

APROXIMACE OBECNÝCH SYSTÉMŮ HROMADNÉ OBSLUHY POMOCÍ EM ALGORITMU

GEJZA DOHNAL, MARTIN MECA

ABSTRACT. It is well known that the main barrier to the explicit solution of even very simple stochastic models is the increasing complexity of the conditional probability distributions that arise in their analysis. One considerable way out consists in approximations by so called phase-type (PH) distributions. It was introduced by Neuts in 1975, a detailed discussion of its properties can be found in [1]. Since then, many authors have used these versatile distributions in several fields of stochastic modeling: in queueing theory (Neuts 1981, Sen Gupta 1989, Asmussen 1992), reliability theory (Bobbio et al. 1980, Johnsson et al. 1994), renewal theory (Kao 1988, Lipsky 1992, Asmussen & Bladt 1996), lifetime analysis (Bhattachajee and Neuts), survival models (Aalen, 1995) and others.

This contribution shows how to fit a phase type distribution to a general distribution using an extended EM algorithm. Some examples are given for Chi-square, log-normal and truncated normal distributions.

Резюме. Общеизвестно, что главным барьером при решении и очень простых упрощений стохастического моделирования является возрастающая сложность входящих условных вероятностных распределений. Одна из возможностей заключается в приближении с помощью так называемого распределения фазового типа. Класс распределений фазового типа определил Нойтс в 1975 году, подробное описание его свойств можно найти в [1]. С этого времени много авторов использовало этот полезное распределение в различных областях стохастического моделирования: в теории массового обслуживания (Нойтс 1981, Сенгупта 1989, Асмуссен 1992), в теории надежности (Боббио и кол. 1980, Джонссон и кол. 1994), теории восстановления (Као 1988, Липскы 1992, Асмуссен и Бладт 1996), в анализе жизнеспособности (Бхаттахае и Нойтс), модели переживания (Аален, 1995) и дальше.

В статье показаем как распределением фазового типа приблизить общее распределение с помощью EM алгоритма и это показано на примерах распределений хи-квадрат, лог-нормальном и отграниченного нормального.

1. ÚVOD

Teorie Markovových procesů patří k pilířům teorie hromadné obsluhy, a to hlavně proto, že její aplikací – i v případě poměrně komplikovaných systémů – dostáváme přehledné výsledky v kompaktní formě, které s výhodou použijeme jak při další analýze, tak při automatickém zpracování dat vznikajících u reálných systémů.

Na druhé straně je mnoho modelů – a často velmi jednoduchých – u nichž nelze přímo předpokládat „markovské chování“. Při analýze takových případů se obvykle dostáváme k velmi složitým distribučním vztahům, které jsou prakticky neřešitelné.

2000 *Mathematics Subject Classification.* Primary 60K20; Secondary 62F99.

Klíčová slova. EM-algoritmus, systémy hromadné obsluhy, PH-rozdělení, Markovské procesy.

Vzniklo za podpory grantu GA AVČR A1075101.

Právě zde se nabízí možnost použít pro aproximaci spojitých rozdělení doby mezi příchody zákazníků či doby trvání obsluhy taková rozdělení, abychom dostali výsledný model, který už lze popsat pomocí Markovova procesu. Tuto možnost nám poskytuje třída distribucí, které se nazývají *rozdělení fázového typu* nebo také *PH-rozdělení*.

2. DEFINICE A VLASTNOSTI PH-ROZDĚLENÍ

Definice: Náhodná veličina X , definovaná na $[0, \infty)$, má rozdělení fázového typu (PH-rozdělení) s parametry $(\boldsymbol{\pi}, \mathbf{T})$, dá-li se reprezentovat jako doba do absorpce Markovova procesu s $p < \infty$ přechodnými stavy (nazývanými *fáze*) a jedním absorpčním stavem, s počátečním rozdělením pravděpodobnosti $(\boldsymbol{\pi}, \boldsymbol{\pi}^0)$, kde $\boldsymbol{\pi}$ je p -rozměrný řádkový vektor. Matice intenzit přechodů tohoto procesu má tvar

$$\begin{pmatrix} \mathbf{T} & \mathbf{T}^0 \\ \mathbf{0} & 0 \end{pmatrix},$$

kde \mathbf{T} je matice typu $p \times p$, pro jejíž prvky platí: $T_{ii} < 0$ pro $i = 1, \dots, p$ a $T_{ij} \geq 0$, $i \neq j$. \mathbf{T}^0 je sloupcový p -rozměrný vektor intenzit přechodu ze stavů $1, \dots, p$ do absorpčního stavu. Označíme-li \mathbf{e} p -rozměrný řádkový vektor samých jedniček, potom musí platit, že $\mathbf{T}\mathbf{e}' + \mathbf{T}^0 = \mathbf{0}$.

Poznámka: Jako absorpční je v předchozí definici zvolen stav jemuž odpovídá poslední řádek a sloupec v matici intenzit přechodů. Chceme-li navíc zaručit, aby všechny ostatní stavy byly přechodné (a náhodná veličina X s pravděpodobností 1 konečná), musí být matice \mathbf{T} regulární ([1] Neuts).

Mezi třídu rozdělení fázového typu patří některá známá a běžně používaná rozdělení, například exponenciální nebo Erlangovo. Libovolná konečná směs nebo konvoluce exponenciálních rozdělení je rozdělením fázového typu. Tak například třída PH-rozdělení pokrývá všechny sériově-paralelní soustavy exponenciálních rozdělení i s případnou zpětnou vazbou. Třída PH-rozdělení je podtřídou tzv. *Coxových distribucí*, to jest distribucí na $[0, \infty)$ s racionální charakteristickou funkcí (Cox, 1953).

Dále uvedeme několik nejdůležitějších vlastností PH-rozdělení. Důkazy zde uvedených vět lze nalézt v citované literatuře.

Věta 1 ([1] Neuts): Má-li náhodná veličina X PH-rozdělení s parametry $(\boldsymbol{\pi}, \mathbf{T})$, potom

- (i) její hustota pravděpodobnosti má tvar

$$f(x) = \boldsymbol{\pi} \exp(x\mathbf{T})\mathbf{T}^0 \quad \text{pro } x \geq 0, \quad f(x) = 0 \quad \text{jinak.}$$

- (ii) distribuční funkce X je dána vztahem

$$F(x) = 1 - \boldsymbol{\pi} \exp(x\mathbf{T})\mathbf{e}' \quad \text{pro } x \geq 0.$$

Distribuční funkce má skok v $x = 0$ o velikosti $\boldsymbol{\pi}^0$.

- (iii) charakteristická funkce $f(s)$ má tvar

$$f(s) = \boldsymbol{\pi}^0 + \boldsymbol{\pi}(s\mathbf{I} - \mathbf{T})^{-1}\mathbf{T}^0 \quad \text{pro } s \geq 0,$$

kde \mathbf{I} je jednotková matice.

- (iv) je-li \mathbf{T} regulární, jsou všechny obecné momenty veličiny X konečné a platí

$$m_k = EX^k = (-1)^k k! \boldsymbol{\pi} \mathbf{T}^{-k} \mathbf{e}', \quad k \in N.$$

Věta 2 ([1] Neuts): Konvoluce dvou PH-rozdělení je opět PH-rozdělením. Libovolná konečná směs PH-rozdělení je opět PH-rozdělením.

Věta 3 ([2] Johnson, Taaffe): Třída PH-rozdělení je hustá ve třídě všech rozdělení náhodných veličin s oborem hodnot v $[0, \infty)$.

Důsledkem poslední vlastnosti je tvrzení, že libovolné rozdělení s nosičem v $[0, \infty)$ je možné, alespoň v principu, aproximovat s libovolnou přesností PH-rozdělením s vhodnými parametry (π, T) .

3. SYSTÉMY M/G/1 A GI/M/1

Jako příklad ukážeme, k čemu vedeme aproximace v systémech M/G/1 a GI/M/1.

Systém M/G/1 nechť je charakterizován těmito vlastnostmi:

- Doba mezi příchody zákazníků je rozdělena exponenciálně s parametrem λ .
- Rozdělení doby obsluhy má rozdělení se spojitou hustotou na intervalu $[0, \infty)$

Pokud aproximujeme rozdělení doby obsluhy PH-rozdělením s parametry π a T , bude možné výsledný systém popsat Markovským procesem s maticí intenzit přechodu

$$\begin{pmatrix} -\lambda & \lambda\pi & \mathbf{0}' & \mathbf{0}' & \dots \\ \mathbf{T}^0 & \mathbf{T} - \lambda\mathbf{I} & \lambda\mathbf{I} & \mathbf{0}' & \dots \\ \mathbf{0} & \mathbf{T}^0\pi & \mathbf{T} - \lambda\mathbf{I} & \lambda\mathbf{I} & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{T}^0\pi & \mathbf{T} - \lambda\mathbf{I} & \dots \\ \vdots & & & & \ddots \end{pmatrix}$$

Systém GI/M/1 je systém hromadné obsluhy, jehož proces příchodů a doba obsluhy mají následující vlastnosti:

- Příchody zákazníků do systému tvoří proces obnovy se spojitou distribuční funkcí časů mezi příchody.
- Doba obsluhy má exponenciální rozdělení s parametrem μ .

Rozdělení doby mezi příchody zákazníků můžeme aproximovat PH-rozdělením s parametry π a T . Matice intenzit přechodu výsledného Markovova procesu potom bude mít tvar

$$\begin{pmatrix} \mathbf{T} & \mathbf{T}^0\pi & \mathbf{0}' & \mathbf{0}' & \dots \\ \mu\mathbf{I} & \mathbf{T} - \mu\mathbf{I} & \mathbf{T}^0\pi & \mathbf{0}' & \dots \\ \mathbf{0} & \mu\mathbf{I} & \mathbf{T} - \mu\mathbf{I} & \mathbf{T}^0\pi & \dots \\ \mathbf{0} & \mathbf{0} & \mu\mathbf{I} & \mathbf{T} - \mu\mathbf{I} & \dots \\ \vdots & & & & \ddots \end{pmatrix}$$

Po aproximaci analyzujeme systémy hromadné obsluhy standardním způsobem. Je-li například systém stabilizovatelný (což můžeme zjistit i před aproximací – např. u systémů s jedním obslužným místem musí být střední doba obsluhy menší než střední doba mezi příchody zákazníků), spočteme stacionární rozdělení a z něho se odvíjející charakteristiky.

4. EM-ALGORITMUS

Jednou z metod nalezení aproximujícího PH-rozdělení je EM-algoritmus (Expectation-Maximization). Ten byl původně navržen pro iterativní zjištění maximálně věrohodného odhadu z pozorovaných dat, která jsou pouze částí rozsáhlejšího experimentu. Předpokládejme, že pozorujeme realizace y náhodné veličiny Y , která má hustotu g_γ , závislou na neznámém parametru γ , jako část experimentu, jehož výsledkem jsou realizace veličiny X s hustotou f_γ . Obě veličiny nechť jsou vázány vztahem $Y = u(X)$. Potom $(n + 1)$ iteraci EM-algoritmu můžeme stručně zapsat:

E-krok: nejprve spočteme hodnotu funkce $L_{n+1}(\gamma) = E_{\gamma_n} [\ln f_\gamma(X) | u(X) = y]$

M-krok: novou hodnotu parametru potom dostaneme z $\gamma_{n+1} = \arg \max L_{n+1}(\gamma)$

Zobecněním EM-algoritmu na nekonečné množství dat – kompletní distribuci – dostaneme iterativní algoritmus pro aproximaci spojitého rozdělení. Označíme-li f původní aproximovanou hustotu, pak jeho aplikací nalezneme rozdělení s hustotou h , které má od f minimální tzv. Kullback-Leiblerovu vzdálenost ([3] Asmussen, Olsson)

$$I(f, h) = \int \frac{f(x)}{h(x)} \mu(dx)$$

Konkrétně v našem případě aproximace PH-rozdělením, máme-li odhady $\pi^{(n)}$ a $T^{(n)}$ parametrů π a T , dostaneme nové hodnoty odhadů v $n + 1$ iteraci:

E-krok: Nejdříve spočítáme funkce

$$\begin{aligned} \mathbf{a}(y) &= \pi^{(n)} \exp(yT^{(n)}) & \mathbf{b}(y) &= \exp(yT^{(n)})T^{0(n)} \\ c(y) &= \int_0^y \mathbf{a}(t)\mathbf{b}(y-t)dt \end{aligned}$$

z nich potom hodnoty

$$\begin{aligned} B_i^{(n+1)} &= \int_0^\infty \frac{\pi_i^{(n)} b_i(y)}{\pi^{(n)} \mathbf{b}(y)} f(y) dy & Z_i^{(n+1)} &= \int_0^\infty \frac{c_{ij}(y)}{\pi^{(n)} \mathbf{b}(y)} f(y) dy \\ N_{ij}^{(n+1)} &= \int_0^\infty \frac{T_{ij} c_{ji}(y)}{\pi^{(n)} \mathbf{b}(y)} f(y) dy, \quad i \neq j, & N_i^{0(n+1)} &= \int_0^\infty \frac{a_i(y) T_i^{0(n)}}{\pi^{(n)} \mathbf{b}(y)} f(y) dy \end{aligned}$$

kde $i, j = 1 \dots p$ a $T^{0(n)} = -T^{(n)}e$

M-krok:

$$\begin{aligned} \pi_i^{(n+1)} &= B_i^{(n+1)} & T_{ij}^{(n+1)} &= \frac{N_{ij}^{(n+1)}}{Z_i^{(n+1)}}, \quad i \neq j \\ & & & i, j = 1, \dots, p \\ T_i^{0(n+1)} &= \frac{N_i^{0(n+1)}}{Z_i^{(n+1)}} & T_{ii}^{(n+1)} &= -T_i^{0(n+1)} - \sum_{j \neq i} T_{ij}^{(n+1)} \end{aligned}$$

EM-algoritmus vždy konverguje, ovšem ne vždy ke globálnímu maximu, ale ve většině případů pouze k některému z lokálních maxim, anebo dokonce k sedlovému bodu. Správná konvergence závisí na volbě počátečního řešení a přeneseně také na přesnosti numerického výpočtu.

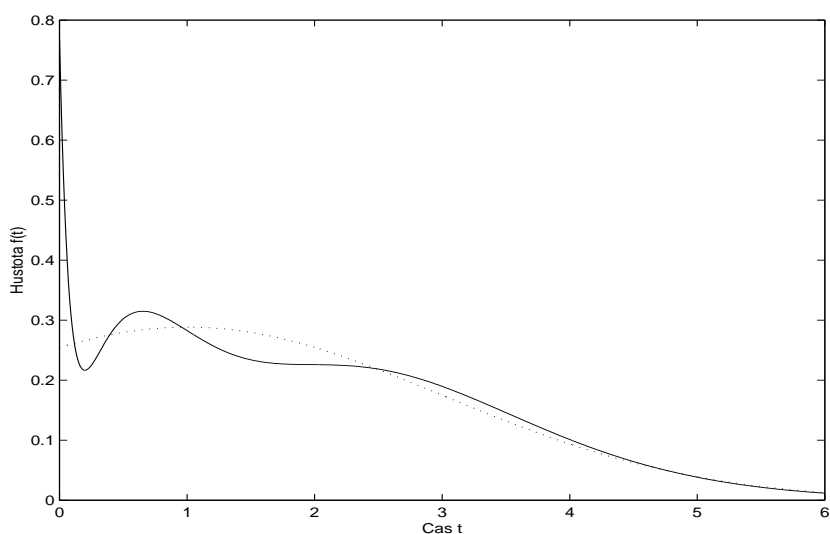
5. NUMERICKÁ IMPLEMENTACE

Největším problémem při numerické implementaci je spočítat integrály z E-kroku. Protože jsou ve tvaru středních hodnot funkcí náhodné veličiny s rozdělením s aproximovanou hustotou f , nabízí se jako jedna z možností použít zákona velkých čísel k přibližnému určení těchto integrálů. Nevýhodou však může být malá přesnost počítaná navíc pouze s určitou pravděpodobností. Výhodou potom jeho rychlost a aplikovatelnost na všechna rozdělení.

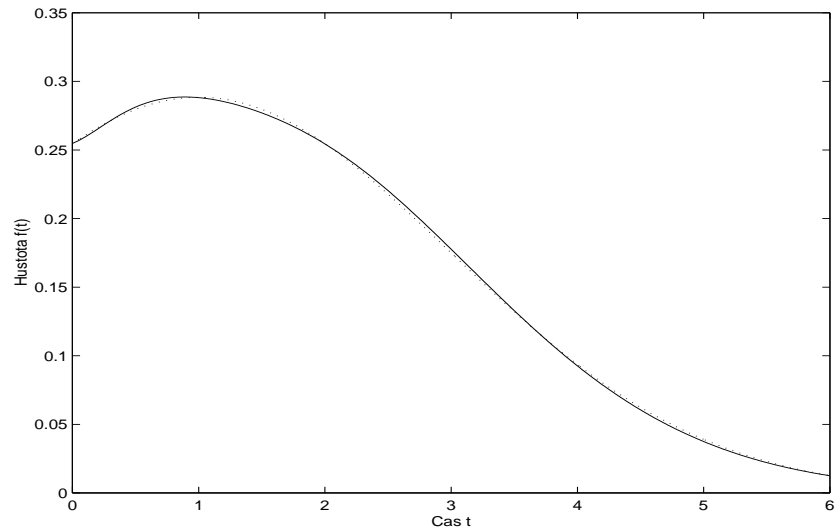
Další numerickou metodou při výpočtů integrálů z E-kroku, kterou jsme v implementaci použili je Simpsnova metoda. Ta bohužel není aplikovatelná na všechna rozdělení, ale pouze na ta, která mají omezenou hustotu (neomezenou má například Weibullovo s parametrem $\beta < 1$, jejíž limita zprava v nule je plus nekonečno). Dalším problémem při použití Simpsnovy metody je neomezený interval přes který se integruje. Bylo proto nutné spočítat horní odhady integrovaných funkcí, tak aby se integrály přes neomezený interval daly s libovolnou přesností aproximovat integrály přes konečný interval. I přes tyto nevýhody EM-algoritmus při použití Simpsnovy metody konverguje po menším počtu kroků a v některých případech se zdá, že je přesnost dokonce rozhodující, k tomu abychom dosáhli globálního maxima.

Představu o rychlosti a přesnosti konvergence po použití obou metod si můžeme udělat z následujících dvou obrázků. V obou případech je aproximováno useknuté normální rozdělení, první obrázek vznikl však použitím metody Monte Carlo a druhý použitím Simpsnovy metody. Číslo za názvem metody znamená počet bodů použitých při aproximaci integrálů. Tečkovaně je vykreslena původní hustota, plnou čarou potom aproximující hustota.

Vidíme, že zatímco v prvním případě EM-algoritmus nekonverguje správně, ve druhém je aproximace po 1000 iteracích již dostatečně přesná.

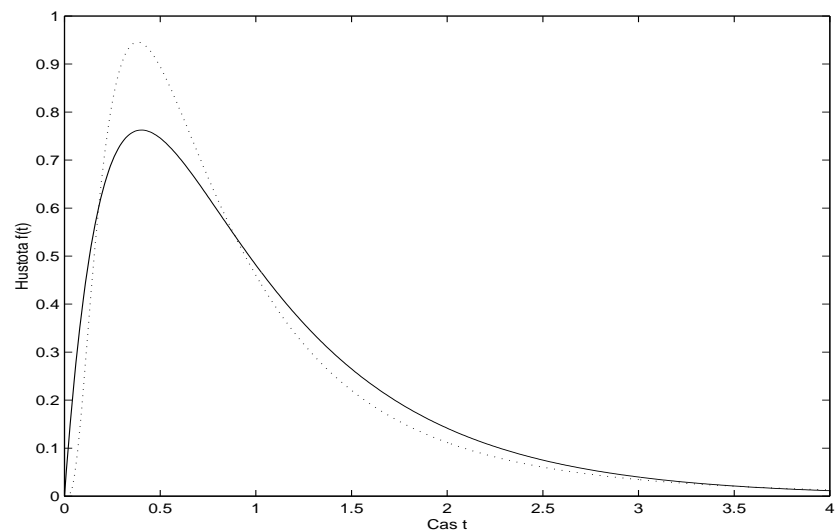


Obr. 1. $TN(1, 2)$, MonteCarlo 500, 1000 iterací, $p=8$, Coxova struktura.

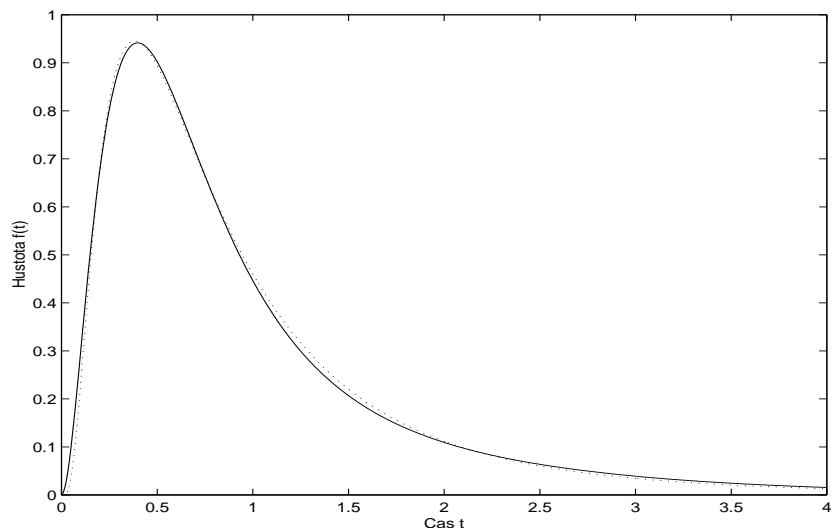


Obr. 2. $TN(1, 2)$ Simpson 500, 1000 iterací, $p=8$, Coxova struktura.

Důležitá je volba dimenze aproximace, tedy volba počtu stavů Markovova řetězce, reprezentujícího aproximující rozdělení. Zvolíme-li malou dimenzi, bude výpočet rychlý, ovšem míra aproximace nemusí být dostačující. S volbou větší dimenze roste zase pravděpodobnost konvergence algoritmu k lokálnímu maximu, anebo sedlovému bodu. Následuje srovnání volby dvou různých dimenzí při aproximaci logaritmicke-normálního rozdělení. Na prvním obrázku je výsledek při použití rozměrů matice T dva krát dva a na druhém čtyři krát čtyři.

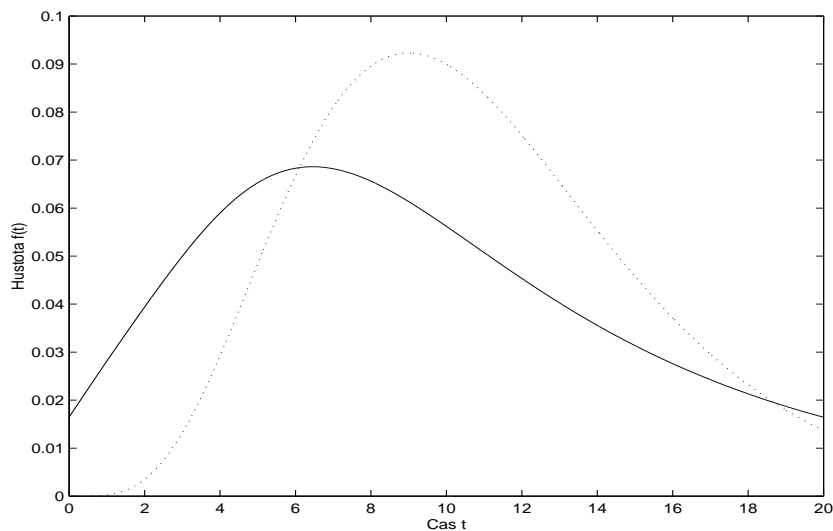


Obr. 3. $LN(-0.32, 0.8)$, Monte Carlo 1000, 1000 iterací, $p=2$.

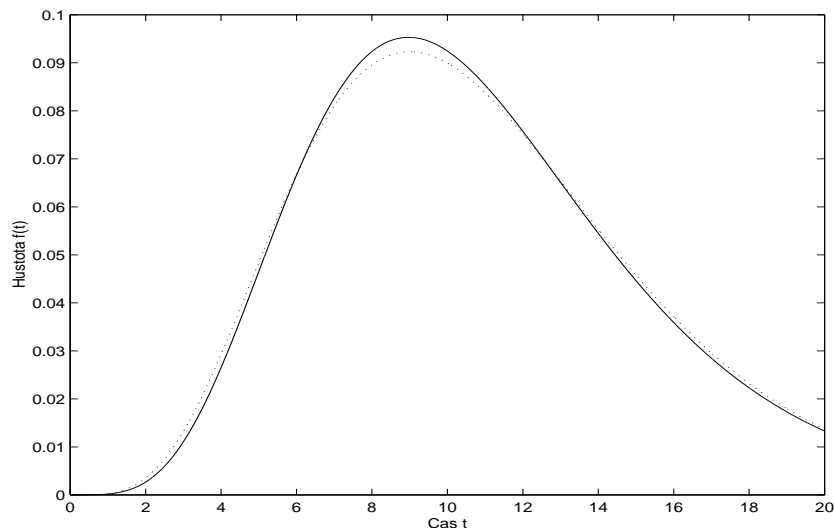


Obr. 4. $LN(-0.32, 0.8)$, Monte Carlo 1000, 2000 iterací, $p=4$.

Klíčovým tématem je volba počátečního řešení. Nulové složky počátečního řešení zůstanou nulové po celou dobu běhu algoritmu. To nám dovoluje již na začátku určit strukturu matice intenzit přechodů konečného řešení. Můžeme chtít například aby výsledný Markovův proces byl acyklický, tj. aby se nemohl vracet do stavů ve kterých již byl (tomu odpovídá horní trojúhelníková počáteční matice T), nebo můžeme chtít aby se Markovův proces mohl přesouvat pouze ze stavu i do stavu $i+1$ (tomu odpovídá matice s nenulovými prvky pouze na diagonále a nad diagonálou) – takzvaná Coxova struktura generátorové matice.



Obr. 5. χ^2 -kvadrát 11, Monte Carlo 500, 500 iterací, $p=8$, Coxova struktura.



Obr. 6. *Chí-kvadrát 11, Monte Carlo 500, 100 iterací, $p=8$, Coxova struktura, počítaná diagonála.*

Výrazně urychlit konvergenci může volba takového počátečního rozdělení, které má střední hodnotu stejnou s aproximovaným rozdělením. Toho například docílíme, pokud jsou všechny složky počátečního rozdělení nenulové, dopočtením diagonálních složek matice \mathbf{T} podle vzorce

$$T_{ij} = -\pi_i \left[\frac{p}{m_i} + \sum_{i \neq j} \frac{T_{ij}}{\pi_j} \right]$$

kde p je dimenze matice \mathbf{T} a m_1 je střední hodnota aproximovaného rozdělení. Na následujících dvou obrázcích vidíme výhodu přepočtu diagonály především pro rozdělení s vysokou střední hodnotou (zde chí-kvadrát s jedenácti stupních volnosti).

LITERATURA

- [1] Neuts, M. F. (1981): Matrix geometric solutions in stochastic models. *Johns Hopkins University Press, Baltimore, MD*
- [2] Johnson, M. A., Taaffe, M. R.: The denseness of phase distributions. *Research Memorandum No. 88-20*, School of Industrial Engineering, Purdue University
- [3] Asmussen, S., Olsson, M., Nerman, O.: Fitting Phase-type Distributions via the EM Algorithm. *Scandinavian Journal of Statistics* **23**

ÚSTAV TECHNICKÉ MATEMATIKY, ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE, KARLOVO NÁM. 13,
128 35 PRAHA 2

E-MAIL: dohnal@fsik.cvut.cz, m.meca@seznam.cz