Třída rozdělení $(a, b, 0)$

Definice

Nechť $p_N(k) = P(N = k)$ je pravděpodobnostní funkce diskrétní náhodné veličiny N. Řekneme, že je členem třídy rozdělení $(a, b, 0)$, jestliže existují reálné konstanty a a b takové, že platí

$$\frac{p_N(k)}{p_N(k-1)} = a + \frac{b}{k} \quad \text{pro } k = 1, 2, 3, \ldots \tag{1}$$

- pravděpodobnost $p_N(0)$ dopočítáme z $\sum_{k=0}^{\infty} p_N(k) = 1$

- do třídy rozdělení $(a, b, 0)$ patří právě Poissonovo rozdělení, negativně binomické a binomické rozdělení
- Pro Poissonovo rozdělení je $a = 0$ a $b = \lambda$.
- Pro binomické je $a = -\frac{p}{1-p} < 0$ a $b = (n+1)\frac{p}{1-p}$.
- Pro NeBi je $a = 1 - p > 0$ a $b = (m-1)(1-p)$.

- pro konkrétní datový soubor s velkým množstvím pozorování lze určit vhodný model pomocí formule (1)

- formuli přepíšeme do tvaru

$$\frac{p_N(k)}{p_N(k-1)} \cdot k = ak + b, \quad k = 1, 2, 3, \ldots \tag{2}$$

- podíl $\frac{p_N(k)}{p_N(k-1)}$ odhadneme na základě pozorovaných četností n_k a n_{k-1} hodnot k a $k-1$

$$\frac{\widehat{p_N(k)}}{\widehat{p_N(k-1)}} \cdot k = \frac{n_k}{n_{k-1}} \cdot k. \tag{3}$$

- graf procházející body $\left[k, k \cdot \frac{n_k}{n_{k-1}}\right]$ by měl přibližně vykazovat lineární průběh

- podle směrnice a dané přímky zvolíme vhodný model

  nulová směrnice $\rightarrow$ Poissonovo rozdělení

  záporná směrnice $\rightarrow$ binomické rozdělení

  kladná směrnice $\rightarrow$ negativně binomické rozdělení

Třída rozdělení $(a, b, 1)$

- rozdělení třídy $(a, b, 0)$ často nepopisují adekvátně data, s nimiž se v praxi setkáváme

  Hlavní příčina:
  - rozdělení třídy $(a, b, 0)$ nejsou s to vystihnout tvar dat v jistých částech rozdělení, zejména hodnotu v nule.

- budeme se věnovat rozložení pravděpodobnosti v nule (pravděpodobnost, že nenastane žádná pojistná událost během stanoveného časového období) – např. u pojištění odpovědnosti, majetku aj. je pravděpodobnost v nule *největší*

- úpravou pravděpodobnosti v nule lze třídu rozdělení $(a, b, 0)$ rozšířit na třídu $(a, b, 1)$

Definice

Nechť $p_N(k)$ je pravděpodobnostní funkce diskrétní náhodné veličiny N. Řekneme, že je členem třídy rozdělení $(a, b, 1)$ za předpokladu, že existují konstanty $a, b \in \mathbb{R}$ takové, že

$$\frac{p_N(k)}{p_N(k-1)} = a + \frac{b}{k} \quad \text{pro } k = 2, 3, \dots \tag{4}$$

- $\sum_{k=1}^{\infty} p_N(k)$ může nabývat libovolných hodnot na $(0, 1\rangle$, zbývající pravděpodobnost je v $k = 0$, jelikož $p_N(0) + \sum_{k=1}^{\infty} p_N(k) = 1$

U třídy $(a, b, 1)$ rozlišujeme dvě podtřídy

$p_N(0) = 0$ ... **rozdělení useknuté v nule** s $p_N^T(k)$

$p_N(0) > 0$ ... **rozdělení modifikované v nule** s $p_N^M(k)$

1. Rozdělení modifikovaná v nule

lze na ně pohlížet jako na směs rozdělení třídy $(a, b, 0)$ a degenerovaného rozdělení se všemi pravděpodobnostmi soustředěnými v nule

- $G_N(s) = \sum_{k=0}^{\infty} p_N(k) \cdot s^k$ je generující funkce rozdělení třídy $(a, b, 0)$
  $G_N^M(s) = \sum_{k=0}^{\infty} p_N^M(k) \cdot s^k$ je generující funkce příslušného v nule modifikovaného rozdělení třídy $(a, b, 1)$
- platí, že $p_N^M(k) = c \cdot p_N(k)$ pro $k = 1, 2, \ldots; c \in \mathbb{R}^+$ a $p_N^M(0)$ je libovolně zvolené z intervalu $(0, 1)$. Musíme *vypočítat hodnotu* c.

- potom

$$
\begin{aligned}
G_N^M(s) = p_N^M(0) + \sum_{k=1}^{\infty} p_N^M(k) \cdot s^k = \\
= p_N^M(0) + c \cdot \sum_{k=1}^{\infty} p_N(k) \cdot s^k = \\
= p_N^M(0) + c \cdot \big(G_N(s) - p_N(0)\big)
\end{aligned}
\tag{5}
$$

- z platnosti $G_N^M(1) = G_N(1) = 1$ plyne $1 = p_N^M(0) + c \cdot \big(1 - p_N(0)\big)$
- odtud $c = \frac{1-p_N^M(0)}{1-p_N(0)}$
- tudíž

$$
p_N^M(k) = \frac{1 - p_N^M(0)}{1 - p_N(0)} \cdot p_N(k) \quad \text{pro } k = 1, 2, \dots \tag{6}
$$

- dostaneme generující funkci modifikovaného rozdělení

$$
\begin{aligned}
G_N^M(s) &= p_N^M(0) + c \cdot \big(G_N(s) - p_N(0)\big) = \\
&= p_N^M(0) + \frac{1 - p_N^M(0)}{1 - p_N(0)} \cdot \big(G_N(s) - p_N(0)\big) = \\
&= \frac{p_N^M(0) - p_N(0)}{1 - p_N(0)} + \frac{1 - p_N^M(0)}{1 - p_N(0)} \cdot G_N(s) = \\
&= \frac{p_N^M(0) - 1 + 1 - p_N(0)}{1 - p_N(0)} + \frac{1 - p_N^M(0)}{1 - p_N(0)} \cdot G_N(s) = \\
&= \left(1 - \frac{1 - p_N^M(0)}{1 - p_N(0)}\right) + \frac{1 - p_N^M(0)}{1 - p_N(0)} \cdot G_N(s)
\end{aligned} \tag{7}
$$

2. Rozdělení useknutá v nule

lze chápat jako speciální typ v nule modifikovaného rozdělení s hodnotou $p_N^M(0) = 0$

- $G_N^T(s)$ je generující funkce v nule useknutého rozdělení
- potom z (6), (7) a $p_N^M(0) = 0$ získáme

$$p_N^T(k) = \frac{p_N(k)}{1 - p_N(0)} \quad \text{pro } k = 1, 2, \dots, \tag{8}$$

$$G_N^T(s) = \frac{G_N(s) - p_N(0)}{1 - p_N(0)} \tag{9}$$

a) rozšířené useknuté negativně binomické (ETNB) rozdělení

množina možných hodnot parametru m je rozšířena z $m > 0$ na $m > -1$, přičemž $m \neq 0$

Pravděpodobnostní funkce:

$$p_N^T(k) = \begin{cases} \frac{\binom{k+m-1}{k}\cdot(1-p)^k}{p^{-m}-1}, & k = 1, 2, \ldots; p \in (0,1), \\ 0, & \text{jinak}, \end{cases} \tag{10}$$

resp.

$$p_N^T(k) = \begin{cases} \frac{\binom{k+m-1}{k}\cdot\left(\frac{\beta}{1+\beta}\right)^k}{(1+\beta)^m-1} = & k = 1, 2, \ldots; \\ = \frac{(k+m-1)\cdot\ldots\cdot(m+1)\cdot m}{k!\cdot((1+\beta)^m-1)} \cdot \left(\frac{\beta}{1+\beta}\right)^k, & \beta = \frac{1-p}{p} > 0, \\ 0, & \text{jinak}. \end{cases} \tag{11}$$

b) logaritmické rozdělení

- je limitním případem ETNB rozdělení pro $m \to 0$
- neexistuje k němu odpovídající rozdělení ve třídě $(a, b, 0)$

Pravděpodobnostní funkce:

$$p_N^T(k) = \begin{cases} -\frac{(1-p)^k}{k \cdot \ln(p)}, & k = 1, 2, \ldots; p \in (0, 1), \\ 0, & \text{jinak}, \end{cases} \tag{12}$$

resp.

$$p_N^T(k) = \begin{cases} \frac{\beta^k}{k(1+\beta)^k \ln(1+\beta)}, & k = 1, 2, \ldots; \beta = \frac{1-p}{p} > 0, \\ 0, & \text{jinak}. \end{cases} \tag{13}$$

Opět lze dokázat že další taková rozdělení *neexistují*.

Další třídy čítacích rozdělení vytvoříme pomocí dvou *operací*

– *Skládání*

– *Míšení* (směsi)

K tomu budeme potřebovat některé nástroje z teorie pravděpodobnosti.

Diskrétní náhodné proměnné

Nechť X je diskrétní náhodná proměnná (náhodná veličina), tedy funkce

$$X : \Omega \to \{x_1, x_2, \dots\} \subseteq \mathbb{R},$$

kde $\{x_1, x_2, \dots\}$ je diskrétní podmnožina $\mathbb{R}$.

Pravděpodobnostní funkci náhodné veličiny X označíme jako

$$f(x) = P(X = x).$$

Definice 2.1. *Distribuční funkce* náhodné veličiny X je

$$F(x) = P(X \leq x).$$

Připomeňme si ještě definici nezávislosti dvou jevů.

Definice 2.2. Jevy A, $B \subseteq \Omega$ jsou nezávislé, jestliže

$$P(A) = \frac{P(A \cap B)}{P(B)},$$

tedy

$$P(A \cap B) = P(A)P(B).$$

Jinak řečeno nastal-li jev B, *nezmění* to pravděpodobnost jevu A.

Definice 2.3. Diskrétní náhodné veličiny X a Y jsou *nezávislé*, jestliže jevy $\{X = x\}$ a $\{Y = y\}$ jsou nezávislé pro všechna x a y.

Jinými slovy, znalost hodnoty X nedává žádnou informaci o hodnotě Y.

Závislost a nezávislost náhodných veličin

Lemma 2.4. *Nechť X a Y jsou nezávislé náhodné veličiny. Potom*

$$E(XY) = E(X)E(Y).$$

Opak obecně neplatí.

Definice 2.5. Říkáme, že náhodné veličiny X a Y jsou *nekorelované*, jestliže platí:

$$E(XY) = E(X)E(Y).$$

Věta 2.6. *Nechť X a Y jsou náhodné veličiny. Pak*
— $Var(aX) = a^2 Var(X)$ pro $a \in \mathbb{R}$.
— Jsou-li X a Y nekorelované (speciálně nezávislé) náhodné veličiny, pak

$$Var(X + Y) = Var(X) + Var(Y).$$

Definice 2.7. ***Kovariance*** náhodných veličin X a Y je definována jako

$$cov(X, Y) = E\left[(X - E(X))(Y - E(Y))\right].$$

Korelační koeficient X a Y je

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{Var(X)Var(Y)}}.$$

Platí:

$$\rho(X,Y) = 0 \Leftrightarrow E(XY) = E(X)E(Y) \Leftrightarrow cov(X,Y) = 0$$

a

$$cov(X,Y) = E(XY) - E(X)E(Y).$$

Dále je

$$|\rho(X,Y)| \leq 1.$$

Jak ověřit nezávislost dvou daných náhodných veličin?

Definice k tomu většinou vhodná není.

Definice 2.8. Nechť X a Y jsou diskrétní náhodné veličiny (na stejném pravděpodobnostním prostoru). *Sdružená distribuční funkce* X a Y je definovaná vztahem

$$F_{X,Y}(x,y) = P(X \leq x \wedge Y \leq y).$$

Definice 2.9. ***Sdružená pravděpodobnostní funkce***:
$f_{X,Y} : \mathbb{R}^2 \to [0,1]$ je definovaná vztahem

$$f_{X,Y}(x,y) = P(X = x \wedge Y = y).$$

Analogicky se definuje sdružená pravděpodobnostní funkce pro více náhodných veličin. Následující lemma dává dobře ověřitelné kriterium nezávislosti.

Lemma 2.10. *Diskrétní náhodné veličiny X a Y jsou nezávislé právě tehdy, když*

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

pro všechna $x, y \in \mathbb{R}$.

Ze znalosti sdružené pravděpodobnostní funkce $f_{X,Y}$ můžeme vypočítat *marginální pravděpodobnostní funkce* f_X a f_Y. Máme

$$f_X(x) = P(X = x) = P(\bigcup_y (\{X = x\} \cap \{Y = y\}))$$

$$= \sum_y P(X = x \wedge Y = y) = \sum_y f_{X,Y}(x, y).$$

Příklad 2.11. Nechť $X : \Omega \to \{1, 2, 3\}$a $Y : \Omega \to \{-1, 0, 2\}$ jsou náhodné veličiny a sdružená pravděpodobnostní funkce je dána tabulkou:

	$y=-1$	$y=0$	$y=2$	f_X
$x=1$	$\frac{1}{18}$	$\frac{3}{18}$	$\frac{2}{18}$	$\frac{6}{18}$
$x=2$	$\frac{2}{18}$	0	$\frac{3}{18}$	$\frac{5}{18}$
$x=3$	0	$\frac{4}{18}$	$\frac{3}{18}$	$\frac{7}{18}$
f_Y	$\frac{3}{18}$	$\frac{7}{18}$	$\frac{8}{18}$	$\frac{18}{18}$

Jsou X a Y nezávislé?

Zřejmě ne, v tom případě by řádky tabulky musely být násobkem jeden druhého.

Vypočteme kovarianci těchto dvou náhodných veličin. Máme

$$XY : \Omega \to \{-1, 0, -2, -3, 2, 4, 6\} .$$

Dále

$$E(X) = \frac{6}{18} + \frac{10}{18} + \frac{21}{18} = \frac{37}{18}, \quad E(Y) = \frac{13}{18}$$

a

$$E(XY) = -1\frac{1}{18} + 2\frac{2}{18} - 2\frac{2}{18} + 4\frac{3}{18} + 6\frac{3}{18} = \frac{29}{18}$$

Celkem tedy

$$cov(X, Y) = E(XY) - E(X)E(Y) = \frac{29}{18} - \frac{481}{324} = \frac{522 - 481}{324} = \frac{41}{324}.$$

Podmíněná pravděpodobnost a podmíněné očekávání

Připomeňme definici podmíněné pravděpodobnosti pro jevy,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

V modelech pojistné a finanční matematiky je obvykle pravděpodobnost podmíněná informací, kterou máme v danou chvíli. Formálně to zachycuje následující definice.

Definice 2.12. Podmíněná pravděpodobnostní funkce náhodné veličiny Y za podmínky $X = x$, kterou budeme označovat $f_{Y|X}(. \mid x)$, je definována jako

$$f_{Y|X}(y \mid x) = P(Y = y \mid X = x),$$

pro každé x takové, že $P(X = x) > 0$.

Z definice máme

$$P(Y = y \mid X = x) = \frac{P(Y = y \wedge X = x)}{P(X = x)} = \frac{f_{X,Y}(x, y)}{f_X(x)},$$

tedy

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)},$$

což je analogický vztah jako platí pro podmíněné pravděpodobnosti jevů.

V předchozím příkladu máme pro $x = 1$

$$f_{Y|X}(y \mid 1) \sim \left(\frac{1}{6}, \frac{3}{6}, \frac{2}{6}\right) = \frac{f_{X,Y}}{f_X}.$$

Víme-li, že $X = x$, pak Y má novou pravděpodobnostní funkci $f_{Y|X}(y \mid x)$ jakožto funkci y (x je pevné).

Očekávání vůči této funkci je podmíněné očekávání Y za podmínky $X = x$, které označíme $\Psi(x) = E(Y \mid X = x)$.

Definice 2.13. Funkce (tj. náhodná veličina)

$$\Psi(x) = E(Y \mid X = x)$$

se nazývá ***podmíněné očekávání*** Y při znalosti X.

V minulém příkladu je:

$$\Psi(1) = \frac{1}{6}(-1) + \frac{3}{6}0 + \frac{2}{6}2 = \frac{1}{2},$$

$$\Psi(2) = \frac{-2}{5} + \frac{6}{5} = \frac{4}{5}$$

a $\Psi(3) = \frac{6}{7}$.

Věta 2.14. *(O celkovém očekávání). Pro podmíněné očekávání* $\Psi(x) = E(Y \mid X = x)$ *platí*

$$E(\Psi(x)) = E(Y),$$

tedy

$$E(Y) = E(E(Y|X)).$$

- Opravdu, střední hodnotu náhodné veličiny $\mathsf{E}(Y|X)$ lze vypočítat

$$\begin{aligned} \mathsf{E}\big[\mathsf{E}(Y|X)\big] &= \sum_x \mathsf{E}(Y|X=x)\cdot f_X(x) = \\ &= \sum_x \sum_y y \cdot f_{Y|X}(y|x)\cdot f_X(x) = \\ &= \sum_y y \sum_x f_{Y|X}(y|x)\cdot f_X(x) = \\ &= \sum_y y\cdot f_Y(y) = \mathsf{E}(Y) \end{aligned} \tag{14}$$