Spojitá nezáporná rozdělení

– Pro modelování velikostí jednotlivých pojistných nároků používáme většinou spojitá rozdělení.

– Normální rozdělení není nezáporné, tedy se pro modelování nehodí

– Někdy se ale používá pro aproximaci v centrální oblasti, t.j. okolo $E(X)$, na základě centrální limitní věty

– Chvosty (konce) rozdělení se ale musí modelovat zvlášť!

Exponenciální rozdělení a jeho vlastnosti

Spojitá náhodná veličina má exponenciální rozdělení, jestliže jeho hustota je dána vztahem

$$f(x) = \lambda e^{-\lambda x} \tag{1}$$

pro $x \geq 0$ a je rovna nule jinak.

Distribuční funkce exponenciálního rozdělení je tedy

$$F(x) = 1 - e^{-\lambda x} \tag{2}$$

pro $x \geq 0$, a rovna nule jinak.

Moment generující funkce exponenciálního rozdělení je dána vztahem

$$E[e^{tX}] = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \frac{\lambda}{\lambda - t} = \frac{1}{1-\theta t} \tag{3}$$

kde $\theta = \frac{1}{\lambda}$. Momenty náhodné veličiny X můžeme získat derivováním moment generující funkce.

Tím dostaneme

$$E[X] = \frac{1}{\lambda} = \theta \tag{4}$$

a

$$Var[X] = \frac{1}{\lambda^2} = \theta^2 \tag{5}$$

Jednou z hlavních vlastností exponenciálního rozdělení je že nemá paměť. Platí totiž

$$P\{X > s + t | X > t\} = P\{X > s\}. \tag{6}$$

Je-li například X životnost daného stroje, pak pravděpodobnost že stroj bude fungovat alespoň $s + t$ hodin, za podmínky že již funguje t hodin, je stejná jako počáteční pravděpodobnost že bude fungovat alespoň s hodin. Stroj si “nepamatuje” svoji minulost.

– Opačná situace: efekt opotřebení nebo zahoření.

Vlastnost absence paměti

Exponenciální rozdělení je jediné rozdělení které nemá paměť.

Dokážeme to následovně. Nechť

$$\tilde{F} = P\{X > x\} \tag{7}$$

je funkce přežití. Pak z předchozího vztahu platí

$$\tilde{F}(s + t) = \tilde{F}(s)\tilde{F}(t) \tag{8}$$

Jinak řečeno, $\tilde{F}$ splnuje funkcionální rovnici

$$h(s+t) = h(s)h(t). \tag{9}$$

Dokážeme ted že jediná zprava polospojitá řešení této rovnice mají tvar exponenciály.

Ze vztahu

$$h(s+t) = h(s)h(t). \tag{10}$$

máme

$$h(\frac{2}{n}) = h(\frac{1}{n} + \frac{1}{n}) = h^2(\frac{1}{n}). \tag{11}$$

Opakováním stejného argumentu dostaneme

$$h(\frac{m}{n}) = h^m(\frac{1}{n}). \tag{12}$$

Dále platí

$$h(1) = h(\frac{1}{n} + \cdots + \frac{1}{n}) = h^n(\frac{1}{n}). \tag{13}$$

Tedy

$$h(\frac{m}{n}) = h(1)^{\frac{m}{n}}. \tag{14}$$

a tedy

$$h(x) = h(1)^x, \tag{15}$$

protože h je pospojitá zprava.

Platí

$$h(1) = h^2(\frac{1}{2}) \geq 0, \tag{16}$$

a tedy

$$h(x) = e^{-\lambda x}, \tag{17}$$

kde $\lambda = -\ln h(1)$.

Musí tedy platit

$$\tilde{F} = e^{\lambda x}, \tag{18}$$

protože distribuční funkce je zprava polospojitá, což je funkce přežití exponenciálního rozdělení.

Míra rizika

– V životním pojištění se používá termín *intenzita úmrtnosti*

Uvažujme spojitou náhodnou veličinu X a distribuční funkcí F a hustotou f.

Definice: Funkce míry rizika je definována jako

$$\lambda(t) = \frac{f(t)}{\tilde{F}(t)} \tag{19}$$

Představme si, že skoumáme životnost nějakého stroje (součástky), a předpokládejme, že stroj již funguje t hodin.

Chceme spočítat pravděpodobnost, že nevydrží další časový úsek dt, tedy

$$P\{X \in (t, t + dt)|X > t\}, \tag{20}$$

kde n.v. X modeluje čas poruchy stroje. Máme

$$P\{X \in (t, t + dt)|X > t\} = \frac{P(X \in (t, t + dt) \wedge X > t)}{P(X > t)} \tag{21}$$

což je rovno

$$\frac{X \in (t, t + dt)}{P(X > t)} \simeq \frac{f(t)dt}{\tilde{F}(t)} \tag{22}$$

$$= \lambda(t)dt. \tag{23}$$

Tedy $\lambda(t)$ reprezentuje *intenzitu pravděpodobnosti*, že t-letý stroj přestane fungovat v čase t.

Je-li rozdělení exponenciální, pak z vlastnosti absence paměti je podmíněné rozdělení stejné jako počáteční, tedy λ je konstantní,

$$\lambda(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda. \tag{24}$$

Tedy míra rizika pro exponenciální rozdělení je konstantní.

Ukážeme ještě že míra rizika naopak jednoznačně určuje pravděpodobnostní rozdělení. Opravdu, máme

$$\lambda(t) = \frac{-\frac{d}{dt}\tilde{F}(t)}{\tilde{F}(t)} = \lambda \tag{25}$$

Integrováním dostaneme

$$\ln \tilde{F}(t) = -\int_0^t \lambda(s)ds + k \tag{26}$$

tedy

$$\tilde{F}(t) = ce^{\int_0^t \lambda(s)ds} \tag{27}$$

kde pro $t = 0$ dosteneme $c = 1$. Celkem

$$\tilde{F}(t) = e^{\int_0^t \lambda(s)ds} \tag{28}$$

Odtud plyne že exponenciální rozdělení je jediné rozdělení s konstantní funkcí rizika.

Zobecnění

– Sčítáním IID náhodných veličin s exponenciálním rozdělením dostaneme Gamma rozdělení

Jsou-li $X_1, ..., X_n$ nezávislé a všechny mají $Exp(\lambda)$. Pak jejich součet má rozdělení $Gamma(n, \lambda)$.

Hustota Gamma rozdělení je

$$f(x) = \frac{\lambda^n e^{-\lambda x} x^{n-1}}{\Gamma(n)} \tag{29}$$

kde $\Gamma(n)$ je Gamma funkce,

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx. \tag{30}$$

Pro celočíselné hodnoty platí

$$\Gamma(n) = (n-1)! \tag{31}$$

Individuální model rizika

S ... náhodná veličina představující **úhrn škod** za období pevně zvolené délky T

n ... počet smluv v pojistném kmeni. Pak **individuální model:**

- pracuje s riziky příslušejícími jednotlivým pojistným smlouvám v pojistném kmeni.
- zabývá se vlastnostmi **individuálních škodních nároků** $X_i, i = 1, ..., n$ připadajících na jednotlivé pojistné smlouvy.
- uplatňuje se v důležité oblasti určování pojistného podle individuálního škodního průběhu.

Konvoluce

- Pomocí modelu pro individuální škodní nárok je možné také modelovat **nárok pojišťovny** S, který je možné chápat jako součet nároků jednotlivých pojištěných $X_i, \ i = 1, 2, ..., n$

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

- Předpokládejme nezávislost jednotlivých nároků.
- Cílem bude určit distribuční funkci, případně pravděpodobnostní funkci či hustotu celkového souhrnu těchto pojistných nároků.

Uvažujme nejprve součet dvou nezávislých n.v.

$$S = X + Y.$$

Pak **distribuční funkce** veličiny S je

$$F_S(s) = \mathrm{P}(S \leq s) = \mathrm{P}(X + Y \leq s).$$

Pro dvě nezávislé, nezáporné **diskrétní** náhodné veličiny

můžeme psát distribuční funkci

$$\begin{aligned} F_S(s) &= \sum_{y \leq s} \mathrm{P}(X + Y \leq s | Y = y)\, \mathrm{P}(Y = y) \\ &= \sum_{y \leq s} \mathrm{P}(X \leq s - y | Y = y)\, \mathrm{P}(Y = y) \\ &= \sum_{y \leq s} F_X(s - y) p_y(y). \end{aligned}$$

Pro **pravděpodobnostní funkci** pak platí

$$p_S(s) = \sum_{y \leq s} p_X(s-y)p_y(y).$$

Pro **spojité** nezáporné náhodné veičiny můžeme psát podobně

$$\begin{aligned} F_S(s) &= \int_0^s \mathrm{P}(X \leq s-y | Y = y) f_Y(y)\,\mathrm{d}y \\ &= \int_0^s F_X(s-y) f_Y(y)\,\mathrm{d}y. \end{aligned}$$

Hustota má tvar

$$f_S(s) = \int_0^s f_X(s-y) f_Y(y)\,\mathrm{d}y.$$

Uvažujeme-li náhodné veličiny, které mohou nabývat záporných hodnot, pak sumy a integrály v předchozích vzorcích budou v rozmezí od $-\infty$ do ∞.

Operaci vyjádřenou rovnicemi

$$
\begin{aligned}
F_S(s) &= \sum_{y \le s} F_X(s-y) p_Y(y) \\
F_S(s) &= \int_0^s F_X(s-y) f_Y(y)\,\mathrm{d}y
\end{aligned}
$$

nazýváme **konvoluce** distribučních funkcí F_X a F_Y a značíme ji $F_X * F_Y$.

Podobně pak pro hustoty nebo pravděpodobnostní funkce

můžeme konvoluci těchto funkcí vyjádřit pomocí rovnic

$$
\begin{aligned}
(f_X * f_Y)(s) &= \int_0^s f_X(s-y) f_Y(y)\,\mathrm{d}y \\
(p_X * p_Y)(s) &= \sum_{y \le s} p_X(s-y) p_y(y).
\end{aligned}
$$

Zobecnění

Budeme-li uvažovat součet více než 2 náhodných proměnných, pak můžeme proces konvoluce použít iterativně.

- Uvažujme náhodnou veličinu S definovanou jako

$$S = X_1 + X_2 + ... + X_n,$$

- Označme F_i distribuční funkci X_i a $F^{(k)}$ distribuční funkci $X_1 + X_2 + ... + X_k$.
- Pak distribuční funkci S můžeme zapsat pomocí iterace

$$\begin{aligned}
F^{(2)} &= F_2 * F^{(1)} = F_2 * F_1 \\
F^{(3)} &= F_3 * F^{(2)} \\
F^{(n)} &= F_n * F^{(n-1)}
\end{aligned}$$

Příklad

Uvažujme nezávislé diskrétní náhodné veličiny X_1, X_2 a X_3. Jejich rozdělení je definováno následující tabulkou

x	$p_1(x)$	$p_2(x)$	$p_3(x)$
0	0.4	0.5	0.6
1	0.3	0.2	0
2	0.2	0.1	0.1
3	0.1	0.1	0.1
4		0.1	0.1
5			0.1

Určete pravděpodobnostní a distribuční funkci veličiny

$$S = X_1 + X_2 + X_3.$$

Příklad

Uvažujme nezávislé spojité náhodné veličiny X_1, X_2 a X_3. Jejich rozdělení je definováno takto

$$\begin{aligned} f_1(x) &= \mathrm{e}^{-x} \quad & x > 0, \\ f_2(x) &= 2\,\mathrm{e}^{-2x} \quad & x > 0, \\ f_3(x) &= 3\,\mathrm{e}^{-3x} \quad & x > 0. \end{aligned}$$

Určete pravděpodobnostní hustotu veličiny

$$S = X_1 + X_2 + X_3.$$

Momentová generující funkce

- Další způsob, jak lze určit rozdělení součtu náhodných veličin, je založen na využití **momentové generující (vytvořující) funkce**.
- Ta je definovaná pomocí vzorce

$$M_X(t) = \mathrm{E}(\mathrm{e}^{tX}).$$

- Je-li $\mathrm{E}(\mathrm{e}^{tX})$ konečná pro každé t na otevřeném intervalu okolo počátku, pak je veličina X touto funkcí jednoznačně určena.

Momentová vytvořující funkce veličiny $S = X_1 + X_2 + ... + X_n$ je pak

$$\begin{aligned} M_S(t) &= \mathrm{E}(\mathrm{e}^{tS}) = \mathrm{E}(\mathrm{e}^{t(X_1+X_2+...+X_n)}) \\ &= \mathrm{E}(\mathrm{e}^{tX_1}\,\mathrm{e}^{tX_2}\cdots\mathrm{e}^{tX_n}). \end{aligned}$$

Jestliže jsou veličiny $X_1, X_2, ..., X_n$ nezávislé, pak

$$\mathrm{E}(\mathrm{e}^{tX_1}\,\mathrm{e}^{tX_2}\cdots\mathrm{e}^{tX_n}) = \mathrm{E}(\mathrm{e}^{tX_1})\,\mathrm{E}(\mathrm{e}^{tX_2})\cdots\mathrm{E}(\mathrm{e}^{tX_n}).$$

Tedy

$$M_S(t) = M_{X_1}(t)M_{X_2}(t)\cdots M_{X_n}(t).$$

Příklad

Předpokládejme nezávislé náhodné veličiny $X_1,\ X_2,\ X_3$ s exponenciálním rozložením, jejichž hustoty jsou

$$\begin{aligned} f_1(x) &= \mathrm{e}^{-x} \quad & x > 0, \\ f_2(x) &= 2\,\mathrm{e}^{-2x} \quad & x > 0, \\ f_3(x) &= 3\,\mathrm{e}^{-3x} \quad & x > 0. \end{aligned}$$

Pomocí momentové vytvořující funkce odvoďte pravděpodobnostní hustotu veličiny

$$S = X_1 + X_2 + X_3.$$