ROBUST'2000, 247 – 256

SROVNÁNí APROXIMAČNíCH METOD V TEORII RIZIKA

MARTIN ROTKOVSKÝ

ABSTRAKT. One of the main terms of the risk theory is so called individual model, which describes for example total aggregate claim from n insurance policies.

$$S^{ind} := \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} C_i D_i$$

For evaluating of distribution of S^{ind} we must calculate convolutions of higher order, what is unpossible in general case and very computationally difficult in case of discrete distributions, so it is not the optimal way to get good (and fast) results. We can approximate this model by various methods. In this contribution I will compare different approximations. First theoretically and later for a concrete example, where X_i are i.i.d. with distribution $C_i \sim Exp(\theta)$ and $D_i \sim A(p)$.

Резюме: Одним из основных понятий теории риска являетцятак называемаяиндивидуалнаямодељ, которая описывает например аккумулированнуюпотерю из числа n цтраховых полисов.

$$S^{ind} := \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} C_i D_i$$

Для опредделения распределения S^{ind} мы должны были бы сосчитатьмногократные конволюции, что, учитывая сложность расчета дискретных случайных величин и почти невозможность расчета общих распределений, не являецця оптимаљным путем. Поетому ета модељ бывает часто аппроксимирована с помощью разных методов. В настоящей статье мы занимаемся сравнением етих методов. Сначала на теоретическом уровне, и потом отдельные методы цравниваютця на конкретном примере, когда X_i н.о.р.с.в. с распределением $C_i \sim Exp(\theta)$ и $D_i \sim A(p)$.

1. Úvod

Jedním ze základních pojmů teorie rizika je takzvaný individuální model, který popisuje kumulovanou ztrátu z n pojistných smluv. Vychází z toho, že i-tá pojistná smlouva s určitou nenulovou pravděpodobností $1 - p_i$ nepřinese ani jednu škodu a s doplňkovou pravděpodobností přinese ztrátu, jejíž rozdělení je dáno distribuční funkcí V_i. Celkovou ztrátu z i-té smlouvy je tedy možno zapsat $X_i = C_i D_i$, kde C_i má d.f. V_i a D_i má alternativní rozdělení s parametrem p_i. Celková škoda z n smluv je tedy

$$S^{ind} := \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} C_i D_i$$

Pro určení rozdělení S^{ind} bychom museli počítat mnohonásobné konvoluce, což je výpočtově náročné pro diskrétní a téměř nemožné pro obecně rozdělené náhodné

2000 *Mathematics Subject Classification.* Primary 62P05.
Klíčová slova. Teorie rizika, individuální model, aproximace, kumulovaná ztráta.
Tato práce vznikla za podpory grantu MSM 113200008.

veličiny. Proto je tento model často aproximován různými metodami. V tomto příspěvku se zabýváme jejich srovnáním. Nejdříve teoreticky a pak jsou jednotlivé metody srovnány pro konkrétní příklad, kdy X_i jsou i.i.d. s rozdělením $V_i \sim Exp(\theta)$ a $D_i \sim A(p)$. V tomto, v praxi hojně používaném, příkladě lze totiž určit poměrně snadno konvoluci libovolného počtu náhodných veličin a tím se dobrat i k poměrně slušným výsledkům. Na závěr je provedeno srovnání aproximace exponenciálním rozdělením vůči reálným datům konkrétní pojišťovny.

0 - 1 Hrubá výpočetní síla

Věta 0 - 1: Jsou-li X a Y diskrétní náhodné veličiny s n a m body nespojitosti má $X + Y$ diskrétní rozdělení s nejméně $n + m - 1$ a nejvíce $n.m$ bodů nespojitosti.

Pokud tedy děláme konvoluci 2^k i.i.d. náhodných veličin s řešetovitým rozdělením, které má body nespojitosti pouze v bodech $0, \dots, m\delta$ dosáhneme u konvoluce přibližně $m2^k$ bodů nespojitosti. Pro výpočet konvoluce dvou veličin s m body nespojitosti potřebujeme udělat řádově m^2 operací. Přibližný celkový počet aproximací při této metodě je dán následující tabulkou:

m/k	**1**	**2**	**3**	**5**	**10**	**15**
2	4	16	64	1024	1048576	1073741824
5	25	100	400	6400	6553600	6710886400
10	100	400	1600	25600	26214400	26843545600
50	2500	10000	40000	640000	655360000	6,71089E+11
100	10000	40000	160000	2560000	2621440000	2,68435E+12

Vzhledem k obvyklému rozsahu počtu smluv $> 2^{13}$ by bylo spočtení konvoluce pro dostatečně jemná dělení nedostupné v reálném čase.

Pozn. 1: Zabýváme se pouze počty $m = 2^k$, pro ostatní m je možné výsledek dosáhnout tak, že

$$m = \sum_{k=0}^{\infty} a_k 2^k, \quad kde \quad a_k = \left\{ \begin{array}{c} 0 \\ 1 \end{array} \right. ,$$

$$X^{*m} = X^{*k_1} \otimes \cdots \otimes X^{*k_l},$$

kde k_1 je odpovídá prvnímu k, pro něž $a_k = 1$ a podle dvojkového rozkladu m odpovídají i další $k_i, \quad i = 1, \dots, l$ výskytům jedničky ve dvojkovém zápise m.

0-2 Jak porovnávat aproximace

Pro určování kvality různých druhů aproximací S^{ind} je potřeba stanovit metriku, která bude stanovovat "vzdálenost" od původního modelu.
Standartní volbou je tzv. stopp-loss metrika řádu 1:

$$d_1(X, Y) := \sup_t \left| \int_t^{\infty} (x - t) d(F_X(x) - F_Y(x)) \right| = \sup_t |E(X - t)_+ - E(Y - t)_+|$$

Ve výše uvedeném příkladu pojistných smluv tato metrika určuje odchylku dvou stopp-loss pojistných odpovídajících jednotlivým modelům (Gerber 1981, str. 71). Existuje zobecnění d_1 na $d_m, m \in \mathbb{N}$, která je takzvanou ideální metrikou řádu m, což znamená:

Definice 0-1: Pro náhodné veličiny X_1, X_2, Z, kde (X_1, X_2) je nezávislý se Z a nenunlovou konstantu c platí:
$(i) \quad d_m(X_1 + Z, X_2 + Z) \leq d_m(X_1, X_2)$ a

(ii) $d_m(cX_1, cX_2) = |c|^m d_m(X_1, X_2)$.
a tím platí i

Lemma 0-2:
(a)Pro $X_1, \ldots, X_n$ nezávislé a $Y_1, \ldots, Y_m$ nezávislé a $c_i > 0$

$$d_m\left(\sum_{i=1}^n c_i X_i, \sum_{i=1}^n c_i Y_i\right) \leq \sum_{i=1}^n c_i^m d_m(X_i, Y_i).$$

(b)Pokud $E(X^j - Y^j) = 0, 1 \leq j \leq m$, potom

$$d_m(X,Y) \leq \frac{k_m(X,Y)}{m!} = \int_{-\infty}^{\infty} \frac{|x|^{m-1}}{(m-1)!}|F_X(x) - F_Y(x)|dx \leq E\left|X|X|^{m-1} - Y|Y|^{m-1}\right|$$

Důkaz :(a)Z trojúhelníkové nerovnosti a ideality. (b) Rachev 1991, str. 320.

2. Teoretická porovnání

Pro aproximaci individuálního modelu se často používá tzv. kolektivní model, který lze zapsat následujícím způsobem:

$$S^{coll} := \sum_{i=1}^N Z_i,$$

kde Z_i jsou i.i.d. s $Z_i \sim V$ a $N \sim Po(\mu)$ a Z_i, N jsou nezávislé.

$$\mu := \sum_{i=1}^n p_i, \quad V := \sum_{i=1}^n \frac{p_i}{\mu} V_i$$

V příkladě pojistných smluv lze N nahlížet jako celkový počet pojistných událostí z n smluv a Z_i jako výši konkrétní škody. S^{coll} můžeme tedy zapsat

$$S^{coll} = \sum_{i=1}^n S_i^{coll}, \quad (+)$$

kde $S_i^{coll} = \sum_{j=i}^{N_i} Z_{ij}$, kde $N_i \sim Po(p_i)$ a $Z_{ij} \sim V_i$.

Pozn. 1: $ES^{ind} = ES^{coll}$, ale $VarS^{ind} = \sum_{i=1}^n p_i Var(C_i) + \sum_{i=1}^n p_i(1-p_i)(EC_i)^2 <$

$$< VarS^{coll} = \sum_{i=1}^n p_i Var(C_i) + \sum_{i=1}^n p_i (EC_i)^2.$$

Pozn. 2: Vyjádření (+) lze upravit tak, aby se shodovaly oba momenty. Tento model budu dále označovat S_{param}^{coll}. Při $a_i := EC_i$ a $b_i := EC_i^2$ vypadá takto:

$$S_i^{coll} = \sum_{j=1}^{N_i} Z_{ij}, \quad Z_{ij} \sim u_i C_i, \quad N_i \sim Po(\mu_i) \quad \mu_i := \frac{p_i b_i}{b_i - p_i a_i^2}, \quad u_i := \frac{p_i}{\mu_i}$$

I - 0 Specifické výpočty pro exponenciální model

Nejprve provedeme několik pomocných výpočtů pro případ výše zmíněného homogeního ($p_i = p$) exponenciálního ($V_i \sim Exp(\theta)$) modelu, ty využijeme později v praktických příkladech

$$d_1(X_1, S_1^{coll}) = \sup_t \left|\int_t^{\infty} (x-t) d(F_{X_1}(x) - F_{S_1^{coll}}(x))\right| \leq \left|\int_0^{\infty} x \sum_{k=0}^{\infty} f_{\Gamma(\theta,k)}(x) c_k dx\right| = \odot$$

kde $f_{\Gamma(\theta,k)}(x)$ je hustota Γ rozdělení s příslušnými parametry (pro $k=0$ uvažujeme $\delta-$ funkci a

$$c_k = \begin{cases} |1-p-e^{-p}| & k=0 \\ p(1-e^{-p}) & k=1 \\ \frac{e^{-p}p^k}{k!} & k \geq 2 \end{cases}$$

což odpovídá rozdílu mezi jednotlivými pravděpodobnostmi k-tých konvolucí C_i tvořících X_1 a S_1^{coll}. Odtuď dostáváme

$$\odot \leq \sum_{k=0}^{\infty} \int_0^{\infty} |xc_k f_{\Gamma(\theta,k)}(x)|dx \stackrel{(a)}{=} p(1-e^{-p})\theta + \sum_{k=2}^{\infty} \frac{e^{-p}p^k}{k!} k\theta = 2p(1-e^{-p})\theta.$$

Při přechodu (a) jsme využili znalost o Γ rozdělení. Podobně spočteme i d_2:

$$d_2(X_1, S_1^{coll}) = \sum_{k=0}^{\infty} \int_0^{\infty} |x^2 c_k f_{\Gamma(\theta,k)}|dx = p(1-e^{-p})2\theta^2 + \sum_{k=2}^{\infty} \frac{e^{-p}p^k}{k!} \theta^2(k+k^2) =$$

$$= 4p\theta^2(1-e^{-p}) + p\sum_{k=1}^{\infty} \frac{e^{-p}p^k}{k!} \theta^2(k+1) = p^2\theta^2 + 4p\theta^2(1-e^{-p})$$

Nyní uvažujme ještě parametrizovaný model S_{param}^{coll}. Vzhledem k tomu, že je formální zápis stejný i pro tento model a liší se pouze parametry, proto spočteme hodnotu těchto parametrů a poté je dosadíme do výsledků z předešlého modelu:

$$\mu\prime := \mu_i = \frac{p_i b_i}{b_i - p_i a_i} = \frac{2p}{2-p} \Longrightarrow u = u_i = \frac{p_i}{\mu_i} = \frac{2-p}{2}$$

Rozdělení zachovávají typ a mění pouze parametr, tzn. $Z_{ij} \sim \frac{2-p}{2} Exp(\theta) \sim Exp(\frac{2-p}{2}\theta)$ a $N_i \sim Po\left(\frac{2p}{2-p}\right)$. Odtud dosazením do výsledků pro jednoduchý model obdržíme:

$$d_1(X_1, S_{param1}^{coll}) = 2p(1-e^{-\frac{2}{2-p}p})\theta, \quad d_2(X_1, S_{param1}^{coll}) = p^2\theta^2 + 4\frac{2-p}{2}p\theta^2(1-e^{-\frac{2}{2-p}p})$$

I - 1 Prosté využití vlastností metriky d_1

Zde i v dalších odstavcích je S^{coll} použito pro teoretické vyjádření obou modelů, protože jejich rozklad je formálně shodný.Pokud použijeme lemma 0-2 b) dostáváme:

$$d_1(S^{ind}, S^{coll}) \leq \sum_{i=1}^{n} d_1(X_i, S_i^{coll})$$

Pro homogení model dostáváme:

$$= nd_1(X_1, S_1^{coll}),$$

pokud bychom neměli informaci o exponencialitě rozdělení V_i mohli bychom rozdíl odhadnout podle lemmatu 0-2 takto:

$$\leq n(E|X_1| + E|S_1^{coll}|) \leq 2np\theta \quad (1)$$

Pokud využijeme jemnější odhad d_1 z bodu I-0 dostaneme. Pro jednoduchý model:

$$d_1(S^{ind}, S^{coll}) \leq 2np(1-e^{-p})\theta \quad (2)$$

Pro parametrizovaný model:

$$d_1(S^{ind}, S_{param}^{coll}) \leq 2np(1-e^{-\frac{2}{2-p}p})\theta \quad (3)$$

I-2 Využití ideality metriky d_2

Poněkud sofistikovanější metodou je využít vztahu mezi d_1 a d_2, kterou popisuje následující lemma:

Lemma I - 1: Pokud $E(X^j - Y^j) = 0$, $j = 1, 2$ platí následující vztah

$$d_1(X,Y) \le \frac{4}{\sqrt{\pi}} d_2(X,Y)$$

Z nutnosti shody prvních dvou momentů lze odhad použít pouze pro parametrizovaný model. Budeme tedy nejprve odhadovat $d_2(S^{ind}, S^{coll})$:

$$d_2(S^{ind}, S^{coll}) \le \sum_{i=1}^{n} d_2(X_i, S_i^{coll})$$

pro homogení model tedy platí $= nd_2(X_1, S_1^{coll})$, bez znalosti rozdělení můžeme opět pouze podle lemmatu 0-2 odhadnout:

$$d_2(X_1, S_1^{coll}) \le \frac{1}{2}(EX_1^2 + E(S_1^{coll})^2) = EX_1^2 = pb$$

Pro parametrizovaný model by tento hrubý odhad odpovídal:

$$d_1(S^{ind}, S^{coll}) \le \frac{4}{\sqrt{\pi}} \sqrt{2np\theta^2} \quad (4)$$

Při použití jemnějšího odhadu z bodu I - 0 dostaneme:

$$d_2(S^{ind}, S^{coll}_{param}) = \frac{4}{\sqrt{\pi}} \sqrt{n \left[p^2\theta^2 + 4\frac{2-p}{2} p\theta^2 (1 - e^{-\frac{2}{2-p}p}) \right]} \quad (5)$$

I - 3 Využití vlastností d_1 a Poissonova rozdělení

V následujících dvou odstavcích budeme využívat k odhadu podobnost rozdělení Z_{ij} s C_i ($Z_{ij} = u_i C_i$) a tím danou podobnost jejich příslušných konvolucí společně s využitím znalostí o Poissonově rozdělení. Následující věta dává tedy jistým způsobem optimální odhad pokud nevíme nic o rozdělení C_i.

Věta I - 2: Pro parametrizovaný model a $C_i \ge 0$ s.j. platí pro $\forall \Delta_i \ge 1$:

$$d_1(S^{ind}, S^{coll}) \le \sum_{i=1}^{n} p_i^2 \tau_i,$$

kde

$$\tau_i := a_i + \Delta_i v_i + max(\Delta_i a_i v_i, 2a_i \tilde{v}_i + (1 + \Delta_i a_i v_i p_i) u_i),$$
$$v_i := \frac{a_i^2}{b_i} \le \frac{1}{p_i}, \quad p_i v_i \le 1 - \Delta_i^{-1}, \quad \tilde{v}_i := \frac{a_i^2}{b_i - p_i a_i^2}$$

Důkaz: Rachev 1991, str. 329-331.

Po dosazení všech parametrů dostaneme pro exponenciální model:

$$d_1(S^{ind}, S^{coll}) \le np^2\theta \left[1 + \frac{2}{2-p} + \frac{p}{2} \right] + np^2 \left[\frac{1}{2-p} + \frac{2-p}{2} \right] (6)$$

I - 4 Využití vlastností d_2 a Poissonova rozdělení

Analogicky jako v odstavci I - 2 odhadneme nejprve d_2 a z něj pak pomocí lemmatu I - 1 i d_1. Postup odhadu d_2 bude analogický odhadu jako v I-3:

Věta I - 3: Pro $C_i \ge 0$ s.j. platí pro parametrizovaný model:

$$d_2(S^{ind}, S^{coll}) \le \frac{1}{2} \sum_{i=1}^{n} p_i^2 \tau_i^*$$

kde $\tau_i^* = b_i + 3a_i^2 + \Delta_i a_i^2 + 2\tilde{v}_i b_i^2 + b_i u_i^2 + \Delta_i a_i p_i$, kde Δ_i, $\tilde{v}_i$ jsou stejné jako ve větě I-2.
Tím máme

$$d_1(S^{ind}, S^{coll}) \leq \frac{4}{\sqrt{2\pi}} \left(\sum_{i=1}^{n} p_i^2 \tau_i^* \right)^{\frac{1}{2}}$$

Důkaz: Rachev 1991, str. 331-333

Po dosazení dostaneme

$$\tau_1^* = 2\theta^2 + 3\theta^2 + \theta^2 \frac{2}{2-p} + \theta^4 \frac{8}{2-p} + 2\theta^2 \left(\frac{2-p}{2}\right)^2 + \theta \frac{2}{2-p} p$$

a tedy

$$d_1(S^{ind}, S^{coll}) \leq \frac{4}{\sqrt{2\pi}} \left[np^2 \left(\frac{8}{2-p}\theta^4 + (7 + \frac{2}{2-p} - 2p + \frac{p^2}{2})\theta^2 + \frac{2p}{2-p}\theta \right) \right]^{\frac{1}{2}} \qquad (7)$$

I - 5 Vyžití všech znalostí o modelu

Postup výpočtu bude podobný jako u neuvedených důkazů z předešlých dvou odstavců. S tím, že na rozdíl od obecného případu jsme schopni přesně napočítat k-té konvoluce $C_i \sim Exp(\theta)$, které mají gama rozdělení $C_1^k \sim \Gamma(\theta, k)$ a tak dostaneme velmi přesný odhad $d_1(S^{ind}, S^{coll})$ pro tento model.

$$d_1(S^{ind}, S^{coll}) = \sup_t | \int_t^{\infty} (x-t) d(F_{S^{ind}}(x) - F_{S^{coll}}(x))| \leq$$

$$| \int_0^{\infty} (x-t) \sum_{k=1}^{\infty} c_k f_{\Gamma(\theta,k)}(x) dx| \leq \sum_{k=1}^{\infty} c_k E_{\Gamma(\theta,k)} X = \sum_{k=1}^{\infty} c_k k\theta \qquad (8)$$

kde pro neparametrizovaný model platí

$$c_k = \left| \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k} - \frac{e^{-np}(np)^k}{k!} \right|$$

I - 6 Aproximace homogeního modelu pomocí normálního rozdělení

V tomto odstavci bude na rozdíl od všech ostatních daleko těžší využít plnou sílu věty při znalosti konkrétního rozdělení C_i a horní mez z věty budeme pro tento typ rozdělení pouze odhadovat.

Věta I - 4: Pokud jsou X_i i.i.d. náhodné veličiny s $a := a_1$ a $\sigma^2 := Var(C_1)$, $p := p_1$, pak platí

$$d_1(S^{ind}, S^{coll}_{param}) \leq 11,5[p\sigma^2 + p(1-p)a^2]^{\frac{1}{2}} \left[\tau_3(X_1, Y) + \tau_3 \left(\sum_{i=1}^{N_1} C_i, Y \right) \right]$$

kde $Y \sim N(0,1)$, $N_1 \sim Po(\mu_1)$ a při označení

$$\tau_3(X, Y) := max((E|\tilde{X}| + E|Y|), \frac{1}{3}(E|\tilde{X}|^3 + E|Y|^3)), \quad \tilde{X} := \frac{X - EX}{\sqrt{Var(X)}}$$

Důkaz: [2], str. 328 a v [5], str. 27 lze nalézt určité zpřesnění (zpřesnění)

Nejprve uveďme absolutní momenty normálního rozdělení.

$$E|Y| = \sqrt{\frac{2}{\pi}}, \quad E|Y|^3 = \sqrt{\frac{8}{\pi}}$$

Pro normalizaci je potřeba znát hodnotu rozptylu, ta je (Gerber 1981, str. 50)

$$VarX_1 = (1-p)p\theta^2 + p\theta^2 = (2-p)p\theta^2$$

Nyní spočteme momenty:

$$E|X_1 - EX_1| = p(1-p)\theta + p\left[\int_{-p\theta}^{0} -x\frac{1}{\theta}e^{-x/\theta}dx + \int_0^\infty x\frac{1}{\theta}e^{-x/\theta}dx\right]$$

$$= p(1-p)\theta + p\theta e^{p\theta}(1-e^p(1-p)) = p\theta\left[1-p+e^{p\theta}(1-e^p(1-p))\right]$$

Spočíst lze i hodnotu $E|\tilde{X}|^3$, ale použjeme odhad:

$$E|X_1 - EX_1|^3 \leq (p\theta)^3 + pEC_1^3 \leq (p\theta)^3 + 6p\theta^3$$

v prostředním členu odpovídá první sčítanec záporné a druhý kladné části rozdělení $X_1 - EX_1$.

Podobným trikem budeme postupovat i pro odhad $\tau_3(S_1^{coll}, Y)$. Vzhledem k parametrizaci zvolené tak, aby se shodovali první dva momenty je rozptyl stejný jako pro X_i

$$E|S_1^{coll} - ES_1^{coll}| \leq p\theta + \sum_{k=1}^{\infty}\frac{e^{-p}p^k}{k!}k\theta \leq 2p\theta$$

a pro výpočet třetího momentu uvedeme nejprve hodnotu třetího momentu $\Gamma(\theta,k)$ rozdělení:

$$E_{\Gamma(\theta,k)}X^3 = \theta^3(k^3 + 3k^2 + 2k)$$

s tímto vyjde

$$E|S_1^{coll} - ES_1^{coll}|^3 \leq (p\theta)^3 + \theta^3\sum_{k=1}^{\infty}\frac{e^{-p}p^k}{k!}(2k+3k^2+k^3) = \theta^3(2p^3+6p^2+6p)$$

Výraz, který by vzniknul po dosazení do věty I - 4 budeme označovat (9).

3. Numerické aproximace

Věta 0 - 1 říká kolik bodů nespojitosti má konvoluce diskrétních n.v. a víme, že minimálního počtu lze dosáhnout v případě řešetovitého rozdělení. Ovšem i pro toto rozdělení nám počet bodů nespojitosti roste velmi rychle. Proto bude vhodné rozdělení konvoluce aproximovat novou náhodnou veličinou, která má méně bodů nespojitosti.

Nejprve zvolme diskrétní náhodnou veličinu K_i aproximující rozdělení C_i, aby se od C_i ve smyslu metriky d_1 příliš nevzdálila. Pokud rozdělení C_i nemá příliš těžký chvost zvolím δ tak, aby pro použitelné n platilo $\int_{n\delta}^{\infty}(x-n\delta)dx \leq \delta$, pak rozdělení K_i volím s následující distribuční funkcí

$$F_{K_i}(x) = \sum_{i=0}^{n} I_{[x\leq i\delta]}P_{i\delta}(\delta), \quad kde \quad P_{i\delta}(\delta) = P(C_1 \in [i\delta,(i+1)\delta[)$$

Vzhledem k tomu, že $C_i \overset{st}{\geq} K_i$ můžeme odhadnout

$$d_1(C_1, K_1) \leq \sum_{k=0}^{n-1} \int_{k\delta}^{(k+1)\delta} xF_{C_1}(dx) - i\delta P_{i\delta}(\delta) + \int_{n\delta}^{\infty} (x - n\delta)F_{C_1}(dx) \leq 2\delta$$

I pokud volíme malý krok $\delta \leq EX/1000$ nebude pro rozdělení s nerostoucí intenzitou poruch problém najít n dostatečně malé ($n \leq 10000$) požadované vlastnosti.
Vlastní algoritmus pro redukci počtu bodů nespojitosti je takový, že při každém kroku vynechám liché násobky délky předešlého kroku a jejich pravděpodobnost přidáme nejbližšímu nižšímu sousedovi. Samozřejmě, že lepším způsobem redukce bodů by bylo rozdělit pravděpodobnost mezi oba sousedy rovným dílem (jak ukazuje tabulka na této stránce), ale pak bychom ztratili jistotu o tom, kterým směrem se vzdalujeme od originální náhodné veličiny.
Tím získáme rozdělení, které má opět pouze n bodů nespojitosti je řešetovité a algoritmus nenabývá na výpočtové náročnosti.

Počet veličin	**k**	d_1	**Očištěná pro** $p = 0,1$	**Oběma sousedům**
2	1	1	0,19	1
8	3	12	4,41373116	8
128	7	448	312,4872656	256
1024	10	5120	4035,898124	2816
8192	13	53248	44575,185	28672

Nyní ještě zbývá popsat metodu jak získat m-té konvoluce pro libovolné $m \in \mathbb{N}$. Nejprve najdu největší l takové, že $2^l | m$ a pro něj najdu aproximaci $m/2^l$-té konvoluce. Tu najdeme tak, že ve všech dříve vytvořených aproximacích 2^i-tých konvolucí vynecháme body a přeneseme jejich pravděpodobnosti do nejbližších nižších sousedů, násobků kroku $2^{m/2^l}\delta$. Toto shluknutí nám způsobí u každé veličiny chybu $2^{m/2^{l-1}}\delta$. Konvoluce takto vzniklých aproximací nám pak dáva aproximaci požadované konvoluce $2^{m/2^l}$, na kterou pak provedeme l kroků původního algoritmu. Při tomto postupu se zvyšuje počet bodů, ale ten nikdy nepřesáhne $2n$.

Záleží tedy pouze na zvoleném δ jak přesná bude tato aproximace. Algoritmus je možné uvažovat za spočitatelný pro n řádu 10^4, neboť má výpočtovou složitost přibližně $(2n)^2 k$.
V případě exponenciálního rozdělení je potřeba uvažovat ještě distanci původní náhodné veličiny $C_i \sim Exp(\theta)$ od její aproximace K_i a jejich příslušných konvolucí

$$d_1(C_i^{*2^k}, K_i^{*2^k}) \leq 2^{k+1}\delta,$$

vzhledem k vlastnostem metriky d_1 platí i pro obecné m vztah $d_1(C_1^{*m}, K_i^{*m}) \leq 2m\delta$.
Škoda $X_i := C_i D_i$ má nenulovou pravděpodobnost $P(X_i = 0)$. Proto při jednotlivých krocích algoritmu bude mít neshluknutá náhodná veličina určitou pravděpodobnost toho, že je rovna nule. S touto pravděpodobností shluknutí nic neudělá. O tuto hodnotu je proto možné odhad očistit (viz. příklad v tabulce).

4. Srovnání aproximačních metod

Pro konkrétní příklady zvolme různé počty smluv n, rozdělení s různou pravděpodobností výskytu škody p, ale její výši ponechme stejnou tak, jak vyjde v příkladu uvedeném ve IV-té kapitole, t.j. $\theta = 0,13$.

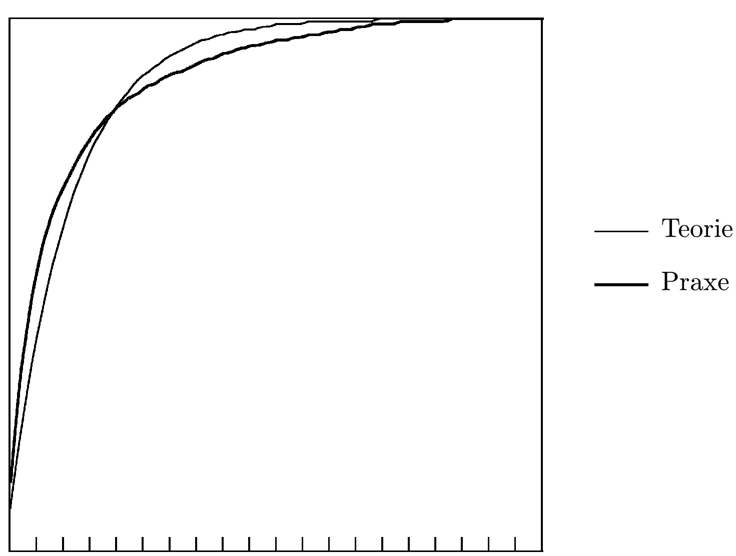
Formulka číslo	**1**	**2**	**3**	**4**
n	10000	10000	1000	1000
p	0,1	0,01	0,1	0,01
θ	0,13	0,13	0,13	0,13
(1)	260,00	26,000	26,000	2,6000
(2)	24,74	0,259	2,474	0,0259
(3)	25,98	0,260	2,598	0,0260
(4)	13,12	4,149	4,149	1,3120
(5)	6,42	0,655	2,032	0,207
(6)	174,97	1,759	17,497	0,1759
(7)	6,13	0,592	1,939	0,187
(8)	6,63	0,065	0,663	0,0065
(9)	4,01	3,244	4,006	3,2445
Numericky	5,19	3,851	0,288	0,1865

Je zřejmé, že neexistuje univerzální odhad nejlepší pro všechny kombinacie n a p. Pro vysoká n a p je nejlepším odhadem aproximace přes normální rozdělení (9), pro malá p je nejlepší odhad přes metriku d_2 (5) a (7), či přímý odhad (8).

5. Příklad z praxe

V předešlých kapitolách jsme se zabývali rozdílem pro exponenciální rozdělení C_i. Nebyla ovšem řešena otázka jak dobrou aproximací je použití tohoto rozdělení místo skutečného rozdělení. Následující graf ukazuje jak vypadá rozdíl distribučních funkcí výše škody získané z dat jedné pojišťovny (z havarijního pojištění aut) vůči exponenciálnímu rozdělení se stejnou střední hodnotou.

Poznamenejme, že hodnota metriky d_1 pro skutečnou výši škod a její exponenciální aproximaci zachovávající střední hodnotu je 0,005183, což představuje 4,4589% střední hodnoty výše škod.



6. Literatura

[1] Gerber, H. (1981): An Induction to Mathematical Risk Theory. Huebner Foundation Monograph.

[2] Rachev, S.T.(1991): Probability Metrics and the Stability of Stochastic Models. John Wiley & Sons

[3] Rotkovský, M.(1999): WDS'99 Proceedings of contributed papers. Matfyzpress

[4] Zolotarev, V. M.(1976): Metric distances in spaces of random variables and their distributions. Math. USSR sb. 30

[5] Zolotarev, V. M.(1986): Contemporary Thery of Summation of Random Variables. Nauka, Moscow. (In Russian) and their distributions. Math. USSR sb. 30

UK MFF, KPMS, Sokolovská 83, 186 75 Praha 8
E-mail address: rotkovsk@karlin.mff.cuni.cz