pro každý interval hodnot příslušné náhodné veličiny. Uvažujme náhodnou veličinu X , jejíž rozdělení pravděpodobnosti je zadáno hustotou pravděpodobnosti

$$f : \mathcal{X} \rightarrow \mathbb{R} : x \mapsto f(x) \quad ,$$

kde \mathcal{X} je obor hodnot, jichž může náhodná veličina X nabývat, tzv. *základní soubor*. Všude v této práci budeme uvažovat pouze náhodné veličiny s reálnými hodnotami, takže $\mathcal{X} \subseteq \mathbb{R}$. Pravděpodobnost, že náhodně vybraná hodnota $x \in \mathcal{X}$ padne do intervalu B , $B \subseteq \mathcal{X}$ vyjádříme obvyklým zápisem $P(x \in B)$.

Potom platí:

$$P(x \in B) = \int_B f(x) dx . \tag{2}$$

Poznámka Integrál v rovnosti (2) je Riemannův integrál, jenž bývá obvykle zapisován ve tvaru

$$\int_a^b f(x) dx ,$$

kde meze a, b jsou krajními body intervalu B , tj. $(a, b) = B$, je-li interval B otevřený. Rovnost (2) by mohla platit i pro složitější množiny B (borelovské množiny), ovšem vyžadovalo by to jisté rozšíření Riemannovy definice integrálu — Lebesgueův integrál. Přesná definice Lebesgueova integrálu je srozumitelně podána např. v [Nag76] (def. 31.2., str. 262), a lze ji též nalézt v každé moderní učebnici teorie pravděpodobnosti (např. v [Tut78]). Podrobně je teorie Lebesgueova integrálu zpracována v [NNV85] nebo v [Jar84].

Rovnost (2) s Lebesgueovým integrálem na pravé straně definuje hustotu pravděpodobnosti f , platí-li pro všechny borelovské množiny $B \subseteq \mathcal{X}$. Lze dokázat, že pokud rovnost (2) platí pro všechny intervaly v \mathcal{X} , platí i pro všechny borelovské množiny $B \subseteq \mathcal{X}$. Dále platí, že pro spojitou funkci je Lebesgueův integrál $\int_B f(x) dx$ roven Riemannovu $\int_a^b f(x) dx$. V této práci bude hustota pravděpodobnosti vždy spojitou funkcí, takže zavádění Lebesgueovy definice je pro účely této práce zbytečné. Rovnost (2) lze proto považovat za definici hustoty pravděpodobnosti f .

Úmluva Termín “rozdělení pravděpodobnosti f ” bude znamenat rozdělení pravděpodobnosti, jehož hustota pravděpodobnosti je popsána funkcí f . Skutečnost, že náhodná veličina X má rozdělení pravděpodobnosti f budeme zapisovat

$$X \sim f. \tag{3}$$

Pomocí právě definovaných pojmů je již možno předběžně vymezit okruh problémů, jejichž řešením se tato práce zabývá.

1.4 Přesnost a pravděpodobnost

Nechť jsou porovnávány dvě náhodné veličiny z hlediska jejich přesnosti. Pokud je pevně dán jeden interval, je přirozené považovat za přesnější tu náhodnou veličinu, pro niž je výskyt její hodnoty v daném intervalu pravděpodobnější. Jiným přirozeným kritériem přesnosti, které není vázáno na jediný interval, je délka intervalu, pro nějž je pravděpodobnost výskytu měřené hodnoty pro obě náhodné veličiny stejná. Toto kritérium ovšem má smysl jen tehdy, pokud je k určité hodnotě pravděpodobnosti jednoznačně přiřazen příslušný interval. Přesnější veličinou je pak samozřejmě ta náhodná veličina, pro niž je interval příslušný dané pravděpodobnosti menší. Toto druhé kritérium má však podobný nedostatek jako první zmíněné kritérium. Závisí na volbě hodnoty pravděpodobnosti, pro kterou jsou délky intervalů porovnávány.

Obě navržená kritéria lze skloubit v jediné, obecnější, které není závislé ani na volbě intervalu, ani na stanovení hodnoty pravděpodobnosti. To je možné tehdy, když pro všechny hodnoty pravděpodobnosti, které mají pro daný účel smysl, jsou příslušné intervaly jedné náhodné veličiny vždy kratší než intervaly druhé z porovnávaných náhodných veličin. Tento předpoklad je splněn např. pro náhodné veličiny s normálním rozdělením pravděpodobnosti. Není však již tak samozřejmý pro jiná (např. asymetrická) rozdělení pravděpodobnosti. Navíc ani toto kritérium není dostatečně obecné, neboť ještě závisí na způsobu přiřazení intervalů jednotlivým hodnotám pravděpodobnosti.

Je vidět, že definování pojmu přesnost není jednoduché, i když je brán v úvahu jen pro porovnávání náhodných veličin (tj. jako relace uspořádání). Podle knihy [Ber77] tvoří náhodné veličiny s takto definovaným atributem přesnosti tzv. ordinální škálu (viz str. 171). Přesnost by však mohla být definována i jako veličina udávající míru rozptylu hodnot náhodné veličiny tak, aby bylo možno náhodné veličiny porovnávat podrobněji, s větším informačním obsahem. Na základě této druhé, pokročilejší definice přesnosti by již bylo možno rozhodnout, zda je jedna veličina např. třikrát přesnější než jiná veličina. To odpovídá tzv. poměrové škále zavedené v [Ber77] na str. 172.

1.5 Statistická indukce

Statistická indukce vytváří, obecně řečeno, pravděpodobnostní model objektivní reality na základě jejího pozorování. Uvažujme část objektivní reality, která je zachycena měřeními a určovanými veličinami a jejich vzájemnými vztahy. Potom lze statistickou indukci zjednodušeně chápat jako obecnou metodu, která umožňuje z hodnot měřených veličin usuzovat na rozdělení pravděpodobnosti určovaných veličin. Statistická indukce tedy zahrnuje problém odhadu hustoty pravděpodobnosti.

Problém statistické indukce lze řešit různými metodami a přístupy (viz např. [Rao78]). V této práci je zkoumán z hlediska bayesovského přístupu.

2 Bayesovský přístup

Bayesovský přístup umožňuje jednoduše a efektivně řešit problém statistické indukce, jak bylo již v bohaté bayesovské literatuře mnohokrát ukázáno (viz např. [BT73], [Koc90]). V této kapitole bude uveden teoretický základ bayesovského přístupu – Bayesův vzorec a jeho zobecnění, tzv. Bayesova věta pro spojité náhodné veličiny. Použití Bayesovy věty bude ilustrováno jednoduchými příklady – vyrovnání přímých měření a zprostředkujících měření. Rovněž bude poukázáno na význam normálního rozdělení v bayesovském přístupu a na souvislost s MNČ.

Než však přikročíme k vysvětlení základních principů bayesovského přístupu, připomeneme si nejprve stručně historické pozadí jeho vzniku. Podrobnější pojednání o historii bayesovského přístupu lze nalézt např. v [Mai74], str. 87 – 100.

2.1 Historie bayesovského přístupu

Historie bayesovského přístupu je těsně spjata s vývojem pravděpodobnostního způsobu uvažování jenž vedl k formulování problému statistické indukce. Vznik statistické indukce byl předznamenán již v 18. stol. v pracích zakladatelů teorie pravděpodobnosti Jacoba Bernoulliho a Pierre Simona Laplaceho. Prvenství však bývá přisuzováno anglickému duchovnímu Thomasi Bayesovi (1702-1761), jehož dílo vyšlo v r. 1763 (viz [Bay63]). Jeho myšlenky však brzy upadly v zapomenutí. Statistická indukce pak byla budována v rámci klasické matematické statistiky v první polovině 20. stol. převážně pod vlivem prací anglického statistika R. A. Fishera. Současně byl rozvíjen i bayesovský přístup; jeho oprávněnost prokázali Keynes (1921), de Finetti (1937), Jeffreys (1939), Cox (1946), Savage (1954). Přesné odkazy na jejich díla jsou uvedeny v [Kul96]. Oba přístupy, klasická matematická statistika a bayesovská statistika, se sice principiálně liší, avšak ve svých důsledcích se shodují, neboť za určitých předpokladů dávají totožné výsledky. Spor těchto dvou statistických teorií je spíše filosofického charakteru a jeho rozbor přesahuje rámec této práce. Vysvětlení odlišností obou přístupů lze nalézt např. v [Rob94], [BS97], [BT73]. Nejvýznamnějším argumentem pro bayesovský přístup je jeho větší obecnost – např. klasickou metodu nejmenších čtverců (MNČ) zahrnuje jako zvláštní případ. Tato pozoruhodná skutečnost bude dokázána v následující kapitole.

2.2 Současnost bayesovského přístupu, dostupná literatura

V současné době se aplikační oblast bayesovského přístupu nebývale rozšiřuje. Našel uplatnění v mnoha oborech, především v teorii řízení, v umělé inteligenci, lékařské diagnostice, ekonometrii, geofyzice a také v geodézii a kartografii. První práce aplikující bayesovský přístup v geodézii byla publikována roku 1972 v USA

([Bos72]) a zájem o bayesovský přístup trvá v geodetické odborné literatuře doposud (viz např. [Sch87], [Koc87], [OK94], [Koc90]).

Ve světové odborné literatuře se za posledních několik desetiletí objevilo množství monografií a učebnic o bayesovském přístupu. Mezi nejvýznamnější patří např. [Lin65a], [Lin65b] [BT73], [FMR90], [Rob94] [BS97]. Rovněž se pravidelně pořádají konference s bayesovskou tematikou, např. Maximum-Entropy and Bayesian Methods in Inverse Problems. Sborníky z těchto konferencí také představují významný zdroj informací o bayesovském přístupu (viz např. [SG85]).

V české a slovenské odborné literatuře je však publikací s bayesovskou tematikou značný nedostatek. Existuje zatím, domnívám se, pouze jediný titul [Huš85] — útlá skripta pojednávající v dosti zhuštěné formě o základních principech bayesovského přístupu při odhadování, rozhodování a při testování hypotéz.

2.3 Podstata bayesovského přístupu — Bayesův vzorec

Bayesovský přístup je založen na tzv. Bayesově vzorci, někdy též označovaném jako Bayesova věta. Tento vzorec je jednoduchým důsledkem základních pravidel pro počítání s pravděpodobnostmi a bývá uveden v téměř každé učebnici teorie pravděpodobnosti (např. v [LM81], str. 30 nebo v [Rén72], str. 83). Bayesův vzorec umožňuje řešit úlohy následujícího typu.

Dáno

Je dán náhodný jev A a dalších n náhodných jevů B_j , $j \in \{1, 2, \dots, n\}$. Přitom současný výskyt dvou různých jevů B_j, B_k ; $j, k \in \{1, 2, \dots, n\}$, $j \neq k$ je nemožný jev, tj. platí $B_j \cap B_k = \emptyset$. Sjednocení všech jevů B_j je jistý jev, tj.

$$P\left(\bigcup_{j=1}^n B_j\right) = \sum_{j=1}^n P(B_j) = 1. \quad (4)$$

Jevy B_j , $j \in \{1, 2, \dots, n\}$ tedy představují systém vzájemně se vylučujících alternativních jevů, jež pokrývají všechny možnosti, které mohou nastat (tzv. úplný systém jevů — viz [Rén72], str. 81). Dále je pro každý jev B_j , $j \in \{1, 2, \dots, n\}$ známa podmíněná pravděpodobnost $P(A|B_j)$, tj. pravděpodobnost výskytu jevu A za podmínky, že nastal jev B_j . Formálně je podmíněná pravděpodobnost definována vztahem

$$P(A|B_j) = \frac{P(A \cap B_j)}{P(B_j)}. \quad (5)$$

Hledá se

Je třeba určit podmíněnou pravděpodobnost $P(B_k|A)$ pro $k \in \{1, 2, \dots, n\}$.

Řešení

Podle definice podmíněné pravděpodobnosti (5) platí:

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)}. \quad (6)$$

S použitím téže definice lze vyjádřit $P(B_j \cap A)$ pomocí podmíněné pravděpodobnosti $P(A|B_j)$:

$$P(B_j|A) = \frac{P(A|B_j) P(B_j)}{P(A)}. \quad (7)$$

Zbývá ještě určit pravděpodobnost $P(A)$. Tu lze vyjádřit pomocí věty o úplné pravděpodobnosti, která bývá rovněž uváděna v úvodních kapitolách učebnic teorie pravděpodobnosti (viz např [LM81], str. 29).

$$P(A) = \sum_{j=1}^n P(A|B_j) P(B_j). \quad (8)$$

Dosazením (8) do (7) vznikne Bayesův vzorec

$$P(B_k|A) = \frac{P(A|B_k) P(B_k)}{\sum_{j=1}^n P(A|B_j) P(B_j)}. \quad (9)$$

Použití Bayesova vzorce názorně vystihuje následující praktický příklad. V geodetické literatuře je tento případ tradičně nazýván vyrovnání přímých měření. Matematická statistika však pro stejný problém používá název “odhad parametru polohy”. Protože současná odborná terminologie přejímá označování pojmů obvyklé v matematické statistice, budeme v této práci používat vedle tradičních názvů také odpovídající statistické termíny.

2.4 Vyrovnání přímých měření v diskrétním případě

Uvažujme nějakou fyzikální veličinu l , jejíž hodnota byla zjišťována opakovaným měřením. Předpokládejme pro jednoduchost, že veličina l může nabývat jen diskrétních hodnot l_i , $i \in \{1, 2, \dots, m\}$, $m \in \mathbb{N}$ s ekvidistantním krokem $h \in \mathbb{R}$. Pro každé $i \in \{1, 2, \dots, m-1\}$ tedy platí

$$l_{i+1} = l_i + h. \quad (10)$$

Náhodná veličina L příslušná k veličině l je definována svým základním soubohem $\mathcal{L} \subseteq \mathcal{T} \triangleq \{l_1, l_2, \dots, l_m\}$.

Dáno

Bylo naměřeno n hodnot náhodné veličiny L : $\tilde{l}_j \in \mathcal{L}$, $j \in \{1, 2, \dots, n\}$, $n \in \mathbb{N}$. Dále je známo rozdělení pravděpodobnosti chyb měření, tj. rozdělení pravděpodobnosti náhodné veličiny $\mathcal{E} \triangleq \theta - L$, kde θ je neznámý parametr – tzv. *parametr polohy*. Toto rozdělení pravděpodobnosti je diskrétní, takže jeho hustotou pravděpodobnosti je tzv. *pravděpodobnostní funkce* $P(\mathcal{E} = \cdot)$

$$P(\mathcal{E} = \cdot) : \mathcal{T} - \mathcal{L} \rightarrow \langle 0, 1 \rangle : \epsilon \mapsto P(\mathcal{E} = \epsilon)$$

(viz [LM81], odst. 9.2.) udávající pravděpodobnost, že náhodná veličina \mathcal{E} nabude hodnoty $\epsilon \in \mathcal{T} - \mathcal{L} = \{\epsilon \in \mathbb{R} \mid \epsilon = \theta - l, l \in \mathcal{L}, \theta \in \mathcal{T}\}$.

Pomocí pravděpodobnostní funkce náhodné veličiny \mathcal{E} je nyní možno určit pravděpodobnostní funkci náhodné veličiny $L = \theta - \mathcal{E}$ pro libovolný parametr polohy $\theta \in \{l_1, l_2, \dots, l_m\}$. Rozdělení pravděpodobnosti náhodné veličiny L je tedy dáno až na parametr polohy θ . Parametr θ může být považován za náhodnou veličinu, označme ji Θ , která nabývá diskrétních hodnot z množiny \mathcal{T} .

Pravděpodobnostní funkcí náhodné veličiny L je podmíněná pravděpodobnostní funkce $P(L = l \mid \Theta = \theta)$ (viz [LM81], odst. 12.1.), která je pro $\forall \theta \in \{l_1, l_2, \dots, l_m\}$ jednoznačně určena funkcí $P(\mathcal{E} = \epsilon)$:

$$P(L = l \mid \Theta = \theta) = P(\mathcal{E} = \theta - l). \quad (11)$$

Hledá se

Je třeba určit podmíněné rozdělení pravděpodobnosti náhodné veličiny Θ za podmínky, že byla provedena měření veličiny l s výsledky \tilde{l}_j , $j \in \{1, 2, \dots, n\}$.

Řešení

Hledané rozdělení pravděpodobnosti má hustotu ve tvaru pravděpodobnostní funkce $P(\Theta = \theta_k \mid (L = \tilde{l}_1) \cap (L = \tilde{l}_2) \cap \dots \cap (L = \tilde{l}_n))$. Její hodnoty vypočteme pomocí Bayesova vzorce (9). Pro jednoduchost a pro uvedení v soulad se značením v (9) označíme

$$A = (L = \tilde{l}_1) \cap (L = \tilde{l}_2) \cap \dots \cap (L = \tilde{l}_n) = \quad (12)$$

$$= \bigcap_{j=1}^n (L = \tilde{l}_j), \quad (13)$$

$$B_i = (\Theta = \theta_i), i \in \{1, 2, \dots, n\}. \quad (14)$$

Podmíněné pravděpodobnosti $P(L = l \mid \Theta = \theta_i)$ jsou dány vztahem (11).

$$P(L = l \mid \Theta = \theta_i) = P(\mathcal{E} = \theta_i - l), \quad (15)$$

který platí pro $\forall l \in \mathcal{L}$. Pravděpodobnost, že bude naměřena hodnota \tilde{l}_j za podmínky $\Theta = \theta_i$ je tedy

$$P(L = \tilde{l}_j | \Theta = \theta_i) = P(\mathcal{E} = \theta_i - \tilde{l}_j). \quad (16)$$

Pravděpodobnost, že budou nezávisle naměřeny hodnoty $\tilde{l}_1, \tilde{l}_2, \dots, \tilde{l}_n$ je pak součinem jednotlivých pravděpodobností:

$$P\left(\bigcap_{j=1}^n (L = \tilde{l}_j) | \Theta = \theta_i\right) = P(A|B_i) = \prod_{j=1}^n P(\mathcal{E} = \theta_i - \tilde{l}_j). \quad (17)$$

Po dosazení tohoto výsledku do Bayesova vzorce (9) dostáváme bayesovské řešení vyrovnání přímých měření v diskrétním případě:

$$P(\Theta = \theta_k | \bigcap_{j=1}^n (L = \tilde{l}_j)) = \frac{\prod_{j=1}^n P(\mathcal{E} = \theta_k - \tilde{l}_j) P(\Theta = \theta_k)}{\sum_{i=1}^m \prod_{j=1}^n P(\mathcal{E} = \theta_i - \tilde{l}_j) P(\Theta = \theta_i)}. \quad (18)$$

Pomocí aposteriorní pravděpodobnostní funkce $P(\Theta = \theta_k | \bigcap_{j=1}^n (L = \tilde{l}_j))$ lze nyní odhadnout libovolnou charakteristiku polohy náhodné veličiny L , např. modus, střední hodnotu nebo medián. Rovněž je možno určit přesnost odhadnuté hodnoty pomocí vhodné charakteristiky přesnosti, kterou nejčastěji bývá směrodatná odchylka (střední kvadratická chyba).

Číselný příklad

Uvedený příklad vyrovnání přímých měření je graficky znázorněn na obrázku 1.

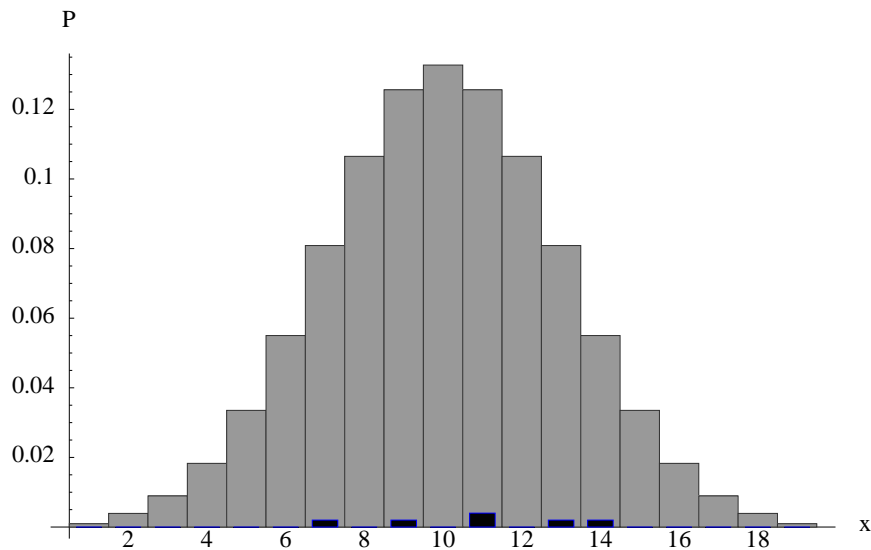
Na obrázku je tučnými obdélníky vyznačeno 6 měřených hodnot.

$$\tilde{l}_1 = 7, \tilde{l}_2 = 9, \tilde{l}_3 = 11, \tilde{l}_4 = 11, \tilde{l}_5 = 13, \tilde{l}_6 = 14.$$

Dále je zde zakreslena pravděpodobnostní funkce $P(L = l | \Theta = 10)$. Její tvar je známý, neboť je dáno rozdělení pravděpodobnosti chyb měření

$$\begin{aligned} & \{P(\mathcal{E} = \epsilon) | \epsilon \in \{-9, \dots, +9\}\} = \\ & = \{0.002, 0.004, 0.009, 0.018, 0.034, 0.055, 0.081, 0.106, 0.125, 0.133, \\ & 0.125, 0.106, 0.081, 0.055, 0.034, 0.018, 0.009, 0.004, 0.002\}. \end{aligned}$$

Poloha pravděpodobnostní funkce $P(L = l | \Theta = \theta)$ nad naměřenými daty však známa není, neboť parametr θ je neznámý a je předmětem odhadování.



Obrázek 1: Odhad parametru polohy v diskretním případě

Pro všechna přípustná rozdělení pravděpodobnosti reprezentovaná různými hodnotami parametru θ je třeba nyní stanovit jejich pravděpodobnosti. Množina \mathcal{T} všech přípustných parametrů polohy θ byla zvolena

$$\mathcal{T} = \{6, \dots, 18\}$$

Na následujícím obrázku 2 jsou přípustná rozdělení pravděpodobnosti $P(L = l | \Theta = \theta)$ znázorněna histogramy odlišenými různými odstíny šedi. Kvůli přehlednosti je vyznačen jen každý třetí histogram.

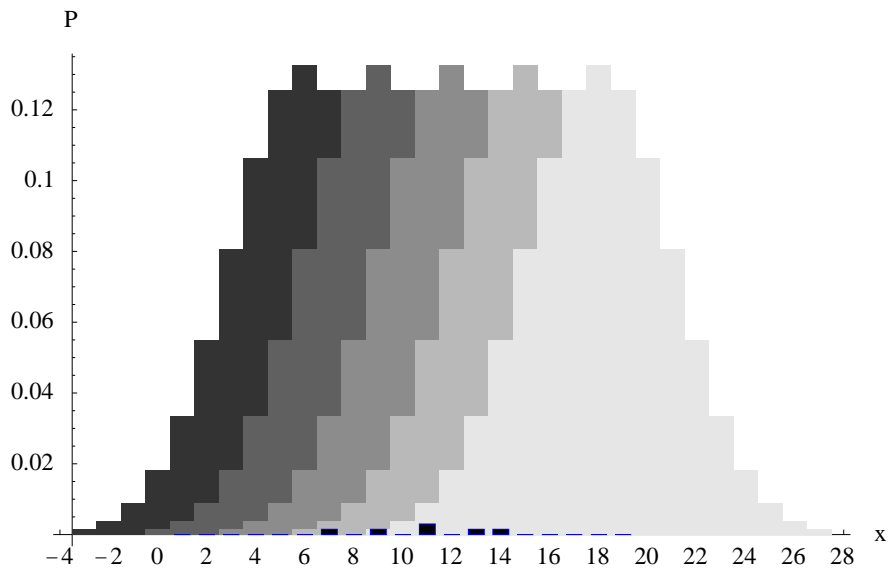
O odhadovaném parametru polohy nejsou k dispozici žádné apriorní informace a proto se předpokládá, že každá hodnota $\theta \in \mathcal{T}$ je očekávána se stejnou pravděpodobností. Apriorní rozdělení pravděpodobnosti parametru Θ je tedy rovnoměrné.

$$P(\Theta = \theta) = \frac{1}{13} \quad \text{pro } \forall \theta \in \mathcal{T}.$$

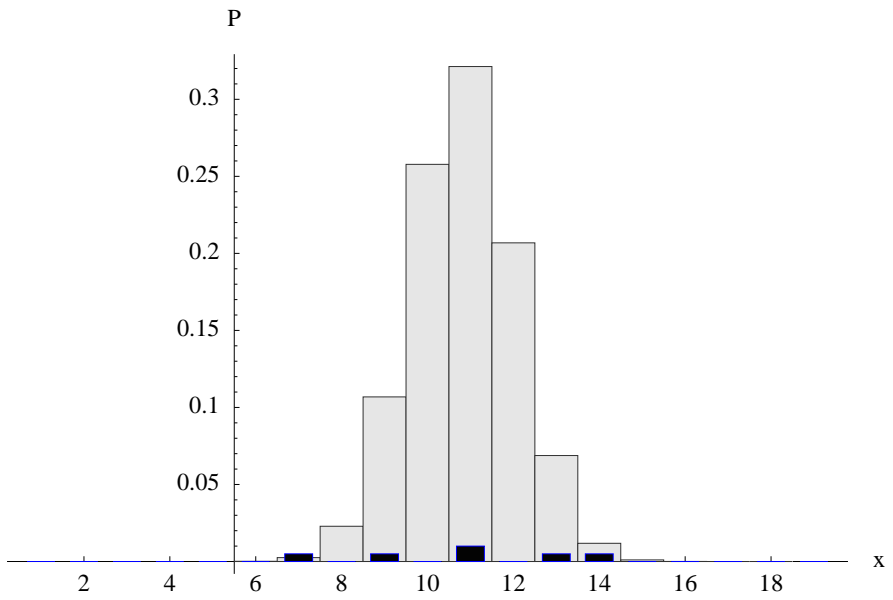
Dosažením zadaných údajů do Bayesova vzorce (18) dostáváme aposteriorní rozdělení pravděpodobnosti

$$\begin{aligned} & \left\{ P \left(\Theta = \theta \mid \bigcap_{j=1}^6 (L = \tilde{l}_j) \right) \mid \theta \in \{6, \dots, 18\} \right\} = \\ & = \{0, 0.003, 0.023, 0.107, 0.258, 0.321, 0.207, 0.069, 0.012, 0.001, \\ & \quad 0, 0, 0\}, \end{aligned}$$

které ukazuje obrázek 3.



Obrázek 2: Přípustná rozdělení pravděpodobnosti parametru polohy



Obrázek 3: Aposteriorní rozdělení pravděpodobnosti parametru polohy

Uvedený příklad vyrovnání přímých měření v diskrétním případě lze snadno zobecnit na spojitý případ pro více určovaných parametrů. Ten bývá tradičně označován jako vyrovnání zprostředkujících měření. Tímto problémem se zabývá odvětví matematické statistiky zvané *regresní analýza*. Bývá proto nazýván lineární, příp. nelineární regrese. Pojednává o něm následující kapitola.

3 Vyrovnání zprostředkujících měření

3.1 Formulace problému

Je třeba určit neznámé parametry $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_m)^T$, $m \in \mathbb{N}$. Tyto parametry jsou určovány pomocí měřených veličin $\mathbf{l} = (l_1, l_2, \dots, l_n)^T$, $n \in \mathbb{N}$ prostřednictvím daných funkcí

$$a_j : \mathcal{T} \rightarrow \mathcal{L}_j : [\theta_1, \theta_2, \dots, \theta_m] \mapsto a_j(\theta_1, \theta_2, \dots, \theta_m),$$

kde $\mathcal{T} \subseteq \mathbb{R}^m$, $\mathcal{L}_j \subseteq \mathbb{R}$, $m \leq n$. Zprostředkujícími vztahy mezi měřenými a určenými veličinami jsou rovnosti

$$a_j(\boldsymbol{\theta}) = l_j, \quad j \in \{1, 2, \dots, n\} \quad (19)$$

platné pro $\forall \boldsymbol{\theta} \in \mathcal{T}$. S použitím vektorového značení

$$\mathbf{a} \triangleq (a_1, a_2, \dots, a_n)^T \quad (20)$$

je definováno zobrazení

$$\mathbf{a} : \mathcal{T} \rightarrow \mathcal{L} : \boldsymbol{\theta} \mapsto \mathbf{a}(\boldsymbol{\theta}) = (a_1(\boldsymbol{\theta}), a_2(\boldsymbol{\theta}), \dots, a_n(\boldsymbol{\theta}))^T, \quad (21)$$

$$\mathcal{L} \triangleq \mathcal{L}_1 \times \dots \times \mathcal{L}_n \subseteq \mathbb{R}^n$$

a soustavu (19) lze přepsat na tvar

$$\mathbf{a}(\boldsymbol{\theta}) = \mathbf{l}. \quad (22)$$

Měřením veličin l_j byly získány hodnoty \tilde{l}_j , $j \in \{1, 2, \dots, n\}$, které lze rovněž uspořádat do aritmetického vektoru $\tilde{\mathbf{l}} \in \mathbb{R}^n$.

$$\tilde{\mathbf{l}} \triangleq (\tilde{l}_1, \tilde{l}_2, \dots, \tilde{l}_n)^T \quad (23)$$

Každá měřená hodnota \tilde{l}_j je zatížena chybou měření ε_j .

$$\varepsilon_j \triangleq l_j - \tilde{l}_j. \quad (24)$$

Vzhledem k nahodilosti chyb měření lze ke každé veličině ε_j , $j \in \{1, 2, \dots, n\}$ přiřadit náhodnou veličinu \mathcal{E}_j . Její hustotu pravděpodobnosti označíme $f_{\mathcal{E}_j}$. Stručně to budeme zapisovat:

$$\varepsilon_j \rightsquigarrow \mathcal{E}_j \sim f_{\mathcal{E}_j}. \quad (25)$$

Aplikací definice (24) v rovnicích (19) vznikne nová soustava rovnic

$$a_j(\boldsymbol{\theta}) - \tilde{l}_j = \varepsilon_j, \quad j \in \{1, 2, \dots, n\}. \quad (26)$$

Variabilita hodnot veličin ε_j , $j \in \{1, 2, \dots, n\}$ musí být s ohledem na rovnici (26) kompenzována variabilitou hodnot vektoru $\boldsymbol{\theta}$. Je proto třeba k vektoru parametrů $\boldsymbol{\theta}$ definovat náhodný vektor $\boldsymbol{\Theta}$.

Soustava rovnic (26) po “znáhodnění” veličin ε_j a vektoru $\boldsymbol{\theta}$ pak dostane podobu

$$a_j(\boldsymbol{\Theta}) - \tilde{l}_j = \varepsilon_j, \quad j \in \{1, 2, \dots, n\}. \quad (27)$$

Chyby měření ε_j můžeme uspořádat do aritmetického vektoru

$$\boldsymbol{\varepsilon} \triangleq (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$$

a k němu pak definovat náhodný vektor

$$\boldsymbol{\mathcal{E}} \triangleq (\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_n)^T.$$

Jeho hustotou pravděpodobnosti je sdružená hustota pravděpodobnosti náhodných veličin \mathcal{E}_j , $j \in \{1, 2, \dots, n\}$; označíme ji $f_{\boldsymbol{\mathcal{E}}}$. Stručně zapsáno:

$$\boldsymbol{\varepsilon} \rightsquigarrow \boldsymbol{\mathcal{E}} \sim f_{\boldsymbol{\mathcal{E}}}.$$

Soustavu rovnic (27) lze pak přepsat do vektorového tvaru

$$\mathbf{a}(\boldsymbol{\Theta}) - \tilde{\mathbf{I}} = \boldsymbol{\varepsilon}. \quad (28)$$

Z rovnice (28) je vidět, že rozdělení pravděpodobnosti náhodného vektoru $\boldsymbol{\Theta}$ závisí na měřených hodnotách \tilde{l}_j , $j \in \{1, 2, \dots, n\}$. Abychom mohli tuto závislost vyjádřit prostřednictvím podmíněného rozdělení pravděpodobnosti, je třeba ještě zavést náhodný vektor \mathbf{L} příslušný vektoru měřených veličin \mathbf{l} .

$$\mathbf{l} \rightsquigarrow \mathbf{L},$$

$$\mathbf{L} \triangleq \tilde{\mathbf{I}} + \boldsymbol{\varepsilon}. \quad (29)$$

Jeho rozdělení pravděpodobnosti je závislé na hodnotě náhodného vektoru $\boldsymbol{\Theta}$. Je jím proto podmíněné rozdělení pravděpodobnosti s hustotou

$$f_{\mathbf{L}}(\cdot | \boldsymbol{\theta}) : \mathcal{L} \rightarrow \mathbb{R} : \mathbf{l} \mapsto f_{\mathbf{L}}(\mathbf{l} | \boldsymbol{\theta}).$$

Tedy

$$\mathbf{L} \sim f_{\mathbf{L}}(\cdot | \boldsymbol{\theta}).$$

Podobně definujeme i podmíněné rozdělení pravděpodobnosti parametrů $\boldsymbol{\Theta}$ za podmínky, že byly naměřeny hodnoty $\tilde{\mathbf{I}}$, tj. že $\mathbf{L} = \tilde{\mathbf{I}}$.

$$\boldsymbol{\Theta} \sim f_{\boldsymbol{\Theta}}(\cdot | \tilde{\mathbf{I}}).$$

Toto tzv. aposteriorní rozdělení pravděpodobnosti lze určit na základě známého rozdělení $f_{\mathbf{L}}(\cdot | \boldsymbol{\theta})$, pokud je dáno tzv. apriorní rozdělení pravděpodobnosti

náhodného vektoru Θ . Apriorní rozdělení pravděpodobnosti představuje souhrn informací o určovaných parametrech θ dostupných ještě před provedením měření. Hustotu apriorního rozdělení pravděpodobnosti označíme p .

$$\Theta \sim p .$$

K náhodné veličině Θ jsou tedy definována dvě různá rozdělení pravděpodobnosti: apriorní p , které je dáno, a aposteriorní $f_{\Theta}(\cdot | \tilde{\mathbf{I}})$, které se má určit.

Nyní lze již zadané předpoklady shrnout a přehledně formulovat problém vyrovnání zprostředkujících veličin (regrese).

Dáno:

- zobrazení \mathbf{a} , viz (20);
- soustava rovnic $\mathbf{a}(\theta) = \mathbf{l}$, viz (22);
- měřené hodnoty $\tilde{\mathbf{I}}$ příslušné veličinám \mathbf{l} ;
- sdružená hustota pravděpodobnosti chyb měření $\dots f_{\varepsilon}$;
- apriorní hustota pravděpodobnosti parametrů $\theta \dots p$.

Hledá se:

- podmíněná hustota pravděpodobnosti parametrů θ za podmínky, že byly naměřeny hodnoty $\tilde{\mathbf{I}}$ veličin $\mathbf{l} \dots f_{\Theta}(\cdot | \tilde{\mathbf{I}})$.

3.2 Obecné řešení problému

Hledanou podmíněnou hustotu pravděpodobnosti $f_{\Theta}(\theta | \tilde{\mathbf{I}})$ lze přímo určit podle Bayesovy věty

$$f_{\Theta}(\theta | \tilde{\mathbf{I}}) = \frac{f_{\mathbf{L}}(\tilde{\mathbf{I}} | \theta) p(\theta)}{\int_{\mathcal{T}} f_{\mathbf{L}}(\tilde{\mathbf{I}} | \mathbf{t}) p(\mathbf{t}) d\mathbf{t}} . \tag{30}$$

Bayesova věta v tomto tvaru je zobecněním Bayesova vzorce (9) pro spojitě náhodné veličiny (jevy B_k v (9) lze chápat jako hodnoty nějaké diskrétní náhodné veličiny). Její důkaz je uveden např. v [And78] na str. 54.

O chybách měření ε předpokládáme, že nejsou závislé na hodnotě náhodného vektoru θ . Pro jejich rozdělení pravděpodobnosti f_{ε} proto platí vzhledem k (29) rovnost

$$f_{\mathbf{L}}(\tilde{\mathbf{I}} | \theta) = f_{\varepsilon}(\mathbf{l} - \tilde{\mathbf{I}}) . \tag{31}$$

Tento přirozený předpoklad umožňuje přejít od hustoty pravděpodobnosti měřených veličin, která se vyskytuje v Bayesově větě (30), k hustotě pravděpodobnosti chyb měření, jež je dána. Pak již zbývá jen dosadit do (31) za \mathbf{l} podle (22) a poté uplatnit rovnost (31) v (30). Tím vznikne hledaný vztah pro aposteriorní hustotu pravděpodobnosti určených parametrů $\boldsymbol{\theta}$:

$$f_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \tilde{\mathbf{l}}) = \frac{f_{\boldsymbol{\varepsilon}}(\mathbf{a}(\boldsymbol{\theta}) - \tilde{\mathbf{l}}) p(\boldsymbol{\theta})}{\int_{\mathcal{T}} f_{\boldsymbol{\varepsilon}}(\mathbf{a}(\mathbf{t}) - \tilde{\mathbf{l}}) p(\mathbf{t}) dt} . \quad (32)$$

Tento vztah platí pro závislá i nezávislá měření. Pokud jsou měření $\tilde{\mathbf{l}}$ vzájemně nezávislá, tzn. že náhodné veličiny ε_j , $j \in \{1, 2, \dots, n\}$ jsou statisticky nezávislé, je možno sdruženou hustotu pravděpodobnosti $f_{\boldsymbol{\varepsilon}}$ vyjádřit jako součin hustot jednotlivých náhodných veličin ε_j .

$$f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}) = \prod_{j=1}^n f_{\varepsilon_j}(\varepsilon_j) . \quad (33)$$

Dosazením tohoto vztahu do (32) vznikne verze Bayesovy věty pro nezávislá měření:

$$f_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \tilde{\mathbf{l}}) = \frac{\prod_{j=1}^n f_{\varepsilon_j}(a_j(\boldsymbol{\theta}) - \tilde{l}_j) p(\boldsymbol{\theta})}{\int_{\mathcal{T}} \prod_{j=1}^n f_{\varepsilon_j}(a_j(\mathbf{t}) - \tilde{l}_j) p(\mathbf{t}) dt} . \quad (34)$$

Porovnáním tohoto vztahu s Bayesovým vzorcem ve tvaru (18) je názorně vidět způsob zobecnění diskrétního případu s pravděpodobnostními funkcemi na spojitý případ s hustotami.

3.3 Bayesovský přístup a teorie odhadu

Bayesova věta ve tvaru (32) umožňuje výpočet aposteriorní hustoty pravděpodobnosti pro dané funkce a_j , $j \in \{1, 2, \dots, n\}$, apriorní hustotu p a sdruženou hustotu pravděpodobnosti chyb měření $f_{\boldsymbol{\varepsilon}}$. Za vyrovnané hodnoty neboli odhad určených veličin neboli vyrovnané hodnoty je možno vzít libovolnou charakteristiku polohy aposteriorního rozdělení, např. střední hodnotu, medián, modus apod. Při každé volbě typu odhadu je současně dáno i jeho rozdělení pravděpodobnosti. Vždy je to aposteriorní rozdělení $f_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \tilde{\mathbf{l}})$, z kterého byl příslušný odhad odvozen. Tím se bayesovský přístup odlišuje od postupu obvyklého v teorii odhadu, kdy se nejprve stanoví určitý odhad jako funkce měřených veličin a k němu se pak odvozuje jeho rozdělení pravděpodobnosti.

Uvedené charakteristiky polohy jsou zvláštními případy tzv. *bayesovských odhadů*. Obecně je bayesovský odhad definován pomocí minimalizace tzv. ztrátové

funkce. Přesná definice bayesovského odhadu je uvedena např. v [Huš85] na str. 50–65 nebo stručněji v [Koc90], str. 33–36.

Pokud je za charakteristiku polohy zvolen modus a apriorní rozdělení je rovnoměrné, příp. neinformativní (s nevlastní hustotou — viz pododdíl 4.2.2 a oddíl 4.3), je tento odhad totožný s maximálně věrohodným odhadem — viz např. [And78], str. 266–273 nebo [Rao78], str. 395–397.

Bayesovský přístup poskytuje rovněž potřebné informace pro statistické testování hypotéz. O tom podrobně pojednává např. [Huš85] na str. 65–85 nebo [Koc90], str. 40–48.

Ve statistické literatuře bývá soustava rovnic (22) nazývána *model* a případ, kdy zobrazení \mathbf{a} je lineární je označován jako *lineární model*. Tohoto názvosloví se přidržíme i v této práci. Někdy bývají pod pojem model zahrnovány i údaje o rozděleních pravděpodobnosti náhodného vektoru \mathbf{L} . Potom by mohl být tento případ označován i jako *lineární a normální model*.

V následující kapitole bude podrobně zpracován bayesovský přístup pro lineární model s normálním rozdělením chyb měření. Podstata následujícího postupu je stručně naznačena v [Ht94] na str. 196, 197.

4 Lineární model s normálním rozdělením chyb měření

Výpočet aposteriorní hustoty je zvláště snadný pro lineární funkce a_j , $j \in \{1, 2, \dots, n\}$ a normální rozdělení chyb měření f_ϵ , pokud je vhodně zvolena apriorní hustota p . Proto je tento případ nejlépe teoreticky propracován a v praxi nejpoužívanější. Je podrobně popsán např. v [Koc90], kap. 31 nebo stručněji v [Ht94], str. 195–199.

4.1 Předpoklady

V této kapitole budeme předpokládat, že funkce a_j , $j \in \{1, 2, \dots, n\}$ jsou lineární. Platí tedy

$$a_j(\boldsymbol{\theta}) = a_j(\tilde{\boldsymbol{\theta}}) + (\mathbf{a}'_j)^T (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}), \quad (35)$$

kde $\tilde{\boldsymbol{\theta}}$ je pevně daný vektor určených parametrů, který přibližně vyhovuje soustavě rovnic (19) pro $l_j = \tilde{l}_j$. Jeho složky bývají proto v geodetické literatuře nazývány *přibližné neznámé*. Vektor \mathbf{a}'_j představuje vektor parciálních derivací funkce a_j , který je vzhledem k linearitě funkcí a_j nezávislý na $\boldsymbol{\theta}$.

$$\mathbf{a}'_j \triangleq \left(\frac{\partial a_j}{\partial \theta_1}(\tilde{\boldsymbol{\theta}}), \dots, \frac{\partial a_j}{\partial \theta_m}(\tilde{\boldsymbol{\theta}}) \right)^T. \quad (36)$$

Dále předpokládáme normalitu chyb měření. Pro nezávislá měření tedy platí:

$$\begin{aligned} \forall j \in \{1, 2, \dots, n\} : \mathcal{E}_j &\sim f_j, \\ f_j(\epsilon) &= \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2}\left(\frac{\epsilon}{\sigma_j}\right)^2\right). \end{aligned} \quad (37)$$

Úmluva Pro normální rozdělení se vžil označení $\mathcal{N}(\mu, \sigma^2)$, kde μ je střední hodnota a σ směrodatná odchylka (střední chyba). Skutečnost, že náhodná veličina X má normální rozdělení $\mathcal{N}(\mu, \sigma^2)$ se zapisuje

$$X \sim \mathcal{N}(\mu, \sigma^2). \quad (38)$$

Podobně se označuje i vícerozměrné normální rozdělení s vektorem středních hodnot $\boldsymbol{\mu} \in \mathbb{R}^m$ a kovarianční maticí \mathbf{Q} : $\mathcal{N}_m(\boldsymbol{\mu}, \mathbf{Q})$, kde m je dimenze příslušného náhodného vektoru \mathbf{X} . Jeho hustota pravděpodobnosti je (viz např. [LM81], str. 123):

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^m \det(\mathbf{Q})}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{Q}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right). \quad (39)$$

Označení (38) budeme používat i v této práci, přestože jsme již symbolu \sim použili pro vztah náhodné veličiny a hustoty pravděpodobnosti. Relační operátor \sim tedy dodefinujeme tak, že pro náhodnou veličinu X s rozdělením pravděpodobnosti D platí

$$X \sim D \Leftrightarrow X \sim f ,$$

pokud f je hustota pravděpodobnosti rozdělení D .

S ohledem na tuto úmluvu lze předpoklad (37) zapsat stručně:

$$\forall j \in \{1, 2, \dots, n\} : \mathcal{E}_j \sim \mathcal{N}(0, \sigma_j^2) . \quad (40)$$

Pro závislá měření má předpoklad normality tvar: $\boldsymbol{\mathcal{E}} \sim \mathcal{N}_n(\mathbf{o}, \mathbf{Q})$, kde \mathbf{o} je nulový vektor prostoru \mathbb{R}^n a \mathbf{Q} kovarianční matice měřených veličin L_j , $j \in \{1, 2, \dots, n\}$.

Aby bylo možno v Bayesově větě (34) využít předpokladů lineariry a normality ((35) a (37), resp. (40)) je třeba ještě vhodně zvolit apriorní rozdělení p . Pro volbu vhodného apriorního rozdělení bylo navrženo několik metod. Těmto metodám je věnována kap. 2 v [Huš85]. Jsou popsány také v [Koc90], příp v dalších knihách zmíněných v oddíle (2.2). Na tomto místě se jimi nebudeme zabývat a použijeme dvě z nich, u nichž je jejich opodstatněnost zřejmá. V našem případě bude vhodným apriorním rozdělením p normální rozdělení (metoda konjugovaných systémů hustot) nebo rovnoměrné rozdělení (princip neurčitosti). Vhodnost této volby bude patrná z úvah na začátku odstavce 4.2.1.

Po zavedení Jacobiho matice \mathbf{A} zobrazení \mathbf{a}

$$\mathbf{A} \triangleq (\mathbf{a}'_1, \mathbf{a}'_2, \dots, \mathbf{a}'_n)^T \quad (41)$$

dostane (35) přehlednější tvar

$$\mathbf{a}(\boldsymbol{\theta}) = \mathbf{a}(\tilde{\boldsymbol{\theta}}) + \mathbf{A}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) . \quad (42)$$

Poznámka: Abychom se vyhnuli případným pozdějším potížím, předpokládáme, že matice \mathbf{A} má plnou hodnost, tzn. její sloupce jsou lineárně nezávislé. Tento předpoklad sice není z teoretického hlediska nutný, neboť lze řešit lineární model i při neúplné hodnosti matice \mathbf{A} , avšak bez něj by se následující výklad zbytečně zkomplikoval. K řešení tohoto problému bylo navrženo několik metod, viz např. [Rao78] str. 333 – 342, [And78], str. 135–141. Z hlediska bayesovského přístupu je lineární model s neúplnou hodností vyřešen v [Koc90] na str. 77–83. V geodézii nastává tento případ při vyrovnání volných sítí.

4.2 Odvození aposteriorní hustoty

Vzhledem k normalitě chyb měření (37) lze určit součin v čitateli (34).

$$\prod_{j=1}^n f_{\mathcal{E}_j}(a_j(\boldsymbol{\theta}) - \tilde{l}_j) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2}\left(\frac{a_j(\boldsymbol{\theta}) - \tilde{l}_j}{\sigma_j}\right)^2\right) =$$

$$= \prod_{j=1}^n \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2} \sum_{k=1}^n \left(\frac{a_k(\boldsymbol{\theta}) - \tilde{l}_k}{\sigma_k}\right)^2\right). \quad (43)$$

Argument exponenciální funkce v předchozím výrazu lze pak s pomocí vektorového značení (20), (23) zapsat v maticovém tvaru, zavedeme-li váhovou matici \mathbf{P} jakožto inverzní kovarianční matici náhodného vektoru $\boldsymbol{\mathcal{E}}$.

$$\mathbf{P} \triangleq (\text{cov}(\boldsymbol{\mathcal{E}}))^{-1}, \text{ tj.}$$

$$\mathbf{P} = \text{diag}\left(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_n^2}\right).$$

Potom

$$\prod_{j=1}^n f_{\mathcal{E}_j}(a_j(\boldsymbol{\theta}) - \tilde{l}_j) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2} \left((\mathbf{a}(\boldsymbol{\theta}) - \tilde{\mathbf{l}})^T \mathbf{P} (\mathbf{a}(\boldsymbol{\theta}) - \tilde{\mathbf{l}})\right)\right). \quad (44)$$

Argument exponenciální funkce na pravé straně předcházející rovnosti lze s pomocí rovnice (42) dále upravit. Jeho podstatnou část (bez násobné konstanty $-\frac{1}{2}$) označíme $\psi(\boldsymbol{\theta})$.

$$\psi(\boldsymbol{\theta}) \triangleq \left(\mathbf{a}(\tilde{\boldsymbol{\theta}}) + \mathbf{A}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) - \tilde{\mathbf{l}}\right)^T \mathbf{P} \left(\mathbf{a}(\tilde{\boldsymbol{\theta}}) + \mathbf{A}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) - \tilde{\mathbf{l}}\right). \quad (45)$$

Zavedeme-li ještě pomocný vektor $\boldsymbol{\lambda}$

$$\boldsymbol{\lambda} \triangleq \mathbf{a}(\tilde{\boldsymbol{\theta}}) - \tilde{\mathbf{l}}, \quad (46)$$

dostane (45) podobu

$$\psi(\boldsymbol{\theta}) = \left(\mathbf{A}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) + \boldsymbol{\lambda}\right)^T \mathbf{P} \left(\mathbf{A}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) + \boldsymbol{\lambda}\right). \quad (47)$$

Rovnost (44) tím dostane přehlednější tvar.

$$\prod_{j=1}^n f_{\mathcal{E}_j}(a_j(\boldsymbol{\theta}) - \tilde{l}_j) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2} \psi(\boldsymbol{\theta})\right). \quad (48)$$

S využitím tohoto výsledku v (34) se zjednoduší i zápis Bayesovy věty.

$$f_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \tilde{\mathbf{l}}) = \frac{\exp\left(-\frac{1}{2} \psi(\boldsymbol{\theta})\right) p(\boldsymbol{\theta})}{\int_{\mathcal{T}} \exp\left(-\frac{1}{2} \psi(\mathbf{t})\right) p(\mathbf{t}) d\mathbf{t}}. \quad (49)$$

Dosadíme-li podle (47) za $\psi(\boldsymbol{\theta})$ a $\psi(\mathbf{t})$ do (49), dostáváme pro aposteriorní hustotu

$$f_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \tilde{\mathbf{l}}) = \frac{\exp\left(-\frac{1}{2} \left((\mathbf{A}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) + \boldsymbol{\lambda})^T \mathbf{P} (\mathbf{A}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) + \boldsymbol{\lambda})\right)\right) p(\boldsymbol{\theta})}{\int_{\mathcal{T}} \exp\left(-\frac{1}{2} \left((\mathbf{A}(\mathbf{t} - \tilde{\boldsymbol{\theta}}) + \boldsymbol{\lambda})^T \mathbf{P} (\mathbf{A}(\mathbf{t} - \tilde{\boldsymbol{\theta}}) + \boldsymbol{\lambda})\right)\right) p(\mathbf{t}) d\mathbf{t}}. \quad (50)$$

Aby bylo možno snadno vypočítat hodnotu integrálu ve jmenovateli (50), je třeba aby integrand byl ve tvaru

$$c \exp\left(-\frac{1}{2}(\mathbf{x}^T \mathbf{C} \mathbf{x})\right)$$

s reálnou konstantou $c \in \mathbb{R}$ a diagonální maticí $\mathbf{C} = \text{diag}(c_1, c_2, \dots, c_m)$, $m \in \mathbb{N}$, $c_i \in \mathbb{R}$, $c_i > 0$ pro $\forall i \in \{1, 2, \dots, m\}$. Potom by integrál

$$\int_{\mathcal{X}} c \exp\left(-\frac{1}{2}(\mathbf{x}^T \mathbf{C} \mathbf{x})\right) d\mathbf{x} \quad (51)$$

bylo možno vyjádřit pomocí chybové funkce erf, pokud by integrační oblast \mathcal{X} byla m -rozměrný interval v \mathbb{R}^m , tj. m -rozměrný kvádr s hranami rovnoběžnými se souřadnicovými osami. Skutečně, pokud je kvádr \mathcal{X} definován jako m -násobný kartézský součin

$$\begin{aligned} \mathcal{X} &\triangleq (x_{1,\min}, x_{1,\max}) \times (x_{2,\min}, x_{2,\max}) \times \dots \times (x_{m,\min}, x_{m,\max}), \\ &\forall i \in \{1, 2, \dots, m\} : x_{i,\min}, x_{i,\max} \in \mathbb{R}, x_{i,\min} < x_{i,\max}, \end{aligned} \quad (52)$$

pak lze integrál (51) rozložit v součin jednoduchých integrálů

$$\begin{aligned} \int_{\mathcal{X}} c \exp\left(-\frac{1}{2}(\mathbf{x}^T \mathbf{C} \mathbf{x})\right) d\mathbf{x} &= c \prod_{i=1}^m \int_{x_{i,\min}}^{x_{i,\max}} \exp\left(-\frac{1}{2} c_i t_i^2\right) dt_i = \\ &= c \prod_{i=1}^m \sqrt{\frac{2\pi}{c_i}} \frac{1}{2} \left(\text{erf}\left(x_{i,\max} \sqrt{\frac{c_i}{2}}\right) - \text{erf}\left(x_{i,\min} \sqrt{\frac{c_i}{2}}\right) \right). \end{aligned} \quad (53)$$

Připomeňme zde definici a základní vlastnosti chybové funkce erf:

$$\text{erf}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt,$$

$$\lim_{x \rightarrow +\infty} \text{erf}(x) = +1, \quad (54)$$

$$\lim_{x \rightarrow -\infty} \text{erf}(x) = -1. \quad (55)$$

V případě, že $\mathcal{X} = \mathbb{R}^m$, se rovnosti (53) dále zjednoduší. Vzhledem k (54) a (55) totiž platí:

$$\begin{aligned} \int_{\mathbb{R}^m} c \exp\left(-\frac{1}{2}(\mathbf{x}^T \mathbf{C} \mathbf{x})\right) d\mathbf{x} &= c \prod_{i=1}^m \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2} c_i t_i^2\right) dt_i = \\ &= c \prod_{i=1}^m \sqrt{\frac{2\pi}{c_i}} = c \frac{(\sqrt{2\pi})^m}{\prod_{i=1}^m \sqrt{c_i}} = c \frac{(\sqrt{2\pi})^m}{\det(\mathbf{C})}. \end{aligned} \quad (56)$$

Aby bylo možno využít výsledku (53), příp. (56), je třeba integrand v (50) upravit na tvar, který by umožňoval nalézt takové p , c , \mathcal{X} , \mathbf{C} , aby platilo:

$$\int_{\mathcal{T}} \exp\left(-\frac{1}{2}\psi(\mathbf{t})\right) p(\mathbf{t}) d\mathbf{t} = \int_{\mathcal{X}} c \exp\left(-\frac{1}{2}(\mathbf{x}^T \mathbf{C} \mathbf{x})\right) d\mathbf{x}. \quad (57)$$

4.2.1 Volba apriorní hustoty pravděpodobnosti

Prvním krokem ke splnění podmínky (57) je stanovení vhodné apriorní hustoty p . Podmínka (57) zůstane zřejmě zachována pro apriorní hustotu rovnoměrnou nebo normální. Dalším důvodem pro tyto dva typy rozdělení je skutečnost, že představují největší neurčitost apriorních informací. Za míru neurčitosti náhodné veličiny s hustotou f se obvykle považuje diferenciální Shannonova entropie

$$H(f) = -k \int_{\mathbb{R}} f(x) \log f(x) dx, \quad (58)$$

kde k je stanovená kladná konstanta. Rozdělením pravděpodobnosti, které má maximální entropii ze všech možných rozdělení definovaných na intervalu (a, b) (tj. jejichž základním souborem je (a, b)), je rovnoměrné rozdělení s hustotou

$$f(\mathbf{t}) = \begin{cases} \frac{1}{b-a} & \Leftarrow \mathbf{t} \in (a, b), \\ 0 & \Leftarrow \mathbf{t} \notin (a, b). \end{cases} \quad (59)$$

Podobnou vlastnost má i normální rozdělení, jež představuje rozdělení pravděpodobnosti s maximální entropií při dané střední hodnotě a varianci. Důkaz těchto vlastností je uveden např. v [Koc90], str. 17–20. Tyto výsledky lze zobecnit i na vícerozměrný případ.

Uvažujme nejprve rovnoměrné apriorní rozdělení pravděpodobnosti, tj. případ, kdy funkce p je konstantní na nějaké, zatím blíže neurčené množině $\mathcal{T} \subset \mathbb{R}^m$ a mimo množinu \mathcal{T} je nulová.

$$p(\mathbf{t}) = \begin{cases} \frac{1}{|\mathcal{T}|} & \Leftarrow \mathbf{t} \in \mathcal{T}, \\ 0 & \Leftarrow \mathbf{t} \notin \mathcal{T}. \end{cases} \quad (60)$$

Symbol $|\mathcal{T}|$ představuje objem množiny \mathcal{T} , tj.

$$|\mathcal{T}| \triangleq \int_{\mathcal{T}} d\mathbf{t}.$$

Potom je možno konstantu $p(\mathbf{t})$ na levé straně rovnosti (57) vytknout před integrál. Dále je třeba vyjádřit $\psi(\mathbf{t})$ jako součet konstanty a kvadratické formy.

Následující úprava

$$\begin{aligned}\psi(\mathbf{t}) &= (\mathbf{A}(\mathbf{t} - \tilde{\boldsymbol{\theta}}) + \boldsymbol{\lambda})^T \mathbf{P} (\mathbf{A}(\mathbf{t} - \tilde{\boldsymbol{\theta}}) + \boldsymbol{\lambda}) = \\ &= (\mathbf{t} - \tilde{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\mathbf{t} - \tilde{\boldsymbol{\theta}}) + 2 (\mathbf{t} - \tilde{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \boldsymbol{\lambda} + \boldsymbol{\lambda}^T \mathbf{P} \boldsymbol{\lambda}\end{aligned}\quad (61)$$

ukazuje, že argumentem hledané kvadratické formy nemůže být ani samotný parametr \mathbf{t} , ani rozdíl $\mathbf{t} - \tilde{\boldsymbol{\theta}}$. Předpokládejme tedy, že argumentem hledané kvadratické formy je $\mathbf{t} - \hat{\boldsymbol{\theta}}$, kde $\hat{\boldsymbol{\theta}}$ je nějaký zatím neurčený pomocný konstantní vektor z \mathcal{T} . S jeho pomocí pak vyjádříme rozdíl $\mathbf{t} - \tilde{\boldsymbol{\theta}}$ z předchozí rovnosti následovně:

$$\mathbf{t} - \tilde{\boldsymbol{\theta}} = (\mathbf{t} - \hat{\boldsymbol{\theta}}) + (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}). \quad (62)$$

Po dosazení (62) do (61) můžeme dále pokračovat v úpravě výrazu $\psi(\mathbf{t})$.

$$\begin{aligned}\psi(\mathbf{t}) &= (\mathbf{t} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\mathbf{t} - \hat{\boldsymbol{\theta}}) + 2 (\mathbf{t} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) + \\ &+ (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) + 2 (\mathbf{t} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \boldsymbol{\lambda} + \\ &+ 2 (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \boldsymbol{\lambda} + \boldsymbol{\lambda}^T \mathbf{P} \boldsymbol{\lambda}.\end{aligned}\quad (63)$$

Výsledek poslední úpravy naznačuje, jak dosáhnout požadovaného vyjádření ve tvaru součtu konstanty a kvadratické formy v $\mathbf{t} - \hat{\boldsymbol{\theta}}$. Je třeba, aby se v (63) nevyskytovaly ty členy, které jsou lineární v $\mathbf{t} - \hat{\boldsymbol{\theta}}$. Součet lineárních členů v (63) tedy musí vymizet, tj.

$$2 (\mathbf{t} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) + 2 (\mathbf{t} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \boldsymbol{\lambda} = 0.$$

Na tuto rovnost lze pohlížet jako na rovnici pro zatím neurčené $\hat{\boldsymbol{\theta}}$. Po vytknutí $(\mathbf{t} - \hat{\boldsymbol{\theta}})^T$ jsou patrné dvě možnosti anulování lineárních členů v $\psi(\mathbf{t})$. První možnost, $\hat{\boldsymbol{\theta}} = \mathbf{t}$ není přípustná, neboť $\hat{\boldsymbol{\theta}}$ má být podle předpokladu konstantní a nemůže tudíž být roven integrační proměnné.

Druhou možnost představuje rovnice

$$\mathbf{A}^T \mathbf{P} \mathbf{A} (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) = -\mathbf{A}^T \mathbf{P} \boldsymbol{\lambda},$$

kteřou lze snadno řešit za podmínky regularity matice $\mathbf{A}^T \mathbf{P} \mathbf{A}$. Tato podmínka je díky předpokladu o plné hodnosti matice \mathbf{A} , který byl učiněn v poznámce za definicí (41), splněn, takže dostáváme jediné řešení

$$\hat{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}} - (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \boldsymbol{\lambda}. \quad (64)$$

Hledané vyjádření je pak

$$\psi(\mathbf{t}) = (\mathbf{t} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\mathbf{t} - \hat{\boldsymbol{\theta}}) + (\mathbf{A}(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) + \boldsymbol{\lambda})^T \mathbf{P} (\mathbf{A}(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) + \boldsymbol{\lambda}). \quad (65)$$

Maticice $\mathbf{A}^T \mathbf{P} \mathbf{A}$ v (65) je maticí hledané kvadratické formy vzhledem k ortonormální bázi aritmetického vektorového prostoru $\mathbb{R}^m - \hat{\boldsymbol{\theta}}$. Příslušnou aditivní konstantu označíme r .

$$r \triangleq (\mathbf{A}(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) + \boldsymbol{\lambda})^T \mathbf{P} (\mathbf{A}(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) + \boldsymbol{\lambda}) . \quad (66)$$

S tímto označením se (65) zjednoduší na tvar:

$$\psi(\mathbf{t}) = (\mathbf{t} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\mathbf{t} - \hat{\boldsymbol{\theta}}) + r . \quad (67)$$

S pomocí této substituce pak již můžeme dosadit za $\psi(\mathbf{t})$ do (57).

$$\frac{e^{-\frac{r}{2}}}{|\mathcal{J}|} \int_{\mathcal{J}} \exp \left(-\frac{1}{2} (\mathbf{t} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\mathbf{t} - \hat{\boldsymbol{\theta}}) \right) d\mathbf{t} = c \int_{\mathcal{X}} \exp \left(-\frac{1}{2} (\mathbf{x}^T \mathbf{C} \mathbf{x}) \right) d\mathbf{x} . \quad (68)$$

Aby bylo možno porovnat integrály v (68), je třeba transformovat kvadratickou formu $(\mathbf{t} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\mathbf{t} - \hat{\boldsymbol{\theta}})$ do jiné báze vektorového prostoru $\mathbb{R}^m - \hat{\boldsymbol{\theta}}$ tak, aby její matice byla již diagonální, tj. diagonalizovat ji. To lze provést pomocí lineární transformace

$$\mathcal{B} : \mathbb{R}^m - \hat{\boldsymbol{\theta}} \rightarrow \mathbb{R}^m : (\mathbf{t} - \hat{\boldsymbol{\theta}}) \mapsto \mathcal{B}(\mathbf{t} - \hat{\boldsymbol{\theta}}) = \mathbf{B}(\mathbf{t} - \hat{\boldsymbol{\theta}}) . \quad (69)$$

Nejčastější a v našem případě nejvýhodnější je ortonormální transformace, kdy \mathbf{B} je ortonormální matice, tj. matice splňující podmínku $\mathbf{B} \mathbf{B}^T = \mathbf{I}$, v níž \mathbf{I} je identická matice řádu m . Potom platí

$$\mathbf{B}^T \mathbf{A}^T \mathbf{P} \mathbf{A} \mathbf{B} = \mathbf{C} . \quad (70)$$

Řádky matice \mathbf{B} jsou vlastní vektory matice $\mathbf{A}^T \mathbf{P} \mathbf{A}$ a diagonální prvky matice \mathbf{C} jsou vlastní čísla matice $\mathbf{A}^T \mathbf{P} \mathbf{A}$ (viz např. [Rao78] str. 62).

Tím jsme získali jedno z řešení rovnice (57).

Zbývající hledané vztahy pro c , \mathcal{X} pak dostaneme po substituci $\mathbf{t} - \hat{\boldsymbol{\theta}} = \mathbf{B} \mathbf{x}$ v integrálu na levé straně rovnosti (68). Tuto substituci provedeme pro přehlednost ve dvou krocích: nejprve $\mathbf{t} - \hat{\boldsymbol{\theta}} = \mathbf{u}$ a pak $\mathbf{u} = \mathbf{B} \mathbf{x}$.

$$\begin{aligned} & \frac{e^{-\frac{r}{2}}}{|\mathcal{J}|} \int_{\mathcal{J}} \exp \left(-\frac{1}{2} (\mathbf{t} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\mathbf{t} - \hat{\boldsymbol{\theta}}) \right) d\mathbf{t} = \\ & = \frac{e^{-\frac{r}{2}}}{|\mathcal{J}|} \int_{\mathcal{J} - \hat{\boldsymbol{\theta}}} \exp \left(-\frac{1}{2} \mathbf{u}^T \mathbf{A}^T \mathbf{P} \mathbf{A} \mathbf{u} \right) d\mathbf{u} = \\ & = \frac{e^{-\frac{r}{2}}}{|\mathcal{J}|} \int_{\mathbf{B}^{-1}(\mathcal{J} - \hat{\boldsymbol{\theta}})} \exp \left(-\frac{1}{2} \mathbf{x}^T \mathbf{B}^T \mathbf{A}^T \mathbf{P} \mathbf{A} \mathbf{B} \mathbf{x} \right) \det(\mathbf{B}^T) d\mathbf{x} . \end{aligned} \quad (71)$$

Nyní můžeme využít ortonormality matice \mathbf{B} . Platí totiž

$$\begin{aligned}\mathbf{B}^{-1} &= \mathbf{B}^T, \\ \det(\mathbf{B}^T) &= \det(\mathbf{B}) = 1.\end{aligned}$$

Použijeme-li ještě vztahu (70), zjednoduší se (71) natolik, že můžeme psát rovnici (68) v konečném tvaru, který přímo ukazuje její řešení.

$$\frac{e^{-\frac{r}{2}}}{|\mathcal{J}|} \int_{\mathbf{B}^T(\mathcal{J}-\hat{\boldsymbol{\theta}})} \exp\left(-\frac{1}{2}\mathbf{x}^T \mathbf{C} \mathbf{x}\right) d\mathbf{x} = c \int_{\mathcal{X}} \exp\left(-\frac{1}{2}(\mathbf{x}^T \mathbf{C} \mathbf{x})\right) d\mathbf{x}. \quad (72)$$

Tedy

$$c = \frac{e^{-\frac{r}{2}}}{|\mathcal{J}|}, \quad (73)$$

$$\mathcal{X} = \mathbf{B}^T(\mathcal{J} - \hat{\boldsymbol{\theta}}). \quad (74)$$

Uplatněním těchto výsledků spolu s již určenou diagonální maticí \mathbf{C} (70) a definicí rovnoměrného apriorního rozdělení (60) v rovnici (57) pak již můžeme určit jmenovatel Bayesova vzorce ve tvaru (50). Abychom přitom mohli využít rozkladu integrálu pomocí chybové funkce jak bylo předesláno v (53), je třeba poněkud vysvětlit rovnost (74). Integrační oblast \mathcal{J} nebyla dosud pevně stanovena. Musí pouze splňovat následující dvě podmínky vycházející z požadavků kladených na apriorní rozdělení pravděpodobnosti a ze vztahů (52) a (74).

- Množina \mathcal{J} musí pokrývat všechny přípustné hodnoty určovaného vektoru $\boldsymbol{\theta}$.
- Množina \mathcal{J} musí být m -rozměrný kvádr v \mathbb{R}^m , který je vůči souřadnicovým osám pootočen transformací \mathcal{B} .

S pomocí (67) a (60) lze též vhodně vyjádřit i čítec (50) a tím celý vztah pro aposteriorní hustotu pravděpodobnosti.

$$f_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \tilde{\mathbf{I}}) = \frac{\frac{e^{-\frac{r}{2}}}{|\mathcal{J}|} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right)}{\frac{e^{-\frac{r}{2}}}{|\mathcal{J}|} \prod_{i=1}^m \sqrt{\frac{2\pi}{c_i}} \frac{1}{2} \left(\operatorname{erf}\left(x_{i,\max} \sqrt{\frac{c_i}{2}}\right) - \operatorname{erf}\left(x_{i,\min} \sqrt{\frac{c_i}{2}}\right) \right)}. \quad (75)$$

Tento vztah však vzhledem k (60) platí pouze pro $\boldsymbol{\theta} \in \mathcal{J}$. Mimo oblast \mathcal{J} je aposteriorní hustota pravděpodobnosti nulová, tj.

$$f_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \tilde{\mathbf{I}}) = 0 \Leftrightarrow \boldsymbol{\theta} \notin \mathcal{J}. \quad (76)$$

Pro omezený objem $|\mathcal{J}| < \infty$ lze výrazem $\frac{e^{-\frac{\tau}{2}}}{|\mathcal{J}|}$ v (75) krátit a vznikne tak výsledný vzorec pro výpočet aposteriorní hustoty pravděpodobnosti.

$$f_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \tilde{\mathbf{I}}) = \frac{\sqrt{\prod_{i=1}^m c_i} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right)}{(2\pi)^m \frac{1}{2} \prod_{i=1}^m \left(\operatorname{erf}\left(x_{i,\max} \sqrt{\frac{c_i}{2}}\right) - \operatorname{erf}\left(x_{i,\min} \sqrt{\frac{c_i}{2}}\right)\right)}. \quad (77)$$

4.2.2 Neinformativní apriorní rozdělení

Pokud nemáme k dispozici žádné apriorní informace o určovaných parametrech $\boldsymbol{\theta}$, musíme připustit, že každý parametr θ_i , $i \in \{1, 2, \dots, m\}$ může nabývat libovolné reálné hodnoty, tj. $\forall i \in \{1, 2, \dots, m\} : \theta_i \in \mathbb{R}$. Apriorní rozdělení pravděpodobnosti odpovídající takovému apriorním informacím bývá nazýváno *neinformativní*. Podrobněji je o neinformativním rozdělení a jeho hustotě pojednáno v oddíle 4.3.

Pro neinformativní rozdělení tedy platí

$$\mathcal{J} = \mathbb{R}^m. \quad (78)$$

což znamená

$$|\mathcal{J}| = \infty. \quad (79)$$

V takovém případě vyplní i kvádr \mathcal{X} celý prostor \mathbb{R}^m (podle (74)) a tedy

$$\forall i \in \{1, 2, \dots, m\} : x_{i,\min} = -\infty, \quad x_{i,\max} = +\infty. \quad (80)$$

Potom je nutné v (77) přejít k limitě

$$\begin{aligned} f_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \tilde{\mathbf{I}}) &= \lim_{|\mathcal{J}| \rightarrow \infty} \frac{\frac{e^{-\frac{\tau}{2}}}{|\mathcal{J}|} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right)}{\frac{e^{-\frac{\tau}{2}}}{|\mathcal{J}|} \prod_{i=1}^m \sqrt{\frac{2\pi}{c_i}} \frac{1}{2} \left(\operatorname{erf}\left(x_{i,\max} \sqrt{\frac{c_i}{2}}\right) - \operatorname{erf}\left(x_{i,\min} \sqrt{\frac{c_i}{2}}\right)\right)} = \\ &= \frac{\sqrt{\prod_{i=1}^m c_i} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \mathbf{A} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right)}{(2\pi)^m \frac{1}{2} \prod_{i=1}^m \left(\lim_{x \rightarrow +\infty} \operatorname{erf}\left(x \sqrt{\frac{c_i}{2}}\right) - \lim_{x \rightarrow -\infty} \operatorname{erf}\left(x \sqrt{\frac{c_i}{2}}\right)\right)}. \quad (81) \end{aligned}$$

4.2.3 Výsledné aposteriorní rozdělení

Uplatněním (54) a (55) ve jmenovateli posledního výrazu a s ohledem na

$$\prod_{i=1}^m c_i = \det(\mathbf{C}) = \det(\mathbf{B}^T \mathbf{A}^T \mathbf{P} \mathbf{A} \mathbf{B}) = \det(\mathbf{A}^T \mathbf{P} \mathbf{A})$$

se vztah pro aposteriorní hustotu určených parametrů $\boldsymbol{\theta}$ zjednoduší na tvar:

$$f_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \tilde{\mathbf{I}}) = \frac{\sqrt{\det(\mathbf{A}^T \mathbf{P} \mathbf{A})}}{\sqrt{(2\pi)^m}} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{A}^T \mathbf{P} \mathbf{A}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right). \quad (82)$$

Porovnejme nyní pravou stranu (82) s výrazem pro hustotou f m -rozměrného normálního rozdělení $\mathcal{N}_m(\boldsymbol{\mu}, \mathbf{Q})$. Z (39) je patrné, že aposteriorní hustota $f_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \tilde{\mathbf{I}})$ je normální s vektorem středních hodnot $\boldsymbol{\mu} = \hat{\boldsymbol{\theta}}$ a kovarianční maticí

$$\mathbf{Q} = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1}. \quad (83)$$

Připomeňme jen, že vektor $\hat{\boldsymbol{\theta}}$ byl již stanoven vztahem (64).

$$\hat{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}} - (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \boldsymbol{\lambda}.$$

Tento vzorec ukazuje, že pro normální rozdělení chyb měření jsou v lineárním případě střední hodnoty určených veličin totožné s vyrovnanými hodnotami metodou nejmenších čtverců (viz např. [BRH90], str. 157, vztah (6.7)). Uvážíme-li ještě výslednou kovarianční matici \mathbf{Q} ve tvaru (83), je vidět, že Bayesovský přístup k vyrovnání zprostředkujících měření dává shodné výsledky jako vyrovnání metodou nejmenších čtverců (viz např. [BRH90], str. 170, vztah (6.29)). Tato pozoruhodná vlastnost normálního rozdělení není výsledkem nějaké šťastné shody okolností, nýbrž vyplývá přímo z definice normálního rozdělení, jak ji podal již roku 1809 geniální německý matematik, astronom, geodet a fyzik K. F. Gauss ve svém slavném díle [Gau09].

4.3 Poznámka o nevlastních hustotách a teorii distribucí

Nyní ještě obraťme pozornost k předpokladu absence jakýchkoliv apriorních informací o určených parametrech. Tento předpoklad (viz (78)) znamená, že množina \mathcal{J} je nekonečná, což zásadním způsobem ovlivňuje definici apriorního rozdělení pravděpodobnosti p danou vztahem (60). Pro $|\mathcal{J}| \rightarrow \infty$ totiž hustota p přestává být funkcí a stává se zobecněnou funkcí neboli tzv. distribucí. Takovými objekty se zabývá teorie distribucí, jejíž podstata je vystižena v [Mar76] na str. 26–34. Podrobnější a důkladnější pojednání o teorii distribucí je dostupné v českém jazyce též v kap. II (str. 71–110) knihy [Sch72].

Ve statistické terminologii se hustoty pravděpodobnosti, které nejsou funkcemi ale distribucemi, nazývají nevlastní hustoty. Příkladem další nevlastní hustoty je tzv. Diracova distribuce, někdy též nesprávně nazývaná Diracova funkce, příp. Diracova δ -funkce. Lze si ji představit jako limitní případ funkce f v (59) pro $a \rightarrow b$. Náhodná veličina s takovouto nevlastní hustotou pak již přestává být náhodnou, neboť může nabývat jen jediné hodnoty $a = b$. Přesná definice Diracovy distribuce je uvedena v [Sch72] na str. 77, příp. v [Mar76] na str. 27.

S distribucemi lze provádět obvyklé operace, které se provádějí s obyčejnými spojitými funkcemi, včetně derivování, integrování, Fourierovy transformace apod. Teorie distribucí přináší formální zjednodušení v mnoha oblastech aplikované matematiky, zvláště v matematické fyzice. Z hlediska těchto zjednodušení mají distribuce v teorii pravděpodobnosti a matematické statistice podobný význam jako komplexní čísla algebry a v geometrii. Vzájemný vztah distribucí a spojitých funkcí lze též názorně přirovnat ke vztahu reálných a racionálních čísel.

5 Normální rozdělení pravděpodobnosti

Normální rozdělení pravděpodobnosti je považováno za základní rozdělení pravděpodobnosti, které mají chyby měření fyzikálních veličin. Tento předpoklad je podepřen teoretickými i empirickými důvody. Jedno z prvních teoretických zdůvodnění normálního rozdělení podal r. 1809 Karl Friedrich Gauss ve svém astronomickém díle [Gau09]. V tradiční geodetické literatuře (viz [Kla43]) je nazýváno “první (Gaussovo) zdůvodnění metody nejmenších čtverců”, neboť podstata metody nejmenších čtverců je přímým důsledkem předpokladu o normalitě měřických chyb. Gaussovo odvození normálního rozdělení je založeno na přirozených vlastnostech měřických chyb a opírá se o aritmetický průměr jakožto nejvěrohodnější hodnotu reprezentující soubor měření.

Je pozoruhodné, že způsob Gaussova usuzování je v podstatě bayesovský, i když Bayesovo jméno není vůbec zmíněno. Bayesův vzorec je v [Gau09] formulován v poněkud neobvyklé, zjednodušené formě a je dokázán pouze kombinatoricky, tj. pro racionální hodnoty pravděpodobností. Thomas Bayes a po něm též nezávisle Pierre-Simon Laplace ve svých pracích dokázali Bayesovu větu pro spojité náhodné veličiny (viz [Bay63], [Rob94], str. 9).

5.1 Způsoby odvození normálního rozdělení

Zmíněné Gaussovo odvození normálního rozdělení není jediné ani nejstarší. K normálnímu rozdělení lze dospět mnoha způsoby. Některé z nich si nyní stručně připomeneme. Podrobnější přehled jednotlivých způsobů je uveden v [Rao78] na str. 190–196.

Zřejmě nejstarší způsob odvození normálního rozdělení spočívá ve zobecnění binomického rozdělení pro nekonečný počet pokusů. Poprvé byl uveřejněn Abrahamem de Moivre v roce 1733, ale v té době se mu nedostalo ohlasu a upadl v zapomenutí. V roce 1812 dospěl k normálnímu rozdělení Laplace, který při něm využil Stirlingova vzorce pro výpočet faktoriálu. Jeho postup bývá uveden ve starších učebnicích teorie pravděpodobnosti nebo vyrovnávacího počtu, např. v [Böh58], str. 42–45.

Nejznámější a teoreticky nejvýznamnější odvození normálního rozdělení je obsaženo v centrální limitní větě. Všechny její verze od původní Laplaceovy až po Ljapunovovu formulaci ukazují, že rozdělení pravděpodobnosti součtu libovolně rozdělených náhodných veličin se za určitých předpokladů při zvyšování počtu sčítanců blíží k normálnímu rozdělení. Velmi přístupnou formou je tento aspekt normálního rozdělení vysvětlen v [Tut78] na str. 82–85.

Další, méně známé odvození normálního rozdělení vychází z tzv. Hagenovy hypotézy. Podle ní je chyba měření způsobena součtem velkého počtu stejně velkých elementárních chyb, jež jsou vzájemně nezávislé a symetricky rozdělené. Limitním přechodem, při němž se zmenšuje velikost elementárních chyb a zvyšuje jejich počet se dospěje k normálnímu rozdělení. Tento postup je stručně vysvětlen

v [Rao78] a jeho původní verze je podána v [Böh64] na str. 246–247. Na Hagenově hypotéze jsou založeny některé způsoby modifikace normálního rozdělení — viz [Rom64], [RG65], [PP95]

Velmi důležitá je metoda maximální entropie, která vychází z požadavku stanovit takové rozdělení pravděpodobnosti, které má ze všech rozdělení s danou střední hodnotou a variancí největší entropii. Tato metoda odvození normálního rozdělení využívá poznatků variačního počtu a je stručně naznačena např. v [Koc90], str. 17–20. Formální odvození bez použití variačního počtu je uvedeno také v [Rao78] na str. 195.

Z hlediska bayesovského přístupu má však zásadní význam Gaussův postup a proto mu bude věnována tato kapitola. Nejdříve bude na základě [Gau09] Gaussův postup přehledně vysvětlen v současné matematicko-statistické terminologii a symbolice a pak budou navrženy různé možnosti jeho modifikace. Výklad Gaussova postupu vychází z ruského překladu původního Gaussova latinského textu [Gau09] (Liber II, Sectio III, str. 208–212). Podstata Gaussova postupu je též stručně nastíněna v [Kla43] na str. 35, 36.

5.2 Gaussovo odvození normálního rozdělení

Postup Gaussova odvození lze rozložit do dvou kroků. V prvním kroku je řešena úloha odhadu parametru polohy náhodné veličiny dané souborem měřených hodnot a rozdělením pravděpodobnosti jejich chyb vzhledem k hledanému parametru polohy. Pomocí metody maximální věrohodnosti je sestavena rovnice, jejímž řešením lze určit odhad parametru polohy libovolného jednovrcholového rozdělení pravděpodobnosti.

Ve druhém kroku Gauss úlohu formuluje obráceně: k danému nejvěrohodnějšímu odhadu parametru polohy hledá odpovídající rozdělení pravděpodobnosti. Za nejvěrohodnější odhad parametru polohy přitom považuje aritmetický průměr měřených hodnot příslušné veličiny. Z rovnice sestavené v prvním kroku pak odvodí diferenciální rovnici pro hustotu hledaného rozdělení pravděpodobnosti. Jejím řešením dospěje k rozdělení pravděpodobnosti, dnes nazývanému normální.

5.2.1 První krok Gaussova postupu

První krok Gaussova postupu spočívá v řešení přeурčené soustavy rovnic (22) pomocí metody maximální věrohodnosti. Odhadovanými parametry jsou zde složky vektoru $\boldsymbol{\theta}$ a věrohodnostní funkcí je funkce

$$\Omega : \mathbb{R} \rightarrow \mathbb{R} : \boldsymbol{\theta} \mapsto \Omega(\boldsymbol{\theta}) \triangleq \prod_{j=1}^n f_{\varepsilon_j}(a_j(\boldsymbol{\theta}) - \tilde{l}_j) . \quad (84)$$

Maximálně věrohodný odhad $\hat{\boldsymbol{\theta}}$ je pak nalezen řešením extrémální úlohy

$$\Omega(\hat{\boldsymbol{\theta}}) = \max_{\boldsymbol{\theta} \in \mathcal{T}} \Omega(\boldsymbol{\theta}) . \quad (85)$$

Řešení této úlohy pomocí věty o lokálních extrémech funkce několika proměnných (viz např. [Jar76], věta 216 na str. 507) vede k soustavě rovnic:

$$\forall i \in \{1, 2, \dots, m\} : \frac{\partial \Omega}{\partial \theta_i}(\hat{\theta}) = 0 \quad (86)$$

a podmínce negativní definitnosti matice:

$$\left[\frac{\partial^2 \Omega}{\partial \theta_i \partial \theta_k}(\hat{\theta}) \right]_{i,k \in \{1, 2, \dots, m\}} . \quad (87)$$

Z hlediska dalšího postupu odvození ve druhém kroku není nutné řešit soustavu (86) v plné obecnosti. Proto byly zavedeny zjednodušující předpoklady:

$$m = 1 \quad (88)$$

$$\theta \triangleq \boldsymbol{\theta}$$

$$\hat{\theta} \triangleq \hat{\boldsymbol{\theta}}$$

$$\forall j \in \{1, 2, \dots, n\} : a_j(\theta) = \theta \quad (89)$$

$$\forall j \in \{1, 2, \dots, n\} : f_{\varepsilon_j} = \varphi . \quad (90)$$

Věrohodnostní funkce Ω pak dostane tvar:

$$\Omega(\theta) = \prod_{j=1}^n \varphi(\theta - \tilde{l}_j) . \quad (91)$$

Po dosazení (91) do (86) a následných zjednodušujících úpravách pak vznikne jediná rovnice:

$$\sum_{j=1}^n \frac{\varphi'(\hat{\theta} - \tilde{l}_j)}{\varphi(\hat{\theta} - \tilde{l}_j)} = 0 . \quad (92)$$

Podmínka negativní definitnosti matice (87) přejde v nerovnost

$$\frac{d^2 \Omega}{d\theta^2}(\hat{\theta}) < 0 . \quad (93)$$

Řešením rovnice (92) za podmínky (93) lze pro libovolné rozdělení pravděpodobnosti chyb měřené veličiny odvodit vzorec pro nejvěrohodnější odhad z n měření.

5.2.2 Druhý krok Gaussova postupu

Rovnici (92) lze interpretovat i jako funkcionální rovnici a pro daný odhad $\hat{\theta}$ hledat funkci φ . Tato funkce by tak měla být hustotou nejvěrohodnějšího rozdělení pravděpodobnosti chyb měřené veličiny. Za odhad, vzhledem k němuž má

být hledané rozdělení pravděpodobnosti nejvěrohodnější, zvolil Gauss aritmetický průměr. Tedy

$$\hat{\theta} \triangleq \frac{1}{n} \sum_{j=1}^n \tilde{l}_j . \quad (94)$$

Zavedením nových veličin — oprav k aritmetickému průměru

$$v_j \triangleq \hat{\theta} - \tilde{l}_j \text{ pro } \forall j \in \{1, 2, \dots, n\} \quad (95)$$

dostanou podmínky (92), (94) přehlednější tvar :

$$\sum_{j=1}^n \frac{\varphi'(v_j)}{\varphi(v_j)} = 0 \quad (96)$$

$$\sum_{j=1}^n v_j = 0 . \quad (97)$$

Poslední rovnost je zřejmě pro každou reálnou nenulovou konstantu K ekvivalentní s rovností:

$$\sum_{j=1}^n K v_j = 0 . \quad (98)$$

Díky tomu platí:

$$\sum_{j=1}^n \frac{\varphi'(v_j)}{\varphi(v_j)} = 0 \Leftrightarrow \forall K \in \mathbb{R}, K \neq 0, \forall j \in \{1, 2, \dots, n\} : \frac{\varphi'(v_j)}{\varphi(v_j)} = K v_j . \quad (99)$$

Předpoklad na pravé straně této implikace je splněn vždy, když platí:

$$\forall K \in \mathbb{R}, K \neq 0, \forall v \in \mathbb{R} : \frac{\varphi'(v)}{\varphi(v)} = K v . \quad (100)$$

Tím je zakončen řetězec obrácených implikací

$$(85) \Leftrightarrow (86) \Leftrightarrow (96) \Leftrightarrow (100) , \quad (101)$$

který vede k obyčejné diferenciální rovnici:

$$\varphi'(v) = K v \varphi(v) \quad (102)$$

s reálnou nenulovou konstantou K . Obecným řešením této diferenciální rovnice je funkce

$$\varphi(v) = C \exp\left(\frac{1}{2} K v^2\right) , \quad (103)$$

kde C je reálná konstanta (viz např. [Kam65], str. 37). Konstanta C musí splňovat normalizační podmínku

$$\int_{\mathbb{R}} C \exp\left(\frac{1}{2}Kv^2\right) dv = 1 . \quad (104)$$

Průběh funkce φ výrazně závisí na znaménku konstanty K . To je dáno podmínkou (93), která po snadné úpravě přejde na tvar:

$$n K \Omega(\theta) < 0 .$$

Výraz $\Omega(\theta)$ představuje součin hustot (viz (91)). Protože hustota φ je všude kladná, je kladný i součin $\Omega(\theta)$. Číslo n je ovšem také kladné, takže aby výraz $n K \Omega(\theta)$ mohl být záporný, musí být záporná konstanta K . Lze ji tudíž vyjádřit ve tvaru:

$$K = -2h^2 , \quad (105)$$

kde h je libovolná reálná nenulová konstanta.

Po výpočtu integrálu v (104) (Laplaceův integrál — viz např. [Jar84] str. 142 – 143, [Bar87] str. 613, vzorce 32) lze snadno odvodit výraz pro konstantu C , po jehož dosazení do (103) vznikne výsledný vzorec pro hledanou hustotu φ :

$$\varphi(v) = \frac{h}{\sqrt{\pi}} \exp(-h^2v^2) . \quad (106)$$

Konstantu h můžeme ještě vyjádřit pomocí variance σ^2 rozdělení φ

$$h^2 = \frac{1}{2\sigma^2} \quad (107)$$

a po jejím dosazení do (106) obdržíme známý vzorec pro hustotu normálního rozdělení:

$$\varphi(v) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{v}{\sigma}\right)^2\right) \quad (108)$$

(viz např. [And78], str. 74).

Vzorcem (106) končí Gaussův postup odvození rozdělení pravděpodobnosti chyb měření, které by mělo být tak přirozené a normální, jako je přirozené pochopitelný a normální aritmetický průměr. Výsledky měření nejrůznějších veličin, a nejen fyzikálních, prováděných během následujících téměř dvou století toto Gaussovo přání bohatě naplnily.

5.3 Rozbor Gaussova odvození

Právě popsany Gaussův postup odvození dovoluje nahlédnout do teoretického zázemí normálního rozdělení z poněkud jiného úhlu pohledu než poskytují výše uvedené tradiční způsoby odvození (např. centrální limitní věty, limitní přechod od binomického rozdělení, metoda maximální entropie aj.). Pozoruhodným rysem tohoto postupu je bayesovský způsob uvažování v prvním kroku.

5.3.1 První krok Gaussova postupu

První krok Gaussova postupu představuje vlastně aplikaci bayesovského přístupu při řešení přeúčené soustavy rovnic, neboť metoda maximální věrohodnosti, kterou Gauss použil, je zvláštním případem aplikace Bayesovy věty. Metodu maximální věrohodnosti Gauss odvodil na základě vzorce, který formuloval jako větu i s důkazem (viz [Gau09], odst. 176). Tento vzorec je v podstatě Bayesův vzorec se stejnými apriorními pravděpodobnostmi. Rozšířením platnosti tohoto vzorce na spojitě hustoty pravděpodobnosti pak Gauss získal Bayesovu větu ve tvaru:

$$f_{\Theta}(\boldsymbol{\theta} | \tilde{\mathbf{I}}) = \frac{\prod_{j=1}^n f_{\mathcal{E}_j}(a_j(\boldsymbol{\theta}) - \tilde{l}_j)}{\int_{\mathcal{T}} \prod_{j=1}^n f_{\mathcal{E}_j}(a_j(\mathbf{t}) - \tilde{l}_j) d\mathbf{t}}, \quad (109)$$

kde

$f_{\Theta}(\cdot | \hat{\mathbf{I}})$... aposteriorní hustota pravděpodobnosti náhodného vektoru $\boldsymbol{\Theta}$,
 $f_{\mathcal{E}_j}$... hustota pravděpodobnosti j -té chyby měření \mathcal{E}_j .

Po zavedení věrohodnostní funkce podle (84) je okamžitě vidět, že úloha (85) je ekvivalentní s úlohou:

$$f_{\Theta}(\hat{\boldsymbol{\theta}} | \tilde{\mathbf{I}}) = \max_{\boldsymbol{\theta} \in \mathcal{T}} f_{\Theta}(\boldsymbol{\theta} | \tilde{\mathbf{I}}). \quad (110)$$

Uvažme nyní obecnou formulaci Bayesovy věty ve tvaru (34).

$$f_{\Theta}(\boldsymbol{\theta} | \tilde{\mathbf{I}}) = \frac{\prod_{j=1}^n f_{\mathcal{E}_j}(a_j(\boldsymbol{\theta}) - \tilde{l}_j) p(\boldsymbol{\theta})}{\int_{\mathcal{T}} \prod_{j=1}^n f_{\mathcal{E}_j}(a_j(\mathbf{t}) - \tilde{l}_j) p(\mathbf{t}) d\mathbf{t}}, \quad (111)$$

kde

p ... apriorní hustota pravděpodobnosti náhodného vektoru $\boldsymbol{\Theta}$.

Gaussův postup lze jednoduše zobecnit dosazením do (110) za aposteriorní hustotu $f_{\Theta}(\cdot | \tilde{\mathbf{I}})$ podle (111) místo podle (109).

Bayesův vzorec ve tvaru (109) má v případě normality měřených veličin velmi výhodné vlastnosti. Za předpokladu linearit zobrazení \mathbf{a} a při volbě

$$\mathcal{T} = \mathbb{R}^m \quad (112)$$

z něj např. přímo vyplývá, že střední hodnota aposteriorního rozdělení $f_{\Theta}(\cdot | \tilde{\mathbf{I}})$ je vždy shodná s odhadem vektoru $\boldsymbol{\theta}$ metodou nejmenších čtverců, jak bylo ukázáno

v kap. 3. Vzhledem k úplné Bayesově formuli (111) však volba (112) znamená, že apriorní hustota p by měla být konstantní na neomezené množině \mathcal{J} . I když ponecháme stranou otázku existence takové hustoty, dostáváme se k dalšímu nereálnému závěru, že o hodnotách určovaných veličin nejsou před provedením měření k dispozici žádné informace. Tento předpoklad je sice velmi obecný, avšak většinou neodpovídá skutečnosti. Před započítáním vlastního měření jsou dostupné přinejmenším údaje o rozsahu měřicích přístrojů a navržené metodice měření, jež představují jistou formu apriorní informace. Aby byla respektována tato skutečnost, je třeba vyjádřit apriorní informaci o určovaných veličinách ve formě rozdělení pravděpodobnosti a dosadit jeho hustotu za p v (111). Metoda nejmenších čtverců tedy nevyužívá všech dostupných informací.

5.3.2 Druhý krok Gaussova postupu

Zatímco první krok Gaussova postupu představuje pouze aplikaci Bayesovy věty, což není z dnešního pohledu nijak nový přístup, druhý krok poskytuje mnohem bohatší možnosti zobecnění a modifikace. Nejnápadnější možnost modifikace spočívá v nahrazení aritmetického průměru jiným typem odhadu. V současné době existuje nepřehledné množství nejrůznějších metod odhadu, z nichž zvláště významnou skupinu tvoří robustní odhady. Experimentování s různými odhady může přinést mnoho zajímavých rozdělení pravděpodobnosti s nečekaně výhodnými vlastnostmi. Ještě pestřejší škála rozdělení může vzniknout v kombinaci s volbou apriorního rozdělení pravděpodobnosti zmíněnou v předchozím odstavci.

5.4 Přípustnost neomezeně velkých chyb měření

Normální rozdělení bylo přijato za rozdělení pravděpodobnosti chyb měření ve většině oborů, kde se provádí fyzikální měření i v mnoha oborech mimofyzikálních. V geodézii byl předpoklad normality chyb měření také mnohokrát prakticky ověřen. Normální rozdělení má však jednu velmi nepraktickou vlastnost, která očividně neodpovídá realitě — připouští libovolně velké chyby. Tato vlastnost odporuje jedné ze tří zákonitostí náhodných chyb, které formuloval již Gauss (symetrie, závislost pravděpodobnosti výskytu chyby na její velikosti, omezenost velikosti chyby — viz [BRH90], str. 105). Tento rozpor lze odstranit v zásadě dvěma způsoby.

Při normálním rozdělení $\mathcal{N}(0, \sigma^2)$ je výskyt velkých chyb (větších než 3σ) natolik nepravděpodobný, že je možno tuto malou pravděpodobnost zanedbat. Kvůli tomuto zanedbání je nutno považovat normální rozdělení za aproximaci skutečného rozdělení chyb měření, které zůstává neznámé.

Druhý způsob řešení problému libovolně velkých chyb spočívá ve stanovení mezí pro velikost chyby měření. Rozdělení pravděpodobnosti, které by respektovalo tyto meze a přitom by vyhovovalo ostatním předpokladům Gaussova postupu, by pak bylo možno považovat za skutečné rozdělení pravděpodobnosti

chyb měření a nikoli pouze jeho aproximaci, jak tomu bylo v předchozím případě. Tento druhý, teoreticky oprávněnější způsob bude v následující kapitole ilustrován několika příklady. Mírně upraveným Gaussovým postupem bude odvozeno useknuté normální rozdělení, von Misesovo rozdělení a Pearsonovo rozdělení.

6 Modifikace Gaussova odvození normálního rozdělení

6.1 Useknuté normální rozdělení

6.1.1 První krok Gaussova postupu

Vzhledem k důvodům zmíněným v závěru předcházející kapitoly je třeba zvolit apriorní rozdělení pravděpodobnosti definované na omezené množině $\mathcal{T} \subset \mathbb{R}^m$. Vzhledem k zjednodušujícím předpokladům (88) – (90) je touto množinou jedno-rozměrný interval $\mathcal{T} \triangleq (\theta_{\min}, \theta_{\max})$.

Jak již bylo zmíněno v předcházející kapitole (před vztahem (59)), rozdělením pravděpodobnosti s největší entropií na omezeném intervalu je rovnoměrné rozdělení. Zvolíme tedy

$$p(\theta) \triangleq \begin{cases} \frac{1}{\theta_{\max} - \theta_{\min}} & \Leftarrow \theta \in \mathcal{T} \\ 0 & \Leftarrow \theta \notin \mathcal{T} \end{cases} . \quad (113)$$

První krok Gaussova postupu pak zůstane stejný jako při odvozování normálního rozdělení, neboť po krácení konstantní hustotou (113) v (111) vznikne opět zjednodušený tvar Bayesovy věty (109). Výsledkem prvního kroku je tedy opět rovnice (92) a nerovnost (93), ke kterým je ovšem třeba připojit podmínku $\hat{\theta} \in (\theta_{\min}, \theta_{\max})$.

6.1.2 Druhý krok Gaussova postupu

Ve 2. kroku je třeba pro daný odhad $\hat{\theta}$ řešit rovnici (92) vzhledem k φ . Za největšohodnější odhad $\hat{\theta}$ zvolíme opět aritmetický průměr, takže Gaussův postup opět vede k diferenciální rovnici (102). Její řešení (103) obsahuje konstantu C , která v tomto případě musí splňovat podmínku

$$\int_{\mathcal{T} - \hat{\theta}} C \exp\left(\frac{1}{2}Kv^2\right) dv = 1 , \quad (114)$$

která se liší od podmínky (104) oborem integrace, neboť opravy v nemohou být libovolně velké. Musí platit

$$v \in \mathcal{T} - \hat{\theta} = (\theta_{\min} - \hat{\theta}, \theta_{\max} - \hat{\theta}) .$$

Podmínka lokálního maxima zůstane nezměněna, takže konstantu K lze vyjádřit ve tvaru $K = -\frac{1}{\sigma^2}$. Pomocí této substituce v (114) již můžeme určit normalizační konstantu C .

$$\begin{aligned}
C &= \frac{1}{\int_{\theta_{\min}-\hat{\theta}}^{\theta_{\max}-\hat{\theta}} \exp\left(-\frac{1}{2}\left(\frac{v}{\sigma}\right)^2\right)} = \\
&= \frac{1}{\frac{\sqrt{2\pi}\sigma}{2} \left(\operatorname{erf}\left(\frac{\theta_{\max}-\hat{\theta}}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{\theta_{\min}-\hat{\theta}}{\sqrt{2}\sigma}\right) \right)}. \tag{115}
\end{aligned}$$

Dosazením konstant C , K do (103) dostáváme konečně hledanou hustotu φ .

$$\varphi(v) = \frac{\exp\left(-\frac{1}{2}\left(\frac{v}{\sigma}\right)^2\right)}{\frac{\sqrt{2\pi}\sigma}{2} \left(\operatorname{erf}\left(\frac{\theta_{\max}-\hat{\theta}}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{\theta_{\min}-\hat{\theta}}{\sqrt{2}\sigma}\right) \right)} \tag{116}$$

pro $v \in (\theta_{\min}-\hat{\theta}, \theta_{\max}-\hat{\theta})$.

To je hustota useknutého normálního rozdělení s parametry σ , $\theta_{\min}-\hat{\theta}$, $\theta_{\max}-\hat{\theta}$. Rozdělení pravděpodobnosti této třídy nejsou obecně symetrická. Symetrie zřejmě nastává pro

$$\theta_{\min}-\hat{\theta} = -(\theta_{\max}-\hat{\theta}). \tag{117}$$

Označme

$$b \triangleq \theta_{\max}-\hat{\theta}. \tag{118}$$

Potom se normalizační konstanta C zjednoduší na tvar

$$C = \frac{1}{\sqrt{2\pi}\sigma \operatorname{erf}\left(\frac{b}{\sqrt{2}\sigma}\right)}.$$

Symetricky useknuté normální rozdělení je tedy dáno vzorcem

$$\varphi(v) = \frac{1}{\sqrt{2\pi}\sigma \operatorname{erf}\left(\frac{b}{\sqrt{2}\sigma}\right)} \exp\left(-\frac{1}{2}\left(\frac{v}{\sigma}\right)^2\right), \tag{119}$$

který platí pro $v \in (-b, b)$. Mimo interval $(-b, b)$ je hustota φ nulová.

6.2 Von Misesovo rozdělení

Useknuté normální rozdělení odvozené v předcházejícím oddíle řeší problém neomezeně velkých chyb měření. V geodézii jej lze použít pro rozdělení pravděpodobnosti chyb měření délek. Používá se též pro chyby měření úhlů a směrů, i když z teoretického hlediska je vhodnější jiný typ rozdělení. Cyklický charakter úhlů a směrů si totiž vynucuje poněkud odlišný přístup k odvození jejich rozdělení pravděpodobnosti. Především není možné ani hypoteticky připustit libovolně

velké chyby měření. Definiční obor rozdělení pravděpodobnosti chyby úhlu nesmí být větší než 2π . Tento předpoklad je v geodézii samozřejmě vždy splněn, neboť i při běžně dosahované přesnosti měření směrů bývá rozpětí přípustných chyb stotisíckrát menší. Pro takovou přesnost je tvar von Misesova rozdělení prakticky totožný s tvarem normálního rozdělení jak dokládá obrázek 4. Analytický tvar vzorce pro hustotu von Misesova rozdělení je však výhodný pro výpočet aposteriorního rozdělení pravděpodobnosti souřadnic při vyrovnání rovinných geodetických sítí, jak ukážeme v následující kapitole. Proto bude prospěšné uvést vzorec pro hustotu von Misesova rozdělení a ukázat jeho odvození Gaussovým postupem (metodou maximální věrohodnosti). Stručně je tento postup naznačen v [Mar72] na str. 64, 65.

Zmíněná kniha [Mar72] je celá věnována statistickému zpracování úhlových měření. Je v ní navrženo několik typů rozdělení pravděpodobnosti úhlů a na jejich základě je vybudována statistická teorie těchto veličin. O rozdělení pravděpodobnosti úhlů je také stručně pojednáno v [Rao78] na str. 208 – 211.

6.2.1 První krok Gaussova postupu

Podobně jako u normálního rozdělení nepředpokládáme žádné apriorní informace o určované veličině Θ (tzv. princip absolutní ignorance). Její rozdělení pravděpodobnosti tedy bude rovnoměrné. Protože veličina Θ představuje úhel, nebude už její hustota nevládní, nýbrž

$$p(\theta) = \frac{1}{2\pi} \quad (120)$$

pro $\forall \theta \in \mathbb{R}$.

Výsledkem prvního kroku je tedy opět rovnice (92) a nerovnost (93), ke kterým tentokrát již není třeba přidávat žádnou další podmínku, neboť apriorní rozdělení (120) platí pro všechna reálná θ .

6.2.2 Druhý krok Gaussova postupu

Druhý krok Gaussova postupu se od předchozích případů liší volbou nejvěrohodnějšího odhadu parametru polohy hledaného rozdělení. U rozdělení pravděpodobnosti úhlů totiž nemůže být tímto odhadem parametru polohy aritmetický průměr, neboť ten již obecně nemá aditivní vlastnost střední hodnoty

$$\forall c \in \mathbb{R} : E(X + c) = E(X) + c, \quad (121)$$

kde X je nějaká náhodná veličina.

Vlastnost (121) však zachovává odhad

$$\hat{\theta} = \arctan \frac{\sum_{j=1}^n \sin \tilde{l}_j}{\sum_{j=1}^n \cos \tilde{l}_j}. \quad (122)$$

Pro nějaké určité $r \in \mathbb{R}$ tedy platí

$$\begin{aligned} r \cos \hat{\theta} &= \sum_{j=1}^n \cos \tilde{l}_j, \\ r \sin \hat{\theta} &= \sum_{j=1}^n \sin \tilde{l}_j. \end{aligned} \quad (123)$$

Zavedeme-li opravy $v_j = \hat{\theta} - \tilde{l}_j$, $j \in \{1, 2, \dots, n\}$, můžeme dále psát

$$\begin{aligned} r \cos \hat{\theta} &= \sum_{j=1}^n \cos(\hat{\theta} + v_j) = \sum_{j=1}^n (\cos \hat{\theta} \cos v_j - \sin \hat{\theta} \sin v_j) = \\ &= \cos \hat{\theta} \sum_{j=1}^n \cos v_j - \sin \hat{\theta} \sum_{j=1}^n \sin v_j, \\ r \sin \hat{\theta} &= \sum_{j=1}^n \sin(\hat{\theta} + v_j) = \sum_{j=1}^n (\sin \hat{\theta} \cos v_j + \cos \hat{\theta} \sin v_j) = \\ &= \sin \hat{\theta} \sum_{j=1}^n \cos v_j + \cos \hat{\theta} \sum_{j=1}^n \sin v_j. \end{aligned} \quad (124)$$

Aby bylo možno vyjádřit odhad $\hat{\theta}$ pomocí jeho oprav v_j , $j \in \{1, 2, \dots, n\}$ podobně jako v případě aritmetického průměru (rovnice (97)), je třeba vyřešit soustavu rovnic (124) pro neznámé $\sum_{j=1}^n \cos v_j$, $\sum_{j=1}^n \sin v_j$. Tím dostaneme vztahy

$$\sum_{j=1}^n \cos v_j = r, \quad (125)$$

$$\sum_{j=1}^n \sin v_j = 0. \quad (126)$$

Vztah (126) je hledanou analogií rovnice (97) a vztah (125) nám poslouží při vyšetřování typu extrému hledané hustoty φ .

Další postup odvození je již stejný jako v oddíle 5.2.2 s tím jediným rozdílem, že rovnici (97) v soustavě rovnic (96), (97) nahradí podmínka (126). S uvážením této záměny dospějeme k diferenciální rovnici

$$\varphi'(v) = K \sin v \varphi(v) \quad (127)$$

s reálnou nenulovou konstantou K . Obecným řešením této diferenciální rovnice je funkce

$$\varphi(v) = C e^{-K \cos v}. \quad (128)$$

Normalizační konstantu C určíme až po stanovení znaménka konstanty K . To je závislé na druhu extrémů věrohodnostní funkce Ω (viz (91)). Aby funkce Ω nabývala maxima, musí být splněna podmínka (93). Tato podmínka vede k nerovnosti

$$\Omega(\hat{\theta}) K \sum_{j=1}^n \cos v_j < 0 ,$$

kteřá vzhledem k (125) a kladné hodnotě $\Omega(\hat{\theta})$ přejde v ekvivalentní nerovnost

$$Kr < 0 .$$

Nyní je třeba určit znaménko konstanty $r = \sum_{j=1}^n \cos v_j$. Z grafu funkce \cos nad intervalem $(-\pi, \pi)$ je patrné, že konstanta r bude kladná, pokud opravy v_j budou vesměs malé, tj. koncentrovány převážně v intervalu $(-\frac{\pi}{2}, \frac{\pi}{2})$. Pro velké opravy ($v_j \in (-\pi, \frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi)$) je konstanta r záporná. Odhad $\hat{\theta}$ tedy odpovídá maximu věrohodnostní funkce Ω při $r > 0$. Proto musí být konstanta K záporná.

Ve shodě s označením v [Mar72] zvolíme kladnou konstantu

$$\kappa \triangleq -K .$$

Rovnice (128) pak dostane tvar

$$\varphi(v) = C e^{\kappa \cos v} , \tag{129}$$

který již umožňuje určit normalizační konstantu C .

Integrál

$$\int_{-\pi}^{\pi} e^{\kappa \cos v} dv$$

lze řešit pomocí substituce $t = -\cos v$, při níž se využije symetrie grafu funkce \cos vzhledem k ose y kartézského souřadnicového systému Oxy . Potom dostáváme pro konstantu C rovnici

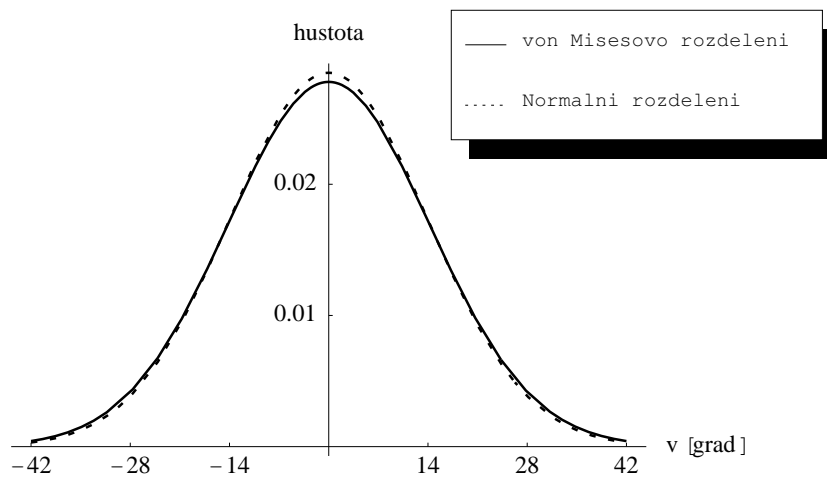
$$\frac{1}{C} = 2 \int_{-1}^1 \frac{e^{-\kappa t}}{\sqrt{1-t^2}} dt .$$

Integrál na pravé straně lze vyjádřit pomocí modifikované Besselovy funkce I_0 , někdy též nesprávně nazývané Besselova funkce ryze imaginárního argumentu. Definice funkce I_0 pomocí nekonečné řady i její integrální vyjádření (tzv. Poissonova integrální reprezentace) je uvedeno např. [Nk77], str. 157, 158. Konstanta má tedy tvar

$$C = \frac{1}{2\pi I_0(\kappa)} .$$

S její pomocí již můžeme psát výsledný vzorec pro hustotu von Misesova rozdělení.

$$\varphi(v) = \frac{e^{\kappa \cos v}}{2\pi I_0(\kappa)} . \tag{130}$$



Obrázek 4: Hustota pravděpodobnosti von Misesova rozdělení a normálního rozdělení při střední chybě $\sigma = 14^g$

Její graf ve srovnání s grafem normálního rozdělení ukazuje obrázek 4. K určení variance von Misesova rozdělení bylo použito vztahu

$$\sigma^2 = 1 - \left(\frac{I_1(\kappa)}{I_0(\kappa)} \right)^2 .$$

Lze jej nalézt v [Rao78] na str. 210. Jeho důkaz však zde nebudeme uvádět, neboť vyžaduje znalost pojmů, které v této práci nebudeme potřebovat (využívá se při něm charakteristické funkce von Misesova rozdělení a tzv. trigonometrických momentů — viz [Mar72], str. 62, 63 a 35, 36).

Těsná blízkost obou rozdělení i pro tak velkou střední chybu ($\sigma = 14^g$) potvrzuje oprávněnost používání normálního rozdělení pro chyby měření úhlů a směrů v geodézii, kde se střední chyba měření směrů běžně pohybuje kolem $\sigma = 0.0008^g$ a při zvlášť přesném měření úhlů lze dosáhnout i několikrát menší střední chyby. Pro tak vysokou přesnost již působí numerické potíže výpočet hodnot hustoty pravděpodobnosti von Misesova rozdělení podle vzorce (130), neboť konstanta κ je příliš velká ($\kappa \approx 10^{10}$). Je proto třeba při něm vyjádřit hodnotu Besselovy funkce $I_0(\kappa)$ pomocí mocninné řady pro $\frac{1}{\kappa}$. Tato řada je uvedena v [AS70] na str. 377 (vzorec 9.7.1) a též v [Mar72] na str. 63.

Z hlediska jeho grafického tvaru nemá von Misesovo rozdělení pro geodézii žádný praktický význam. Jeho význam spočívá, jak již bylo předesláno, ve výhodném analytickém tvaru jeho hustoty, který lze využít při souřadnicových výpočtech. O tom bude pojednávat následující kapitola.

Než přejdeme k dalšímu, nejzajímavějšímu příkladu použití Gaussova postupu odvození rozdělení pravděpodobnosti chyb měření, nemůžeme pominout jednu

významnou vlastnost von Misesova rozdělení.

Podobně jako normální rozdělení má ze všech rozdělení na přímce se stejnou variancí největší entropii, má i von Misesovo rozdělení největší entropii ze všech rozdělení na kružnici (opět při konstantní varianci). To lze dokázat analogicky jako u normálního rozdělení prostředky variačního počtu. Stručnější důkaz využívající jedné nerovnosti teorie informace je podán v [Mar72] na str. 65, 66.

6.3 Pearsonovo rozdělení

6.3.1 První krok Gaussova postupu

Podobně jako v případě useknutého normálního rozdělení zvolíme apriorní rozdělení ve tvaru (113). Také je možno konstantní hustotou (113) krátit v (111) a získat tak zjednodušený tvar Bayesovy věty (109). Výsledkem prvního kroku je tedy opět rovnice (92) a nerovnost (93) spolu s podmínkou $\hat{\theta} \in (\theta_{\min}, \theta_{\max})$.

6.3.2 Druhý krok Gaussova postupu

Při řešení rovnice nezadáme nyní odhad $\hat{\theta}$ explicitně jako v (94), nýbrž implicitně pomocí následujících podmínek.

Useknuté normální rozdělení sice respektuje všechny tři Gaussem stanovené zákonitosti náhodných chyb, přesto však jedna jeho vlastnost je nepřirozená — v koncových bodech intervalu $(\theta_{\min}, \theta_{\max})$ dochází ke skokovým změnám hustoty. Tento nedostatek odstraňuje následující podmínka:

$$\lim_{v \rightarrow \theta_{\max} - \hat{\theta}_-} \varphi(v) = \lim_{v \rightarrow \theta_{\min} - \hat{\theta}_+} \varphi(v) = 0 . \quad (131)$$

Druhou podmínkou je rovnice (96), která zůstává nezměněna. Rovnost (98) však již nemůže být splněna, neboť ve spojení s podmínkou (131) způsobuje neřešitelnost diferenciální rovnice (102). Proto je potřeba požadavek (98) poněkud oslabit. To lze zařídit různými způsoby. Nejjednodušším způsobem je zřejmě prosté nahrazení konstanty K funkcí

$$K : \mathcal{T} - \hat{\theta} \rightarrow \mathbb{R} : v \mapsto K(v) . \quad (132)$$

Podmínku (98) pak nahradí podmínka

$$\sum_{j=1}^n K(v_j) v_j = 0 , \text{ kde } v_j = y_j - \hat{\theta} . \quad (133)$$

Tím však odhad $\hat{\theta}$ přestává být aritmetickým průměrem, pokud funkce K není na intervalu $(-b, b)$ konstantní. Aby se tento nový, zatím blíže neurčený odhad $\hat{\theta}$ od aritmetického průměru příliš nelišil, je třeba zvolit takovou funkci K , aby rovnost (98) platila alespoň přibližně. Přitom je přirozené požadovat, aby podmínka (98)

platila přesněji pro malé chyby v než pro velké, tzn. aby funkce K byla přibližně konstantní především ve střední části intervalu $(-b, b)$.

Pokud tedy v úvahách odstavce 5.2.2 všude nahradíme konstantu K funkcí (132), dostaneme diferenciální rovnici (102) v poněkud pozměněném tvaru:

$$\varphi'(v) = K(v) v \varphi(v) . \quad (134)$$

Obecným řešením této diferenciální rovnice (viz např. [Kam65], str. 37) je funkce

$$\varphi(v) = C \exp\left(\int K(v) v dv\right) , \quad (135)$$

kde C je konstanta, jejíž hodnotu určíme normováním podobně jako v (114). Abychom se vyhnuli zbytečně komplikovaným výpočtům jako v (115), které bychom pak zjednodušovali zavedením požadavku symetrie (117), učiníme jej již nyní. Požadujeme tedy

$$\theta_{\min} - \hat{\theta} = -(\theta_{\max} - \hat{\theta}) .$$

a podobně jako v (118) označíme

$$b \triangleq \theta_{\max} - \hat{\theta} . \quad (136)$$

Požadavek (131) tedy lze vyjádřit ve tvaru:

$$\lim_{|v| \rightarrow b} \int K(v) v dv = -\infty . \quad (137)$$

Zřejmě existuje celá třída funkcí vyhovujících podmínce (137). Aby odhad $\hat{\theta}$ byl určen jednoznačně, je třeba ještě zvolit konkrétní funkci K . Na základě jednoduché analogie se speciální teorií relativity lze dospět k funkci

$$K(v) \triangleq \frac{\kappa}{1 - \left(\frac{v}{b}\right)^2} , \quad (138)$$

kteřá podmínku (137) splňuje, pokud konstanta κ je záporná. Lze se o tom snadno přesvědčit výpočtem integrálu v (137). Tato volba je výhodná z důvodů, které budou diskutovány v oddíle 6.3.3.

Pomocí právě definované funkce K můžeme nyní provést integraci naznačenou v (135). Pomocí rozkladu racionální funkce na parciální zlomky a po dalších jednoduchých úpravách získáme vztah

$$\int K(v) v dv = \kappa \int \frac{v}{1 - \left(\frac{v}{b}\right)^2} dv = -\kappa \frac{b^2}{2} \ln \left(1 - \left(\frac{v}{b}\right)^2\right) + c . \quad (139)$$

Integrační konstantu c není třeba dále uvažovat, neboť po dosazení (139) do (135) ovlivní pouze multiplikační konstantu C , která bude později určena normalizací (142).

Podobně jako ve druhém kroku Gaussova postupu (pododdíl 5.2.2, str. 40) dokážeme pomocí (93), že $\kappa < 0$. Analogicky k substitucím (105), (107) pak dostaneme vztah

$$\kappa = -\frac{1}{\sigma^2}.$$

Jeho dosazením do posledního výrazu v (139) tak dostáváme vztah

$$\int K(v) v dv = \ln \left(1 - \left(\frac{v}{b} \right)^2 \right)^{\frac{1}{2} \left(\frac{b}{\sigma} \right)^2},$$

pomocí něhož lze upravit vzorec (135) na tvar:

$$\varphi(v) = C \left(1 - \left(\frac{v}{b} \right)^2 \right)^{\frac{1}{2} \left(\frac{b}{\sigma} \right)^2}. \quad (140)$$

Pro exponent v posledním výrazu zavedeme označení

$$s \triangleq \frac{1}{2} \left(\frac{b}{\sigma} \right)^2. \quad (141)$$

Konstantu C určíme z normalizační podmínky

$$\int_{-b}^b C \left(1 - \left(\frac{v}{b} \right)^2 \right)^s dv = 1. \quad (142)$$

Integrál v normalizační podmínce lze po vytknutí konstanty C dále upravit:

$$\begin{aligned} \int_{-b}^b \left(1 - \left(\frac{v}{b} \right)^2 \right)^s dv &= 2 \int_0^b \left(1 - \left(\frac{v}{b} \right)^2 \right)^s dv = b \int_0^1 t^s (1-t)^{-\frac{1}{2}} dt = \\ &= b B\left(s+1, \frac{1}{2}\right). \end{aligned}$$

Přitom jsme využili toho, že integrovaná funkce je sudá, dále substituce $t = 1 - \left(\frac{v}{b} \right)^2$ a nakonec integrální definice funkce beta (viz např. [Jar84] str. 236, 689). Použitím tohoto výsledku v normalizační podmínce (142) dostaneme hledaný vztah pro konstantu C :

$$C = \frac{1}{b B\left(s+1, \frac{1}{2}\right)}.$$

Hustotu hledaného rozdělení pravděpodobnosti pak můžeme vyjádřit vzorcem:

$$\varphi(v) = \frac{\left(1 - \left(\frac{v}{b} \right)^2 \right)^s}{b B\left(s+1, \frac{1}{2}\right)}. \quad (143)$$

Číslo s je vhodné pro parametrizaci třídy rozdělení daných vzorcem (143). Nové rozdělení pravděpodobnosti proto budeme nadále označovat symbolem φ_s . Pearsonovo rozdělení typu II. je tedy definováno vztahem:

$$\varphi_s(v) \triangleq \frac{\left(1 - \frac{\frac{1}{2}\left(\frac{v}{\sigma}\right)^2}{s}\right)^s}{\sqrt{2s}\sigma B\left(s+1, \frac{1}{2}\right)} \Leftrightarrow |v| < \sqrt{2s}\sigma \quad (144)$$

$$\varphi_s(v) \triangleq 0 \Leftrightarrow |v| \geq \sqrt{2s}\sigma . \quad (145)$$

Pro celočíselné s lze hodnotu beta funkce B snadno určit podle vztahu

$$B\left(s+1, \frac{1}{2}\right) = 2 \prod_{i=1}^s \frac{2i}{2i+1} \quad (146)$$

Čítec posledního zlomku v (144) konverguje k exponenciální funkci, tj.

$$\lim_{s \rightarrow \infty} \left(1 - \frac{\frac{1}{2}\left(\frac{v}{\sigma}\right)^2}{s}\right)^s = \exp\left(-\frac{1}{2}\left(\frac{v}{\sigma}\right)^2\right) \quad (147)$$

(viz např. [Jar84] str. 142, [Bar87] str. 485). Funkce φ_s se tedy se zvětšující šířkou intervalu $(-b, b)$ při konstantním σ blíží k normálnímu rozdělení. To ozřejmuje význam parametru σ jakožto směrodatné odchylky normálního rozdělení, ke kterému rozdělení φ_s konverguje při $s \rightarrow \infty$. Parametr σ je tedy kladný a vzhledem k definici (141) je kladný i parametr s .

Vztahy (144), (145) spolu s podmínkami

$$\sigma > 0, s > 0, \quad (148)$$

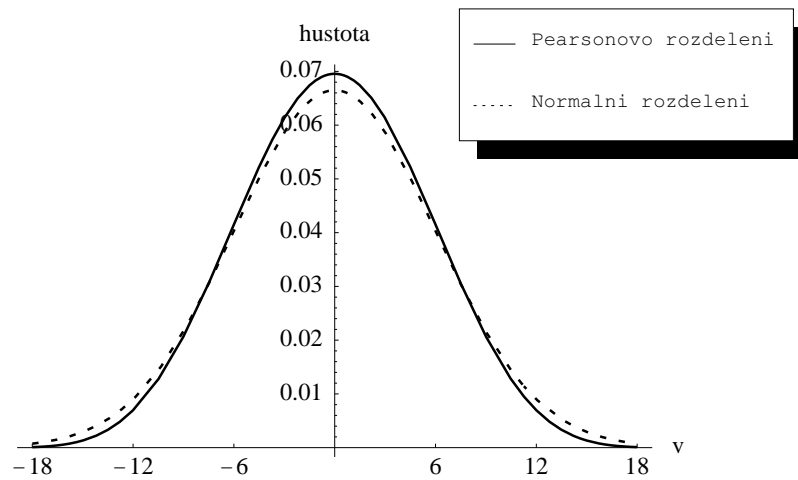
tedy definují nové rozdělení pravděpodobnosti.

Na základě studia starší statistické literatury (např. [Jos64], [EJ69], [Jan37]) lze zjistit, že rozdělení pravděpodobnosti s hustotou (144), (145) je jedno z rozdělení takzvaného Pearsonova systému. Podle klasifikace této třídy rozdělení uvedené v [EJ69] je to typ II. Kniha [EJ69] obsahuje nejobsáhlejší a nejpodrobnější pojednání o Pearsonových rozděleních z dosud zmíněných publikací.

6.3.3 Vlastnosti Pearsonova rozdělení

Právě odvozené Pearsonovo rozdělení φ_s má kromě vlastností požadovaných na začátku oddílu 6 (především podmínky (113), (131)) ještě další výhodné vlastnosti.

1. Je podobné normálnímu rozdělení, pokud je parametr s dostatečně velký ($s \geq 8$) — viz obr. 5. Ostatní vlastnosti normálního rozdělení (centrální limitní věta, maximální entropie při pevné varianci ap.) by tedy měly zůstat přibližně zachovány.



Obrázek 5: Hustota pravděpodobnosti Pearsonova rozdělení φ_s a normálního rozdělení při $\sigma = 6$

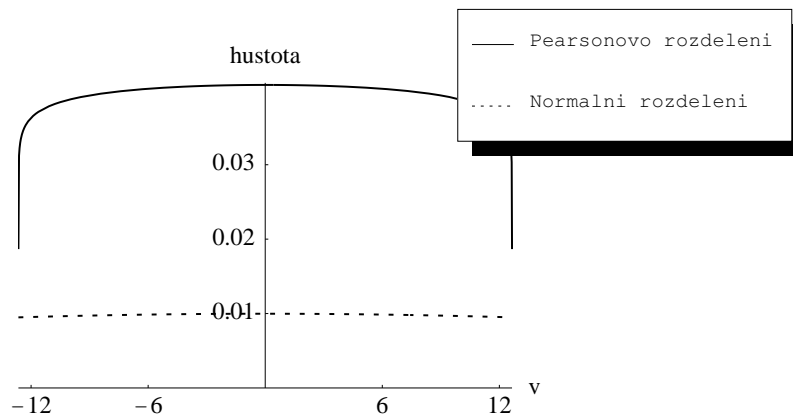
2. Je podobné rovnoměrnému rozdělení, pokud parametr s je dostatečně malý ($s < 0.1$) — viz obr. 6. Pomocí vzorce (143) lze snadno dokázat, že rovnoměrné rozdělení je limitním případem φ_s při pevném b a $\sigma \rightarrow \infty$.
3. Je podobné Studentovu rozdělení. Tato podobnost není jen důsledkem podobnosti s normálním rozdělením. Vzorec (143) totiž silně připomíná výraz pro hustotu Studentova rozdělení (viz např. [And78] str. 87). Pro dostatečně velký počet stupňů volnosti lze výraz pro hustotu Studentova rozdělení upravit na tvar, v němž vynikne souvislost počtu stupňů volnosti (počtu opakování měření) s parametrem b (rozsah přípustných měřených hodnot) modifikovaného normálního rozdělení.
4. Lze jej snadno zobecnit tak, aby aproximovalo i jiná, nesymetrická rozdělení. Toho lze dosáhnout excentrickým umístěním odhadu $\hat{\theta}$ v intervalu \mathcal{T} a zobecněním funkce K na tvar:

$$K(v) \triangleq \frac{\kappa}{\left[1 + \left(\frac{v}{a}\right)\right] \left[1 - \left(\frac{v}{b}\right)\right]},$$

kde a, b jsou kladné reálné konstanty, jejichž součet udává šířku intervalu \mathcal{T} . Tak lze odvodit Pearsonovo rozdělení typu I.

5. Hustotu φ_s lze symbolicky integrovat s pomocí hypergeometrické funkce ${}_2F_1$. Pokud je parametr s nebo jeho dvojnásobek celočíselný, není k integraci hustoty φ_s potřeba žádné speciální funkce.

Poslední vlastnost je velmi užitečná z teoretického i praktického hlediska. Zaslouží si tedy podrobnější vysvětlení.



Obrázek 6: Hustota pravděpodobnosti Pearsonova rozdělení $\varphi_{0.05}$ a normálního rozdělení při $\sigma = 40$

6.3.4 Integrace hustoty Pearsonova rozdělení

Celočíselný parametr s Pro celočíselné s je hustota φ_s polynomem a proto je polynomem i její integrál. Jeho výpočet lze usnadnit použitím rekurentního vzorce

$$\int \varphi_s(v) dv = \frac{v}{2s+1} \varphi_s(v) + \int \varphi_{s-1}(v) dv. \quad (149)$$

Lze jej snadno odvodit pomocí substituce $t = \sqrt{1 - \left(\frac{v}{b}\right)^2}$ a s následným použitím integrace per partes a dalších operací s neurčitými integrály.

Parametr s jako polovina lichého čísla Vzorec (149) lze použít i při neceločíselném s , kdy je celočíselný pouze výraz $2s$. V takovém případě použijeme pro výpočet $\int \varphi_{\frac{1}{2}}(v) dv$ rovnosti

$$\int \sqrt{1 - \left(\frac{v}{b}\right)^2} dv = \frac{1}{2} \left(v \sqrt{1 - \left(\frac{v}{b}\right)^2} + b \arcsin \left(\frac{v}{b}\right) \right) + c,$$

která platí pro všechna $v \in (-b, b)$ (viz např. [Bar87], str. 587, vzorec 3 po drobné úpravě). Integrační konstanta c má zde opět pouze formální význam, neboť při přechodu k určitému integrálu vymizí.

Reálný kladný parametr s V nejobecnějším případě, kdy parametr s je kladné reálné číslo je možno vyjádřit integrál hustoty Pearsonova rozdělení pomocí hypergeometrické funkce (Gaussovy hypergeometrické řady) ${}_2F_1$.

$$\int \varphi_s(v) dv = \frac{v}{b} \frac{{}_2F_1\left(\frac{1}{2}, -s; \frac{3}{2}; \left(\frac{v}{b}\right)^2\right)}{B\left(s+1, \frac{1}{2}\right)} \quad (150)$$

Funkce ${}_2F_1$ je definována nekonečnou řadou

$${}_2F_1(a, b; c; z) = \sum_{k=0}^{\infty} \prod_{i=1}^k \frac{(a+i-1)(b+i-1)}{c+i-1} \frac{z^k}{k!}$$

pro $\forall z \in \mathbb{C}, |z| < 1$. Argument z může být tedy i komplexní, ovšem v našem případě nabývá pouze kladných reálných hodnot. Podrobně je hypergeometrická funkce ${}_2F_1$ a její vlastnosti popsána např. v [Leb56] na str. 237–253. Souvisí také s ortogonálními polynomy (Legendreovými, Čebyševovými, Gegenbauerovými, Jacobiovými) a dalšími speciálními funkcemi (neúplnou beta funkcí, sférickými funkcemi, eliptickými integrály a dalšími, méně známými funkcemi). Vztahy, které zachycují tyto souvislosti jsou uvedeny v [AS70] na str. 561 a také v [Leb56] a [Erd54]. Vztah (150) lze odvodit pomocí vzorce

$${}_2F_1(\alpha, \beta; \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)} \int_0^1 t^{\beta-1} (1-t)^{\gamma-\beta-1} (1-tz)^{-\alpha} dt \quad (151)$$

uvedeného pod číslem 15.3.1 v [AS70] na str. 558. (viz též [Leb56], str. 238). Při úpravě integrálu

$$\begin{aligned} \int \varphi_s(v) dv &= \frac{1}{B(s+1, \frac{1}{2})} \int_0^v \left(1 - \left(\frac{v}{b}\right)^2\right)^s dx = \\ &= \frac{1}{B(s+1, \frac{1}{2})} \frac{v}{2} \int_0^1 t^{-\frac{1}{2}} (1-t)^0 \left(1 - t \left(\frac{v}{b}\right)^2\right)^s dt \end{aligned} \quad (152)$$

bylo použito substituce $t = \left(\frac{x}{v}\right)^2$. Po aplikaci vzorce (151) na poslední výraz v (152) vznikne po drobných úpravách rovnost (150).

6.3.5 Závěrečné poznámky k Pearsonovu rozdělení

Výhodné vlastnosti Pearsonova rozdělení uvedené v předchozím oddílu byly získány za cenu porušení Gaussova požadavku, aby nejvěrohodnějším odhadem souboru měřených hodnot byl aritmetický průměr. Tento požadavek byl porušen nahrazením rovnosti (98) rovností (133). Vzhledem k volbě funkce K (vzorec (138)) splňující podmínku (137) se odhad $\hat{\theta}$ definovaný v pododdíle 6.3.2 odlišuje od aritmetického průměru jen málo. Jejich odchylku lze díky asymptotické vlastnosti (147) učinit tak malou, aby byla srovnatelná s variabilitou aritmetického průměru.

Vzhledem k uvedeným vlastnostem může být Pearsonovo rozdělení použito v Bayesově vzorci (111) v těch případech, kdy normální rozdělení ztrácí své výhodné vlastnosti. Mezi tyto případy patří zejména:

1. Je dáno apriorní rozdělení pravděpodobnosti, které nemusí být rovnoměrné nebo normální.

2. Rozdělení pravděpodobnosti měřených veličin není normální, např. je kontaminováno jinými rozděleními (směs rozdělení).
3. Zobrazení \mathbf{a} z (22) není lineární.

V takovýchto případech se uplatní analytická integrovatelnost Pearsonova rozdělení především tehdy, když jsou apriorní či kontaminující rozdělení nebo funkce a_j vyjádřeny pomocí polynomů. V případě jiných elementárních funkcí nelze samozřejmě symbolickou integrovatelnost zaručit, ale ani vyloučit. To závisí na konkrétním zadání příslušných funkcí (hustot pravděpodobnosti, zobrazení \mathbf{a}). Takové případy lze samozřejmě řešit polynomickou aproximací, která je v matematické literatuře velmi podrobně rozpracována.

Tyto výhody mají však zatím pouze teoretický význam. Při odvozování aposteriorního rozdělení je totiž třeba integrovat součin polynomů, jenž je velmi vysokého řádu již při malém počtu měření. Při praktickém výpočtu by tedy bylo nutné výsledný polynom po integraci aproximovat polynomem nižšího řádu. Rovněž omezení definičního oboru (suportu) hustoty pravděpodobnosti f_{ε_j} na konečný interval vyžaduje další dodatečné výpočty oproti rozdělení s nekonečným suportem. Tyto potíže sice nejsou principiálního charakteru, avšak k jejich efektivnímu řešení by bylo potřeba důkladného rozboru numerického chování integrovaných polynomů při konkrétních funkcích a_j a pro různá apriorní rozdělení pravděpodobnosti. Patrně by bylo možné přitom využít vlastností hypergeometrické funkce ${}_2F_1$. To by však vyžadovalo další podrobný výzkum ve spolupráci s odborníky na matematickou analýzu, což je již mimo rámec této práce.

6.4 Význam modifikace normálního rozdělení pro bayesovské výpočty

Uvedené příklady modifikace normálního rozdělení ukazují, že pro použití určitého rozdělení pravděpodobnosti na chyby měření není rozhodující analytický tvar vzorce pro výpočet hustoty pravděpodobnosti, nýbrž tvar a rozměry frekvenční křivky $y = f_{\varepsilon_j}(x)$ (v kartézském souřadnicovém systému Oxy).

Gaussův postup i ostatní zmíněné metody odvození normálního rozdělení dokazují, že normální rozdělení má seriózní teoretické opodstatnění. Teoretické důvody byly také mnohokrát experimentálně ověřeny, avšak hlavním důvodem pro jeho použití bývá velmi často výpočetní jednoduchost při odhadování parametrů lineárního modelu. Tento důvod je tak významný, že mnohdy je přijat předpoklad normality chyb měření, přestože evidentně neodpovídá skutečnosti. Kdyby byl totiž respektován skutečný tvar hustoty pravděpodobnosti chyb měření, znamenalo by to opustit výhodný analytický tvar výrazu pro hustotu pravděpodobnosti normálního rozdělení, což téměř vždy neúměrně zkomplikuje výpočty.

Tyto výpočetní komplikace přitom mohou nastat i tehdy, když se skutečně rozdělení pravděpodobnosti liší od normálního jen nepatrně. Zvláště názorně to

dokumentuje právě odvozené Pearsonovo rozdělení (144). To se liší od normálního rozdělení tím méně, čím větší je exponent s , tj. poloviční řád polynomu vyjadřujícího hustotu pravděpodobnosti Pearsonova rozdělení. Se zvyšujícím se řádem polynomu však vzrůstá výpočetní náročnost. Výpočet je paradoxně tím složitější, čím blíže je Pearsonovo rozdělení k normálnímu. Výsledná aposteriorní hustota přitom musí v obou případech vyjít prakticky stejně, je-li parametr s dostatečně velký ($s \geq 8$).

Tohoto aspektu lze využít i v obráceném smyslu. Při výpočtu parametrů nelineárního modelu ztrácí normální rozdělení své výhodné vlastnosti. Mohlo by proto být přínosné mírně jej modifikovat tak, aby se výpočty zjednodušily a přitom aby výsledné modifikované rozdělení bylo stále přípustným rozdělením pravděpodobnosti chyb měření. V následující kapitole se proto zaměříme na takovou modifikaci normálního rozdělení, která by umožnila použít lineárních metod i v nelineárním případě.

7 Nelineární model

7.1 Přehled existujících metod

Řešení nelineárního modelu je ve statistické literatuře věnováno mnoho pozornosti. Kromě množství článků roztroušených v odborných časopisech (Journal of the American Statistical Association, Statistical Science, Metrika, Biometrika, Computational Statistics and Data Analysis apod.) existují i specializované monografie o nelineárním odhadování — viz např. [Páz93], [BW88]. Rovněž v geodetické odborné literatuře se vyskytují články o nelineárních problémech, viz např. [Teu90]. Zvláštní postavení má nelineární model v bayesovském pojetí.

Podle Bayesovy věty je možno vyjádřit aposteriorní rozdělení pravděpodobnosti určených parametrů pro libovolné zprostředkující funkce. Problémem je však výpočet integrálu aposteriorní hustoty, který je potřebný k určení normalizační konstanty (jmenovatele v Bayesově vzorci), k výpočtu pravděpodobnosti výskytu určité hodnoty v dané oblasti nebo ke stanovení konfidenčních intervalů při různých statistických testech. Integrál aposteriorní hustoty lze přímo explicitně vypočítat jen v některých speciálních případech. Kromě již zmíněného lineárního modelu s normálním rozdělením chyb měření, jemuž byla věnována kapitola 4 existuje i několik málo dalších případů, kdy lze využít vlastností některých typů rozdělení. Např. v článku [Dic68] jsou popsány tři takové případy, v nichž se využívá mnoharozměrného normálního a Studentova rozdělení a Dirichletova rozdělení. Jednoduchý ilustrativní příklad přesného výpočtu aposteriorní hustoty v nelineárním případě si ukážeme v následujícím oddíle.

Ve většině případů, kdy zprostředkující funkce nejsou lineární nebo rozdělení pravděpodobnosti měřených veličin není normální, je k výpočtu integrálu aposteriorní hustoty nutné použít přibližných numerických metod. Podrobný přehled metod numerické integrace v bayesovské statistice podávají články [ES95], [SSSN87], [Smi91], [SSS⁺85]. Některé standardní metody jsou stručně popsány též v [Rob94] na str. 329–368 a také v [Koc90] na str. 52–60. V zásadě je možno tyto metody zhruba rozdělit do dvou skupin. Do první skupiny patří metoda Monte Carlo a další příbuzné metody založené na simulaci náhodných procesů. Druhá skupina zahrnuje metody založené na analytické aproximaci funkcí vyskytujících se v integrandu určeného integrálu. Nejčastěji se aproximuje marginální aposteriorní hustota. Jako příklad tohoto druhého přístupu je možno uvést velmi efektivní algoritmus založený na Laplaceově metodě přibližné integrace, jehož podstata je popsána v článku [TK86] a dále rozvíjena v následných publikacích [TKK89a], [TKK89b].

Přístup k řešení nelineárního modelu, který byl zvolen v této práci, náleží ke druhé skupině metod. Jeho podstata bude vysvětlena v oddíle 7.4.

7.2 Přesný výpočet aposteriorní hustoty

V některých případech lze integrál v Bayesově větě pro normální rozdělení chyb měření určit přesně i v nelineárním případě. Ukážeme si to na jednoduchém příkladě odhadu jednoho parametru.

Ilustrativní příklad Uvažujme nelineární model s jedním určovaným parametrem θ a pouze jednou zprostředkující funkcí a . O funkci a předpokládáme, že je vzájemně jednoznačným zobrazením množiny \mathcal{T} na množinu \mathcal{L} , tj.

$$a : \mathcal{T} \leftrightarrow \mathcal{L} .$$

Vztah měřených a určovaných veličin je tedy dán soustavou zprostředkujících rovnic (19), zjednodušenou podmínkami

$$m = 1, \quad \theta_1 = \theta, \quad (153)$$

$$a_j = a, \quad \forall j \in \{1, 2, \dots, n\} . \quad (154)$$

Platí proto

$$a(\theta) = \tilde{l}_j + \varepsilon_j, \quad \forall j \in \{1, 2, \dots, n\} . \quad (155)$$

Předpokládejme dále, že všechny měřené veličiny mají normální rozdělení a jsou vzájemně nezávislé. Pro chyby měření ε_j tedy platí

$$\varepsilon_j \rightsquigarrow \mathcal{E}_j \sim \mathcal{N}(0, \sigma_j^2), \quad \forall j \in \{1, 2, \dots, n\} .$$

Hustoty pravděpodobnosti chyb měření jsou

$$f_j(v) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2}\left(\frac{v}{\sigma_j}\right)^2\right), \quad \forall j \in \{1, 2, \dots, n\} . \quad (156)$$

Apriorní rozdělení pravděpodobnosti p je dáno prostřednictvím předběžného měření \tilde{l}_0 . Toto předběžné měření bylo provedeno s přesností odpovídající střední chybě σ_0 . Předpokládáme, že chyby tohoto měření mají také normální rozdělení $\mathcal{N}(0, \sigma_0^2)$. Skutečnou chybu předběžného měření označíme ε_0 .

$$\varepsilon_0 \triangleq a(\theta) - \tilde{l}_0, \quad (157)$$

$$\varepsilon_0 \rightsquigarrow \mathcal{E}_0 \sim f_0 \quad (158)$$

Apriorní hustotu náhodné veličiny Θ určíme pomocí hustoty náhodné veličiny \mathcal{E}_0 na základě transformačního vztahu

$$\Theta = a^{-1}(\mathcal{E}_0 + \tilde{l}_0), \quad (159)$$

který vychází z rovnice (157). Podle věty o transformaci náhodných veličin (viz např. [LM81], str. 69) platí

$$p(\theta) = f_0(a(\theta) - \tilde{l}_0) \left| \frac{da}{d\theta}(\theta) \right|. \quad (160)$$

Aposteriorní hustotu určované veličiny θ obdržíme po dosažení funkce a , hustot f_j , p a měřených hodnot \tilde{l}_j , $j \in \{1, 2, \dots, n\}$ do Bayesova vzorce (34).

$$\begin{aligned} f_{\Theta}(\theta | \tilde{\mathbf{l}}) &= \frac{\prod_{j=1}^n f_{\varepsilon_j}(a_j(\theta) - \tilde{l}_j) p(\theta)}{\int_{\mathcal{T}} \prod_{j=1}^n f_{\varepsilon_j}(a_j(t) - \tilde{l}_j) p(t) dt} = \\ &= \frac{\prod_{j=1}^n f_j(a(\theta) - \tilde{l}_j) f_0(a(\theta) - \tilde{l}_0) \left| \frac{da}{d\theta}(\theta) \right|}{\int_{\mathcal{T}} \prod_{j=1}^n f_j(a(t) - \tilde{l}_j) f_0(a(t) - \tilde{l}_0) \left| \frac{da}{d\theta}(t) \right| dt}. \end{aligned} \quad (161)$$

Vzhledem k tomu, že zprostředkující funkce je jen jedna, lze integrál ve jmenovateli zlomku v (161) dále upravit pomocí substituce $l = a(t)$.

$$\begin{aligned} &\int_{\mathcal{T}} \prod_{j=1}^n f_j(a(t) - \tilde{l}_j) f_0(a(t) - \tilde{l}_0) \left| \frac{da}{d\theta}(t) \right| dt = \\ &= \pm \int_{a(\mathcal{T})} \prod_{j=0}^n f_j(l - \tilde{l}_j) dl = \\ &= \pm \int_{a(\mathcal{T})} \prod_{j=0}^n \frac{1}{\sqrt{2\pi} \sigma_j} \exp\left(-\frac{1}{2} \sum_{j=0}^n \left(\frac{l - \tilde{l}_j}{\sigma_j}\right)^2\right) dl. \end{aligned} \quad (162)$$

Symbol \pm představuje kladné, resp. záporné znaménko podle toho, zda funkce a je na množině \mathcal{T} rostoucí, resp. klesající. Tedy

$$\pm \in \{+, -\}.$$

Poslední výraz v (162) je ve tvaru, kdy je možno argument exponenciální funkce vyjádřit jako součet konstanty a kvadrátu lineární funkce v l . Využijeme přitom postupu popsaného v kapitole 4, který se zde ovšem výrazně zjednoduší. Po zavedení pomocné konstanty

$$\hat{l} \triangleq \frac{\sum_{j=0}^n \frac{\tilde{l}_j}{\sigma_j^2}}{\sum_{j=0}^n \frac{1}{\sigma_j^2}} \quad (163)$$

a prozatímním vypuštění násobné konstanty $\pm \prod_{j=0}^n \frac{1}{\sqrt{2\pi} \sigma_j}$ pak můžeme pokračovat v úpravě jmenovatele.

$$\int_{a(\mathcal{J})} \exp \left(-\frac{1}{2} \sum_{j=0}^n \left(\frac{l - \tilde{l}_j}{\sigma_j} \right)^2 \right) dl = \quad (164)$$

$$\begin{aligned} &= \int_{a(\mathcal{J})} \exp \left(-\frac{1}{2} \sum_{j=0}^n \left(\left(\frac{\hat{l} - \tilde{l}_j}{\sigma_j} \right)^2 + \left(\frac{l - \hat{l}}{\sigma_j} \right)^2 \right) \right) dl = \\ &= \exp \left(-\frac{1}{2} \sum_{j=0}^n \left(\frac{\hat{l} - \tilde{l}_j}{\sigma_j} \right)^2 \right) \int_{a(\mathcal{J})} \exp \left(-\frac{1}{2} \left(\frac{l - \hat{l}}{\hat{\sigma}} \right)^2 \right) dl, \end{aligned} \quad (165)$$

kde

$$\hat{\sigma} \triangleq \sqrt{\frac{1}{\sum_{j=0}^n \frac{1}{\sigma_j^2}}}. \quad (166)$$

Jmenovatel v (161) je tedy

$$\pm \prod_{j=0}^n \frac{1}{\sqrt{2\pi} \sigma_j} \exp \left(-\frac{1}{2} \sum_{j=0}^n \left(\frac{\hat{l} - \tilde{l}_j}{\sigma_j} \right)^2 \right) \int_{a(\mathcal{J})} \exp \left(-\frac{1}{2} \left(\frac{l - \hat{l}}{\hat{\sigma}} \right)^2 \right) dl.$$

Úpravy (164) až (165) můžeme pro $l = a(\theta)$ aplikovat i na čítec zlomku (161) a po krácení konstantou $\pm \prod_{j=0}^n \frac{1}{\sqrt{2\pi} \sigma_j} \exp \left(-\frac{1}{2} \sum_{j=0}^n \left(\frac{\hat{l} - \tilde{l}_j}{\sigma_j} \right)^2 \right)$ dostáváme aposteriorní hustotu ve tvaru

$$f_{\Theta}(\theta | \tilde{\mathbf{I}}) = \frac{\exp \left(-\frac{1}{2} \left(\frac{a(\theta) - \hat{l}}{\hat{\sigma}} \right)^2 \right) \frac{da}{d\theta}(\theta)}{\int_{a(\mathcal{J})} \exp \left(-\frac{1}{2} \left(\frac{l - \hat{l}}{\hat{\sigma}} \right)^2 \right) dl}. \quad (167)$$

Je-li navíc $a(\mathcal{J}) = \mathbb{R}$, zjednoduší se (167) na tvar připomínající hustotu normálního rozdělení.

$$f_{\Theta}(\theta | \tilde{\mathbf{I}}) = \frac{1}{\sqrt{2\pi} \hat{\sigma}} \exp \left(-\frac{1}{2} \left(\frac{a(\theta) - \hat{l}}{\hat{\sigma}} \right)^2 \right) \frac{da}{d\theta}(\theta) \quad (168)$$

Tento vzorec je řešením daného příkladu.

Povšimněme si, že pomocná konstanta \hat{l} definovaná vztahem (163) je vlastně váženým průměrem pro váhy měření $\frac{1}{\sigma_j^2}$, $j \in \{0, 1, 2, \dots, n\}$. Vztah (166) potom vyjadřuje tzv. zákon hromadění vah. Pokud by funkce a byla identická, tj.

$$\forall \theta \in \mathcal{T} : a(\theta) = \theta ,$$

zjednodušil by se daný příklad na již zmíněné vyrovnání přímých měření o různé přesnosti. Aposteriorní hustota (168) by pak byla normální $\mathcal{N}(\hat{l}, \hat{\sigma}^2)$.

Kdyby funkce a byla jen mírně nelineární, byla by její derivace poblíž $a^{-1}(\hat{l})$ téměř konstantní a hustota (168) by tudíž byla velmi blízká hustotě normálního rozdělení. Tento poznatek se vztahuje i na obecný nelineární model a využijeme jej v oddíle 7.4.

7.3 Návrh spolehlivého odhadu aposteriorní hustoty

Většinou nelze najít takovou substituci, která by umožňovala analyticky vypočítat integrál aposteriorní hustoty jako v předchozím příkladě. Pro obecné zadání zprostředkujících funkcí a_j , $j \in \{1, 2, \dots, n\}$ je proto třeba vypracovat univerzální postup, který by nebyl závislý na zvláštních vlastnostech zprostředkujících funkcí.

Velmi často se stává, že zprostředkující funkce jsou jen mírně nelineární. V geodetických problémech bývá tento předpoklad téměř vždy splněn. Nabízí se proto tuto nelinearitu zanedbat a použít lineárních metod. V praxi, a nejen geodetické, se tak také obvykle postupuje. Výsledné aposteriorní rozdělení bude tudíž normální, avšak bude se od skutečně správného aposteriorního rozdělení lišit. V obecném případě je velmi obtížné nějak odhadnout míru této odlišnosti. Bez tohoto odhadu však není možné považovat výsledky zpracování nelineárního modelu lineárními metodami za spolehlivé. Proto je třeba nalézt takovou metodu zpracování měření, která by v nelineárním případě poskytovala spolehlivé výsledky i při použití lineárních metod. Taková metoda bude navržena v následujícím oddíle.

Při návrhu metody integrace aposteriorní hustoty určených veličin θ budeme vycházet z této hypotézy:

Aposteriorní hustota $f_{\Theta}(\theta | \mathbf{1})$ je jen málo odlišná od hustoty normálního rozdělení, pokud:

1. zprostředkující funkce jsou lineární a rozdělení pravděpodobnosti chyb měření je blízké normálnímu,
2. zprostředkující funkce jsou mírně nelineární a rozdělení pravděpodobnosti chyb měření je normální.

Použité vágní výrazy “málo”, “blízké”, “mírně” budou později upřesněny tak, aby bylo možno je kvantifikovat pomocí statistických testů a informačních vzdáleností.

Je zřejmé, jak lze dosáhnout toho, aby bylo možno analyticky integrovat aposteriorní hustotu v případě 1. Nejjednodušší by bylo nahradit nenormální rozdělení chyb měření normálním a použít obvyklých lineárních metod. Pokud se obě rozdělení nebudou příliš lišit, je možné považovat výslednou aposteriorní hustotu za spolehlivý odhad hustoty pravděpodobnosti určovaných veličin. Je-li k dispozici dostatečné množství měřených dat, je obvyklé testovat jejich normalitu pomocí vhodného kritéria (Pearsonova, Kolmogorovova-Smirnovova, excessu, asymetrie apod. — viz např. [BRH90], str. 361–374 nebo [And78], str. 191–196). Pak lze s určitým rizikem daným hladinou významnosti testu považovat naměřené hodnoty za výběr z normálního rozdělení. Toto riziko může také sloužit jako míra spolehlivosti výsledného aposteriorního rozdělení.

Podobně lze postupovat i v případě 2. Problémem je, jak modifikovat normální rozdělení měřených veličin tak, aby pro daný nelineární model bylo možné snadno vyčíslit integrál aposteriorní hustoty. Modifikace musí být přitom pouze mírná, tak, aby modifikované rozdělení nebylo v rozporu s měřenými daty. Řešením tohoto problému se budeme zabývat v této kapitole. Po nalezení vhodného rozdělení měřených veličin však vyvstane další problém, a sice testování shody měřených dat s nově navrženým rozdělením. Tímto problémem se v této práci zabývat nebudeme. Metodika testování shody náhodného výběru s určitým, obecně zadaným rozdělením (tzv. testy dobré shody) je ve statistické literatuře velmi důkladně zpracována (viz např. [BRH90], str. 372–374 ; [LM88], str. 125–127,132,133 ; [And78] apod.)

7.4 Modifikace rozdělení pravděpodobnosti chyb měření

V tomto oddíle se zaměříme na konstrukci rozdělení pravděpodobnosti, které by umožnilo efektivní výpočet integrálu aposteriorní hustoty. Nejjednodušší způsob bude zřejmě takový, který využívá výhodného analytického tvaru hustoty normálního rozdělení.

Předpokládejme pro jednoduchost, že chyby měření ε_j , $j \in \{1, 2, \dots, n\}$ jsou vzájemně nezávislé. Pak je třeba nalézt taková rozdělení f_{ε_j} , $j \in \{1, 2, \dots, n\}$, aby po jejich dosazení do Bayesova vzorce (34) platilo

$$f_{\boldsymbol{\theta}}(\boldsymbol{\theta} | \tilde{\mathbf{I}}) = c \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{Q}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right), \quad (169)$$

kde c je vhodná (normalizační) konstanta a \mathbf{Q} symetrická pozitivně definitní matice. Znamená to vlastně požadovat, aby aposteriorní rozdělení bylo normální. Jak bylo ukázáno v kapitole 4, tvaru (169) lze dosáhnout tehdy, když

$$f_{\varepsilon_j}(a_j(\boldsymbol{\theta}) - \tilde{l}_j) = c_j \exp\left(-\frac{1}{2}\left(b_j(\tilde{l}_j, \boldsymbol{\theta})\right)^2\right), \quad j \in \{1, 2, \dots, n\} \quad (170)$$

a b_j je lineární funkce v $\boldsymbol{\theta}$ pro $\forall j \in \{1, 2, \dots, n\}$. Symbol c_j představuje konstantu, která bude později určena normováním.

Místo abychom se při výpočtu integrálu aposteriorní hustoty snažili zachovat normální rozdělení měřických chyb a aproximovali zprostředkující funkce a_j , $j \in \{1, 2, \dots, n\}$ lineárními funkcemi, modifikujeme normální rozdělení tak, aby při platnosti rovnosti (170) mohly funkce a_j zůstat nezměněny. Zprostředkující funkce a_j se obvykle nahrazují jejich Taylorovým rozvojem v okolí nějakého bodu $\tilde{\boldsymbol{\theta}}$, takže pro normální rozdělení s hustotou $\tilde{f}_{\varepsilon_j}$ platí přibližná rovnost

$$\tilde{f}_{\varepsilon_j}(a_j(\boldsymbol{\theta}) - \tilde{l}_j) \cong \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} \left(\frac{a_j(\tilde{\boldsymbol{\theta}}) + (\mathbf{a}'_j(\tilde{\boldsymbol{\theta}}))^T (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) - \tilde{l}_j}{\sigma_j} \right)^2 \right). \quad (171)$$

Vektor $\mathbf{a}'_j(\tilde{\boldsymbol{\theta}})$ představuje vektor parciálních derivací funkce a_j v bodě $\tilde{\boldsymbol{\theta}}$.

$$\mathbf{a}'_j(\tilde{\boldsymbol{\theta}}) \triangleq \left(\frac{\partial a_j}{\partial \theta_1}(\tilde{\boldsymbol{\theta}}), \dots, \frac{\partial a_j}{\partial \theta_n}(\tilde{\boldsymbol{\theta}}) \right)^T, \quad j \in \{1, 2, \dots, n\}. \quad (172)$$

Porovnáním pravých stran v (171) a (170) dostáváme

$$b_j(\tilde{l}_j, \boldsymbol{\theta}) = \frac{a_j(\tilde{\boldsymbol{\theta}}) + (\mathbf{a}'_j(\tilde{\boldsymbol{\theta}}))^T (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) - \tilde{l}_j}{\sigma_j} \beta_1 + \beta_0, \quad (173)$$

přičemž konstanty β_0 , β_1 jsou volitelné a později je využijeme k tomu, aby se hustota modifikovaného rozdělení f_{ε_j} příliš nelišila od skutečného rozdělení chyb měření. Z hlediska jejich interpretace je lepší místo konstant β_0 , β_1 zvolit jinou dvojici konstant M_0 , M_1 tak, aby ekvivalentně s (173) platilo

$$b_j(\tilde{l}_j, \boldsymbol{\theta}) = \frac{a_j(\tilde{\boldsymbol{\theta}}) + (\mathbf{a}'_j(\tilde{\boldsymbol{\theta}}))^T (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) - \tilde{l}_j - M_0}{M_1}. \quad (174)$$

Konstanty M_0 , M_1 lze interpretovat jako analogie střední hodnoty a střední chyby, čehož využijeme v následujících úvahách.

Dosazením (174) do (170) vznikne rovnice pro neznámou funkci f_{ε_j} .

$$f_{\varepsilon_j}(a_j(\boldsymbol{\theta}) - \tilde{l}_j) = c_j \exp \left(-\frac{1}{2} \left(\frac{a_j(\tilde{\boldsymbol{\theta}}) + (\mathbf{a}'_j(\tilde{\boldsymbol{\theta}}))^T (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) - \tilde{l}_j - M_0}{M_1} \right)^2 \right). \quad (175)$$

7.4.1 Jednorozměrný případ

Rovnici (175) lze snadno řešit pro jednu určovanou veličinu

$$\theta = \boldsymbol{\theta}$$

za pomoci substituce

$$\theta = a_j^{-1}(\tilde{l}_j + \varepsilon_j). \quad (176)$$

Pro přehlednost zápisu vypustíme prozatím index j .

$$f_{\varepsilon}(\varepsilon) = c \exp \left(-\frac{1}{2} \left(\frac{a(\tilde{\theta}) + a'(\tilde{\theta}) (a^{-1}(\tilde{l} + \varepsilon) - \tilde{\theta}) - \tilde{l} - M_0}{M_1} \right)^2 \right). \quad (177)$$

Konstanty M_0 , M_1 zvolíme tak, aby rozdělení pravděpodobnosti f_{ε} mělo stejnou střední hodnotu μ_{ε} a rozptyl σ_{ε}^2 jako původní normální rozdělení chyb měření, které modifikujeme. Musí tedy být splněny podmínky

$$\int_{\mathcal{J}} \varepsilon f_{\varepsilon}(\varepsilon) d\varepsilon = \mu_{\varepsilon} = 0, \quad (178)$$

$$\int_{\mathcal{J}} \varepsilon^2 f_{\varepsilon}(\varepsilon) d\varepsilon = \sigma_{\varepsilon}^2, \quad (179)$$

které po dosazení za $f_{\varepsilon}(\varepsilon)$ podle (177) přejdou v soustavu dvou rovnic pro dvě neznámé M_0 , M_1 . Konstantu c je přitom třeba vyjádřit ve tvaru

$$c = \frac{1}{\int_{\mathcal{J}} \exp \left(-\frac{1}{2} \left(\frac{a(\tilde{\theta}) + a'(\tilde{\theta}) (a^{-1}(\tilde{l} + \varepsilon) - \tilde{\theta}) - \tilde{l} - M_0}{M_1} \right)^2 \right) d\varepsilon}.$$

Z první rovnice lze konstantu c eliminovat vydělením. Potom

$$\int_{\mathcal{J}} \varepsilon \exp \left(-\frac{1}{2} \left(\frac{a(\tilde{\theta}) + a'(\tilde{\theta}) (a^{-1}(\tilde{l} + \varepsilon) - \tilde{\theta}) - \tilde{l} - M_0}{M_1} \right)^2 \right) d\varepsilon = 0, \quad (180)$$

$$\frac{\int_{\mathcal{J}} \varepsilon^2 \exp \left(-\frac{1}{2} \left(\frac{a(\tilde{\theta}) + a'(\tilde{\theta}) (a^{-1}(\tilde{l} + \varepsilon) - \tilde{\theta}) - \tilde{l} - M_0}{M_1} \right)^2 \right) d\varepsilon}{\int_{\mathcal{J}} \exp \left(-\frac{1}{2} \left(\frac{a(\tilde{\theta}) + a'(\tilde{\theta}) (a^{-1}(\tilde{l} + \varepsilon) - \tilde{\theta}) - \tilde{l} - M_0}{M_1} \right)^2 \right) d\varepsilon} = \sigma_{\varepsilon}^2. \quad (181)$$

Přibližné řešení soustavy rovnic (180), (181). Soustavu rovnic (180), (181) lze přibližně řešit, využijeme-li podobnosti hustoty $f_{\varepsilon}(\varepsilon)$ s hustotou normálního rozdělení. Později uvedeme i přesnější řešení a proto přibližné řešení označíme zvláštními symboly \tilde{M}_0 , \tilde{M}_1 , takže

$$M_0 \cong \tilde{M}_0, \quad M_1 \cong \tilde{M}_1.$$

Podobnost f_{ε} s normálním rozdělením vyjadřuje tato přibližná rovnost:

$$f_{\varepsilon}(\varepsilon) \cong \frac{1}{\sqrt{2\pi}\sigma_{\varepsilon}} \exp \left(-\frac{1}{2} \left(\frac{\varepsilon - \mu_{\varepsilon}}{\sigma_{\varepsilon}} \right)^2 \right). \quad (182)$$

Abychom této rovnosti mohli využít, je třeba v (177) linearizovat funkci a^{-1} v bodě \tilde{l} .

$$\begin{aligned} f_{\varepsilon}(\varepsilon) &\cong c \exp \left(-\frac{1}{2} \left(\frac{a(\tilde{\theta}) + a'(\tilde{\theta}) \left(a^{-1}(\tilde{l}) + \frac{da^{-1}}{d\tilde{l}}(\tilde{l}) \varepsilon - \tilde{\theta} \right) - \tilde{l} - \tilde{M}_0}{\tilde{M}_1} \right)^2 \right) = \\ &= c \exp \left(-\frac{1}{2} \left(\frac{\varepsilon - \frac{a'(a^{-1}(\tilde{l}))}{a'(\tilde{\theta})} \left(\tilde{M}_0 + \tilde{l} - a(\tilde{\theta}) - a'(\tilde{\theta}) \left(a^{-1}(\tilde{l}) - \tilde{\theta} \right) \right)}{\tilde{M}_1 \frac{a'(a^{-1}(\tilde{l}))}{a'(\tilde{\theta})}} \right)^2 \right) \end{aligned}$$

Porovnáním posledního výrazu s pravou stranou (182) dostáváme

$$\sigma_{\varepsilon} = \tilde{M}_1 \frac{a'(a^{-1}(\tilde{l}))}{a'(\tilde{\theta})}, \quad (183)$$

$$\mu_{\varepsilon} = \frac{a'(a^{-1}(\tilde{l}))}{a'(\tilde{\theta})} \left(\tilde{M}_0 + \tilde{l} - a(\tilde{\theta}) - a'(\tilde{\theta}) \left(a^{-1}(\tilde{l}) - \tilde{\theta} \right) \right), \quad (184)$$

odkud s ohledem na předpoklad $\mu_{\varepsilon} = 0$ po elementárních úpravách plyne

$$\tilde{M}_1 = \sigma_{\varepsilon} \frac{a'(\tilde{\theta})}{a'(a^{-1}(\tilde{l}))}, \quad (185)$$

$$\tilde{M}_0 = \tilde{l} - a(\tilde{\theta}) - a'(\tilde{\theta}) \left(a^{-1}(\tilde{l}) - \tilde{\theta} \right). \quad (186)$$

Přesnější řešení soustavy rovnic (180), (181). Abychom mohli soustavu rovnic (180), (181) řešit přesněji než v předchozím odstavci, je třeba ve všech integrálech změnit integrační proměnnou ε substitucí

$$\varepsilon = a(\theta) - \tilde{l}.$$

Tím dostaneme v argumentu exponenciální funkce kvadrát lineární funkce, což umožní rozvoj integrálu v nekonečnou řadu. Pro integrál z první rovnice (180) tedy máme

$$\begin{aligned} &\int_{\mathcal{J}} \varepsilon \exp \left(-\frac{1}{2} \left(\frac{a(\tilde{\theta}) + a'(\tilde{\theta}) \left(a^{-1}(\tilde{l} + \varepsilon) - \tilde{\theta} \right) - \tilde{l} - \tilde{M}_0}{\tilde{M}_1} \right)^2 \right) d\varepsilon = \\ &= \int_{a(\mathcal{J}) - \tilde{l}} (a(\theta) - \tilde{l}) a'(\theta) \exp \left(-\frac{1}{2} \left(\frac{a(\tilde{\theta}) + a'(\tilde{\theta}) \left(\theta - \tilde{\theta} \right) - \tilde{l} - \tilde{M}_0}{\tilde{M}_1} \right)^2 \right) d\theta = \end{aligned}$$

$$\begin{aligned}
&= \int_{a(\mathcal{T})-\tilde{l}} (a(\theta) - \tilde{l}) a'(\theta) \exp \left(-\frac{1}{2} \left(\frac{\theta - \left(\tilde{\theta} + \frac{M_0 + \tilde{l} - a(\tilde{\theta})}{a'(\tilde{\theta})} \right)}{\frac{M_1}{a'(\tilde{\theta})}} \right)^2 \right) d\theta = \\
&= \int_{a(\mathcal{T})-\tilde{l}} (a(\theta) - \tilde{l}) a'(\theta) \exp \left(-\frac{1}{2} \left(\frac{\theta - \mu}{\sigma} \right)^2 \right) d\theta .
\end{aligned}$$

Před poslední úpravou jsme použili zjednodušujícího označení

$$\mu = \tilde{\theta} + \frac{M_0 + \tilde{l} - a(\tilde{\theta})}{a'(\tilde{\theta})}, \quad \sigma = \frac{M_1}{a'(\tilde{\theta})}. \quad (187)$$

Aby bylo možno soustavu (180), (181) přesně řešit, je třeba ještě předpokládat

$$a(\mathcal{T}) = \mathbb{R}.$$

První rovnici (180) lze tedy psát ve tvaru

$$\int_{\mathbb{R}} (a(\theta) - \tilde{l}) a'(\theta) \exp \left(-\frac{1}{2} \left(\frac{\theta - \mu}{\sigma} \right)^2 \right) d\theta = 0. \quad (188)$$

Stejným způsobem upravíme i druhou rovnici (181).

$$\frac{\int_{\mathbb{R}} (a(\theta) - \tilde{l})^2 a'(\theta) \exp \left(-\frac{1}{2} \left(\frac{\theta - \mu}{\sigma} \right)^2 \right) d\theta}{\int_{\mathbb{R}} a'(\theta) \exp \left(-\frac{1}{2} \left(\frac{\theta - \mu}{\sigma} \right)^2 \right) d\theta} = \sigma_{\varepsilon}^2. \quad (189)$$

Nyní mají všechny integrály v rovnicích (188), (189) jednotný tvar

$$\int_{\mathbb{R}} b(\theta) \exp \left(-\frac{1}{2} \left(\frac{\theta - \mu}{\sigma} \right)^2 \right) d\theta, \quad (190)$$

kde pomocný výraz $b(\theta)$ zastupuje výrazy $a'(\theta)$, $a'(\theta) (a(\theta) - \tilde{l})$, $a'(\theta) (a(\theta) - \tilde{l})^2$, tedy

$$b(\theta) \in \{a'(\theta), a'(\theta) (a(\theta) - \tilde{l}), a'(\theta) (a(\theta) - \tilde{l})^2\}.$$

Funkci b lze rozvinout v Taylorovu řadu v okolí nějakého daného bodu θ' , který nemusí nutně splývat s již stanovenou hodnotou $\tilde{\theta}$. Zvolíme

$$\theta' = a^{-1}(\tilde{l}), \quad (191)$$

neboť to později výrazně zjednoduší výpočty.

Integrál (190) lze tedy zapsat pomocí nekonečné řady

$$\begin{aligned}
& \int_{\mathbb{R}} b(\theta) \exp\left(-\frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2\right) d\theta = \\
& = \int_{\mathbb{R}} \sum_{i=0}^{\infty} \frac{b^{(i)}(\theta')}{i!} (\theta - \theta')^i \exp\left(-\frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2\right) d\theta = \\
& = \int_{\mathbb{R}} \sum_{i=0}^{\infty} \sum_{j=0}^i \frac{b^{(i)}(\theta')}{i!} \binom{i}{j} (\mu - \theta')^{i-j} (\theta - \mu)^j \exp\left(-\frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2\right) d\theta.
\end{aligned} \tag{192}$$

Před poslední úpravou jsme k umocnění rozdílu $(\theta - \theta')$ použili binomické věty

$$(\theta - \theta')^i = ((\theta - \mu) - (\mu - \theta'))^i = \sum_{j=0}^i \binom{i}{j} (\mu - \theta')^{i-j} (\theta - \mu)^j.$$

S využitím identity

$$\sum_{i=0}^k \sum_{j=0}^i x_{ij} = \sum_{j=0}^k \sum_{i=j}^k x_{ij},$$

kteřá platí pro libovolný výraz x_{ij} a každé $k \in \mathbb{N}$, můžeme zaměnit pořadí sumací v posledním integrálu a dále pokračovat v jeho úpravách.

$$\begin{aligned}
& \int_{\mathbb{R}} b(\theta) \exp\left(-\frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2\right) d\theta = \\
& = \int_{\mathbb{R}} \sum_{j=0}^{\infty} \left(\sum_{i=j}^{\infty} \frac{b^{(i)}(\theta')}{i!} \binom{i}{j} (\mu - \theta')^{i-j} \right) (\theta - \mu)^j \exp\left(-\frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2\right) d\theta = \\
& = \sum_{j=0}^{\infty} \frac{1}{j!} \sum_{i=j}^{\infty} \frac{b^{(i)}(\theta')}{(i-j)!} (\mu - \theta')^{i-j} \int_{\mathbb{R}} (\theta - \mu)^j \exp\left(-\frac{1}{2}\left(\frac{\theta-\mu}{\sigma}\right)^2\right) d\theta. \tag{193}
\end{aligned}$$

Integrál v (193) lze vyjádřit pomocí tzv. základního integrálu teorie chyb — viz [BRH90], str 17, 18 nebo [Bar87], str. 613, vzorec 32.

$$\int_0^{\infty} x^i e^{-h^2 x^2} dx = \frac{1}{2h^{i+1}} \Gamma\left(\frac{i+1}{2}\right) \tag{194}$$

Pro sudý exponent i , tj. $i = 2j$, $j \in \mathbb{N}$ s pomocí (194) dostáváme

$$\int_{-\infty}^{+\infty} x^i e^{-h^2 x^2} dx = 2 \int_0^{\infty} x^i e^{-h^2 x^2} dx = \frac{(2j-1)!!}{h^{2j+1} 2^j} \sqrt{\pi}; \tag{195}$$

(viz též [AS70], str. 935, vzorec 26.2.46). Pro liché i není třeba vzorce použít, neboť $\int_{-\infty}^{+\infty} x^i e^{-h^2 x^2} dx = 0$.

Abychom mohli použít výsledku (195), musíme položit

$$h = \frac{1}{\sqrt{2}\sigma}.$$

Pro stručnost zápisu zavedeme ještě označení

$$\beta_j(N, \mu) \triangleq \sum_{k=0}^{N-j} \frac{b^{(j+k)}(\theta')}{k!} (\mu - \theta')^k. \quad (196)$$

Pak již můžeme integrál (190) vyjádřit ve tvaru nekonečné řady

$$\begin{aligned} \int_{\mathbb{R}} b(\theta) \exp\left(-\frac{1}{2}\left(\frac{\theta - \mu}{\sigma}\right)^2\right) d\theta &= \sum_{i=0}^{\infty} \beta_{2i}(\infty, \mu) \frac{(2i-1)!!}{(2i)!} \sqrt{2\pi} \sigma^{2i+1} = \\ &= \sqrt{2\pi} \sigma \lim_{N \rightarrow \infty} \sum_{i=0}^N \frac{\beta_{2i}(N, \mu)}{(2i)!!} \sigma^{2i}. \end{aligned} \quad (197)$$

Abychom mohli tohoto výsledku využít k řešení soustavy rovnic (188), (189), je třeba jej aplikovat na výrazy

$$b(\theta) \in \{a'(\theta), a'(\theta)(a(\theta) - \tilde{l}), a'(\theta)(a(\theta) - \tilde{l})^2\}.$$

a poté dosadit do příslušných rovnic. Sestavíme tak soustavu rovnic pro neznámé $\mu - \theta'$, σ , z kterých pak můžeme snadno určit hledané konstanty M_0 , M_1 pomocí (187).

$$\lim_{N \rightarrow \infty} \sum_{i=0}^N \sum_{k=0}^{N-2i} \frac{d^{2i+k}(a'(\theta')(a(\theta') - \tilde{l}))}{d\theta^{2i+k}} \frac{(\mu - \theta')^k}{k!} \frac{\sigma^{2i}}{(2i)!!} = 0 \quad (198)$$

$$\lim_{N \rightarrow \infty} \frac{\sum_{i=0}^N \sum_{k=0}^{N-2i} \frac{d^{2i+k}(a'(\theta')(a(\theta') - \tilde{l})^2)}{d\theta^{2i+k}} \frac{(\mu - \theta')^k}{k!} \frac{\sigma^{2i}}{(2i)!!}}{\sum_{i=0}^N \sum_{k=0}^{N-2i} \frac{d^{2i+k} a'(\theta')}{d\theta^{2i+k}} \frac{(\mu - \theta')^k}{k!} \frac{\sigma^{2i}}{(2i)!!}} = \sigma_{\varepsilon}^2 \quad (199)$$

V tomto tvaru bychom mohli soustavu řešit pomocí nekonečných řad. V první rovnici je mocninná řada pro σ^2 , ve druhé je podíl dvou takových řad, který lze vyjádřit jako jednu mocninnou řadu v σ^2 . Z první rovnice bychom mohli vyjádřit σ^2 inverzí řady (viz např. [Vyš48], str. 97, 98) a dosadit do druhé rovnice. Tím by se levá strana druhé rovnice stala dosti složitou kompozicí řad mocninami $(\mu - \theta')$, ale i tu by teoreticky bylo možné upravit do jediné mocninné řady v $(\mu - \theta')$. Inverzí této řady v druhé rovnici bychom získali $(\mu - \theta')$, pomocí kterého bychom vyjádřili i σ^2 . Tím by vznikly dvě nekonečné mocninné řady pro známé σ_{ε}^2 , s jejichž pomocí bychom mohli určit μ a σ s libovolnou přesností.

Oprávněnost tohoto postupu by však nejprve musela být prokázána rozбором podmínek konvergence zúčastněných řad. To zde nebudeme provádět. Uvedeme pouze zjednodušenou verzi tohoto postupu.

Postup popsany v předchozím odstavci není třeba provádět v plné obecnosti. Dostatečně přesné řešení obdržíme i při velmi malém počtu členů v uvedených nekonečných řadách. Velmi snadno lze soustavu rovnic (198), (199) řešit pro $N = 1$ (po odstranění symbolu limity). V takovém případě můžeme z první rovnice přímo vyjádřit σ^2 a dosadit do druhé rovnice. Tím vznikne kvadratická rovnice pro $(\mu - \theta')$, kterou snadno vyřešíme.

Po výpočtu konstant M_0, M_1 řešením rovnic (198), (199), (187) a jejich dosazení do vzorce (177) získáme hustotu pravděpodobnosti chyb měření f_ε , která má předepsanou střední hodnotu μ_ε a střední chybu σ_ε .

Nyní je již potřeba opět používat index j pro rozlišení různých měřených veličin l_j a jejich atributů $\tilde{l}_j, \mu_{\varepsilon_j}, \sigma_{\varepsilon_j}, f_{\varepsilon_j}, a_j$ pro $j \in \{1, 2, \dots, n\}$.

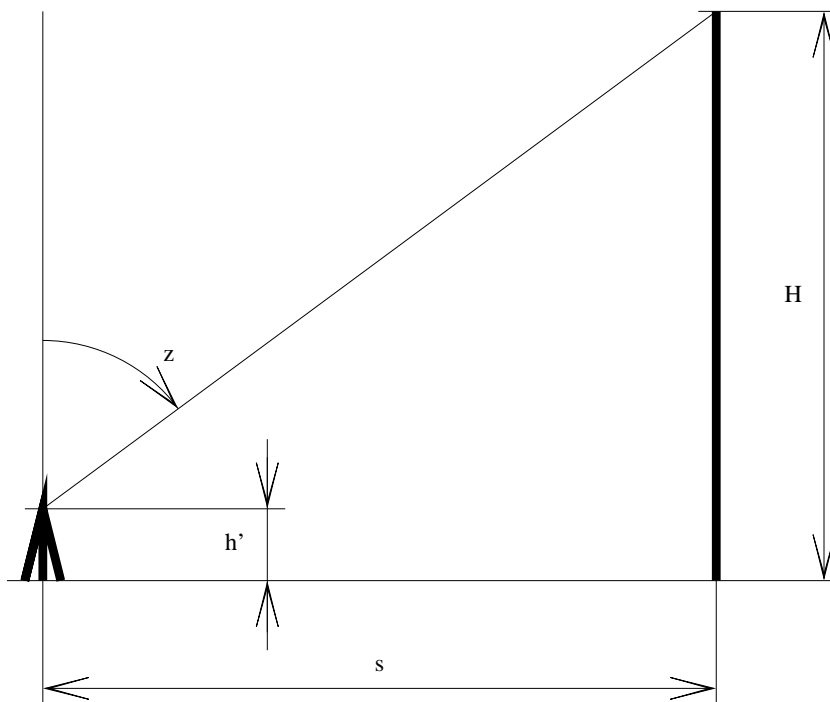
7.4.2 Závěrečné poznámky k navrženému postupu modifikace

Dosazením hustot f_{ε_j} , získaných právě popsáním procesem modifikace pro všechna $j \in \{1, 2, \dots, n\}$ do Bayesova vzorce (34) získáme aposteriorní hustotu f_θ , která by se neměla příliš lišit od aposteriorní hustoty vypočtené pro původní normální rozdělení pravděpodobnosti chyb měření. Při určování hustoty f_θ se využije platnosti požadované podmínky (169), takže je možné použít některou vhodnou metodu pro řešení lineárního modelu. Podmínka (169) vlastně představuje požadavek normality aposteriorního rozdělení. Výsledkem navrhované metody je tudíž aproximace skutečného aposteriorního rozdělení normálním rozdělením. Výhodou této metody je možnost kontroly provedené aproximace pomocí modifikovaných rozdělení chyb měření.

Jak již bylo naznačeno v oddíle (7.3), míra odlišnosti výsledného aposteriorního rozdělení od skutečného aposteriorního rozdělení by mohla být stanovena na základě statistického testu, který by nezamítl hypotézu, že měřená data mají modifikovaná rozdělení pravděpodobnosti. Hladina významnosti, případně jiné charakteristiky použitého testu by mohly sloužit jako kritérium spolehlivosti výsledného aposteriorního rozdělení pravděpodobnosti.

Navržený postup modifikace je možné použít i když původní rozdělení pravděpodobnosti chyb měření není přesně normální. Mohlo by se dokonce stát, že modifikované rozdělení pravděpodobnosti lépe vystihuje měřená data než normální rozdělení. Bylo by také možné zvolit jiné kritérium podobnosti původního a modifikovaného rozdělení než shodu střední hodnoty a střední chyby. Muselo by však být zvoleno tak, aby rovnice pro konstanty M_0, M_1 , které by nahradilo rovnice (198), (199), byly snadno řešitelné.

Účinnost navrženého postupu je patrná z následujícího příkladu.



Obrázek 7: Trigonometrické určení výšky

7.4.3 Příklad — trigonometrické určení výšky

Předpokádejme, že je třeba trigonometricky určit relativní výšku nějakého objektu. Ze stanoviska ve výšce h' ve vzdálenosti s od objektu byl změřen zenitový úhel z spojnice stanovisko — vrchol objektu. Situace je znázorněna na obrázku 7

Pro určovanou výšku H objektu tedy platí jednoduchý vztah:

$$H = h' + \frac{s}{\tan(z)}. \quad (200)$$

Pro jednoduchost předpokládejme, že hodnoty veličin s , h' jsou známé (byly změřeny s mnohem vyšší přesností než zenitový úhel). Určovanou veličinou je tedy jen výška H .

Zprostředkující rovnice mají potom tvar

$$z_j = \arctan\left(\frac{s_j}{H - h'_j}\right), \quad (201)$$

a v němž z_j představuje zenitový úhel měřený ze stanoviska ve vzdálenosti s_j a ve výšce h'_j nad vztážnou rovinou.

Při běžně dosahované přesnosti měření zenitových úhlů odpovídající střední chybě 0.0010^g až 0.0020^g je chyba linearizace zprostředkujících rovnic (201) zanedbatelná. Např. při určování výšky kolem 300 m ze vzdálenosti $s_j = 400$ m

($h'_j \approx 0$) s uvedenou přesností měření dosahuje chyba linearizace zprostředkující funkce a_j hodnot maximálně několika tisícín střední chyby. Aby hodnota chyby linearizace byla srovnatelná s velikostí chyb měření, musely by chyby měření být několik gradů (g). Pro účely demonstrace byl proto zvolen fiktivní případ měření zenitových úhlů se střední chybou $\sigma = 3^g$. Potom by v nejnepříznivějším případě, kdy přibližná hodnota $\hat{\theta}$ určované výšky padne na okraj intervalu odpovídajícího maximální přípustné chybě měření úhlů (3σ), mohla chyba linearizace dosáhnout hodnoty až 4^g .

Dané hodnoty

$$\begin{aligned} s_j &= 300\text{m} , & \text{pro } j \in \{1, 2, \dots, 7\} , \\ s_j &= 375\text{m} , & \text{pro } j \in \{8, 9, \dots, 14\} , \\ s_j &= 225\text{m} , & \text{pro } j \in \{15, 16, \dots, 21\} , \\ h'_j &= 1.72\text{m} , & \text{pro } j \in \{1, 2, \dots, 7\} , \\ h'_j &= 1.63\text{m} , & \text{pro } j \in \{8, 9, \dots, 14\} , \\ h'_j &= 1.74\text{m} , & \text{pro } j \in \{15, 16, \dots, 21\} . \end{aligned}$$

Měřené úhly byly simulovány pomocí generátoru pseudonáhodných čísel programového systému *Mathematica* tak, aby měřené hodnoty měly normální rozdělení se střední hodnotou odpovídající výšce $H = 317$ m (výška Eifelovy věže v Paříži) a střední chybou 3^g . Tak byl získán soubor 21 měření

$$\begin{aligned} &\{\tilde{z}_j \mid j \in \{1, 2, \dots, 21\}\} = \\ &= \{52.7^g, 48.7^g, 43.2^g, 43.7^g, 49.9^g, 49.8^g, 52.2^g, 56.7^g, 55.4^g, 54.0^g, \\ &54.2^g, 58.5^g, 55.2^g, 52.1^g, 39.0^g, 39.2^g, 44.5^g, 35.5^g, 36.9^g, 42.3^g, 36.4^g\} . \end{aligned} \quad (202)$$

Chyby měření vyjádřené v radiánech by tedy měly představovat náhodný výběr z rozdělení $\mathcal{N}\left(0, \left(\frac{3\pi}{200}\right)^2\right)$, tj. $\mathcal{N}(0, 0.0022207)$. Předpokládáme rovněž nezávislost měřených veličin. Střední chybu měření vyjádřenou v radiánech označíme σ_ε ,

$$\sigma_\varepsilon = \frac{3.0 \pi}{200} = 0.0471 .$$

Řešení

Zadaná data (202) byla zpracována třemi různými postupy:

1. bayesovským přístupem pro nelineární model,
2. metodou nejmenších čtverců (MNC) pro lineární model.
3. navrhovanou metodou modifikace rozdělení chyb měření.

Všechny tři případy představují aplikace bayesovského přístupu, avšak liší se způsobem modifikace vstupních údajů tak, aby byly vhodné pro výpočet. Vycházíme tedy z obecného tvaru Bayesovy věty pro nezávislá měření (34). Kvůli možnosti srovnání s MNČ zvolíme neinformativní apriorní hustotu danou vztahem (60) pro $\mathcal{T} = \mathbb{R}$. Použijeme-li ve vzorci (34) limitní přechod pro $|\mathcal{T}| \rightarrow \infty$, můžeme vyjádřit hledanou aposteriorní hustotu výšky H vztahem

$$f_H(\theta | \tilde{\mathbf{I}}) = \frac{\prod_{j=1}^n f_{\varepsilon_j}(a_j(\theta) - \tilde{l}_j)}{\int_{\mathcal{T}} \prod_{j=1}^n f_{\varepsilon_j}(a_j(t) - \tilde{l}_j) dt} . \quad (203)$$

Do tohoto vztahu je třeba dosadit počet měření

$$n = 21$$

a simulovaná data

$$\tilde{l}_j = \tilde{z}_j, \quad j \in \{1, 2, \dots, n\}$$

ze souboru (202). Za zprostředkující funkce a_j a hustoty pravděpodobnosti f_{ε_j} , $j \in \{1, 2, \dots, n\}$ budeme dosazovat rozdílně podle toho, jakou metodou budeme data zpracovávat.

Bayesovský přístup pro nelineární model. Tento případ představuje striktní aplikaci bayesovského přístupu, při němž nejsou dané zprostředkující funkce a_j ani hustoty f_{ε_j} , $j \in \{1, 2, \dots, n\}$ nijak upravovány. Je tedy

$$a_j(\theta) = \arctan\left(\frac{s_j}{\theta - h'_j}\right),$$

$$f_{\varepsilon_j}(\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp\left(-\frac{1}{2}\left(\frac{\varepsilon}{\sigma_\varepsilon}\right)^2\right), \quad j \in \{1, 2, \dots, n\} .$$

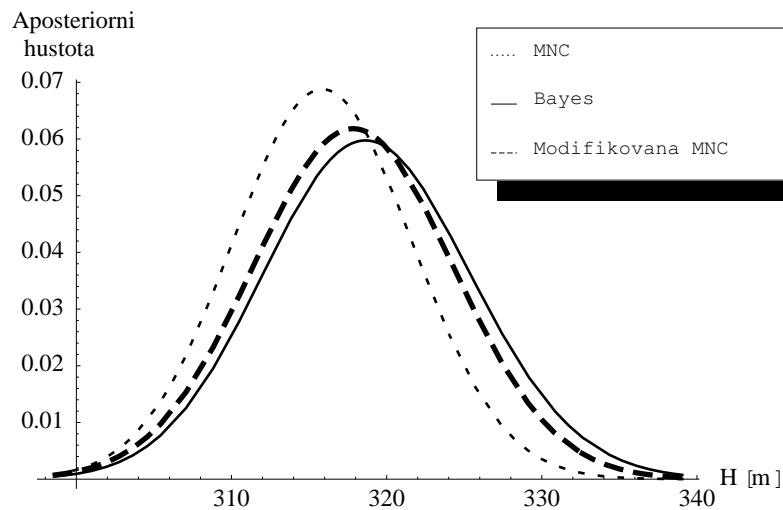
Jmenovatel (203) je třeba vypočítat numerickou integrací.

Výsledná aposteriorní hustota f_H je graficky znázorněna v obrázku 8 plnou čarou. Odhadem určené výšky může být modus aposteriorní hustoty jehož hodnota \hat{H} byla určena iteračním algoritmem

$$\hat{H} = 318.6 \text{ m} .$$

MNČ pro lineární model. V tomto případě je třeba zprostředkující funkce a_j v (204) linearizovat. Rozdělení pravděpodobnosti chyb měření bude stejné jako v předcházejícím případě, tj. (204). Jak bylo dokázáno v kapitole 4, výsledná aposteriorní hustota je normální. Je zobrazena v obrázku 8 tenkou čárkovanou čarou. Její střední hodnota \hat{H} a střední chyba $\hat{\sigma}$ byla výpočtena MNČ.

$$\hat{H} = 315.9 \text{ m}, \quad \hat{\sigma} = 5.8 \text{ m}$$



Obrázek 8: Porovnání různých metod trigonometrického určení výšky

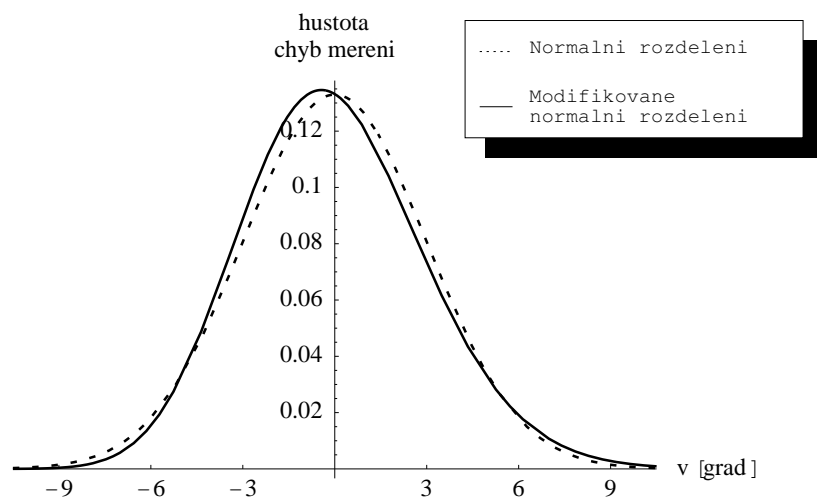
Modifikace rozdělení chyb měření. V posledním případě byla použita metoda navržená v této kapitole v oddíle 7.4. Přestože byly do Bayesova vzorce (203) dosazeny původní nelineární zprostředkující funkce a_j v (204), bylo možno parametry aposteriori hustoty opět určit MNČ.

$$\hat{H} = 317.8 \text{ m}, \quad \hat{\sigma} = 6.5 \text{ m}.$$

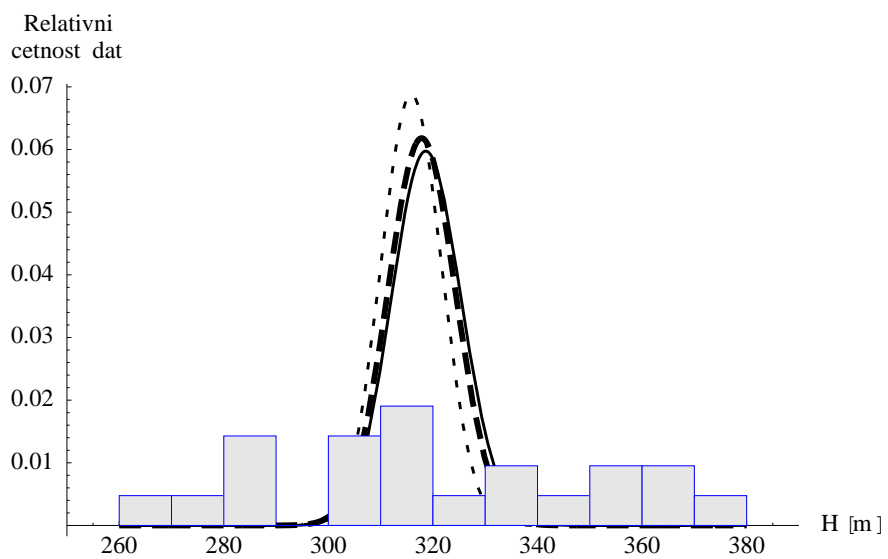
Její graf se nyní více přimyká ke grafu přesné aposteriori hustoty než ke Gaussově křivce získané standardní MNČ, jak ukazuje obrázek 8 (silná čárkovaná čára).

Modifikace původního rozdělení, kterou bylo nutno provést je skutečně jen mírná. Obě rozdělení jsou porovnány v obrázku 9.

Simulovaná data (202) jsou znázorněna formou histogramu na obrázku 10. V pozadí histogramu jsou pro srovnání vykresleny všechny tři aposteriori hustoty z tohoto příkladu.



Obrázek 9: Porovnání modifikovaného rozdělení chyb měření s původním normálním



Obrázek 10: Histogram simulovaných dat

8 Shrnutí

Bayesovský přístup představuje univerzální metodu zpracování experimentálních dat. Zahrnuje jako svůj speciální případ i obvykle používanou metodu nejmenších čtverců. Hlavními výhodami bayesovského přístupu jsou:

- možnost přesně určit rozdělení pravděpodobnosti určovaných veličin i v nelineárním případě,
- možnost zpracování měření s libovolným rozdělením pravděpodobnosti,
- možnost zahrnout do zpracovatelského procesu apriorní informace o určovaných veličinách.

Námětem této práce je využití těchto výhod při zpracování geodetických měření.

Hlavní překážkou využití prvních dvou výhod je velká výpočetní náročnost existujících metod nelineárního bayesovského odhadování i při malém porušení předpokladu linearit a normality. Byla proto navržena metoda, která platí pro mírně nelineární model a malou odlišnost rozdělení pravděpodobnosti chyb měření od normálního rozdělení. Podstata této metody spočívá v modifikaci analytického tvaru hustoty pravděpodobnosti chyb měření, která by nelinearit korigovala. Skutečný tvar hustoty pravděpodobnosti chyb měření přitom zůstane přibližně zachován. Modifikace rozdělení je navržena tak, aby nově navržené rozdělení mělo stejnou střední hodnotu a rozptyl jako původní rozdělení. Potom je možno použít standardních lineárních metod.

Výsledkem tohoto postupu je aposteriorní hustota, která se jen mírně liší od hustoty získané aplikací bayesovského přístupu na původní nelineární model při původním rozdělení pravděpodobnosti chyb měření. I přes tuto mírnou odlišnost lze výslednou aposteriorní hustotu považovat za spolehlivý odhad rozdělení pravděpodobnosti určovaných veličin, pokud připustíme hypotézu, že chyby měření mají modifikované rozdělení pravděpodobnosti. Mírou spolehlivosti tohoto odhadu může být hladina významnosti statistického testu, který nezamítne hypotézu o modifikovaném rozdělení pravděpodobnosti chyb měření.

Zmíněný postup je použitelný při vyrovnání geodetických měření v případech, kdy je chyba linearizace významná. Takové případy mohou nastat např. při nedostatečně přesném stanovení přibližných neznámých nebo při uvažování hrubých chyb měření. Umožňuje také aplikaci geodetických výpočetních postupů v jiných oborech, v nichž je přesnost měření úhlů a délek natolik nízká, že se chyba linearizace výrazně projeví (např. meteorologie, geologie, oceánografie apod.).

Poslední výhoda bayesovského přístupu (zahrnutí apriorních informací) může být využita k návrhu nových rozdělení s požadovanými vlastnostmi. Na základě Gaussova zdůvodnění normálního rozdělení byla navržena metodika odvození rozdělení pravděpodobnosti náhodné veličiny, jsou-li dány apriorní informace o jejích hodnotách a jejíž nejvěrohodnější odhad má předepsané vlastnosti. Jako příklad

aplikace této metodiky je odvozeno Pearsonovo rozdělení typu II, které lze použít pro rozdělení pravděpodobnosti chyb měření, neboť je blízké normálnímu (pro určité hodnoty jeho parametrů). Přitom bylo požadováno, aby velikost chyby měření byla omezena předem stanovenou hodnotou maximálně přípustné chyby, aby se nejvěrohodnější odhad souboru měření s tímto rozdělením příliš nelišil od aritmetického průměru a aby hustota pravděpodobnosti měla hladký tvar na celé reálné ose. Ve statistické literatuře jsou Pearsonova rozdělení vyvozována z jiných principů (na základě hypergeometrického rozdělení).

Využití Pearsonova rozdělení v bayesovské statistice je předmětem dalšího výzkumu. Rovněž řešení nelineárního modelu s modifikovaným rozdělením chyb měření vyžaduje další rozpracování, zejména při posuzování spolehlivosti aposteriorního rozdělení, měření odlišnosti modifikovaného rozdělení od původního rozdělení a stanovení vztahu těchto kritérií ke stupni nelinearity. Toto rozpracování bude možno provést z hlediska teorie informace. Je třeba také odůvodnit konvergenci navrhané aposteriorní hustoty ke skutečné aposteriorní hustotě získané přímo z Bayesovy věty bez modifikace rozdělení chyb měření a stanovit podmínky této konvergence.

Použitá literatura

- [And78] Jiří Anděl. *Matematická statistika*. SNTL, Praha, 1978.
- [AS70] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions*. Dover publications, Inc., 1970.
- [Bar87] Hans-Jochen Bartsch. *Matematické vzorce*. SNTL, Praha, 1987.
- [Bay63] Thomas Bayes. *Essay Towards Solving a Problem in the Doctrine of Chances*. (1702-1761), 1763. znovu publikováno v *Biometrika*, Vol. 45 (1958), pp 298–315.
- [Ber77] Karel Berka. *Měření; pojmy/teorie/problémy*. Academia, 1977.
- [Böh58] Josef Böhm. *Vyrovňovací počet*. Učební texty vysokých škol, ČVUT Praha, fakulta zeměměřická. SNTL, Praha, 1958.
- [Böh64] Josef Böhm. *Vyrovňovací počet*. SNTL, 1964.
- [Bos72] J. D. Bossler. *Bayesian inference in geodesy*. PhD thesis, The Ohio State University, 1972.
- [BRH90] Josef Böhm, Vladimír Radouch, and Miroslav Hampacher. *Teorie chyby a vyrovňovací počet*. Geodetický a kartografický podnik, Praha, 1990.
- [BS97] José M. Bernardo and Adrian F. M. Smith. *Bayesian Theory*. Wiley series in probability and mathematical statistics. John Wiley & Sons, 1997.
- [BT73] G. E. P. Box and G. C. Tiao. *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, Massachusetts, 1973.
- [BW88] Douglas M. Bates and Donald G. Watts. *Nonlinear Regression Analysis and Its Applications*. John Willey & Sons, 1988.
- [Dic68] James M. Dickey. Three multidimensional-integral identities with Bayesian applications. *The Annals of Mathematical Statistics*, 39(5):1615–1628, 1968.
- [EJ69] W. P. Elderton and N. L. Johnson. *Systems of Frequency Curves*. Cambridge University Press, 1969.
- [Erd54] Arthur Erdélyi. *Tables of integral transforms*. Bateman Manuscript Project, 1954.

- [ES95] Michael Evans and Tim Swartz. Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. *Statistical Science*, 10(3):254–272, 1995.
- [FMR90] Jean Pierre Florens, Michel Mouchart, and Jean-Marie Rolin. *Elements of Bayesian Statistics*, volume 134 of *Monographs and textbooks in pure and applied mathematics*. Marcel Dekker, New York, 1990.
- [Gau09] Carolo Friderico Gauss. *Theoria motus corporum coelestium in sectionibus conicis Solem ambientium*. Frid. Perthes et I. H. Besser, Hamburg, 1809. Ruský překlad: C. F. Gauss, Izbrannyje geodězičeskije sočiněnija, str. 92–96, Moskva, 1957.
- [Gle96] James Gleick. *Chaos. Vznik nové vědy*. Ando Publishing, 1996.
- [Ht94] Vladimír Havlena and Jan Štecha. *Moderní teorie řízení*. Vydavatelství ČVUT, 1994.
- [Huš85] Marie Hušková. *Bayesovské metody*. Univerzita Karlova, 1985.
- [Jan37] J. Janko. *Základy statistické indukce*. SUS, Praha, 1937.
- [Jar76] Vojtěch Jarník. *Diferenciální počet II*. Academia, Praha, 1976.
- [Jar84] Vojtěch Jarník. *Integrální počet II*. Academia, Praha, 1984.
- [Jos64] Marcel Josifko. *Matematická statistika*. MFF UK, Praha, 1964.
- [Kam65] E. Kamke. *Spravočnick po obyknověnnym diferenciaalnym uravněnijam*. Nauka, Moskva, 1965.
- [KKK82] Ludmila Kubáčková, Lubomír Kubáček, and Ján Kukuča. *Pravdepodobnost a štatistika v geodézii a geofyzike*. Veda, Bratislava, 1982.
- [Kla43] B. Kladio. *Měřické chyby a jejich vyrovnávání*. Jednota českých matematiků a fyziků, Praha, 1943.
- [Koc87] Karl-Rudolf Koch. Bayesian inference for variance components. *Manuscripta Geodaetica*, 12(4):309–313, 1987.
- [Koc90] Karl-Rudolf Koch. *Bayesian Inference with Geodetic Applications*, volume 31 of *Lecture Notes in Earth Sciences*. Springer-Verlag, 1990.

- [Kov90] Pavel Kovanic. Gnostická teorie neurčitých dat. doktorská disertační práce, březen 1990.
- [Kub83] Lubomír Kubáček. *Základy teorie odhadu*. Veda, Bratislava, 1983.
- [Kul96] Rudolf Kulhavý. *Recursive Nonlinear Estimation*, volume 216 of *Lecture notes in Control and Information Sciences*. Springer-Verlag, 1996.
- [Leb56] N. N. Lebeděv. *Speciální funkce a jejich použití*. SNTL, Praha, 1956.
- [Lin65a] D. V. Lindley. *Introduction to Probability and Statistics from a Bayesian Viewpoint*, volume Part 1: Probability. Cambridge, University Press, 1965.
- [Lin65b] D. V. Lindley. *Introduction to Probability and Statistics from a Bayesian Viewpoint*, volume Part 2: Inference. Cambridge, University Press, 1965.
- [LM81] Jiří Likeš and Josef Machek. *Počet pravděpodobnosti*. SNTL, Praha, 1981.
- [LM88] Jiří Likeš and Josef Machek. *Matematická statistika*. SNTL, Praha, 1988.
- [Mac80] Josef Machek. *Teorie odhadu*. Universita Karlova, Praha, 1980.
- [Mai74] L. E. Maistrov. *Probability Theory. A Historical Sketch*. Probability and Mathematical Statistics. Academic Press, 1974.
- [Mar72] K. V. Mardia. *Statistics of Directional Data*. Academic Press, London, New York, 1972. též ruský překlad: Statistčeskij analiz uglovych nabljuděnij, Nauka, Moskva 1978.
- [Mar76] Solomon Marcus. *Matematická analýza čtená podruhé*. Academia, 1976.
- [Nag76] Jozef Nagy. *Vybrané partie z moderní matematiky*. SNTL, Praha, teoretická knihnice inženýra edition, 1976.
- [Nk77] Jiří Nečas and kol. *Aplikovaná matematika*, volume I., II. of *Oborové encyklopedie*. SNTL, 1977.

- [NNV85] J. Nagy, E. Nováková, and M. Vacek. *Lebesgueova míra a integrál*, volume XXII of *Matematika pro vysoké školy technické*. SNTL, 1985.
- [Nov90] Vilém Novák. *Fuzzy množiny a jejich aplikace*. Number 23 in *matematický seminář*. SNTL, 1990.
- [Nov92] Mirko Novák. *Neuronové sítě a neuropočítače*. Výběr, 1992.
- [OK94] Ziqiang Ou and Karl-Rudolf Koch. Analytical expressions for Bayes estimates of variance components. *Manuscripta Geodaetica*, 19(5):284–293, 1994.
- [PP95] C. Platt and Z. Paprzycki. About some properties of the Romanowski distributions. *Manuscripta Geodaetica*, 20(6):408–412, 1995.
- [Páz93] Andrej Pázman. *Nonlinear Statistical Models*. Kluwer Academic Publishers, 1993.
- [Rao78] Radhakrishna C. Rao. *Lineární metody statistické indukce a jejich aplikace*. Academia, Praha, 1978.
- [RG65] M. Romanowski and E. Green. Practical applications of the modified normal distribution. *Bulletin Géodésique*, (76):155–171, 1965.
- [Rén72] Alfréd Rényi. *Teorie pravděpodobnosti*. Academia, Praha, 1972.
- [Rob94] Christian P. Robert. *The Bayesian Choice*. Springer Texts in Statistics. Springer-Verlag, 1994.
- [Rom64] M. Romanowski. On the normal law of errors. *Bulletin Géodésique*, (73):195–216, 1964.
- [Sch72] Laurent Schwartz. *Matematické metody ve fyzice*. SNTL, 1972.
- [Sch87] B. Schaffrin. Approximating the Bayesian estimate of the standard deviation in a linear model. *Bulletin Geodesique*, 61:276–280, 1987.
- [SG85] C. Ray Smith and W. T. Grandy Jr., editors. *Maximum-Entropy and Bayesian Methods in Inverse Problems*. D. Riedel Publishing Company, 1985.
- [Smi91] Adrian F. M. Smith. Bayesian computational methods. *Phil. Trans. R. Soc. London*, 337:369–386, 1991.

- [SSS⁺85] A. F. M. Smith, A. M. Skene, J. E. H. Shaw, J. C. Naylor, and M. Dransfield. The implementation of the Bayesian paradigm. *Communications in Statistics*, 14(5):1079–1102, 1985.
- [SSSN87] A. F. M. Smith, A. M. Skene, J. E. H. Shaw, and J. C. Naylor. Progress with numerical and graphical methods for practical Bayesian statistics. *The Statistician*, 36:75–82, 1987.
- [Teu90] P. J. T. Teunissen. Nonlinear least squares. *Manuscripta Geodaetica*, 15:137–150, 1990.
- [TK86] Luke Tierny and Joseph B. Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81:82–86, 1986.
- [TKK89a] Luke Tierny, Robert E. Kass, and Joseph B. Kadane. Approximate marginal densities of nonlinear functions. *Biometrika*, 76(3):425–433, 1989.
- [TKK89b] Luke Tierny, Robert E. Kass, and Joseph B. Kadane. Fully exponential laplace approximation to expectation and variances of non-positive functions. *Journal of the American Statistical Association*, 84(407):710–716, 1989.
- [Tut78] V. N. Tutubalin. *Teorie pravděpodobnosti*. Teoretická knižnice inženýra. SNTL, 1978.
- [Vie96] Reinhard Viertl. *Statistical Methods for Non-Precise Data*. CRC Press, 1996.
- [Vyš48] Jan Vyšín. *O nekonečných řadách*. Jednota československých matematiků a fyziků, Praha, 1948.