Pokroky matematiky, fyziky a astronomie

Jitka Zichová

Thorvald Nicolai Thiele - dánský statistik a aktuár

Pokroky matematiky, fyziky a astronomie, Vol. 55 (2010), No. 1, 30--42

Persistent URL: http://dml.cz/dmlcz/141935

Thorvald Nicolai Thiele — dánský statistik a aktuár

Jitka Zichová, Praha

Životní zápasy, které podstupujeme,
nás občas nutí poradit se s věštci.
Ale moderní proroctví musí být vědecká . . .
T. N. Thiele, 1903

I. Život a dílo

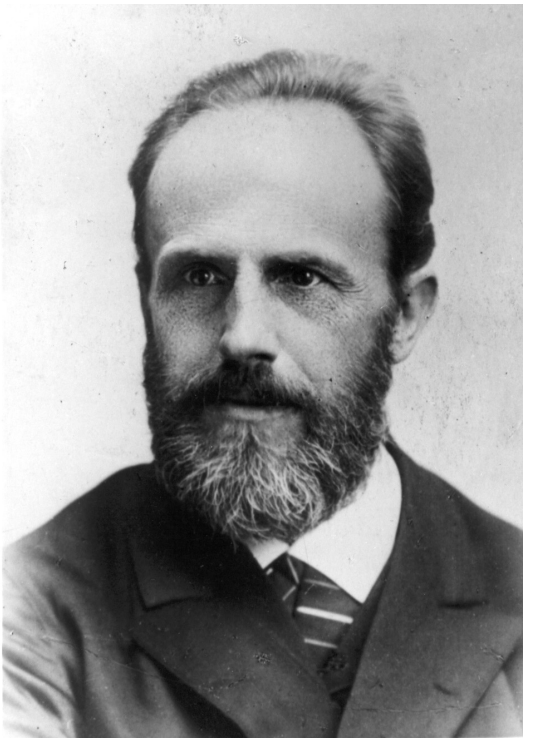
V letošním roce si připomínáme sté výročí úmrtí jednoho z průkopníků matematické statistiky a pojistné matematiky. Je jím dánský astronom a matematik Thorvald Nicolai Thiele. O uznání, kterého se mu dostává v rodné zemi, svědčí například existence Thieleho centra (viz [27]), mezinárodního vědeckého pracoviště univerzity v Aarhusu. Bylo založeno v roce 2004 a zaměřuje se na výzkum v oblasti stochastického modelování, jehož základy pomáhal T. N. Thiele budovat. Centrum navštěvují i čeští odborníci, a proto si ten, po němž bylo pojmenováno, jistě zaslouží vzpomínku v naší zemi.

Thorvald Nicolai Thiele se narodil na Štědrý den roku 1838 v Kodani v rodině, jejíž příslušníci byli tiskaři a výrobci fyzikálních přístrojů. Vyrůstal v intelektuálním a kulturně bohatém prostředí. Jeho otec, Just Mathias Thiele (1795–1874), byl velmi vzdělaný a uměnímilovný muž. Působil jako knihovník dánského krále Christiana VIII., byl ředitelem Královské sbírky tisků a tajemníkem Královské akademie krásných umění. Psal dramata, poezii a zajímal se o lidovou tvorbu. Měl přátele v uměleckých kruzích, jedním z nich byl slavný sochař Bertel Thorvaldsen (1770–1844). Ten se stal kmotrem Justova syna, který na jeho počest dostal jméno Thorvald. J. M. Thiele napsal o svém příteli Thorvaldsenovi knihu, která se záhy dočkala anglického překladu [15].

T. N. Thiele studoval astronomii na univerzitě v Kodani, řádné studium ukončil v roce 1860. V letech 1860–1870 byl vědeckým asistentem u profesora Heinricha Louise d'Arresta na astronomické observatoři zmíněné univerzity, v roce 1866 obhájil doktorát prací o oběžných drahách dvojhvězd systému Gamma Virginis [16]. V roce 1875 byl ustanoven profesorem astronomie a ředitelem observatoře. Tyto funkce zastával až do odchodu do penze v roce 1907. Na univerzitě rovněž vyučoval a v letech 1900 až 1906 byl jejím rektorem. Byl postižen silným astigmatismem, a proto nemohl konat

RNDr. Jitka Zichová, Dr., Katedra pravděpodobnosti a matematické statistiky, Matematicko-fyzikální fakulta UK, Sokolovská 83, 186 75 Praha 8, zichova@karlin.mff.cuni.cz

astronomická pozorování, obrátil tedy svou pozornost k matematice. Podle doložených pramenů byl ke konci života téměř slepý. Kromě rozsáhlých odborných zájmů lze zmínit i jeho zálibu ve hře v šachy, kterou výtečně ovládal jako aktivní člen prvního dánského šachového klubu existujícího od roku 1865. V roce 1867 se oženil s Marií Martine Trolle. Měli šest dětí, dvě zemřely v prvním roce života. V roce 1889 opustila Thieleho ve věku 48 let i manželka.



Thorvald Nicolai Thiele

Thiele byl velmi aktivní a měl organizační schopnosti. Na univerzitě v Kodani neúspěšně usiloval o zřízení profesorského místa pro výpočetní matematiku. Jako první odborník v Dánsku zakoupil a používal počítací stroj. Má zásluhu na vzniku dvou vědeckých společností. V roce 1873 spolu s H. G. Zeuthenem a P. C. Petersenem zakládá Dánskou matematickou společnost a v roce 1901 je ustavena z jeho iniciativy Dánská aktuárská společnost, jíž předsedá až do své smrti. Do oblasti pojišťovnictví však vstoupil o třicet let dříve. V letech 1870–1871 se věnuje vypracování pojistně matematické koncepce životní pojišťovny Hafnia, která je jako první dánská soukromá instituce tohoto typu založena v roce 1872. Thiele byl do konce života jejím matematickým ředitelem, dnes bychom řekli vrchním aktuárem či odpovědným pojistným matematikem. Byl také členem Dánské královské společnosti od roku 1879, členem korespondentem Institutu aktuárů v Londýně od roku 1895 a členem rady Stálého výboru mezinárodních kongresů aktuárů od roku 1895. Zemřel 26. září 1910 v Kodani.

Seznam jeho publikací obsahuje téměř padesát titulů, z toho dvě monografie věnované statistické inferenci [19], [20] a jednu na téma interpolační teorie [25]. Další jsou články o astronomii, statistice, pojistné matematice a numerické analýze. Většina z nich byla inspirována praktickými problémy. Kniha [20] byla v roce 1903 přeložena do angličtiny [23] a získala mezinárodní věhlas. V roce 1931 byla v plném rozsahu přetištěna v časopise Annals of Mathematical Statistics [26]. Kniha [19] čekala na anglický překlad až do roku 2002. Provedl jej profesor statistiky na univerzitě v dánském Aalborgu Steffen L. Lauritzen, který dnes působí jako vedoucí katedry statistiky univerzity v Oxfordu. Anglickou verzi Thieleho stěžejního díla zařadil do monografie [13].

Pamětní medaile k Thieleho sedmdesátinám

Thieleho kolega a přítel J. P. Gram uvádí v nekrologu [7], že T. N. Thiele do hloubky promýšlel všechna témata, na kterých pracoval. Preferoval návrhy vlastních řešení před studiem metod a postupů jiných autorů. Jeho význam spočívá v originálních myšlenkách, kterými mnohdy předběhl svou dobu. Bohužel je nedokázal vždy srozumitelně písemně a ústně formulovat. To mohlo způsobit, že neměl mnoho následovníků z řad svých studentů, že jeho publikační činnost nebyla rozsáhlejší a že se dříve nedočkal uznání, jehož by si býval zasloužil.

V dalším textu se zaměříme na Thieleho odborný přínos pro pojistnou matematiku a statistiku. V pojistné matematice formuloval diferenciální rovnici pro rezervu pojistného v závislosti na úmrtnosti a úrokových sazbách a pracoval na problematice modelování úmrtnosti a úmrtnostních tabulek. Ve statistice se věnoval teorii pravděpodobnostních rozdělení, lineárního modelu, analýzy rozptylu, filtrování a Gramových–Charlierových řad. Zabýval se i procesem Brownova pohybu, nejvíce jej však proslavila jím vytvořená teorie kumulantů. Jeden z velikánů matematické statistiky první poloviny 20. století R. A. Fisher jej ve svém díle [5] zařadil k osobnostem, které nejvíce přispěly k rozvoji této vědní disciplíny, po bok Thomase Bayese, P. S. Laplace, C. F. Gausse, Karla Pearsona a Williama Sealy Gosseta publikujícího pod pseudonymem Student.

II. Pojistná matematika

V roce 1871 sestrojil Thiele úmrtnostní tabulku na základě dat dánské státní životní pojišťovny. Tyto tabulky jsou jedním ze základních nástrojů pro výpočty prováděné v rámci životního pojištění. Obsahují pravděpodobnosti úmrtí ve věku x, počty zemřelých ve věku x a další veličiny, $x = 0, 1, \ldots, \omega$, kde ω představuje věk ω let a více. Závislost úmrtnosti na věku vyjadřuje funkce zvaná intenzita úmrtnosti. Jejímu modelování jsou věnovány práce [22] a [24]. Thiele vycházel z následující hypotézy: příčiny smrti lze klasifikovat do čtyř skupin. Tři z nich obsahují nejčastější důvody úmrtí v dětském, dospělém a seniorském věku. Čtvrtá skupina zahrnuje příčiny smrti vyskytující se zhruba stejnou měrou ve všech věkových kategoriích, ty lze považovat za nevýznamné. Intenzitu úmrtnosti můžeme pak vyjádřit pomocí sčítanců odpovídajících prvním třem skupinám vztahem

$$\mu(x) = a_1 \exp(-b_1 x) + a_2 \exp\Big(-b_2^2 \frac{(x-c)^2}{2}\Big) + a_3 \exp(-b_3 x),$$

kde x je věk a $a_1, a_2, a_3, b_1, b_2, b_3, c$ jsou parametry modelu, přičemž Thiele popisuje metody jejich odhadu. Idea rozkladu intenzity úmrtnosti podle příčin smrti byla později použita mnoha autory, zmiňme například text [2].

Největší Thieleho zásluhou v oblasti pojišťovnictví je diferenciální rovnice pro netto rezervu pojistného v životním pojištění. Je jedním ze základních vztahů v matematice životního pojištění se spojitým časem. Abychom ji mohli formulovat, zaveďme nejprve některé pojmy z finanční a pojistné matematiky (viz např. [4]). Efektivní úroková míra i představuje zhodnocení částky 1 za jeden rok. To je stejně velké jako zhodnocení, které poskytne nominální úroková míra $i_{(p)}$, připisuje-li se úrok výše $i_{(p)}/p$ každou p-tinu roku. Zmíněné úrokové míry jsou svázány vztahem

$$1 + i = \left(1 + \frac{i_{(p)}}{p}\right)^p.$$

Limitou nominální úrokové míry $i_{(p)}$ při frekvenci úročení p rostoucí nade všechny meze je intenzita úročení

$$\delta = \lim_{p\to\infty} i_{(p)} = \ln(1 + i).$$

Hovoříme pak o spojitém úročení, což znamená průběžné zhodnocování dané částky.

Finanční tok je posloupnost plateb $P_1, \ldots, P_n$ v časových okamžicích $1, \ldots, n$. Spojitý finanční tok je charakterizován intenzitou plateb $\pi(t)$ ve spojitém čase. V intervalu $(t, t + dt)$ se realizuje platba souhrnné výše $\pi(t)dt$.

Pojistné představují platby, které hradí klient pojišťovně, pojistné plnění je poté v případě pojistné události vyplaceno pojišťovnou klientovi. Vzhledem k tomu, že okamžik vzniku pojistné události je nejistý, vytvářejí pojišťovny rezervy za účelem plnění budoucích závazků.

Symbolem ${}_tp_x$ označme pravděpodobnost, že se jedinec, který je naživu ve věku x, dožije věku $x + t$. Intenzita úmrtnosti ve věku $x + t$ za podmínky, že se jedinec dožil

věku x, je definována předpisem

$$\mu_{x+t} = -\frac{\mathrm{d}}{\mathrm{d}t} \ln({}_tp_x).$$

Uvažujme klienta ve věku x let, který uzavřel pojištění pro případ smrti. Zavázal se hradit pojišťovně spojité pojistné s konstantní intenzitou plateb π, pojistná částka S bude vyplacena v případě jeho úmrtí. Thieleho diferenciální rovnici pro rezervu V_t v čase t let od sjednání pojištění lze zapsat ve tvaru

$$\frac{\mathrm{d}}{\mathrm{d}t} V_t = \pi + \delta V_t - \mu_{x+t}(S - V_t). \tag{2.1}$$

Thiele rovnici (2.1) nikdy nepublikoval, ale seznámil s ní J. P. Grama, který ji uveřejnil v textu [7]. Později byla zařazena do prací [11], [3], [8]. Thieleho činnost v oblasti pojistné matematiky je zmíněna například v článku [10]. Dodejme, že výpočty rezerv pro různé typy pojištění jsou dnes jedním z klíčových úkolů pojistné matematiky a tvorba rezerv je pro pojišťovny předepsána zákonem.

III. Matematická statistika

III. 1. Kumulanty

V 19. století hrálo dominantní roli při modelování náhodných dějů normální rozdělení definované Gaussovou křivkou hustoty, která má v normované podobě tvar

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big), \quad x \in \mathbb{R}. \tag{3.1}$$

Připomeňme, že náhodná veličina X mající pravděpodobnostní rozdělení s hustotou $f = f(x)$ nabude hodnoty z intervalu $[a, b]$ s pravděpodobností

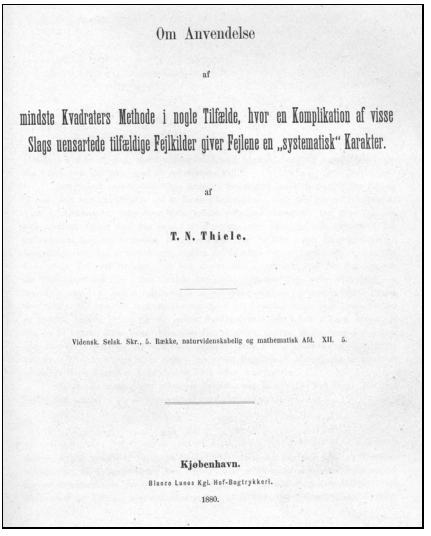
$$P(a \leqq X \leqq b) = \int_a^b f(x)\,\mathrm{d}x, \quad -\infty \leqq a < b \leqq \infty.$$

Matematici, kteří pracovali s daty z oblasti ekonomie a pojišťovnictví, si však začali uvědomovat, že veličiny tam používané mají rozdělení s nesymetrickými hustotami. Proto vznikla potřeba vybudovat širší teorii pravděpodobnostních rozdělení, k čemuž přispěl i T. N. Thiele. Rozdělení, z něhož pocházejí pozorovaná data, charakterizuje posloupnost odhadnutých momentů. *Střední hodnota* náhodné veličiny X s hustotou f je

$$\mathrm{E}X = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x,$$

r-tý moment je

$$\mu'_r = \mathrm{E}X^r = \int_{-\infty}^{\infty} x^r f(x)\,\mathrm{d}x,$$


Om Anvendelse

af

mindste Kvadraters Methode i nogle Tilfælde, hvor en Komplikation af visse Slags uensartede tilfældige Fejlkilder giver Fejlene en „systematisk" Karakter.

af

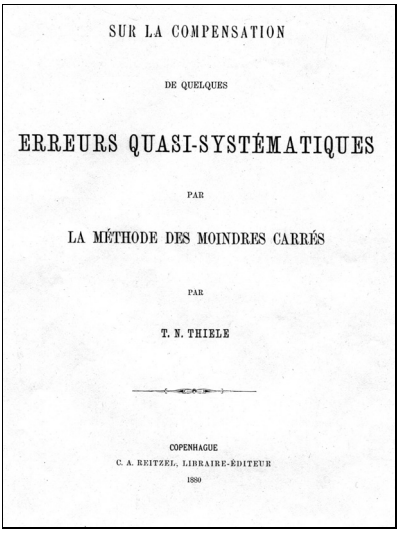
T. N. Thiele.

Vidensk. Selsk. Skr., 5. Række, naturvidenskabelig og mathematisk Afd. XII. 5.

Kjøbenhavn.

Bianco Lunos Kgl. Hof-Bogtrykkeri.

1880.


SUR LA COMPENSATION

DE QUELQUES

ERREURS QUASI-SYSTÉMATIQUES

PAR

LA MÉTHODE DES MOINDRES CARRÉS

PAR

T. N. THIELE

COPENHAGUE

C. A. REITZEL, LIBRAIRE-ÉDITEUR

1880

Titulní strany prací [17], [18]

r-tý centrální moment je

$$\mu_r = \mathrm{E}(X - \mathrm{E}X)^r,$$

μ_2 se nazývá *rozptyl*. Základním odhadem střední hodnoty $\mathrm{E}X$ je aritmetický průměr pozorování $x_1, \ldots, x_n$ náhodné veličiny X, průměrem mocnin $x_1, \ldots, x_n$ odhadujeme momenty vyšších řádů.

Vzhledem k tomu, že odhadnuté momenty rychle rostou s rostoucím r, je vhodné charakterizovat rozdělení jinou posloupností funkcí, která nebude rychle rostoucí. To přivádí Thieleho k zavedení tzv. polovičních invariantů. Dnes se používá název kumulanty zmíněný poprvé v práci [6]. Thiele v knize [19] definuje kumulanty κ_i, $i = 1, 2, \ldots$, pomocí momentů jako koeficienty vystupující v rekurentním vyjádření

$$\mu'_{r+1} = \sum_{i=0}^{r} \binom{r}{i} \mu'_{r-i} \kappa_{i+1}, \quad r = 0, 1, \ldots \tag{3.2}$$

Ukazuje, že $\kappa_1 = \mu'_1$ (střední hodnota), $\kappa_2 = \mu_2$ (rozptyl), $\kappa_3 = \mu_3$, $\kappa_4 = \mu_4 - 3\mu_2^2$ a že třetí a čtvrtý kumulant charakterizují zešikmení a zašpičatění hustoty. Dále pracuje mimo jiné s Gramovým–Charlierovým rozdělením typu A, jehož hustotu lze zapsat ve tvaru

$$f(x) = k_0\varphi(x) - k_1\varphi^{(1)}(x) + k_2\varphi^{(2)}(x)\frac{1}{2!} - k_3\varphi^{(3)}(x)\frac{1}{3!} + k_4\varphi^{(4)}(x)\frac{1}{4!} - \ldots,$$

kde $\varphi(x)$ je dána vztahem (3.1) a horní index značí derivaci příslušného řádu. K určení koeficientů $k_0, k_1, \ldots$ používá kumulanty. Popisuje také způsob odhadu kumulantů z pozorovaných dat.

V článku [21] pak formuluje obecnou definici kumulantů náhodné veličiny s hustotou f, která je dána rovnostmi

$$\exp\Big(\kappa_1 t + \kappa_2 \frac{t^2}{2!} + \dots\Big) = 1 + \mu_1' t + \mu_2' \frac{t^2}{2!} + \dots = \int_{-\infty}^{\infty} \mathrm{e}^{tx} f(x)\,\mathrm{d}x. \tag{3.3}$$

Derivováním (3.3) a porovnáním koeficientů u mocnin t dostává rekurentní předpis (3.2). Pro dvě náhodné veličiny s hustotami f, f^* a s kumulanty κ_i, κ_i^*, $i = 1, 2, \dots$, pro něž platí $\kappa_1 = \kappa_1^*$ (rovnost středních hodnot) a $\kappa_2 = \kappa_2^*$ (rovnost rozptylů), nakonec uvádí vztah

$$f(x) = \exp\Big(-(\kappa_3 - \kappa_3^*)\frac{f^{*(3)}(x)}{3!} + (\kappa_4 - \kappa_4^*)\frac{f^{*(4)}(x)}{4!} - \dots\Big)$$

a diskutuje konvergenci tohoto rozvoje.

III. 2. Brownův pohyb a metoda nejmenších čtverců

První Thieleho práce věnovaná problematice metody nejmenších čtverců vyšla v roce 1880 v dánské a francouzské verzi (viz [17], [18]), což svědčí o důležitosti, které jí autor přikládal. Je v ní vyšetřován model pro časovou řadu Brownova pohybu, inspirovaný problémem z oblasti astronomické geodézie, a to určením vzdálenosti z Kodaně do švédského Lundu. Thiele navrhuje rekurentní postup pro odhad parametrů modelu. Text byl ve své době odbornou veřejností přehlížen. Jedním z důvodů mohla být jeho malá srozumitelnost, autor předpokládá u čtenáře hlubokou znalost Gaussovy metody nejmenších čtverců (viz [1]). V pozdější době se rozvoj statistiky přesunul do Anglie a USA, kde patrně nikdo neměl povědomí o přínosu dánského astronoma z roku 1880. Thieleho metoda tak byla připomenuta pouze v Helmertově knize [9], využívané po dlouhá léta jako základní učebnice statistiky pro geodety. Z dnešního pohledu je zajímavý Thieleho přístup spočívající ve velmi podrobném vyšetřování jednoho konkrétního modelu časové řady. V současné analýze časových řad se spíše uplatňuje studium širokých tříd modelů, někdy i bez přímé souvislosti s praxí a způsobem získání dat.

Podívejme se na základní výsledky v práci [17]. Autor uvažuje posloupnost náhodných veličin $Z_0, Z_1, \dots$ Přírůstky $Z_{i+1} - Z_i$ jsou nezávislé a mají normální rozdělení s nulovou střední hodnotou a s rozptylem $\sigma_i^2 = \sigma^2(t_{i+1} - t_i)$. Jedná se tedy o stochastický proces Brownova pohybu (viz např. [14]). Jeho realizace $z_0, z_1, \dots z_n$ v časech $t_0, t_1, \dots t_n$ nejsou pozorovatelné a lze je nahradit pozorováními $y_0, y_1, \dots y_n$ veličin Y_i takových, že

$$Y_i = Z_i + e_i. \tag{3.4}$$

Náhodné chyby e_i jsou nezávislé na Z_i a mají normální rozdělení s nulovou střední hodnotou a s rozptylem ω^2. Podmíněné rozdělení veličin Y_i při pevném $Z_i = z_i$ je normální, má střední hodnotu z_i a rozptyl ω^2. Je třeba odhadnout z_i, $i = 0, 1, \dots n$, a parametry rozptylu σ^2, ω^2.

Thiele řeší problém odhadu středních hodnot minimalizací součtu

$$\sum_{i=0}^{n-1} \frac{(z_{i+1}-z_i)^2}{\sigma^2 k_i^2} + \sum_{i=0}^{n} \frac{(y_i-z_i)^2}{\omega^2}, \tag{3.5}$$

kde $k_i = \sqrt{t_{i+1}-t_i}$. Předpokládáme-li známé σ^2 a ω^2, dostáváme z (3.5) tzv. soustavu normálních rovnic

$$\begin{aligned} y_0\,\omega^{-2} &= \hat{z}_0(\omega^{-2}+\sigma_0^{-2}) - \hat{z}_1\sigma_0^{-2}\,, \\ y_i\,\omega^{-2} &= -\hat{z}_{i-1}\sigma_{i-1}^{-2} + \hat{z}_i(\sigma_{i-1}^{-2}+\omega^{-2}+\sigma_i^{-2}) - \hat{z}_{i+1}\sigma_i^{-2},\ i=2,\dots,n-1\,, \\ y_n\,\omega^{-2} &= -\hat{z}_{n-1}\sigma_{n-1}^{-2} + \hat{z}_n(\sigma_{n-1}^{-2}+\omega^{-2}) \end{aligned} \tag{3.6}$$

o neznámých $\hat{z}_0, \hat{z}_1, \dots, \hat{z}_n$. K jejímu řešení navrhuje Thiele rekurentní postup dnes známý jako Kalmanův filtr (viz [12]) a ilustruje jej na numerickém příkladu se 74 pozorováními.

Dále se zabývá odhadováním σ^2 a ω^2. Odhady lze získat z kvadratických forem

$$Q_1 = \sum_{i=0}^{n-1} \frac{1}{t_{i+1}-t_i}(\hat{z}_{i+1}-\hat{z}_i)^2, \quad Q_2 = \sum_{i=0}^{n}(y_i-\hat{z}_i)^2$$

s hodnotami $\hat{z}_i$ získanými řešením soustavy (3.6) při použití vhodných počátečních hodnot parametrů σ^2, ω^2. Konkrétně Thiele navrhuje tento postup:

$$\hat{\sigma}^2 = \frac{Q_1}{f_1}, \quad \hat{\omega}^2 = \frac{Q_2}{f_2}, \tag{3.7}$$

kde

$$\begin{aligned} f_1 &= n - \sum_{i=0}^{n-1} \sigma_i^{-2}\,\mathrm{E}[(Z_{i+1}-Z_i)-(\hat{z}_{i+1}-\hat{z}_i)]^2, \\ f_2 &= n+1 - \sum_{i=0}^{n} \omega^{-2}\,\mathrm{E}(Z_i-\hat{z}_i)^2. \end{aligned}$$

V textu [17] je dán návod na výpočet $\mathrm{E}[(Z_{i+1}-Z_i)-(\hat{z}_{i+1}-\hat{z}_i)]^2$ a $\mathrm{E}(Z_i-\hat{z}_i)^2$. S odhady (3.7) se znovu řeší soustava (3.6) a určí se další iterace odhadů $\hat{\sigma}^2$, $\hat{\omega}^2$. Procedura se opakuje tak dlouho, až se získají dostatečně stabilní hodnoty odhadů parametrů rozptylu, podle Thieleho k tomu stačí tři až čtyři kroky iteračního algoritmu.

V závěru zmíněného textu rozšiřuje Thiele model (3.4) přidáním regresního členu na tvar

$$Y_i = Z_i + \sum_{k=1}^{r} \alpha_k f_k(t_i) + e_i,$$

kde $f_k(t_i)$ jsou známé funkce a α_k neznámé parametry, a uvádí rekurentní odhadové procedury.

III. 3. Lineární model

Lineární model pro náhodný vektor Y je definován předpisem

$$\mathrm{E}Y = \mu = X\beta, \tag{3.8}$$

kde X je pevná matice typu $n \times m$ s hodností $m < n$, β je m-rozměrný vektor neznámých parametrů a rozptyl složek $Y_1, \ldots, Y_n$ vektoru Y nezávisí na β (viz [1]). Významným přínosem Thieleho pro teorii lineárního modelu je formulace tzv. kanonické formy lineární hypotézy v práci [19]. Nechť jsou složky $Z_1, \ldots, Z_n$ náhodného vektoru Z nezávislé normálně rozdělené se středními hodnotami $\mathrm{E}Z_i = \eta_{i0}$, $i = 1, \ldots, n-m$ a $\mathrm{E}Z_i = \eta_i$, $i = n-m+1, \ldots n$, přičemž první sada středních hodnot je známá, druhou sadu neznáme. Předpokládejme stejný neznámý rozptyl σ_Z^2 u všech veličin $Z_1, \ldots, Z_n$ a mějme k dispozici jejich pozorování $z_1, \ldots, z_n$. Parametr σ_Z^2 můžeme odhadnout podle vzorce

$$s_Z^2 = \frac{1}{n-m} \sum_{i=1}^{n-m} (z_i - \eta_{i0})^2.$$

Za odhady středních hodnot η_i, $i = n-m+1, \ldots n$ lze vzít pozorování $z_{n-m+1}, \ldots, z_n$. Thiele se ve zmíněné knize dále zabývá problémem převedení lineárního modelu (3.8) s nezávislými normálně rozdělenými veličinami $Y_1, \ldots, Y_n$ s rozptylem σ_Y^2 do výše popsané kanonické formy. Uvedeme na tomto místě jeho úvahy v dnes používaném maticovém značení.

Rozdělme vektor μ na dva podvektory délek $n-m$ a m tak, že

$$\mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} = \begin{pmatrix} X^{(1)}\,\beta \\ X^{(2)}\,\beta \end{pmatrix} = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \cdot \beta, \quad \text{kde} \quad X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$$

a čtvercová matice $X^{(2)}$ má plnou hodnost m. Definujme lineární transformaci

$$Z = \begin{pmatrix} Z^{(1)} \\ Z^{(2)} \end{pmatrix} = \begin{pmatrix} A^T Y \\ B^T Y \end{pmatrix} = \begin{pmatrix} A^T \\ B^T \end{pmatrix} \cdot Y, \tag{3.9}$$

kde

$$A^T = \left(I_{n-m}, -X^{(1)} {X^{(2)}}^{-1} \right) \quad \text{a} \quad B = X,$$

I_{n-m} je jednotková matice rozměru $n-m$. Zřejmě je

$$\beta = {X^{(2)}}^{-1} \mu^{(2)}, \quad \mu^{(1)} = X^{(1)} {X^{(2)}}^{-1} \mu^{(2)}$$

a odtud plyne

$$\mathrm{E}Z^{(1)} = A^T \mathrm{E}Y = A^T \mu = 0.$$

Transformací (3.9) vektoru Y jsme přešli ke kanonickému modelu pro vektor Z, v němž $\mathrm{E}Z_i = \eta_{i0} = 0$ pro $i = 1, \ldots, n-m$.

Odhad střední hodnoty $\mathrm{E}Z^{(2)} = X^T\mathrm{E}Y = X^TX\beta$ metodou nejmenších čtverců je X^Ty při pozorované hodnotě y náhodného vektoru Y. K získání odhadu vektoru parametrů β máme soustavu normálních rovnic

$$X^TX\beta = X^Ty. \tag{3.10}$$

Při použití Gaussova algoritmu k nalezení jejího řešení násobíme (3.10) dolní trojúhelníkovou maticí G s vlastností $GX^TXG^T = D = \mathrm{diag}\{d_1, \ldots, d_m\}$. Dostáváme

$$GX^TX\beta = GX^Ty. \tag{3.11}$$

Složky U_i náhodného m-rozměrného vektoru $U = GX^TY$ jsou nezávislé náhodné veličiny s rozptyly $\sigma_Y^2 d_i$. K dispozici máme vektor jejich pozorování $u = GX^Ty$. Thiele zavádí nový vektor parametrů λ, pro který platí $\beta = G^T\lambda$. Soustavu (3.11) pak lze psát ve tvaru

$$D\lambda = u$$

s řešením

$$\hat{\lambda} = D^{-1}u = D^{-1}GX^Ty,$$

přičemž složky náhodného vektoru $D^{-1}U$ jsou nezávislé s rozptyly σ_Y^2/d_i. V publikaci [19] Thiele odvozuje vlastnosti odhadu $\hat{\beta} = G^T\hat{\lambda}$ a odhadu $X\hat{\beta}$ střední hodnoty $\mathrm{E}Y$ pomocí vlastností $\hat{\lambda}$. Zabývá se též zobecněními modelu, například případem matice X, která nemá plnou hodnost m, což vede k soustavě normálních rovnic s nejednoznačným řešením.

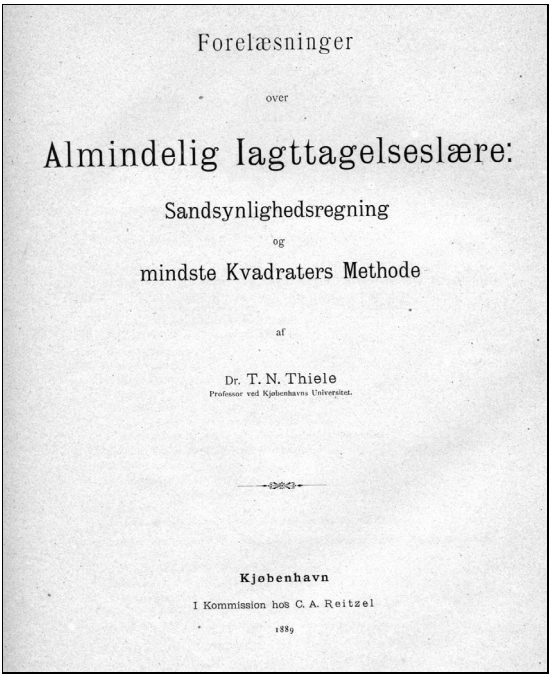
III. 4. Analýza rozptylu

Základním modelem analýzy rozptylu (viz [1]) je jednoduché třídění umožňující vyšetřit vliv jednoho faktoru na nějakou měřenou veličinu. Sledovaný faktor má k úrovní, posuzujeme tedy například, zda typ použitého hnojiva ovlivní hektarový výnos určité plodiny. Aplikujeme k druhů hnojiva, i-tý druh na n_i pokusných polích. Nechť Y_{ij} označuje hektarový výnos při i-tém druhu hnojiva na j-tém poli a předpokládejme, že veličiny Y_{ij} jsou nezávislé normálně rozdělené s rozptylem σ^2. Celkový počet $n = n_1 + \ldots + n_k$ pozorování lze uspořádat do k výběrů

$$\begin{aligned} &y_{11}, \ldots, y_{1n_1}, \\ &y_{21}, \ldots, y_{2n_2}, \\ &\quad\vdots \\ &y_{k1}, \ldots, y_{kn_k}. \end{aligned}$$

Střední hektarový výnos je

$$\mathrm{E}Y_{ij} = \mu_i, \quad i = 1, \ldots, k,\ j = 1, \ldots, n_i,$$


Forelæsninger

over

Almindelig Iagttagelseslære:

Sandsynlighedsregning

og

mindste Kvadraters Methode

af

Dr. T. N. Thiele

Professor ved Kjøbenhavns Universitet.

Kjøbenhavn

I Kommission hos C. A. Reitzel

1889

Titulní strana práce [19]

průměrný výnos ze všech polí je

$$\overline{y} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij},$$

výběrové průměry jsou

$$\overline{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}, \quad i = 1, \ldots, k.$$

Variabilitu měřené veličiny Y lze rozložit na variabilitu mezi výběry

$$\frac{1}{k-1} \sum_{i=1}^{k} n_i (\overline{y}_i - \overline{y})^2$$

a variabilitu uvnitř výběrů

$$\frac{1}{n-k} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_i)^2.$$

Hypotéze neexistence vlivu faktoru na měřenou veličinu odpovídá shoda středních hodnot $\mu_i, i = 1, \ldots, k$, a s tím související zanedbatelná hodnota variability mezi výběry. Model jednoduchého třídění je podrobně diskutován v knize [20] na příkladu s 20 výběry a 25 pozorováními v každém z výběrů.

V monografii [19] se Thiele zabývá dvojným tříděním, které spočívá ve vyšetření vlivu dvou faktorů na měřenou veličinu. Tuto teorii pojímá jako speciální případ lineárního modelu a odvozuje odhady parametrů ve formě používané i dnes. Motivací je mu problém měření v astronomii. Pozorujeme průchody k hvězd přes m paralelních vláken nitkového kříže v astronomickém měřicím přístroji. Střední hodnoty pozorování lze psát ve tvaru (viz [13, s. 173])

$$\mathrm{E}Y_{ij} = \mu_{ij} = \alpha_i + \frac{\beta_j}{h_i}, \quad i = 1, \ldots, k,\ j = 1, \ldots, m. \tag{3.12}$$

Symbol α_i označuje čas průchodu i-té hvězdy středovým vláknem nitkového kříže známou rychlostí h_i. Parametry β_j označují vzdálenost j-tého vlákna od vlákna středového. O veličinách Y_{ij} předpokládáme, že jsou nezávislé normálně rozdělené s rozptylem σ^2.

Vztah (3.12) je rovnicí lineárního modelu (3.8) se speciálním tvarem matice X s neúplnou hodností. Thiele se snaží odhadnout neznámé parametry α_i, β_j pomocí soustavy normálních rovnic. Aby byla řešitelná, je nutné přidat do modelu fiktivní pozorování z. Označíme-li

$$\overline{y}_i = \frac{1}{m}\sum_{j=1}^{m} y_{ij}, \quad w = \sum_{i=1}^{k} h_i^{-2},$$

mají odhady tvar

$$\hat{\alpha}_i = \overline{y}_i - \frac{z}{mh_i},$$
$$\hat{\beta}_j = \frac{1}{w}\sum_{i=1}^{k} h_i^{-1}(y_{ij} - \overline{y}_i) + \frac{z}{m}\,.$$

Thiele poznamenává, že bez přidání z nelze sestrojit odhady parametrů α_i, β_j, je ale možné odhadnout lineární parametrické funkce typu $a_1\alpha_1 + a_2\alpha_2 + \ldots + a_k\alpha_k$ a $b_1\beta_1 + b_2\beta_2 + \ldots + b_m\beta_m$, platí-li reparametrizační podmínky $a_1 + a_2 + \ldots + a_k = 0$ a $b_1 + b_2 + \ldots + b_m = 0$, a také střední hodnoty μ_{ij}. Thiele odvozuje odhady tvaru

$$\hat{\mu}_{ij} = \overline{y}_i + \frac{1}{h_i w}\sum_{r=1}^{k} h_r^{-1}(y_{rj} - \overline{y}_r)$$

a ukazuje, že platí

$$\mathrm{E}\sum_{i=1}^{k}\sum_{j=1}^{m}(Y_{ij} - \hat{\mu}_{ij})^2 = \sigma^2(k-1)(m-1),$$

což umožňuje odhadnout rozptyl σ^2. Model dvojného třídění je zmíněn i v publikacích [20] a [23].

Poděkování. Autorka článku děkuje doc. RNDr. Aleně Šolcové, Ph. D., za laskavou pomoc s překladem astronomického problému demonstrujícího dvojné třídění z anglické verze v [13, s. 173] do češtiny.

Literatura

[1] Anděl, J.: *Základy matematické statistiky.* Matfyzpress, Praha 2005.
[2] Barnett, H. A. R.: *Experiments in mortality graduation and projection using a modification of Thiele's formula.* Journal of the Institute of Actuaries *84* (1958), 212–229.
[3] Berger, A.: *Mathematik der Lebensversicherung.* Julius Springer, Wien 1939.
[4] Cipra, T.: *Finanční a pojistné vzorce.* Grada, Praha 2006.
[5] Fisher, R. A.: *Statistical methods for research workers.* Oliver and Boyd, Edinburgh 1932.
[6] Fisher, R. A., Wishart, J.: *The derivation of the pattern formulae of two–way partitions from those of simpler patterns.* Proc. London Math. Soc., Series 2, *33* (1931), 195–208.
[7] Gram, J. P.: *Professor Thiele som Aktuar.* Dansk Forsikrings–Aarbog, 1910, 26–37.
[8] Hansen, C.: *Om Thiele's Differentialligning for Præmiereserver i Livsforsikring.* H. Hagerups Forlag, Copenhagen 1946.
[9] Helmert, F. R.: *Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate.* Teubner, Leipzig 1907.
[10] Hoem, J. M.: *The reticent trio: Some little-known early discoveries in life insurance mathematics by L. H. F. Oppermann, T. N. Thiele and J.P. Gram.* International Statistical Review, *51* (1983), 213–221.
[11] Jørgensen, N. R.: *Grundzüge einer Theorie der Lebensversicherung.* Fischer, Jena 1913.
[12] Kalman, R. E., Bucy, R.: *New results in linear filtering and prediction.* Journal of Basic Engineering, *83 D* (1961), 95–108.
[13] Lauritzen, S. L.: *Thiele: Pioneer in Statistics.* Oxford University Press, 2002.
[14] Mandl, P.: *Pravděpodobnostní dynamické modely.* Academia, Praha 1985.
[15] Thiele, J. M.: *The life of Thorvaldsen.* Chapman and Hall, London 1865.
[16] Thiele, T. N.: *Undersøgelse af Omløbsbevaegelsen i Dobbeltstjernesystemet Gamma Virginis.* Univerzita Copenhagen, 1866.
[17] Thiele, T. N.: *Om Anvendelse af mindste Kvadraters Methode i nogle Tilfælde, hvor en Komplikation af visse Slags uensartede tilfældige Fejlkilder giver Fejlene en „systematisk" Karakter.* Det kongelige danske Videnskabernes Selskabs Skrifter, 5. Række, naturvidenskabelig og mathematisk Afdeling *12* (1880), 381–408.
[18] Thiele, T. N.: *Sur la compensation de quelques erreurs quasi-systématiques par la méthode des moindres carrés.* C. A. Reitzel, Copenhagen 1880.
[19] Thiele, T. N.: *Almindelig Iagttagelseslære: Sandsynlighedsregning og mindste Kvadraters Methode.* C. A. Reitzel, Copenhagen 1889.
[20] Thiele, T. N.: *Elementær Iagttagelseslære.* Gyldendal, Copenhagen 1897.
[21] Thiele, T. N.: *Om Iagttagelseslærens Halvinvarianter.* Oversigt over det kongelige danske Videnskabernes Selskabs Forhandlinger, *3* (1899), 135–141.
[22] Thiele, T. N.: *Om Dødelighedstavlers Beregning.* Oversigt over det kongelige danske Videnskabernes Selskabs Forhandlinger, 1900, 139–142.
[23] Thiele, T. N.: *Theory of observations.* Layton, London 1903.
[24] Thiele, T. N.: *Adjustment of tables of mortality.* Aktuaren, 1904, 1–10.
[25] Thiele, T. N.: *Interpolationsrechnung.* Teubner, Leipzig 1909.
[26] Thiele, T. N.: *Theory of observations.* Annals of Mathematical Statistics, *2* (1931), 165–307.
[27] `http://www.thiele.au.dk`