Směsi rozdělení

Vedle *skládání* je důležitou operací s diskrétními rozděleními *mísení*

Konečné směsi

– Studovaný jev = ve skutečnosti více jevů

– každý z nich má své rozdělení

Např: pojistný nárok ze zdravotního zubního pojištění může být pravidelná kontrola, plomba, vytržení zubu. Každá má jiné rozdělení.

Definice: Náhodná veličina Y je n-bodová směs náhodných veličin $X_1, \ldots, X_n$ jestliže její distribuční funkce je dána vztahem

$$F_Y(y) = a_1 F_{X_1}(y) + a_2 F_{X_2}(y) + \cdots + a_n F_{X_n}(y)$$

kde $a_j > 0$ a $\sum_{j=1}^{n} a_j = 1$.

a_j ... pravděpodobnost, že Y je realizací n.v. X_j.

– Stejný vztah platí pro hustotu nebo pravděpodobnostní funkci

– Obecněji, počet rozdělení n může být také parametr.

– Směs může popsat vlastnosti které jednoduché rozdělení nemá.
Např. 2-bodová směs Gamma rozdělení může být bimodální

Smíšená Poissonova rozdělení

Příklad: Řidiče lze rozdělit na dobré a špatné. Každá skupina má Poissonovo rozdělení s parametry $\lambda_1 < \lambda_2$. Rozdělení (smíšené 2-bodové) je pak

$$P(Y = k) = p\frac{e^{-\lambda_1}\lambda_1^k}{k!} + (1-p)\frac{e^{-\lambda_2}\lambda_2^k}{k!}$$

– Tři parametry, λ_1, λ_2, p odhadneme např. metodou maximální věrohodnosti

Obecné směsi rozdělení

Motivace

pojištění automobilů

- každý řidič je v něm zařazen do jedné z tarifních tříd za účelem stanovení výše pojistného
- při klasifikaci bere pojistitel v potaz tarifní proměnné – věk, zkušenosti s řízením, historie nehodovosti atd.
- řidiči v rámci jedné tarifní třídy si nejsou úplně stejní – odlišné řidičské schopnosti, rozdílné reakce na silnici apod.

– Každý řidič má své rozdělení, popisující jeho rizikovost (typicky Poissonovo).

- směsi rozdělení
  (i) využíváme k modelování heterogenity uvnitř jednotlivých tarifních tříd
  (ii) jejich parametry můžeme chápat jako náhodné veličiny
- rozdělení náhodných parametrů nazýváme mísícím rozdělením

Zavedení rizikového parametru θ

- předpokládejme, že úroveň rizika každého klienta lze charakterizovat nezáporným rizikovým parametrem θ
- zahrnuje v sobě skryté faktory jednotlivých pojištěnců, jež pojišťovna není s to pozorovat, tj. je neznámý
- je realizací (buď diskrétní, nebo spojité) náhodné veličiny Θ

- pro každou tarifní třídu jsme schopni určit pravděpodobnostní funkci, resp. hustotu pravděpodobnosti $\pi_\Theta(\theta)$ náhodné veličiny Θ
- pravděpodobnost, že úroveň rizika náhodně vybraného klienta nepřekročí hodnotu θ udává distribuční funkce $F_\Theta(\theta) = \mathsf{P}(\Theta \leq \theta)$ náhodné veličiny Θ
- parametrem θ může být např. střední hodnota Poissonova rozdělení.

Pravděpodobnostní a generující funkce směsi rozdělení

- pravděpodobnost, že nastane právě k pojistných událostí je vážený průměr pravděpodobností pro jednotlivé θ:

$$\begin{aligned} p_N(k) &= \int p_{N|\Theta}(k|\theta)\mathrm{d}F_\Theta(\theta) = \mathsf{E}\big(p_{N|\Theta}(k|\theta)\big) = \\ &= \mathsf{E}\big(P(N=k|\Theta=\theta)\big) \end{aligned} \tag{1}$$

  - pro spojité Θ:
    $p_N(k) = \int p_{N|\Theta}(k|\theta)\cdot\pi_\Theta(\theta)\mathrm{d}\theta$
  - pro diskrétní Θ:
    $p_N(k) = \sum_j p_{N|\Theta}(k|\theta_j)\cdot\pi_\Theta(\theta_j)$

– $p_{N|\Theta}(k|\theta)$, resp. $p_{N|\Theta}(k|\theta_j)$ představuje podmíněnou pravděpodobnostní funkci náhodné veličiny N za podmínky $\Theta = \theta$, resp. $\Theta = \theta_j$

- $G_{N|\Theta}(s|\theta)$ značí podmíněnou generující funkci počtu událostí, je-li θ rizikový parametr

  nepodmíněná generující funkce počtu událostí je dána vztahem

  $$G_N(s) = \int G_{N|\Theta}(s|\theta)\mathrm{d}F_\Theta(\theta), \tag{2}$$

  integrujeme přes obor hodnot θ (obvykle od 0 do ∞ pro $\mathrm{Po}(\theta)$).

  - pro spojité Θ:
    $G_N(s) = \int G_{N|\Theta}(s|\theta) \cdot \pi_\Theta(\theta)\mathrm{d}\theta$
  - pro diskrétní Θ:
    $G_N(s) = \sum_j G_{N|\Theta}(s|\theta_j) \cdot \pi_\Theta(\theta_j)$

- mísící rozdělení určené distribuční funkcí $F_\Theta(\theta)$ může být jak diskrétního, tak spojitého typu

  (i) je-li diskrétního typu (Θ je diskrétní náhodná veličina) $\rightarrow$ diskrétní směs rozdělení (speciálně konečná)

  (ii) je-li spojitého typu (Θ je spojitá náhodná veličina) $\rightarrow$ spojitá směs rozdělení

Směsi Poissonových rozdělení

– jsou oblíbeným nástrojem pro modelování nehomogenních portfolií

– Poissonovo rozdělení samo o sobě špatně fituje data pozorovaná na heterogenním pojistném kmeni

- uvažujme podmíněnou pravděpodobnostní funkci náhodné veličiny N vzhledem k $\Theta = \theta$ v následujícím tvaru

$$p_{N|\Theta}(k|\theta) = \mathrm{e}^{-\lambda\theta} \cdot \frac{(\lambda\theta)^k}{k!}$$

  (počet nehod daného klienta má $Po(\lambda\theta)$)

- z (1) odvodíme nepodmíněnou pravděpodobnostní funkci smíšeného Poissonova rozdělení

Definice

Diskrétní náhodná veličina N se řídí smíšeným Poissonovým rozdělením s parametrem $\lambda > 0$ a mísící distribucí $F_\Theta(\theta)$, značíme $N \sim MPo(\lambda, F_\Theta(\theta))$, jestliže

$$p_N(k) = \int \mathrm{e}^{-\lambda\theta} \cdot \frac{(\lambda\theta)^k}{k!} \mathrm{d}F_\Theta(\theta), \quad k = 0, 1, \ldots \tag{3}$$

- λ je **intenzita nehodovosti**, udává průměrnou nehodovost jisté tarifní skupiny
- θ je **rizikový parametr konkrétního klienta**

  $\theta > 1$ ... špatný řidič
  $\theta = 1$ ... průměrný řidič
  $\theta < 1$ ... dobrý řidič

Generující funkce:

$$
\begin{aligned}
G_N(s) &= \int \mathrm{e}^{\lambda\theta\cdot(s-1)}\mathrm{d}F_\Theta(\theta) = \int \left[\mathrm{e}^{\lambda\cdot(s-1)}\right]^\theta \mathrm{d}F_\Theta(\theta) = \\
&= \mathsf{E}\Big(\left[\mathrm{e}^{\lambda\cdot(s-1)}\right]^\theta\Big) = G_\Theta\left(\mathrm{e}^{\lambda\cdot(s-1)}\right) = M_\Theta\left(\lambda(s-1)\right)
\end{aligned}
\tag{4}
$$

$G_\Theta(s) = \mathsf{E}(s^\Theta)$ je generující funkce mísícího rozdělení

$M_\Theta(t) = \mathsf{E}(\mathrm{e}^{t\Theta})$ je moment generující funkce mísícího rozdělení

Střední hodnota a ***rozptyl***:

$$\mathrm{E}(N) = G_N'(1) = \lambda \cdot M_\Theta'\big(\lambda(s-1)\big)\Big|_{s=1} = \lambda \cdot M_\Theta'(0) = \lambda \cdot \mathrm{E}(\Theta) \quad (5)$$

$$\begin{aligned}
\mathrm{Var}(N) &= G_N''(1) + G_N'(1) - \big(G_N'(1)\big)^2 = \\
&= \lambda^2 \cdot M_\Theta''\big(\lambda(s-1)\big)\Big|_{s=1} + G_N'(1) - \big(G_N'(1)\big)^2 = \\
&= \lambda^2 \cdot M_\Theta''(0) + \mathrm{E}(N) - \big(\mathrm{E}(N)\big)^2 = \\
&= \lambda^2 \cdot \mathrm{E}(\Theta^2) + \lambda \cdot \mathrm{E}(\Theta) - \lambda^2 \cdot \big(\mathrm{E}(\Theta)\big)^2 = \\
&= \lambda \cdot \mathrm{E}(\Theta) + \lambda^2 \cdot \big[\mathrm{E}(\Theta^2) - \big(\mathrm{E}(\Theta)\big)^2\big] = \\
&= \lambda \cdot \mathrm{E}(\Theta) + \lambda^2 \cdot \mathrm{Var}(\Theta)
\end{aligned} \quad (6)$$

- parametr rizika vybíráme tak, aby $\mathrm{E}(\Theta) = 1$
- na rozdíl od Poissonova rozdělení s equidispersion má smíšené Poissonovo rozdělení overdispersion, tj. $\mathrm{Var}(N) > \mathrm{E}(N)$

Jak popsat smíšená Poissonova rozdělení?

Definice

Nechť X je náhodná veličina definovaná na pravděpodobnostním prostoru $(\Omega, \mathcal{A}, P)$. Pak funkce $\varphi : \mathbb{R} \to \mathbb{C}$ daná vztahem

$$\varphi_X(s) = \mathsf{E}(\mathrm{e}^{isX}) = \mathsf{E}\big(\cos(sX) + i \cdot \sin(sX)\big), \quad s \in \mathbb{R}, i = \sqrt{-1}, \tag{7}$$

se nazývá charakteristická funkce náhodné veličiny X.

- charakteristická funkce existuje pro každé pravděpodobnostní rozdělení, zatímco moment generující funkci nelze pro některá rozdělení s těžkým chvostem získat

Lemma

Existuje-li generující funkce diskrétní náhodné veličiny N, pak platí

$$G_N(s) = \varphi_N(-i \cdot \ln(s)) \quad \text{a} \quad \varphi_N(s) = G_N(\mathrm{e}^{is}). \tag{8}$$

Definice:

Rozdělení náhodné veličiny X je nekonečně dělitelné, jestliže pro všechny hodnoty $n = 1, 2, \dots$ lze jeho charakteristickou funkci $\varphi_X(s)$ zapsat

$$\varphi_X(s) = \left[\varphi_n(s)\right]^n, \tag{9}$$

kde $\varphi_n(s)$ je charakteristická funkce nějaké náhodné veličiny.

- normální, Poissonovo, negativně binomické a gamma rozdělení jsou nekonečně dělitelnými rozděleními
- binomické rozdělení není nekonečně dělitelné, jelikož jeho nosič (množina $\mathrm{supp} f = \{x \in \mathbb{R} : f(x) \neq 0\}$) je konečný

Věta:

Nechť $G(s)$ je generující funkce smíšeného Poissonova rozdělení s nekonečně dělitelným mísícím rozdělením. Pak $G(s)$ je též generující funkcí složeného Poissonova rozdělení a lze ji vyjádřit ve tvaru $G(s) = \mathrm{e}^{\lambda\left(G^*(s)-1\right)}$, kde $G^*(s)$ je generující funkce sekundárního rozdělení. Požadujeme-li, aby $G^*(0) = 0$, potom $G^*(s)$ je určena jednoznačně.

- vybereme-li jakékoliv nekonečně dělitelné mísící rozdělení, pak lze odpovídající smíšené Poissonovo rozdělení zapsat jako složené Poissonovo rozdělení

- Pak pro výpočet můžeme použít např. Panjerovu rekurzi

Příklady směsí Poissonových rozdělení

a) Negativně binomické rozdělení

- je smíšené Poissonovo rozdělení s mísícím gamma rozdělením
- pravděpodobnostní funkce negativně binomického rozdělení je ve tvaru

$$p_{N'}(k) = \binom{k+m-1}{k}\left(\frac{1}{1+\beta}\right)^m\left(\frac{\beta}{1+\beta}\right)^k,$$
$$m > 0, \beta > 0, k = 0, 1, \dots$$

- ukážeme, že je shodná s pravděpodobnostní funkcí smíšeného Poissonova rozdělení, kde mísícím rozdělením je gamma rozdělení

$$p_N(k) = \mathsf{E}(p_{N|\Theta}(k|\theta)) = \int_0^\infty p_{N|\Theta}(k|\theta)\mathrm{d}F_\Theta(\theta) =$$
$$= \int_0^\infty \mathrm{e}^{-\theta}\cdot\frac{\theta^k}{k!}\mathrm{d}F_\Theta(\theta)$$

- mějme $\Theta \sim Gam(\alpha, \beta)$ (kde $\alpha = n, \beta = \frac{1}{\lambda}$)

$$p_N(k) = \int_0^\infty \mathrm{e}^{-\theta} \cdot \frac{\theta^k}{k!} \cdot \frac{\theta^{\alpha-1} \cdot \mathrm{e}^{-\frac{\theta}{\beta}}}{\Gamma(\alpha) \cdot \beta^\alpha} \mathrm{d}\theta =$$

$$= \frac{1}{k!\Gamma(\alpha) \cdot \beta^\alpha} \int_0^\infty \mathrm{e}^{-\theta \cdot \left(1+\frac{1}{\beta}\right)} \cdot \theta^{k+\alpha-1} \mathrm{d}\theta$$

$$\left| \text{subst.: } \theta \cdot \left(1 + \frac{1}{\beta}\right) = z; \mathrm{d}\theta = \frac{\beta}{1+\beta} dz \right| =$$

$$= \frac{1}{k!\Gamma(\alpha) \cdot \beta^\alpha} \int_0^\infty \mathrm{e}^{-z} \cdot z^{k+\alpha-1} \left(\frac{\beta}{1+\beta}\right)^{k+\alpha} =$$

$$= \frac{\Gamma(k+\alpha)}{k!\Gamma(\alpha) \cdot \beta^\alpha} \left(\frac{\beta}{1+\beta}\right)^{k+\alpha} = \frac{\Gamma(k+\alpha)}{k!\Gamma(\alpha)} \cdot \frac{\beta^k}{(1+\beta)^k \cdot (1+\beta)^\alpha} =$$

$$= \binom{k+\alpha-1}{k} \left(\frac{\beta}{1+\beta}\right)^k \left(\frac{1}{1+\beta}\right)^\alpha$$

- pro $m = \alpha$ se $p_{N'}(k) = p_N(k)$

b) Další směsi Poissonových rozdělení

Název směsi rozdělení	**Neymanovo rozdělení typu A**
Mísící rozdělení	Poissonovo rozdělení
Název směsi rozdělení	**Poisson-inverzní Gaussovo rozdělení**
Mísící rozdělení s parametry	inverzní Gaussovo rozdělení $\mu = 1, \nu > 0$
Název směsi rozdělení	**Poisson-log-normální rozdělení**
Mísící rozdělení s parametry	log-normální rozdělení $\mu = -\frac{\sigma^2}{2}, \sigma^2 > 0$

Tabulka: Příklady směsí Poissonových rozdělení

- hustota inverzního Gaussova rozdělení s parametry $\mu > 0$ a $\nu > 0$:

$$f_X(x) = \begin{cases} \sqrt{\frac{\nu}{2\pi x^3}} \cdot \mathrm{e}^{-\frac{\nu(x-\mu)^2}{2x\mu^2}}, & x > 0, \\ 0, & \text{jinak.} \end{cases} \tag{10}$$

- hustota log-normálního rozdělení s parametry $\mu \in \mathbb{R}$ a $\sigma^2 > 0$:

$$f_X(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \cdot \mathrm{e}^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}, & x > 0, \\ 0, & \text{jinak.} \end{cases} \tag{11}$$