Rekurentné triedy diskrétnych rozdelení pravdepodobnosti a odhadovanie ich parametrov

Gábor Szűcs

e-mail: Gabor.Szucs@fmph.uniba.sk

Fakulta matematiky, fyziky a informatiky, Univerzita Komenského v Bratislave

ROBUST 2016 (12.09.2016)

Obsah prezentácie

1. Úvod, motivácia výskumu
2. Rekurentné triedy diskrétnych rozdelení
3. Odhad parametrov rekurentných tried
4. Praktická aplikácia
5. Záver

Motivácia

Model kolektívneho rizika

$$S = \sum_{i=1}^{N} X_i$$

$$\begin{aligned} S &= \text{ celková výška poistných plnení,} \\ N &= \text{ počet poistných plnení,} \\ X_i &= \text{ výška poistných plnení.} \end{aligned}$$

Motivácia

Model kolektívneho rizika

$$S = \sum_{i=1}^{N} X_i$$

$$\begin{aligned} S &= \text{ celková výška poistných plnení,} \\ N &= \textbf{ počet poistných plnení,} \\ X_i &= \text{ výška poistných plnení.} \end{aligned}$$

Motivácia

Ciele výskumu

* Nájsť vhodný model pre počet poistných plnení z triedy rekurentných rozdelení,
* odhadnúť parametre modelu pre počet poistných plnení,
* overiť kvalitu odhadnutého modelu,
* odhadnúť rozdelenie celkovej výšky poistných plnení.

Publikácie

→ Baradlaiová, Z. 2016. *Rekurentné triedy počtu poistných plnení a zovšeobecnené rekurzie v modeli kolektívneho rizika.* Diplomová práca, Fakulta matematiky, fyziky a informatiky, Univerzita Komenského v Bratislave.

→ Szűcs, G. 2015. Metódy odhadu parametrov rekurentných tried diskrétnych rozdelení pravdepodobnosti. *Forum Statisticum Slovacum*, č. 6/2015, s. 154-159. ISSN 1336-7420.

→ Szűcs, G. 2013. Schröterova trieda rozdelení. *Forum Statisticum Slovacum*, IX. ročník, č. 7/2013, s. 227-232. ISSN 1336-7420.

$\mathcal{R}_k$-trieda diskrétnych rozdelení

Definícia.

Uvažujme diskrétnu náhodnú premennú N s nekumulatívnym pravdepodobnostným rozdelením $\{p_n\}_{n=0}^{\infty}$, kde $p_n = \Pr(N = n)$; $n = 0, 1, 2, \ldots$ Hovoríme, že rozdelenie $\{p_n\}_{n=0}^{\infty}$ patrí do rekurzívnej triedy $\mathcal{R}_k$, ak je splnený vzťah

$$p_n = \sum_{i=1}^{k} \left(a_i + \frac{b_i}{n} \right) p_{n-i} \qquad \text{pre } n = 1, 2, 3, \ldots,$$

kde k je prirodzené číslo, a_i, b_i sú reálne parametre pre $i = 1, 2, \ldots, k$ a $p_n = 0$ pre všetky $n < 0$.

Špeciálne prípady $\mathcal{R}_k$-triedy

Ak $k = 1$, tak dostávame tzv. **Panjerovu triedu**: $\quad \mathcal{R}_1 \quad \Leftrightarrow \quad Pan(a_1, b_1)$.

$$p_n = \left(a_1 + \frac{b_1}{n}\right) p_{n-1} \qquad \text{pre } n = 1, 2, 3, \ldots$$

Špeciálne prípady $\mathcal{R}_k$-triedy

Ak $k = 1$, tak dostávame tzv. **Panjerovu triedu**: $\mathcal{R}_1 \quad \Leftrightarrow \quad Pan(a_1, b_1)$.

$$p_n = \left(a_1 + \frac{b_1}{n}\right) p_{n-1} \qquad \text{pre } n = 1, 2, 3, \ldots$$

<u>Rozdelenia patriace do Panjerovej triedy</u>:

$$N \sim Pan(0, \lambda) \iff N \sim Pois(\lambda),$$

$$N \sim Pan\left(-\frac{p}{1-p},\ (m+1)\frac{p}{1-p}\right) \iff N \sim Bin(m, p),$$

$$N \sim Pan(1-p,\ (1-p)(r-1)) \iff N \sim NegBin(r, p),$$

$$N \sim Pan(1-p,\ 0) \iff N \sim Geom(p).$$

Panjerova trieda rozdelení

Panjerova trieda rozdelení (1981):

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1} \qquad \text{pre } n = 1, 2, 3, \ldots$$

Katzova trieda rozdelení (1945):

$$\frac{p_n}{p_{n-1}} = \frac{\alpha + \beta(n-1)}{n} \qquad \text{pre } n = 1, 2, 3, \ldots, \quad \alpha \geq 0, \quad \beta < 1.$$

Panjerova trieda rozdelení

Panjerova trieda rozdelení (1981):

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1} \qquad \text{pre } n = 1, 2, 3, \ldots$$

Panjerova trieda useknutých rozdelení: $\overline{Pan}_u^v(a, b)$.

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1} \qquad \text{pre } n = u + 1, u + 2, \ldots, v.$$

Špeciálne prípady $\mathcal{R}_k$-triedy

Ak $k = 2$ a $a_2 = 0$, tak dostaneme tzv. **Schröterovu triedu**:

$\mathcal{R}_2 \supset Schr(a_1, b_1, b_2)$.

$$p_n = \left(a_1 + \frac{b_1}{n}\right) p_{n-1} + \frac{b_2}{n} p_{n-2} \qquad \text{pre } n = 1, 2, 3, \ldots$$

$p_{-1} = 0$.

Špeciálne prípady $\mathcal{R}_k$-triedy

Schröterova trieda: $Schr(a, b, c)$.

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1} + \frac{c}{n} p_{n-2} \qquad \text{pre } n = 1, 2, 3, \dots$$

$$p_{-1} = 0.$$

Schröterova trieda useknutých rozdelení: $\overline{Schr}_u^v(a, b, c)$.

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1} + \frac{c}{n} p_{n-2} \qquad \text{pre } n = u+1, u+2, \dots, v.$$

$$p_n = 0 \qquad \text{pre } n \in \mathbb{Z} \setminus \{u, u+1, u+2, \dots, v\}.$$

Predpoklady

- → Predpokladajme, že máme k dispozícii historické dáta: $n_1, n_2, \ldots, n_d$ (napríklad počty poistných plnení v uplynulom období).
- → Zostrojíme empirické rozdelenie pravdepodobnosti $\pi = (\pi_0, \pi_1, \pi_2, \ldots, \pi_m)$, kde $m = \max\{n_1, n_2, \ldots, n_d\}$.

Predpoklady

- → Predpokladajme, že máme k dispozícii historické dáta: $n_1, n_2, \ldots, n_d$ (napríklad počty poistných plnení v uplynulom období).
- → Zostrojíme empirické rozdelenie pravdepodobnosti $\pi = (\pi_0, \pi_1, \pi_2, \ldots, \pi_m)$, kde $m = \max\{n_1, n_2, \ldots, n_d\}$.

Úloha: nájsť také rozdelenie patriace do rekurentnej triedy diskrétnych rozdelení (napr. do triedy $Schr(a, b, c)$), ktoré "je najbližšie" k nášmu empirickému rozdeleniu.

Inak povedané: potrebujeme odhadnúť parametre zvolenej rekurentnej triedy (napr. v prípade Schröterovej triedy: parametre a, b, c).

Metódy odhadu parametrov

Metódy odhadu parametrov

- metóda kalibračných pätíc (`pentad`),
- optimalizačné metódy

Metódy odhadu parametrov

Metódy odhadu parametrov

- metóda kalibračných pätíc (`pentad`),
- optimalizačné metódy

  minimalizácia zvolenej účelovej funkcie:

  - ▷ testovej štatistiky χ^2-testu,
  - ▷ modifikovanej testovej štatistiky χ^2-testu,
  - ▷ testovej štatistiky K-S testu,
  - ▷ testovej štatistiky C-vM testu,
  - ▷ ...

Minimalizácia testovej štatistiky K-S testu

Metóda založená na minimalizácii hodnoty testovacej štatistiky Kolmogorovovho-Smirnovovho testu:

$$D = \sup_{x} |F_0(x) - F_{data}(x)|,$$

kde $F_0(x)$, pre $x \in \{0, 1, \ldots, m\}$ je kumulatívna distribučná funkcia teoretického rozdelenia a $F_{data}(x)$ je empirická distribučná funkcia získaná z pozorovaných dát.

Minimalizácia testovej štatistiky K-S testu

Viazaná optimalizačná úloha (*napríklad v prípade Schröterovho rozdelenia*):

$$(\tilde{p}_0, \tilde{a}, \tilde{b}, \tilde{c}) = \operatorname*{arg\,min}_{p_0 \in (0,1),\ a,b,c \in \mathbb{R}} D, \tag{1}$$

pri platnosti podmienok

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1} + \frac{c}{n} p_{n-2} \quad \text{pre všetky } n = 1, 2, \ldots, \tag{2}$$

$$\sum p_n = 1, \tag{3}$$

$$p_n \in (0, 1) \ \text{ pre všetky } n = 1, 2, \ldots \tag{4}$$

Minimalizácia testovej štatistiky K-S testu

Viazaná optimalizačná úloha (*napríklad v prípade Schröterovho rozdelenia*):

$$(\tilde{p}_0, \tilde{a}, \tilde{b}, \tilde{c}) = \underset{p_0 \in (0,1),\ a,b,c \in \mathbb{R}}{\arg\min} \sup_x |F_0(x) - F_{data}(x)|,$$

pri platnosti podmienok

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1} + \frac{c}{n} p_{n-2} \quad \text{pre všetky } n = 1, 2, \ldots,$$

$$\sum p_n = 1,$$

$$p_n \in (0, 1) \ \text{ pre všetky } n = 1, 2, \ldots$$

Minimalizácia testovej štatistiky K-S testu

Viazaná optimalizačná úloha (*napríklad v prípade Schröterovho rozdelenia*):

$$(\tilde{p}_0, \tilde{a}, \tilde{b}, \tilde{c}) = \underset{p_0 \in (0,1),\ a,b,c \in \mathbb{R}}{\arg\min} \sup_x \left| \sum_{n=0}^{x} p_n - F_{data}(x) \right|,$$

pri platnosti podmienok

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1} + \frac{c}{n} p_{n-2} \quad \text{pre všetky } n = 1, 2, \ldots,$$

$$\sum p_n = 1,$$

$$p_n \in (0,1) \ \text{ pre všetky } n = 1, 2, \ldots$$

Minimalizácia testovej štatistiky K-S testu

Viazaná optimalizačná úloha (*napríklad v prípade Schröterovho rozdelenia*):

$$(\tilde{p}_0, \tilde{a}, \tilde{b}, \tilde{c}) = \underset{p_0 \in (0,1),\ a,b,c \in \mathbb{R}}{\arg\min} \sup_x \left| \sum_{n=0}^{x} \left[\left(a + \frac{b}{n} \right) p_{n-1} + \frac{c}{n} p_{n-2} \right] - F_{data}(x) \right|,$$

pri platnosti podmienok

$$p_n = \left(a + \frac{b}{n} \right) p_{n-1} + \frac{c}{n} p_{n-2} \quad \text{pre všetky } n = 1, 2, \ldots,$$

$$\sum p_n = 1,$$

$$p_n \in (0, 1) \;\; \text{pre všetky } n = 1, 2, \ldots$$

Numerické riešenie úlohy

Numerické riešenie minimalizačnej úlohy pomocou *penalizačnej metódy*.

Pomocná účelová funkcia

$$H = sup_x \left| \sum_{n=0}^{x} p_n - F_{data}(x) \right| + h_1 \max \left\{ 0; 1 - \sum_{n=0}^{m} p_n \right\} +$$

$$+ h_2 \max \left\{ 0; \sum_{n=0}^{m} p_n - 1 \right\} - h_3 \min \left\{ 0; p_0; p_1; \ldots; p_m \right\}.$$

Numerické riešenie úlohy

Numerické riešenie minimalizačnej úlohy pomocou *penalizačnej metódy*.

Pomocná účelová funkcia

$$H = sup_x \left| \sum_{n=0}^{x} p_n - F_{data}(x) \right| + h_1 \max \left\{ 0; 1 - \sum_{n=0}^{m} p_n \right\} + \\ + h_2 \max \left\{ 0; \sum_{n=0}^{m} p_n - 1 \right\} - h_3 \min \left\{ 0; p_0; p_1; \ldots; p_m \right\}.$$

Metóda simulovaného žíhania

Zadanie

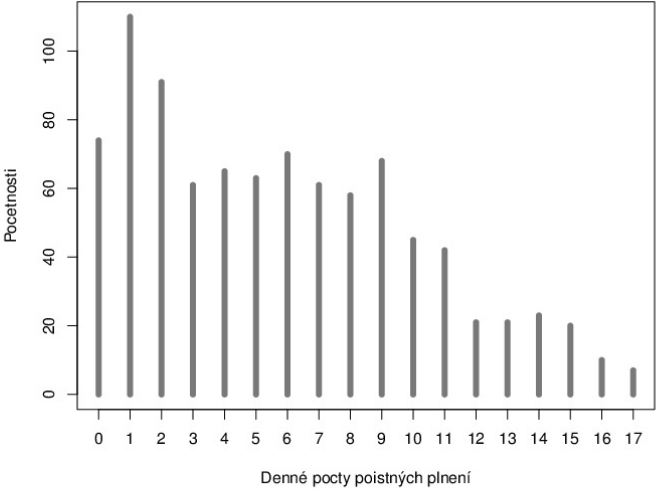
- Uvažujme určitý typ *neživotného poistenia* ,
- *dáta*: denné počty poistných plnení,
- *dĺžka dátového súboru*: $d = 910$.

- Výberové charakteristiky polohy dátového súboru:

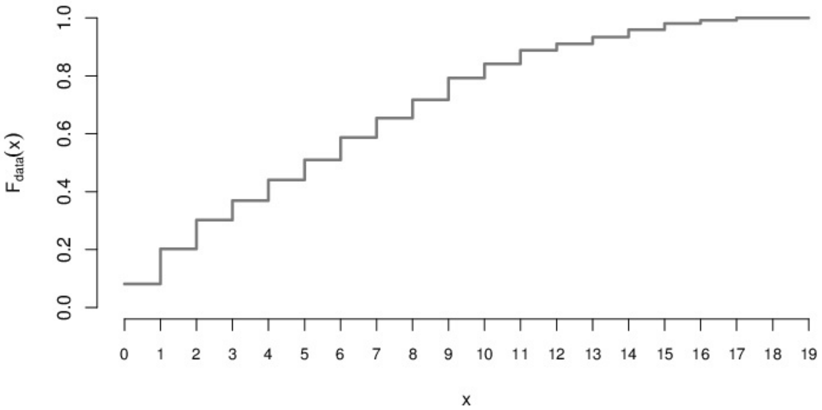
Min.	Prvý kvar.	Medián	Priemer	Tretí kvar.	Max.
0,00	2,00	5,00	5,836	9,00	17,00

- Maximum dátového súboru: $m = 17$.

Histogram historických počtov poistných plnení



Empirická distribučná funkcia



Odhadnuté Panjerovo rozdelenie

Fitovanie počtu poistných plnení pomocou Panjerovho rozdelenia

Odhadnuté parametre Panjerovho rozdelenia:

$$\hat{a} = 0{,}81700975; \quad \hat{b} = 0{,}44461343.$$

Posúdenie kvality fitu odhadnutého Panjerovho rozdelenia

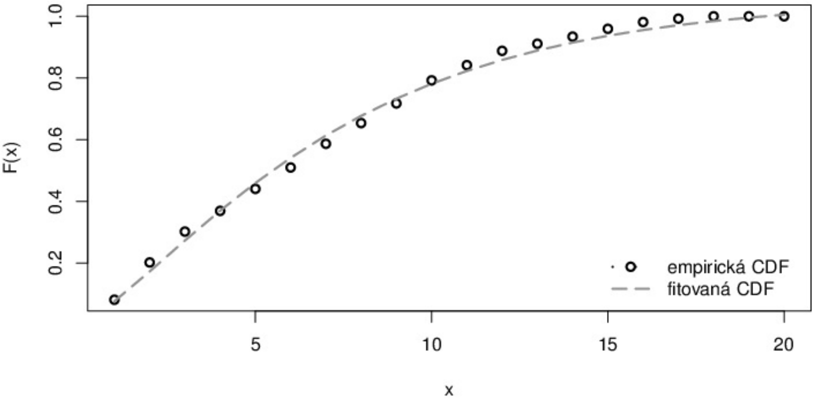
Kolmogorovov-Smirnovov test dobrej zhody:

$$D = 0{,}0311; \quad \text{p-hodnota K-S testu} = 0{,}3426.$$

Cramérov-von Misesov test dobrej zhody:

$$W^2 = 0{,}4356; \quad \text{p-hodnota C-vM testu} = 0{,}05855.$$

Kvalita fitu odhadnutého Panjerovho rozdelenia



Odhadnuté Schröterovo rozdelenie

Fitovanie počtu poistných plnení pomocou Schröterovho rozdelenia

Odhadnuté parametre Schröterovho rozdelenia:

$$\tilde{a} = 0{,}78159847; \quad \tilde{b} = 0{,}08567724; \quad \tilde{c} = 0{,}56850511.$$

Posúdenie kvality fitu odhadnutého Schröterovho rozdelenia

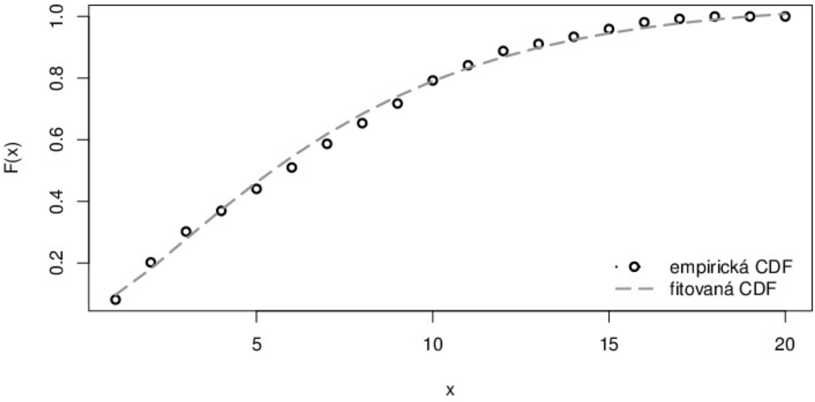
Kolmogorovov-Smirnovov test dobrej zhody:

$$D = 0{,}0346; \quad \text{p-hodnota K-S testu} = 0{,}2257.$$

Cramérov-von Misesov test dobrej zhody:

$$W^2 = 0{,}4070; \quad \text{p-hodnota C-vM testu} = 0{,}06955.$$

Kvalita fitu odhadnutého Schröterovho rozdelenia



Riešenie ilustračnej úlohy

Porovnanie kvality fitu

Porovnanie kvality fitu pre počet poistných plnení medzi odhadnutým Panjerovým rozdelením a odhadnutým Schröterovým rozdelením

	K-S D	C-vM W^2	RSS*	AIC*	BIC*
$Pan(\hat{a}, \hat{b})$	0,0311	0,4356	0,008675216	−7,543017	−148,8689
$Schr(\tilde{a}, \tilde{b}, \tilde{c})$	0,0346	0,4070	0,007010002	−7,756150	−153,1315

Ďalšie ciele výskumu

* Skúmať rekurentné triedy vyššieho rádu,
* zefektívniť numerický algoritmus odhadovania parametrov rekurentných tried,
* zahrnúť do programového aparátu aj iné typy účelových funkcií,
* automatizovať proces hľadania najvhodnejšieho rozdelenia z $\mathcal{R}_k$-triedy pre modelovanie diskrétnej náhodnej premennej N.

Literatúra

→ Arnold, T. B. - Emerson, J. W. 2011. Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions. *The R Journal*, č. 3/2, s. 34 - 39. ISSN 2073-4859.

→ Baradlaiová, Z. 2016. *Rekurentné triedy počtu poistných plnení a zovšeobecnené rekurzie v modeli kolektívneho rizika.* Diplomová práca, FMFI, Univerzita Komenského v Bratislave, 66 s.

→ Belisle, C. J. P. 1992. Convergence theorems for a class of simulated annealing algorithms on Rd. *Journal of Applied Probability*, roč. 29, č. 4, s. 885–895. ISSN 0021-9002.

→ Buchanan, J. I. - Turner, P. R. 1992. *Numerical methods and analysis*. Springer, New York. ISBN-13: 978-0070087170.

Literatúra

→ R Core Team. 2015. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

→ Szűcs, G. 2013. Schröterova trieda rozdelení. *Forum Statisticum Slovacum*, IX. ročník, č. 7/2013, s. 227-232. ISSN 1336-7420.

→ Szűcs, G. 2015. Metódy odhadu parametrov rekurentných tried diskrétnych rozdelení pravdepodobnosti. *Forum Statisticum Slovacum*, č. 6/2015, s. 154-159. ISSN 1336-7420.

Ďakujem vám za pozornosť.

Tento výskum bol podporený grantom VEGA č. 2/0047/15.