TESTY HYPOTÉZ O PARAMETRU OBECNÉHO EXPONENCIÁLNÍHO MODELU

Lucie Fajfrová

Klíčová slova: Stochastické procesy, Rényiho vzdálenosti, exponenciální model.

Abstrakt: Příspěvek se věnuje problému testování hypotéz o parametru v rodinách stochastických procesů, které lze popsat hustotou exponenciálního typu vzhledem k nějaké referenční míře. Zaměříme se na testové statistiky odvozené z informačních vzdáleností nazývaných Rényiho. Cílem příspěvku je především ilustrovat problematiku na příkladech známých procesů.

1 Definice a motivace

Klasické výsledky o exponenciálních rodinách pravděpodobnostních rozdělení můžeme nalézt v řadě monografií, jmenujme například [1], [2]. Tyto knihy se věnují standardnímu statistickému parametrickému modelu s nezávislými pozorováními pocházejícími z rozdělení tzv. exponenciálního typu. Poznamenejme, že mezi tato rozdělení patří například normální, lognormální, gamma, beta, Weibullovo či Maxwellovo, z diskrétních pak jmenujme Poissonovo, negativně binomické či multinomické. Podle definice: rodina $\{P_\theta : \theta \in \Theta\}$ pravděpodobnostních rozdělení na $\{\mathfrak{X}, \mathcal{A}\}$ lišících se v parametru θ z nějaké otevřené $\Theta \subseteq \mathbb{R}^k$ se nazývá *exponenciální rodina*, jestliže existuje σ-konečná míra μ, vůči níž je každá P_θ absolutně spojitá s hustotou $f_\theta = \frac{dP_\theta}{d\mu}$ mající tvar

$$f_{\theta,t}(x) = \exp\{\theta^{\mathrm{T}} T(x) - C(\theta)\} \qquad \forall x \in \mathfrak{X}.$$

Přirozeným rozšířením takového modelu jsou exponenciální rodiny stochastických procesů majících nezávislé stacionární přírůstky. Stochastický proces $(X_t : t \in \mathbb{R}^+)$ s nezávislými a stacionárními přírůstky, jehož trajektorie jsou zprava spojité a mají konečné limity zleva (dále jen *càdlàg* trajektorie), nazýváme *Lévyho proces*. Lévyho procesy jsou ze statistického pohledu důkladně studovány v knize [4]. Sami autoři v souvislosti s Lévyho procesy hovoří o *přirozených exponenciálních rodinách procesů*, nazývejme je v dalším tedy také tak.

Definice 1.1 *Buď $\Theta \subseteq \mathbb{R}^k$ otevřená. Rodina pravděpodobnostích měr na kanonickém filtrovaném prostoru $(\mathfrak{X}, \mathcal{A}, (\mathcal{F}_t)_{t\geq 0})$ (tzn. $\mathfrak{X}$ je množina càdlàg funkcí $\varphi : \mathbb{R}^+ \to \mathbb{R}$) se nazývá* přirozená exponenciální rodina procesů, *jestliže existuje σ-konečná míra μ na $\{\mathfrak{X}, \mathcal{A}\}$ taková, že pro každé $\theta \in \Theta$*

$$P_\theta\restriction_{\mathcal{F}_t} \ll \mu\restriction_{\mathcal{F}_t} \qquad \forall t \geq 0$$

a hustota $f_\theta = \frac{dP_{\theta,t}}{d\mu_t}$ má při označení $P_{\theta,t} = P_\theta\restriction_{\mathcal{F}_t}$, $\mu_t = \mu\restriction_{\mathcal{F}_t}$ tvar

$$f_{\theta,t}(\mathbb{X}_t) = \exp\{\theta^{\mathrm{T}} T(X_t) - t\kappa(\theta)\} \qquad \forall t > 0, \tag{1}$$

kde $\mathbb{X}_t \in \mathfrak{X}_t = \mathfrak{X}\restriction_{\mathcal{F}_t}$ *značí trajektorii do času* t*, na rozdíl od* X_t *označujícího hodnotu trajektorie přímo v čase* t*.* $T = (T_1, ..., T_k)$ *jsou zobrazení z* $\mathbb{R}$ *do* $\mathbb{R}$*.*

Z následujících příkladů uvidíme, že typ exponenciální rodiny procesů určuje především funkce κ z (1). Vzhledem k tomu, že

$$\kappa(\theta) = \frac{1}{t} \ln \mathbb{E}_\theta \exp\{\theta^{\mathrm{T}} T(X_t)\}, \tag{2}$$

kde $\mathbb{E}_\theta$ symbolizuje střední hodnotu vzhledem k míře P_θ, budeme tuto funkci κ nazývat *kumulantová funkce.*

Typickým příkladem přirozené exponenciální rodiny je jednak rodina Wienerových procesů s driftem $\theta \in \mathbb{R}$, zastupující procesy se spojitými trajektoriemi, a jednak rodina Poissonových procesů s intenzitou $\lambda > 0$, která zastupuje skokové procesy. První jmenovanou rodinu zkonstruujeme následovně. Uvažujme jako referenční míru μ rozdělení obyčejného Wienerova procesu $(W_t)_{t\geq 0}$ a definujme rodinu $\{P_\theta : \theta \in \mathbb{R}\}$ pomocí hustot (1), kde za kumulantovou funkci uvažujeme $\kappa(\theta) = \frac{\theta^2}{2}$ a $T(X_t) = X_t$. Pomocí Laplaceovy transformace lze ukázat, že potom pro každé $\theta \in \mathbb{R}$ je P_θ rozdělní procesu $X_t = \theta t + W_t, t \geq 0$. Referenční míra je tedy přímo P_0.

Podobně můžeme postupovat u rodiny Poissonových procesů. Uvažujme nejprve změnu parametru λ, odpovídajícího intenzitě, na parametr $\theta = \ln \lambda$ probíhající všechna reálná čísla. Jako referenční zvolíme rozdělení Poissonova procesu s intenzitou $\lambda = 1$. Opět definujeme rodinu $\{P_\theta : \theta \in \mathbb{R}\}$ pomocí hustot (1), kde tentokrát za kumulantovou funkci zvolíme $\kappa(\theta) = e^\theta - 1$, $T(X_t) = X_t$. Lze ukázat, že potom pro každé $\theta \in \mathbb{R}$ je P_θ rozdělní Poissonova procesu s intenzitou $\lambda = e^\theta$.

Majíce data pocházející z přirozené exponenciální rodiny procesů do nějakého času $t > 0$, můžeme testovat hypotézy o parametru θ. Klasickým nástrojem je maximálně věrohodný odhad parametru θ odvozený v [4]. Citujme zde stručně tento důležitý výsledek:

Mějme přirozenou exponenciální rodinu dle definice 1.1. Je-li $\operatorname{dom}\kappa \subseteq \mathbb{R}$ *otevřená, pak maximálně věrohodný odhad* (MLE) $\hat{\theta}_t$ *parametru* θ *založený na pozorování do času* t *existuje a je dán předpisem* $\hat{\theta}_t = \dot{\kappa}^{-1}\left(\frac{T(X_t)}{t}\right)$ *právě tehdy, když* $\frac{T(X_t)}{t} \in \operatorname{int}(\overline{\mathrm{co}}(\operatorname{supp} T(X_t)))$*. Zde* $\dot{\kappa}^{-1}$ *značí inverzi k derivaci kumulantove funkce. Navíc pak při* P_θ

$$\hat{\theta} \stackrel{t\to\infty}{\longrightarrow} \theta \qquad \text{s.j.} \qquad \& \qquad \sqrt{t}(\hat{\theta} - \theta) \stackrel{t\to\infty}{\longrightarrow} N(0, \left(\ddot{\kappa}(\theta)\right)^{-1}) \quad \text{v distribuci} \tag{3}$$

a testová statistika podílem věrohodností $Q_t = \frac{L_t(\theta_0)}{\sup_{\theta\in\Theta} L_t(\theta)}$ *pro test jednoduché hypotézy* $H : \theta = \theta_0$ *má následující asymptotické rozdělení při* P_{θ_0}

$$-2 \ln Q_t \stackrel{t\to\infty}{\longrightarrow} \chi^2_k \qquad \text{v distribuci,} \tag{4}$$

kde χ^2_k *značí náhodnou veličinu s* χ^2 *rozdělením s* k *stupni volnosti a* $L_t(\theta) = \exp\{\theta^{\mathrm{T}} T(X_t) - t\kappa(\theta)\}$ *je věrohodnostní funkce.*

Pro statistiku zobecněným podílem věrohodností $Q_t^* = \frac{L_t(\theta_t^*)}{\sup_{\theta\in\Theta} L_t(\theta)}$ *pro test složené hypotézy* $H : h(\theta) = 0$ *platí* $-2\ln Q_t^* \xrightarrow{t\to\infty} \chi^2_{k-m}$ *v distribuci, při* P_θ*, pro všechna* $\theta : h(\theta) = 0$*. Zde* θ_t^* *je zúžený maximálně věrohodný odhad, maximalizující* $L_t(\theta)$ *přes množinu* $\{\theta : h(\theta) = 0\}$ *a funkce* $h : \mathbb{R}^k \to \mathbb{R}^m$*, pro nějaké* $1 \le m < k$*, je dvakrát diferencovatelná a* $\dot{h}$ *má plnou hodnost.*

Obecnější přístup k problému testování hypotéz zvolili autoři článků [5], [6]. K testování použili statistiky odvozené z různých informačních vzdáleností. Jako nejzajímavější se ukázaly tzv. Rényiho vzdálenosti. Pro rodinu $\{P_\theta : \theta \in \Theta\}$ pravděpodobnostích rozdělení procesů s dominující mírou μ a pro konstantu $a \in \mathbb{R} \setminus \{0, 1\}$ je *Rényiho vzdálenost* $D_{a,t}$ mezi rozděleními $P_{\theta_1,t}$ a $P_{\theta_0,t}$ do času t daná vztahem

$$D_{a,t}(\theta_1, \theta_0) = \frac{1}{a(a-1)} \ln \int_{\mathfrak{X}_t} \left(\frac{dP_{\theta_1,t}}{d\mu_t}\right)^a \left(\frac{dP_{\theta_0,t}}{d\mu_t}\right)^{1-a} d\mu. \tag{5}$$

Rényiho vzdálenost pro $a = 1$ je speciální případ nazývaný *Kullbackova vzdálenost*:

$$K_t(\theta_1, \theta_0) = D_{1,t}(\theta_1, \theta_0) = \int_{\mathfrak{X}_t} \frac{dP_{\theta_1,t}}{d\mu_t} \ln \frac{dP_{\theta_1,t}/d\mu_t}{dP_{\theta_0,t}/d\mu_t} \; d\mu$$

a pro $a = 0$ jde o inversní Kullbackovu vzdálenost $D_{0,t}(\theta_1, \theta_0) = K_t(\theta_0, \theta_1)$.

Článek [6] se věnuje užití Rényiho vzdáleností pro statistiku právě přirozených exponenciálních rodin procesů, viz definice 1.1. Ukázalo se totiž, že Rényiho vzdálenosti D_a mají v těchto modelech jednoduchý tvar, platí

$$\begin{aligned} D_{1,t}(\theta_1, \theta_0) &= t\Big(\dot\kappa(\theta_1)\,(\theta_1 - \theta_0) + \kappa(\theta_0) - \kappa(\theta_1)\Big) \\ D_{a,t}(\theta_1, \theta_0) &= \frac{t}{a(a-1)}\Big(\kappa(\theta_a) - a\kappa(\theta_1) - (1-a)\kappa(\theta_0)\Big), \end{aligned}$$

kde $\theta_a = a\theta_1 + (1-a)\theta_0$. Vztahy snadno ověříme dosazením (1) do předpisů pro $D_{a,t}$. Pro testování jednoduché hypotézy $H : \;\theta = \theta_0$ lze užít testovou statistiku $2D_{a,t}(\hat\theta_t, \theta_0)$ odvozenou z Rényiho vzdálenosti mezi rozděleními s hypotetickým parametrem θ_0 a maximálně věrohodným odhadem $\hat\theta_t$ parametru z pozorování do času t. Věta 2 v [6] pak říká, že v přirozených exponenciálních modelech má pro každé $a \in \mathbb{R}$ tato testová statistika asymptoticky χ^2 rozdělení s k stupni volnosti, tedy při P_{θ_0}

$$2D_{a,t}(\hat\theta_t, \theta_0) \xrightarrow{t\to\infty} \chi^2_k \qquad \text{v distribuci.} \tag{6}$$

Poznamenejme, že testová statistika $2D_{1,t}(\hat\theta_t, \theta_0)$ odvozená od Kullbackovy vzdálenosti odpovídá statistice $-2\ln Q_t$ odvozené z podílu věrohodností, viz (4). Zmiňme také, že výsledek (6) lze opět zobecnit - užitím zúženého maximálně věrohodného odhadu - také pro složenou hypotézu $H : \; h(\theta) = 0$. Pak

$$T_{a,t} = 2D_{a,t}(\hat\theta_t, \theta_t^*) \xrightarrow{t\to\infty} \chi^2_{k-m} \quad \text{v distribuci, při } P_\theta,\ \forall\theta : h(\theta) = 0, \tag{7}$$

kde h a θ_t^* jsou jak v předchozím. Lze nalézt v [7].

Význam použití Rényiho statistik je mimo jiné v tom, že můžeme na základě nějaké optimalizace vycházející z výpočtu přesné síly a hladiny testů pro jednotlivé řády a vybrat statistiku pro daný model nejvhodnější.

Účelem tohoto příspěvku je jednak přiblížit použití Rényiho statistik na konkrétním příkladě přirozené exponenciální rodiny procesů (sekce 2) a jednak pokusit se rozšířit tuto metodu testování na širší třídu exponenciálních rodin procesů, zahrnující mnohem více než jen Lévyho procesy (sekce 3).

2 Přirozené exponenciální modely

V této části se soustředíme na konkrétní příklad přirozené exponenciální rodiny. Uvažujme dvourozměrný náhodný proces (X_t, Y_t) se spojitým časem, kde $X_t = \exp(\zeta t + W_t)$ je geometrický Brownův pohyb a Y_t na něm nezávislý Poissonův proces s intenzitou $\lambda = e^\eta$. Jak vypadá hustota rodiny Poissonových procesů jsme již zmínili dříve a tvar hustoty rodiny geometrického Brownova pohybu vyplývá z faktu, že proces X_t je Itôovův diffusní proces, konkrétně řešení rovnice $dX_t = (\zeta + \frac{1}{2})X_t dt + X_t dW_t$, $X_0 = 1$, a pro takové je tvar hustoty znám, viz [3]. Tedy

$$f_{(\zeta,\eta),t} = \exp\Big(\zeta \ln(X_t) + \eta Y_t - t(\frac{\zeta^2}{2} + e^\eta - 1)\Big)$$

je hustota rozdělení procesu (X_t^ζ, Y_t^η) vzhledem k referenční míře $P_{0,t} \times Q_{0,t}$, což je rozdělení procesu (X_t^0, Y_t^0). Testovat bychom chtěli hypotézu, že drift ζ je roven intenzitě λ, tj. $H : \zeta - e^\eta = 0$.

Ze zmíněných výsledků pro přirozené rodiny vyplývá, že odhad metodou maximální věrohodnosti má být ve tvaru

$$(\hat\zeta, \hat\eta) = \Big(\tfrac{\ln(X_t)}{t}, \ln(\tfrac{Y_t}{t})\Big).$$

Jde ovšem pouze o tzv. asymptoticky maximálně věrohodný odhad, neboť podle klasické teorie o exponenciálních rodinách MLE parametru v Poissonově procesu neexistuje z důvodu, že $Q_{\eta,t}(Y_t = 0) > 0$. Nicméně tato pravděpodobnost jde do 0, když $t \to \infty$.

Rényiho statistika řádu a má tvar

$$T_{a,t} = t(e^{\eta_t^*} - \hat\zeta_t)^2 + \tfrac{2t}{a(a-1)}\Big[\exp(\eta_t^* + a(\hat\eta_t - \eta_t^*)) - ae^{\hat\eta_t} - (1-a)e^{\eta_t^*}\Big]$$

pro $a \neq 1, 0$, kde $\eta_t^* = \operatorname{argmax}\{\zeta \ln(X_t) + \eta Y_t - t(\frac{\zeta^2}{2} + e^\eta - 1)\ :\ \zeta = e^\eta\}$, a

$$T_{1,t} = t(e^{\eta_t^*} - \hat\zeta_t)^2 + 2t[e^{\hat\eta_t}(\hat\eta_t - \eta_t^* - 1) + e^{\eta_t^*}].$$

Z výsledků článku [7] shrnutých v (7) plyne, že asymptotické rozdělení těchto statistik je χ_1^2.

Majíce data do nějakého času t z této rodiny, uvažujme test hypotézy H s asymptotickou hladinou α_0. Pro rozhodnutí o platnosti hypotézy pak použijeme tu testovou statistiku $T_{a,t}$, pro níž srovnání přesných hladin testu $\alpha = \alpha(\alpha_0, \theta_0, a) = P_{\theta_0,t}(T_{a,t} > \chi_1^2(1-\alpha_0))$ a přesných sil testu $\beta = \beta(\alpha_0, \theta_1, a) =$

$P_{\theta_1,t}(T_{a,t} > \chi_1^2(1-\alpha_0))$ vyjde lépe v následujícím smyslu. Nejprve zvolíme konečnou množinu A všech řadů, pro něž chci toto srovnání provést. Vhodným způsobem také zvolíme konečnou množinu parametrů θ_0 zastupujících hypotézu a množinu parametrů θ_1 zastupujících alternativu. Pokud nelze přímo vypočítat α a β pro tyto hodnoty, lze použít simulaci. V tomto případě bylo použito 5 hodnot parametru z hypotézy, v okolí každé 8 hodnot z alternativy, a pro každé $\theta = (\zeta, \eta)$ z této množiny bylo nasimulováno 10^4 realizací n.v. $Z_{50} \sim N(0, 50)$ a n.v. $Y_{50} \sim Po(50e^{\eta})$, z nichž se určila hladina α, respektive síla β, pro $t = 50$. Metodou 'relative inefficiency', která je popsaná v [5], vyšla porovnáním těchto hodnot jako nejlepší Rényiho statistika řádu $a = 0.5$.

3 Zakřivené exponenciální modely

Definice 3.1 *Buď $\Theta \subseteq \mathbb{R}^k$ otevřená. Rodina pravděpodobnostích měr na kanonickém filtrovaném prostoru $(\mathfrak{X}, \mathcal{A}, (\mathcal{F}_t)_{t\geq 0})$ se nazývá* obecná exponenciální rodina procesů, *jestliže hustota $f_\theta = \frac{dP_{\theta,t}}{dP_{0,t}}$ má (při stejném značení jako v definici 1.1) tvar:*

$$f_{\theta,t} = \exp\{\theta^{\mathrm{T}} A_t - \kappa(\theta) S_t\} \qquad \forall t > 0, \tag{8}$$

kde $\kappa : \Theta \to \mathbb{R}$ je nějaká funkce pouze parametru, $A_t = (A_t^{(1)}, ..., A_t^{(k)})$ jsou $\mathcal{F}_t$-adaptované stochastické procesy s càdlàg trajektoriemi a S_t je neklesající stochastický proces takový, že $S_0 = 0$ a $\lim_{t\to\infty} S_t = +\infty$ s.j.(P_θ) $\forall \theta \in \Theta$.

Rodiny procesů z definice 3.1 jsou speciálním případem tzv. *zakřivených exponenciálních rodin.* Je-li však S_t nenáhodné, dostáváme se zpět k přirozeným rodinám procesů. Proto budeme-li chtít zdůraznit, že jde o rodinu takto nedegenerovanou, tedy že S_t je náhodné, budeme hovořit o rodině *zakřivené.*

Ač na první pohled by se mohl zdát rozdíl mezi definicemi 1.1 a 3.1 malý, o opaku svědčí jednak množství procesů, které nám druhou definicí přibyly, a jednak řada nových problémů, které je nyní potřeba řešit.

Mezi obecné exponenciální rodiny procesů patří například všechny Itôovy diffusní procesy, jestliže příslušná stochastická diferenciální rovnice

$$dX_t^\theta = a_t(X_t^\theta)dt + \theta b_t(X_t^\theta)dt + c_t(X_t^\theta)dW_t$$

má pro danou počáteční podmínku jediné (silné) řešení $(X_t^\theta : t \geq 0)$ pro každé $\theta \in \Theta$. Odvození přesného tvaru exponenciální hustoty pro Itôovy diffusní procesy může čtenář nalézt v [3].

Příklad 3.2 *Ornstein-Uhlenbeckův proces je řešením rovnice*

$$dX_t = \theta X_t dt + dW_t.$$

Uvažujeme-li počáteční podmínku $X_0 = 0$, potom explicitní vyjádření procesu je $X_t = \int_0^t e^{\theta(t-s)} dW_s$. Jde o exponenciální rodinu danou souborem hustot:

$$\frac{dP_{\theta,t}}{dP_{0,t}} = \exp\{\tfrac{\theta}{2}(X_t^2 - t) - \tfrac{\theta^2}{2}\int_0^t X_s^2 ds\}$$

pro každé $t > 0$, kde P_0 je rozdělení Wienerova procesu.

Další velikou skupinou procesů, které jsou pokryty definicí 3.1, jsou čítací (skokové) procesy. Tím míníme zprava spojitý proces N_t, který nabývá celých čísel a jehož na čase závislá intenzita λ_t má tvar $\lambda_t = \lambda H_t$, kde H_t je do času t adaptovaný positivní proces s konečnou střední hodnotou pro každé $t > 0$. Intenzitou myslíme intenzitu ve smyslu Markovského procesu se stavy v $\mathbb{Z}$, který se v jednom kroku může změnit pouze o jedničku. Označíme-li $\theta = \ln \lambda$, rodina procesů N_t je zadaná hustotou

$$\tfrac{dP_{\theta,t}}{dP_{0,t}} = \exp\{\theta(N_t - N_0) - (e^\theta - 1)\int_0^t H_s ds\},$$

kde P_0 je rozdělení daného čítacího procesu s parametrem 0, tj. s intenzitou $\lambda = 1$. Speciálním degenerovaným případem je jistě Poissonův proces s $H_t \equiv 1$. Dalším speciálním případem, tentokrát již zakřivené exponenciální rodiny, je *jednoduchý proces zrodu*, kde $H_t = N_{t_-}$. Budeme se tímto procesem zabývat detailně později.

Vraťme se nyní k tomu, o které vlastnosti z přirozených exponenciálních rodin jsme přišli a které nám zůstaly zachovány. Uvědomme si nejprve, že mnoho vlastností v přirozených rodinách plynulo z faktu, že kumulantovou funkci bylo možné vyjádřit vztahem (2), což v zakřivených rodinách možné není. Jedním ze způsobů, jak s těmito rodinami dále pracovat, je nahlédnout, že vhodnou náhodnou časovou transformací lze obecnou exponenciální rodinu převést na přirozenou. Jen stručně uveďme, že jde o transformaci definovanou časem zastavení $\tau_u = \inf\{t : S_t > u\}$ a pokud definujeme filtraci $(\mathcal{F}_{\tau_u})_{u\geq 0}$ σ-algeber do času zastavení τ_u, pak lze ukázat, že

$$\tfrac{dP_{\theta,\tau_u}}{dP_{0,\tau_u}} = \exp\{\theta^{\mathrm{T}} A_{\tau_u} - \kappa(\theta)u\},$$

kde $P_{\theta,\tau_u} = P_\theta\lceil_{\mathcal{F}_{\tau_u}}$. Pro více detailů odkažme čtenáře do [4], kde je také jako důsledek této transformace odvozen následující výsledek pro maximálně věrohodný odhad parametru θ v obecném exponenciálním modelu. Citujme:

Věta 3.3 *(Theorem 5.2.1-2 v [4]) Buď κ v (8) taková, že* $\operatorname{dom}\kappa$ *je otevřená podmnožina $\mathbb{R}^k$, a nechť $S_t > 0$ $\forall t > 0$ P_θ-s.j $\forall\theta \in \Theta$, kde Θ je otevřená. Pak*

- *MLE $\hat\theta_t$ založený na pozorování do času t existuje a je dán jednoznačně vztahem $\hat\theta_t = \dot\kappa^{-1}\left(\frac{A_t}{S_t}\right)$ právě tehdy, když $\frac{A_t}{S_t} \in \operatorname{int}(\overline{\mathrm{co}}(\operatorname{supp} A_{\tau_1}))$,*
- $\hat\theta \stackrel{t\to\infty}{\longrightarrow} \theta \qquad \text{s.j.}(P_\theta)$ *pro všechny $\theta \in \Theta$,*
- *jestli existuje pro každé $\theta \in \Theta$ rostoucí, kladná nenáhodná funkce $\varphi_\theta(t)$, že*

$$\tfrac{S_t}{\varphi_\theta(t)} \stackrel{t\to\infty}{\longrightarrow} Y_\theta \quad \text{v pravděpodobnosti, při } P_\theta,$$

a že $P_\theta(Y_\theta > 0) > 0$, kde Y_θ je nějaká n.v. s rozdělením F_θ, potom při P_θ

$$\left(\sqrt{S_t}(\hat\theta - \theta), \tfrac{S_t}{\varphi_\theta(t)}\right) \stackrel{t\to\infty}{\longrightarrow} N(0, (\ddot\kappa(\theta))^{-1}) \times F_\theta \qquad \& \qquad -2\ln Q_t \stackrel{t\to\infty}{\longrightarrow} \chi^2_k$$

v distribuci, za podmínky $\{Y_\theta > 0\}$, kde Q_t je jako v (4).

Nějaký analogický výsledek v obecných exponenciálních modelech výsledkům (6) a (7) v přirozených modelech pro testování pomocí Rényiho statistik zatím k disposici není. Problém je v nalezení vhodného tvaru Rényiho vzdáleností (5), které pro obecný exponenciální model (8) mají mít tvar:

$$D_{1,t}(\theta_1,\theta_0) = \int_{\mathfrak{X}_t} \Big(A_t(\theta_1-\theta_0) - S_t(\kappa(\theta_1)-\kappa(\theta_0))\Big) dP_{\theta_1,t} \tag{9}$$

a pro $a \neq 1,0$: $D_{a,t}(\theta_1,\theta_0) =$ (10)

$$\frac{1}{a(a-1)} \ln \int_{\mathfrak{X}_t} \exp\Big\{A_t\Big(a\theta_1+(1-a)\theta_0\Big) - S_t\Big(a\kappa(\theta_1)+(1-a)\kappa(\theta_0)\Big)\Big\} dP_{0,t}.$$

Věnujme se ve zbytku tomuto problému v konkrétním příkladě zakřivené exponenciální rodiny, a to rodiny jednoduchých procesů zrodu.

Příklad 3.4 *Buď $(X_t)_{t\geq 0}$ jednoduchý proces zrodu a nechť $X_0 = 1$. To znamená, že $X_t \in \{1,2,...\}$ značí velikost populace v čase t, jestliže v čase 0 populace čítala jediného jedince, který má dát po exponenciálním čase se střední hodnotou $1/\lambda$ život novému jedinci. Přičemž každý jedinec této populace se bude chovat stejně a jejich chování bude navzájem nezávislé. Při parametrizaci $\theta = \ln\lambda \in \mathbb{R}$ je rodina jednoduchých procesů zrodu zakřivenou exponenciální rodinou s hustotami do času t:*

$$f_{\theta,t} = \exp\{\theta(X_t-1) - (e^\theta - 1)\int_0^t X_s ds\}$$

vzhledem k referenční míře $P_{0,t}$, rozdělení procesu zrodu s intenzitou $\lambda = 1$.

Odvoďme tvar Rényiho vzdáleností speciálně pro proces zrodu. Zafixujme $t > 0$. Začneme se snadnější Kullbackovou vzdáleností, která (dle (9)) má tvar $D_{1,t}(\theta,\theta_0) = (\mathbb{E}_\theta X_t - 1)(\theta-\theta_0) - \mathbb{E}_\theta \int_0^t X_s ds\ (e^\theta - e^{\theta_0})$. Vidíme, že stačí znát $\mathbb{E}_\theta X_t\ \forall t > 0$, tedy střední počet jedinců v čase t v populaci s intenzitou zrodu $\lambda = e^\theta$. Lze odvodit, že $\mathbb{E}_\theta X_t = \exp\{te^\theta\}$ a tedy

$$D_{1,t}(\theta,\theta_0) = (\exp\{te^\theta\} - 1)(e^{\theta_0-\theta} + \theta - \theta_0 - 1). \tag{11}$$

Rényiho vzdálenost pro $a \neq 1,0$ má podle (10) mít tvar $D_{a,t}(\theta,\theta_0) =$

$$\frac{1}{a(a-1)} \ln \mathbb{E}_{a\theta+(1-a)\theta_0} \exp\Big\{-\int_0^t X_s ds\Big(ae^\theta + (1-a)e^{\theta_0} - e^{a\theta+(1-a)\theta_0}\Big)\Big\}.$$

Označíme-li si

$$\nu = \nu(a,\theta,\theta_0) = ae^\theta + (1-a)e^{\theta_0} - e^{a\theta+(1-a)\theta_0},$$

potom potřebujeme znát $\mathbb{E}_{\theta_a} \exp\Big\{-\nu\int_0^t X_s ds\Big\}$, kde $\theta_a = a\theta + (1-a)\theta_0$. Lze odvodit, že tato střední hodnota je konečná pro všechna $t > 0$ právě tehdy, když $\nu \geq 0$, a potom

$$\mathbb{E}_{\theta_a} \exp\Big\{-\nu\int_0^t X_s ds\Big\} = \frac{e^{\theta_a}+\nu}{e^{\theta_a} + \nu\exp\{t(e^{\theta_a}+\nu)\}}.$$

Vzhledem k tomu, že $\nu \geq 0$ právě tehdy, když $a \in [0,1]$, má dále smysl zabývat se pouze Rényiho vzdálenostmi pro $a \in [0,1]$. Pro $a \in (0,1)$ pak:

$$D_{a,t}(\theta,\theta_0) = \frac{1}{a(1-a)} \ln\Big(1 + \nu(a,\theta,\theta_0)\frac{\exp\{t(ae^\theta+(1-a)e^{\theta_0})\} - 1}{ae^\theta + (1-a)e^{\theta_0}}\Big). \tag{12}$$

Lemma 3.5 *Rényiho statistiky $T_{a,t} = 2D_{a,t}(\hat{\theta}_t, \theta_0)$ odvozené z (11) a (12) mají pro každé $a \in [0,1]$ Fisher-Snedecorovo $F(1,2)$ rozdělení, tj. vzhledem k rozdělení P_{θ_0} procesu zrodu*

$$T_{a,t} \stackrel{t\to\infty}{\longrightarrow} F(1,2) \qquad \text{v distribuci.}$$

Důkaz. Metodou Taylorova rozvoje pro funkci $f(\theta) := D_{a,t}(\theta, \theta_0)$ v bodě θ_0 zjistíme, že $D_{a,t}(\theta, \theta_0) = \frac{1}{2}\psi_{a,t}(\theta)(\theta - \theta_0)^2$, kde $\psi_{a,t}(\cdot)$ jsou spojité funkce takové, že $\psi_{a,t}(\theta_0) = \exp\{te^{\theta_0}\} - 1$ pro každé $a \in [0,1], t > 0$. A vzhledem k tomu, že proces $S_t = \int_0^t X_s ds$, kde X_s je proces zrodu s intenzitou $\lambda = e^{\theta}$, se při $t \to \infty$ chová následovně (lze zjistit z Laplaceovy transformace):

$$\lambda e^{-t\lambda} S_t = e^{\theta} \exp\{-te^{\theta}\}\ S_t \stackrel{t\to\infty}{\longrightarrow} \chi^2_2 \quad \text{s.j.}(P_\theta),$$

použitím věty 3.3 obdržíme: $2D_{a,t}(\hat{\theta}_t, \theta_0) \stackrel{t\to\infty}{\longrightarrow} F(1,2)$ v distribuci při P_{θ_0}. $\square$

Závěrem tedy můžeme říct, že Rényiho statistiky, pokud je umíme spočítat, půjdou alespoň pro hodnoty řádu a z intervalu $[0,1]$ použít i v zakřivených exponenciálních modelech pro testy hypotéz, s toutéž výhodou jako v přirozených exponenciálních modelech.

Reference

[1] Barndorff-Nielsen O.E. (1978) *Information and Exponential Families.* J.Wiley, Gluchester.

[2] Brown L.D. (1986) *Fundamentals of Statistical Exponential Families of Stochastic Processes.* Lectures Notes Vol.9. Inst. of Math. Statistics, Hayward, California.

[3] Liptser R.Sh., Shiryaev A.N. (1989) *Statistika slučajnych procesov.* Springer, New York.

[4] Küchler U., Sørensen M. (1997) *Exponential Families of Stochastic Processes.* Springer, Berlin.

[5] Morales D., Pardo L., Vajda I. (1997) *Some new statistics for testing hypothesis in parametric models.* Journal of Multivariate Analysis, **62**, 137 – 168.

[6] Morales D., Pardo L., Vajda I. (1999) *Rényi stistics in directed families of exponential experiments.* Statistics, **34**, 151 – 174.

[7] Morales D., Pardo L., Pardo M.C., Vajda I. (2004) *Rényi stistics for testing composite hypothesis in general exponential models.* Statistics, **38**, No.2, 133 – 147.

Poděkování: Tato práce byla podporována grantem GA ČR 201/06/1323 a výzkumným centrem MŠMT ČR 1M0572.

Adresa: ÚTIA AV ČR, Pod vodárenskou věží 4, 182 08 Praha 8

E-mail: `fajfrova@utia.cas.cz`