evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
UNIVERZITA OBRANY
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Poissonův proces a lineární proces zrodu a zániku

1 Poissonův proces

Budeme se zabývat matematickým modelem časového sledu vzniků rizikových jevů daného typu v dané oblasti. Vyjdeme z následujících předpokladů.

P1 Počty rizikových jevů, které vzniknou v disjunktních časových intervalech, jsou nezávislé náhodné veličiny.

P2 Čas počítáme od nuly a předpokládáme, že v čase $t = 0$ není žádná riziková událost pozorována. Dále předpokládáme, že pro libovolný časový okamžik $t \geq 0$ je pravděpodobnost vzniku rizikového jevu v intervalech $(t, t + \Delta t\rangle$ rovna $\lambda\Delta t + o(\Delta t)$ pro časový přírůstek $\Delta t \to 0$. Jinak řečeno, je tato pravděpodobnost až na nekonečně malé veličiny vyššího řádu, úměrná délce tohoto intervalu, $\lambda > 0$ je konstanta úměrnosti vyjadřující intenzitu vzniku rizikových jevů.

P3 Pro $t \geq 0$ je pravděpodobnost vzniku dvou nebo více rizikových jevů v intervalu $(t, t + \Delta t\rangle$ „nekonečně malá“ a je tedy řádu $o(\Delta t)$ pro $\Delta t \to 0$. Tento předpoklad vylučuje současný vznik několika rizikových jevů.

Počet rizikových jevů v čase $t \geq 0$ označíme $X(t)$. Zřejmě $X(t)$ je náhodný proces, $T = \langle 0, \infty)$ a jeho množina stavů je tvořena možnými počty rizikových událostí v daných časech, tedy $E = \{0, 1, 2, \ldots\}$. Cílem je zjistit rozdělení pravděpodobností procesu $X(t)$, tedy stanovit pravděpodobnosti

$$p_k(t) = P(X(t) = k) \text{ pro } k = 0, 1, 2, \ldots \text{ a } t \geq 0.$$

Snadno zjistíme, že pro $k = 0$ platí

$$p_0(t + \Delta t) = P(X(t + \Delta t) = k) = P(X(t) = 0 \wedge X(t + \Delta t) - X(t) = 0).$$

Z P1 vyplývá, že $P(X(t) = 0 \wedge X(t + \Delta t) - X(t) = 0) = P(X(t) = 0)P(X(t + \Delta t) - X(t) = 0)$, protože časové intervaly $\langle 0, t\rangle$ a $(t, t + \Delta t\rangle$ jsou disjunktní. Když nyní dosadíme do výrazu pro $p_0(t + \Delta t)$ a použijeme předpoklad P2, dostaneme

$$p_0(t + \Delta t) = P(X(t) = 0)P(X(t + \Delta t) - X(t) = 0) = p_0(t)(1 - \lambda\Delta t - o(\Delta t)).$$

Odtud po jednoduché úpravě máme

$$\frac{p_0(t + \Delta t) - p_0(t)}{\Delta t} = -\lambda p_0(t) - p_0(t)\frac{o(\Delta t)}{\Delta t}$$

a po limitním přechodu $\Delta t \to 0$, protože $\frac{o(\Delta t)}{\Delta t} \to 0$, dostaneme diferenciální rovnici pro $p_0(t)$

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t).$$

Podobně odvodíme, že pro $k = 1, 2, \ldots$ platí

$$\frac{dp_k(t)}{dt} = \lambda p_{k-1}(t) - \lambda p_k(t).$$

Operační program Vzdělávání pro konkurenceschopnost
Název projektu: Inovace magisterského studijního programu Fakulty ekonomiky a managementu
Registrační číslo projektu: CZ.1.07/2.2.00/28.0326

PROJEKT JE SPOLUFINANCOVÁN EVROPSKÝM SOCIÁLNÍM FONDEM A STÁTNÍM ROZPOČTEM ČESKÉ REPUBLIKY.

Dále můžeme uvedenou soustavu diferenciálních rovnic postupně řešit a s přihlédnutím k počátečním podmínkám
$p_0(0) = P(X(0) = 0) = 1$,
$p_k(0) = P(X(0) = k) = 0$ pro $k = 1, 2, \ldots$
dostaneme řešení ve tvaru

$$p_k(t) = P(X(t) = k) = \frac{(\lambda t)^k}{k!}\, e^{-\lambda t},\;\; k = 0, 1, 2, \ldots,\;\; t \geq 0.$$

Protože rozdělení počtu rizikových jevů $X(t)$, které jsou zaznamenány v čase t, je rozdělení Poissonovo, nazývá se náhodný proces $X(t)$ odvozený za předpokladů P1, P2 a P3 **Poissonův proces**. Parametr λ se nazývá **intenzita** tohoto procesu. Významnou vlastností Poissonova procesu je, že délky intervalů mezi vznikem sousedních rizikových událostí jsou mezi sebou nezávislé náhodné veličiny a všechny mají stejné exponenciální rozdělení pravděpodobností s parametrem λ.

Poissonův proces lze definovat jako proces s nezávislými přírůstky a se stejným exponenciálním rozdělením dob mezi vznikem rizikových jevů.

Z vlastností Poissonova procesu lze odvodit, že pro přírůstek počtu rizikových jevů na intervalech $(t_1, t_2\rangle$ platí

$$P(X(t_2) - X(t_1) = k) = p_k(t_2 - t_1) = \frac{(\lambda(t_2 - t_1))^k}{k!}\, e^{-\lambda(t_2 - t_1)},\;\; k = 0, 1, \ldots$$

Tedy přírůstek $X(t_2) - X(t_1)$ má opět Poissonovo rozdělení s parametrem $\lambda(t_2 - t_1)$. Odtud plyne interpretace intenzity λ. Protože střední hodnota Poissonova rozdělení $Po(\lambda)$ je rovna λ, platí, že pro střední počet rizikových jevů na intervalu $(t_1, t_2\rangle$ pro libovolné $0 \leq t_1 < t_2 < \infty$ platí

$$E(X(t_2) - X(t_1)) = \lambda(t_2 - t_1).$$

Je tedy λ rovno střednímu počtu vzniklých rizikových jevů na intervalu jednotkové délky $(t, t+1\rangle$ pro libovolný čas $t \geq 0$.

1.1 Odhad intenzity Poissonova procesu

Vyjdeme z procesu, který popisuje vznik dopravních nehod v daném městě. Předpokládejme, že tento proces byl sledován pro časový interval $(0, t\rangle$ a během tohoto intervalu nastalo právě $N(t)$ nehod. Předpokládejme dále, že $N(t)$ je Poissonův proces s intenzitou λ. Je-li na konci časového intervalu $(0, t\rangle$ pozorováno právě n nehod, tj. $N(t) = n$, pak lze odhad $\widetilde{\lambda}$ intenzity λ najít tzv. metodu maximální věrohodnosti ve tvaru

$$\widetilde{\lambda} = \frac{n}{t}\,.$$

Uvedený odhad $\widetilde{\lambda}$ udává počet dopravních nehod za jednotku času. Lze odvodit, že uvedený odhad je nestranný, tzn., že $E\widetilde{\lambda} = \lambda$ a pro jeho rozptyl dostaneme vyjádření $D\widetilde{\lambda} = \frac{\lambda}{t}$. Protože $N(t)$ je Poissonův proces, platí $N(t) \sim P_0(\lambda t)$ a odtud lze odvodit $100(1-\alpha)$%ní interval spolehlivosti pro λ ve tvaru

$$\left(\frac{1}{2t}\chi^2_{\frac{\alpha}{2}}(2n);\;\; \frac{1}{2t}\chi^2_{1-\frac{\alpha}{2}}(2n+2)\right), \tag{1}$$

kde $\chi^2_\alpha(\nu)$ značí α-kvantil Pearsonova χ^2 rozdělení o ν stupních volnosti.

evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
UNIVERZITA OBRANY
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

1.2 Srovnání intenzit dvou nezávislých Poissonových procesů

Předpokládejme, že počet dopravních nehod na dané křižovatce tvoří Poissonův proces $N_1(t)$ s intenzitou λ_1. Po určité době byly na této křižovatce provedeny stavební úpravy a po těchto úpravách tvoří počet nehod opět Poissonův proces $N_2(t)$ s intenzitou λ_2. Úlohou je zjistit, zda se změnila intenzita výskytu poruch. Budeme předpokládat, že oba Poissonovy procesy jsou nezávislé . Cílem je testovat hypotézu $H_0\colon \lambda_1 = \lambda_2$ proti alternativní hypotéze $H_1\colon \lambda_1 \neq \lambda_2$.

Dále předpokládejme, že máme k dispozici pozorování procesu $N_1(t)$ po dobu t_1 a počet nehod v čase t_1 na křižovatce před úpravou byl $N_1(t_1) = n_1$, a podobně pozorování procesu $N_2(t)$ na křižovatce po úpravě proběhlo po dobu t_2 a po tu dobu bylo zjištěno n_2 nehod, tedy $N_2(t_2) = n_2$. Pak vhodnou testovací statistikou pro testování H_0 proti H_1 (tedy pro srovnání obou intenzit) je statistika

$$U = \frac{|n_1 t_2 - n_2 t_1|}{\sqrt{(n_1 + n_2) t_1 t_2}}\,. \tag{2}$$

Statistika U má přibližně normální rozdělení $N(0, 1)$, a proto pro velké hodnoty $n_1 + n_2$ lze H_0 zamítnout na hladině významnosti α, když $U > u_{1-\alpha/2}$, kde u_α značí α-kvantil normálního rozdělení $N(0, 1)$.

2 Lineární proces růstu

Budeme uvažovat populaci jedinců, kteří nezanikají a rozmnožují se např. dělením nebo štěpením. Takový model je typický pro jednobuněčné organismy jako jsou třeba bakterie. Při velkém nárůstu této populace je velké riziko epidemiologické pandemie.

Budeme předpokládat, že velikost populace v čase $t = 0$ je a jedinců. Každý jedinec může dát v intervalu $(t, t+\Delta t)$ vznik dalšímu jedinci s pravděpodobností $\lambda\Delta t + o(\Delta t)$ pro $\Delta t \to 0$ a pravděpodobnost, že dá v tomto intervalu vznik více než jednomu jedinci, je $o(\Delta t)$. Jedinci se vzájemně neovlivňují. Označme $X(t)$ rozsah populace v čase t. Z počáteční podmínky plyne, že $X(0) = a$. $X(t)$ je náhodný proces a nazývá se **lineární proces růstu** nebo též **Yuleův proces**. Naším cílem bude stanovit pravděpodobnost $p_k(t) = P(X(t) = k)$, že v čase $t > 0$ bude rozsah populace k. Z uvedených předpokladů užitím binomického rozdělení zjistíme, že pravděpodobnost, že populace, která má v čase t počet jedinců k, se v časovém intervalu $(t, t + \Delta t\rangle$ rozroste o jednoho jedince, je $k\lambda\Delta t + o(\Delta t)$ a pravděpodobnost, že se rozroste o více než jednoho jedince, je $o(\Delta t)$.

Pak pro $t > 0$ a $\Delta t > 0$ dostaneme, že

$$\begin{aligned} p_a(t + \Delta t) &= P(X(t + \Delta t) = a) = P(X(t) = a \wedge X(t + \Delta t) - X(t) = 0) = \\ &= P(X(t) = a)P(X(t + \Delta t) - X(t) = 0) = p_a(t)(1 - \lambda a\Delta t - o(\Delta t)). \end{aligned}$$

Odtud

$$\frac{1}{\Delta t}(p_a(t + \Delta t) - p_a(t)) = -a\lambda p_a(t) - \frac{o(\Delta t)}{\Delta t} p_a(t)$$

a pro $\Delta t \to 0$ dostaneme diferenciální rovnici pro $p_a(t)$

$$\frac{dp_a(t)}{dt} = -a\lambda p_a(t). \tag{3}$$

Analogicky zjistíme, že pro $k > a$ platí

$$\begin{aligned} p_k(t+\Delta t) &= P(X(t+\Delta t)=k) = P(X(t)=k \wedge X(t+\Delta t)-X(t)=0)+ \\ &+ P(X(t)=k-1 \wedge X(t+\Delta t)-X(t)=1) + o(\Delta t) = \\ &= p_k(t)(1-\lambda k\Delta t - o(\Delta t))+ \\ &+ p_{k-1}(t)(\lambda(k-1)\Delta t + o(\Delta t)) + o(\Delta t)). \end{aligned}$$

Odtud

$$\frac{1}{\Delta t}(p_k(t+\Delta t)-p_k(t)) = -\lambda k p_k(t) + \lambda(k-1)p_{k-1}(t) + \frac{o(\Delta t)}{\Delta t}(1-p_k(t)+p_{k-1}(t))$$

a limitním přechodem pro $\Delta t \to 0$ dostaneme

$$\frac{dp_k(t)}{dt} = \lambda(k-1)p_{k-1}(t) - \lambda k p_k(t) \text{ pro } k = a+1, a+2, \dots. \tag{4}$$

Soustavu diferenciálních rovnic (3) a (4) lze řešit pomocí tzv. vytvořujících funkcí a při počáteční podmínce $p_a(0) = 1$ dostaneme řešení

$$p_k(t) = \binom{k-1}{a-1} e^{-a\lambda t}\left(1-e^{-\lambda t}\right)^{k-a} \quad \text{pro } k \geq a.$$

Označme $N(t) = X(t) - a$ nárůst populace v čase t. Pak

$$\begin{aligned} P(N(t)=n) = P(X(t)-a=n) &= P(X(t)=n+a) = p_{n+a}(t) = \\ &= \binom{n+a-1}{a-1} e^{-a\lambda t}\left(1-e^{-\lambda t}\right)^{n}, \quad n = 0,1,2,\dots. \end{aligned}$$

Tudíž nárůst populace $N(t)$ má v čase t negativně binomické rozdělení $NB(\gamma, \theta)$ s parametry $\gamma = a$ a $\theta = e^{-\lambda t}$.

Dále z vlastností negativně binomického rozdělení lze snadno odvodit, že střední rozsah populace v čase t je $m(t) = EX(t) = ae^{\lambda t}$, tedy střední hodnota rozsahu populace roste exponenciálně rychle.

Pro rozptyl rozsahu populace v čase t dostaneme

$$\sigma^2(t) = DX(t) = E(X(t)-m(t))^2 = ae^{\lambda t}\left(e^{\lambda t}-1\right).$$

3 Lineární proces zrodu a zániku

Vyjdeme ze stejných předpokladů jako u lineárního procesu růstu a budeme navíc předpokládat, že jedinci dané populace mohou také zanikat. Budeme tedy předpokládat, že pro každý časový okamžik t platí:

- libovolný jedinec uvažované populace může dát vznik dalšímu jedinci v intervalu $(t, t+\Delta t\rangle$ s pravděpodobností $\lambda\Delta t + o(\Delta t)$, $\lambda > 0$ je konstanta,
- libovolný jedinec této populace může v intervalu $(t, t+\Delta t)$ zaniknout s pravděpodobností $\mu\Delta t + o(\Delta t\rangle$, $\mu > 0$ je konstanta,
- jedinci se vzájemně neovlivňují,
- v čase $t = 0$ obsahuje populace a jedinců.

Proces $X(t)$, který popisuje rozsah populace v čase t, se nazývá **lineární proces zrodu a zániku**.

Lze snadno nahlédnout, že

– pravděpodobnost vzniku 1 jedince v intervalu $(t, t+\Delta t\rangle$ je $\lambda X(t)\Delta t + o(\Delta t)$,

– pravděpodobnost zániku 1 jedince v intervalu $(t, t+\Delta t\rangle$ je $\mu X(t)\Delta t + o(\Delta t)$,

– pravděpodobnost, že dojde ke změně rozsahu populace o více než 1 jedince je $o(\Delta t)$,

– že nedojde ke změně rozsahu populace je $1 - (\lambda+\mu)X(t)\Delta t + o(\Delta t)$.

Z těchto předpokladů pak lze odvodit následující soustavu diferenciálních rovnic pro pravděpodobnosti $p_k(t) = P(X(t) = k)$:

$$\frac{dp_0(t)}{dt} = \mu p_1(t),$$
$$\frac{dp_k(t)}{dt} = \lambda(k-1)p_{k-1}(t) - (\lambda+\mu)kp_k(t) + \mu(k+1)p_{k+1}(t), \quad k \geq 1.$$

Řešení uvedené soustavy za počáteční podmínky $p_a(0) = 1$ je v případě, že $\lambda \neq \mu$, tvaru

$$p_k(t) = \sum_{j=0}^{\min(a,k)} \binom{a}{j}\binom{a+k-j-1}{a-1}\alpha^{a-i}\beta^{k-j}(1-\alpha-\beta)^j \text{ pro } k > 0,$$
$$p_0(t) = \alpha^a,$$

přičemž

$$\alpha = \alpha(t) = \frac{\mu\left(e^{(\lambda-\mu)t}-1\right)}{\lambda e^{(\lambda-\mu)t}-\mu} \qquad \text{a} \qquad \beta = \beta(t) = \frac{\lambda\left(e^{(\lambda-\mu)t}-1\right)}{\lambda e^{(\lambda-\mu)t}-\mu}.$$

Pro střední hodnotu tohoto procesu pak dostaneme

$$m(t) = EX(t) = ae^{(\lambda-\mu)t}$$

a pro rozptyl

$$\sigma^2(t) = DX(t) = \frac{a(\lambda+\mu)}{(\lambda-\mu)}\, e^{(\lambda-\mu)t}\left(e^{(\lambda-\mu)t}-1\right).$$

V případě, že $\lambda = \mu$ dostaneme pro střední hodnotu

$$m(t) = EX(t) = a$$

a pro rozptyl

$$\sigma^2(t) = DX(t) = 2a\lambda t.$$

Je-li rozsah populace v čase 0 roven $a = 1$, pak lze pro $\lambda = \mu$ vyjádřit pravděpodobnosti $p_k(t)$ ve tvaru

$$p_0(t) = \frac{\lambda t}{1+\lambda t},$$
$$p_k(t) = \frac{(\lambda t)^{k-1}}{(1+\lambda t)^{k+1}} \quad \text{pro } k = 1, 2, \ldots .$$

Příklady k procvičení

1. Při práci balicí linky došlo během 2456 hodin provozu v 254 případech k poruše spočívající ve roztrhání balicího papíru. Předpokládejme, že časové okamžiky vzniku poruch tvoří Poissonův proces intenzitou λ. Nalezněte odhad parametru λ a stanovte 95% interval spolehlivosti pro λ.

2. Předpokládejme, že na dané křižovatce došlo během 24 měsíců k 25 dopravním nehodám. Potom byla křižovatka opravena a nahrazena kruhovým objezdem. Na tomto kruhovém objezdu následně došlo během 12 měsíců k deseti nehodám. Předpokládejme, že procesy vzniku nehod lze modelovat nezávislými Poissonovy procesy s intenzitami λ_1 a λ_2. Má se zjistit, zda přestavba křižovatky měla významný vliv na snížení nehodovosti.

3. Předpokládejme, že počet jedinců dané populace lze popsat pomocí lineárního procesu zrodu a zániku. Každý jedinec může dát vznik novému jedinci s konstantní intenzitou $\lambda = 0{,}02$ každý jedinec může zahynout s konstantní intenzitou $\mu = 0{,}005$. Počáteční počet jedinců populace je $a = 30$. Stanovte:

   a) Pravděpodobnost, že v čase t bude mít populace právě k jedinců pro $k \in \{1, 2, 3\}$ a graficky ji v závislosti na čase znázorněte.

   b) Střední hodnotu počtu jedinců v populaci v čase t a graficky ji znázorněte v závislosti na čase.

   c) Rozptyl počtu jedinců v čase t a graficky jej znázorněte v závislosti na čase.

   K řešení můžete využít software R nebo MATLAB.