# Pravděpodobnostní metody

## 1 Úvod

### 1.1 Zákl. definice

**Definice** *(Jevy)*

Pro diskrétní pozorování: $\Omega$ – lib. konečná nebo spočetná množ. – *pravděpodobnostní prostor*,
$\omega_i \in \Omega$ – *elementární jevy*
$A, B, \dots \subseteq \Omega$ – *jevy*, odp. nějakým reálným otázkám pokusu

**Definice** *(Pravděpodobnostní prostor)*

*Pravděpodobnostní prostor* – $(\Omega, \mathcal{A}, P)$, kde $\mathcal{A}$ je mn. spočetných sjednocení a průniků jevů a $P$ zobrazení: $P(A) : \mathcal{A} \to [0,1]$, t.ž. $O \leq P(A) \leq 1$, $P(\Omega) = 1$, $P(\cup_i A_i) = \sum_i P(A_i)$ pro $A_i \cap A_j = \emptyset$, pro nedisjunktní používám PIE.

**Definice** *(Podmíněná pravděpodobnost)*

$P(A|B) = \frac{P(A \cap B)}{P(B)}$, pokud $P(B) \neq 0$. Zmenším si tak soubor výsledků, zjednoduším úlohu.

**Definice** *(Nezávislost jevů)*

V prostoru $(\Omega, \mathcal{A}, P)$ jsou $A$ a $B$ nezávislé, pokud $P(A|B) = P(A)$, tj. $P(A \cap B) = P(A)P(B)$ – $A$ nastane se stejnou pravděpodobností, ať už $B$ nastalo, nebo ne.

$A_1, \dots, A_n$ jsou *sdruženě nezávislé*, pokud $\forall i_1, \dots, i_k$ $(k = 2, \dots, n)$ platí $P(A_{i_1} \cap \dots \cap A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j})$, tj. jakákoliv podmnožina je nezávislá.

**Definice** *(Úplná pravděpodobnost)*

Uvažujeme dělení prostoru $(\Omega, \mathcal{A}, P)$ na „mřížku“, tj. $\Omega := \cup_{i=1}^{n} H_i$, kde $H_i \cap H_j = \emptyset$ $(i \neq j)$; pak $H_i$ je *úplný systém jevů*. $H_i$ jsou disjunktní, ale nemusí být nezávislé. Potom platí:

$$P(A) = P(A \cap \Omega) = P(A \cap (\cup_i H_i)) = \sum_i P(A \cap H_i) = \sum_i P(A|H_i)P(H_i)$$

**Věta** *(Bayesova věta)*

$$P(H_i|A) = \frac{P(H_i \cap A)}{P(A)} = \frac{P(A|H_i)P(H_i)}{\sum_j P(A|H_j)P(H_j)}$$

**Definice** *(Náhodná veličina)*

je zobrazení $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}_1)$ (kde $\mathcal{B}_1$ značí borellovské množiny), náhodná funkce. Zjednodušuje počítání, konstrukce prostoru pravděpodobnosti může v někt. případech dělat potíže.

### 1.2 Rozdělení

**Definice** *(Rovnoměrné)*

$P(X = i) = \frac{1}{n}$ $\forall i \in \{1, \dots, n\}$ – žádná z možností není pravděpodobnější než ostatní.

**Definice** *(Poissonovo($\lambda$))*

$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ pro $k = 0, 1 \dots$ – docela umělé, ale občas se objevuje jako limitní rozdělení, faktoriál rychle dotlačí pravděpodobnosti k 0 – „rozdělení řídkých jevů“.

**Definice** *(Binomické rozdělení($p, n$))*

$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ – např. opakované tahání kuliček z osudí s vracením – součet nezávislých jevů z alt. rozdělení.

**Definice** *(Hypergeometrické rozdělení$(N, M, n)$)*
$P(X=k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$ – tahání bez vracení, tj. jevy jsou závislé.

**Definice** *(Zobecněné binomické)*
$P(n_0 = 0, n_1 = 1, \ldots, n_k = k) = \frac{n!}{n_0! n_1! \ldots n_k!} p_0^{n_0} \cdot \cdots \cdot p_k^{n_k}$ – provádím nezávislých $n$ pokusů, které mohou dopadnout s hodnotou v rozmezí $\{0, \ldots, k\}$. Zobecnění na vícerozměrné, tj. vektor.

**Definice** *(Geometrické rozdělení, negativně binomické rozdělení)*
– *geometrické* – $P(X=n) = p(1-p)^{n-1}$ – čekání na 1. zdar pro alternativní rozdělení.
– *negativně binomické* – $P(X=r) = \binom{n-1}{r-1}(1-p)^{n-r}p^r$ – zobecnění, $r$-tý zdar v $n$ pokusech pro alt. rozdělení.

## 1.3 Vlastnosti rozdělení

**Definice** *(Distribuční funkce)*
je funkce $F(x) : F(x) = P(X \leq x)$ por nějakou náh. veličinu $x$. Platí: $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to +\infty} F(x) = 1$, funkce je neklesající a zprava spojitá. V diskrétním rozdělení lze definovat tabulkou.

**Definice** *(Vytvořující funkce)*
$A(x) = \sum_i p_i x^i$ – má jednoznačný vztah k distribuční funkci, ve spojitém případě je to její integrál, v diskrétním platí, že znám-li $p_i$ (např. při zadání distribuční funkce tabulkou), pak znám i vytv. funkci a naopak: $\frac{A^{(n)}(0)}{n!} = p_i$.

**Definice** *(Střední hodnota)*
$EX = \sum_i x_i p_i = \sum_\omega X(\omega)P(\omega)$ – vážený průměr hodnot, z hlediska 1 opakování testu ale nic neříká.
– platí $E(a + bX + cY) = a + bEX + cEY$ a to i když $X, Y$ jsou závislé

**Definice** *(Kovariance, korelace)*
– *kovariance* – $\mathbf{cov}(X,Y) = E((X-EX)(Y-EY))$ (a pro lineárně nezávislé jevy je nulová)
– *korelace* – popis lin. závislosti $X, Y$: $\mathbf{cor}(X,Y) = \frac{\mathbf{cov}(X,Y)}{\sqrt{\mathbf{var}(X)}\sqrt{\mathbf{var}(Y)}}$

**Definice** *(Rozptyl)*
$\mathbf{var}(X) = E(X-EX)^2$ – doplnění o další informaci, stř. hodnota čtverce odchylky od stř. hodnoty
– platí $\mathbf{var}(a+bX) = b^2\mathbf{var}(X)$
– odmocnina z rozptylu je *směrodatná odchylka*
– platí $\mathbf{var}(a+bX+cY) = b^2\mathbf{var}(X) + c^2\mathbf{var}(Y) + 2bc \cdot \mathbf{cov}(x,Y)$

**Poznámka** *(Rozdělení součtů)*
Přesný výpočet – pomocí konvolucí: $P(X+Y=i) = P(X=j \& Y=i-j) = \sum_{j=0}^{i} P(X=j)P(Y=i-j)$

Pomocí vytvořujících funkcí – $X$ odp. $A(X)$, $Y$ odp. $B(X)$ nezávislé, pak $(X+Y)$ odpovídá vytv. funkce $A(x)B(x)$ (pro důkaz stačí porovnat koeficienty u jednotlivých $x^n$, indukcí platí i pro součty více veličin).

**Poznámka** *(Náhodný součet náhodných veličin)*
Tj. mám-li náhodný počet náh. (stejně rozdělených) veličin, chci jejich součet ($S_N = X_1 + \cdots + X_N$) – vytvořující funkce vyjde $B(A(x))$ (přes $P(S_N = k)$ odpovídá $k$-tému členu $n$-té konvoluce nez. vel. $X_i$). Z toho $ES_N = EX \cdot EN$.

Příklady použití – přežití hadích vajíček ($X_i \sim Alt(p)$, $N \sim Po(X)$), $n$-tá generace / štěpění jader – $P(P(x))$ je druhá generace.

## 1.4 Věty

**Věta** *(Čebyševova nerovnost)*
$P(|X-EX| \geq \varepsilon) \leq \frac{\mathbf{var}(X)}{\varepsilon^2}$
– plyne z ní „konvergence v pravděpodonosti“, tj. $\bar{X_n} \to (p)\mu$ (při opakování pokusů a průměrování výsledku se nemění stř. hodnota, ale rozptyl klesá s $\frac{1}{n}$)

**Věta** *(Markovova nerovnost)*
$P(|X - EX|^2 \geq \varepsilon) \leq \frac{E|X|^2}{\varepsilon^2}$

**Věta** *(Centrální limitní věta)*
Platí pro $X_1, X_2 \ldots$ nezávislé a stejně rozdělené, $EX_i = \mu$, $\mathbf{var}(X_i) = \sigma^2$ (konečný rozptyl!). Potom:

$$P(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathbf{d}t$$

Tedy rozdělení součtu nezávislých stejně rozdělených náh. veličin mohu aproximovat Gaussovým rozdělením. Aproximace je dobrá už od $n = 20$, pro binom. rozdělení od $n = 50$, $p \in [0.1, 0.9]$.

## 2 Rekurentní jevy

**Motivace**
Mějme náh. veličinu, nabývající hodnot $-1, +1$ s pravděpodobnostmi $p, 1-p$. V tomhle případě chceme simulovat součet $n$ hodnot $S_n$ (např. stav konta po prohrách/výhrách). To nabývá hodnot po celém $\mathbb{Z}$. Chceme zjistit, kdy se dostaneme zpět na 0, jednotlivé hodnoty jsou závislé – $S_n$ přímo závisí na $S_{n-1}$.

**Definice** *(Rekurentní jev)*
Mějme náhodný pokus, s postupnými výsledky $E_0, E_1, \ldots$ (stav konta) dodanými pomocnými pokusy $E_{j0}, E_{j1}, \ldots$ (např. přičítání $+1, -1$). Řekneme, že *jev $\xi$ nastal v $n$-tém pokusu* (jev $\xi$ je nad $E_0, E_1 \ldots$, tj. např. stav konta se vrátil na 0 po $n$-tém pokusu), pokud nastal na konci posloupnosti $\{E_{j1}, E_{j2} \ldots, E_{jn}\}$. Řeknu, že *jev $\xi$ nastal v $m$-tém a $n+m$-tém pokusu*, pokud nastal na konci posloupnosti $\{E_{j1}, E_{j2} \ldots, E_{jm}\}$ a $\{E_{j(m+1)}, E_{j(m+2)} \ldots, E_{j(m+n)}\}$.

Pokud pro takový jev platí $P(E_{j1}, \ldots E_{j(m+n)}) = P(E_{j1}, \ldots E_{jm}) \cdot P(E_{j(m+1)}, \ldots E_{j(m+n)})$, nazveme ho *rekurentní jev*. Neříkám tu nic o závislosti bezprostředně po sobě jdoucích výsledků.

**Definice** *(Pravděpodobnosti nastání)*
Označím pravděpodobnosti

1. $u_n = P(\xi$ nastal v kroku $n)$, vytvořující funkce $U(x)$
2. $f_n = P(\xi$ nastal v kroku $n$ POPRVÉ $)$, funkce $F(x)$
3. $f_n^r = P(\xi$ nastal v kroku $n$ PO $r$-TÉ $)$

**Poznámka** *(Vztah $u_n$ a $f_n$)*
Když jev $\xi$ nastal v kroku $n$, tak někdy musel nastat poprvé. Pokud to bylo v $k$-tém kroku, pak lze uvažovat, že $\xi$ nastal na konci posloupnosti délky $n - k$. Podle věty o úplné pravděpodobnosti platí pro každé $n \geq 1$:

$$u_n = f_1 u_{n-1} + f_2 u_{n-2} + \cdots + f_n u_0$$

Takže dostávám konvoluci bez 1 členu, dodefinuji $u_0 = 1$, $f_0 = 0$ (potom $f_n$ tvoří rozdělení pravděpodobnosti,ale $u_n$ ne).

**Věta** *(Vztah vytv. funkcí rek. jevů)*
Platí: $U(x) = \frac{1}{1-F(x)}$.
**Důkaz**
Z předch. výpočtů dostanu rovnice tvaru $u_n = f_0 u_n + f_1 u_{n-1} + \cdots + f_n u_0$ (pro každé $n > 0$, pro $n = 0$ neplatí!), vynásobím je $x^n$ a sečtu po sloupcích, takže dostanu $U(x) = u_0 x^0 + U(x)F(x)$.

**Věta** *(Vytv. funkce pro $r$-tý návrat)*
Pro rekurentní jevy spočteme vytv. funkci $r$-tých návratů jako $r$-tou mocninu vytv. funkce prvních návratů.

**Důkaz**

Pro $r = 2$ podobně jako v předchozí větě, uvažuju jenom, že $f_n^2 = f_0 f_n + f_1 f_{n-1} + \cdots + f_n f_0$, tj. dostávám úplnou konvoluci pro každé $n \geq 0$, z ní podobně vynásobením rovnic $x^n$ a sečtením po sloupcích dostávám výsledek. Pro $r$ vyšší než 2 plyne zjevně indukcí.

**Poznámka**

Chová se podobně jako součet nezávislých náhodných veličin – jde vlastně o nezávislé náhodné veličiny $T_1, T_2, \ldots$, které představují *doby mezi návraty* jevu. Po nastání jevu se jede „odznova“, zapomínám předchozí pokusy, tj. veličiny jsou nezávislé. Proto se doba čekání na $n$-tý výskyt jevu $\xi$ dá popsat náh. veličinou $T_1 + T_2 \cdots + T_n$, kde $T_i$ jsou nezávislé, stejně rozdělené a popisují dobu do prvního výskytu jevu.

# 3 Markovovy řetězce

Přechod z rekurentních jevů, změna značení: $f_i \to f_{ii}$, $u_i \to p_{ii}$.

## 3.1 Definice

**Definice** *(Markovův řetězec s diskréním časem a množinou stavů $S$)*

Je dána posloupnost náh. veličin $\{X_n, n \in \mathbb{N}_0\}$ s celočíselnými hodnotami a stavový prostor $S \subseteq \mathbb{Z}$, definovaný předpisem $i \in S \equiv \exists n \in \mathbb{N} : P(X_n = i) \geq 0$. Potom posloupnost $\{X_n\}$ nazveme *Markovův řetězec*, pokud splňuje „Markovovu vlastnost“:

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots X_0 = i_0) = P(X_{n+1} = j | X_n = i) \quad \forall n \in \mathbb{N}_0$$

přičemž musí platit $\forall i, j, i_{n-1}, \ldots, i_0 \in S : P(X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) > 0$.

**Definice** *(Pravděpodobnost přechodu)*

Označíme $P_{ij}(n, m+n) := P(X_{m+n} = j | X_n = i)$, $P_{ij}(n) := P(X_{n+1} = j | X_n = i)$.

**Definice** *(Pravděpodobnost a stř. hodnota podmíněná stavem)*

Označíme $P_j(\cdot) = P(\cdot | X_0 = j)$, kde $\cdot$ znamená libovolný jev, tj. pravděpodobnost nastání mého jevu, za podmínky, že se nacházím ve stavu $j$.

Označíme $E_j(\cdot) = E(\cdot | X_0 = j)$ střední hodnotu nějakého jevu za předpokladu, že vycházím ze stavu $j$.

**Definice** *(Homogenní Markovův řetězec)*

Nezávisí-li $P_{ij}(n, m+n)$, potažmo $P_{ij}(n) := p_{ii}$ na $n$, nazývá se řetězec *homogenní.*

**Poznámka**

Průchod počátečním stavem v Markovově řetězci je rekurentní jev. Průchod nějakým jiným stavem je rekurentní jev *se zpožděním*, způsobeným nutností poprvé dosáhnout onoho stavu z počátečního. Markovův řetězec lze reprezentovat maticí $\mathcal{P} = (p_{ij})$ spolu s „rozdělením počátečních pravděpodobností“ (udáním počátečního stavu) $a = \{a_i\}$.

## 3.2 Klasifikace stavů

**Definice** *(Čas návratu)*

Označíme minimální čas návratu do stavu $j$ – $\tau_j(1) = \inf\{n > 0, n \in \mathbb{N}, X_n = j\}$ (s $\inf\{\emptyset\} = \infty$) jako *čas prvního návratu.* $k$-tý návrat definujeme rekurentně $\tau_j(k+1) = \inf\{n > \tau_j(k) : X_n = j\}$ a 0-tý návrat $\tau_j(0) = 0$.

Definuji:

1. $P_i(\tau_j(1) = n) := f_{ij}^{(n)}$, $f_{ij}^{(0)} = 0$ (pravděpodobnost prvního přechodu z $i$ do $j$ v čase $n$)
2. $P_i(\tau_j(1) < \infty) = \sum_{n=1}^{\infty} f_{ij}^{(n)} := f_{ij}$ (pravd. vůbec nějakého přechodu z $i$ do $j$)

**Definice** *(Trvalý a přechodný stav)*
Řeknu, že stav $j$ je:

1. *trvalý*, když $f_{jj} = 1$ (v konečném čase se do něj jistě vrátím)
2. *přechodný*, když $f_{jj} < 1$ (existuje možnost, že čas návratu je nekonečný)

**Definice** *(Nulový a nenulový stav)*
Řeknu, že trvalý stav $j$ je:

1. *nenulový*, když $\mu_j := E_j\tau_j(1) < +\infty$ (tj. stř. hodnota času 1. návratu je konečná)
2. *nulový*, když $\mu_j = \infty$

**Poznámka** *(O konvolucích v homogenních mark. řetězcích)*

1. $p_{jj}^{(n)} = \sum_{k=0}^{n} f_{jj}^{(k)} p_{jj}^{(n-k)}$ $(n \geq 1)$ (pravd. nějakého návratu = 1. návrat + zbytek)
2. $p_{ij}^{(n)} = \sum_{k=0}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}$ $(n \geq 0, i \neq j)$ (1. přechod do cíl. stavu + zbytek)

**Věta** *(O trvalém stavu)*
Stav v homogením M. ř. je trvalý, právě když $\sum_{n=0}^{\infty} p_{jj}^{(n)} = \infty$ (pravděpodobnosti návratu pro čas jdoucí k nekonečnu jdou k 1, tj. sčítám řadu jedniček).

Podobně lze definovat, že stav je přechodný, když tato suma konverguje.

**Poznámka**
Do trvalého stavu vstoupí nějaká nekonečná markovovská posloupnost stavů nekonečně mnohokrát s pravděpodobností 1. Stav je přechodný, když počet vstupů do něj v takové posloupnosti je konečný s pravděpodobností 1.

**Definice** *(Periodický a neperiodický stav)*
Nechť $d_j$ je největší společný dělitel čísel $n \geq 1$ $(n \in \mathbb{N})$ takových, že $p_{jj}^{(n)} > 0$. Potom

1. jestliže $d_j = 1$, řekneme, že $j$ je *neperiodický* stav,
2. jestliže $d_j > 1$, řekneme, že $j$ je *periodický* stav.

**Poznámka** *(Zjišťování stavů na základě konvergence $p_{ij}^{(n)}$)*
V případě, že je stav $j$ přechodný nebo trvalý a nulový, platí $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ $\forall i \in S$. Jen je-li stav $j$ trvalý a nenulový, je tato limita pro nějaké $i$ nenulová.

**Příklady**

1. Symetrická náh. procházka po přímce – začínám z 0, s pravd. 1/2 jdu doleva a s 1/2 doprava z každého stavu. Pak je stav 0 trvalý, nulový a periodický s periodou 2.
2. Nesymetrická procházka – doleva jdu s pravd. $q$, doprava s $p$. Pak stav 0 je přechodný! Používá se v generátorech náh. čísel.

## 3.3 Stacionární rozdělení

**Definice** *(Rozložitelný homogenní Markovův řetězec)*
Stav $E_j$ je *dosažitelný* ze stavu $E_i$, když existuje $n$ (čas) takové, že $p_{ij}^{(n)} > 0$. Markovův řetězec je *nerozložitelný*, jestliže všechny stavy jsou navzájem dosažitelné.

**Věta** *(O rozložitelném řetězci)*

Máme-li homogenní M. řetězec s konečně mnoha stavy, potom je rozložitelný, právě když matice přechodu se dá vyjádřit jako

$$P = \begin{pmatrix} P_1 & \emptyset \\ A & B \end{pmatrix}$$

po případných (simultánních!) permutacích sloupců a řádků, a $P_1$ a $B$ jsou čtvercové.

**Věta** *(O dosažitelnosti a typu stavu)*

1. Je-li stav $j$ dosažitelný z $i$ a opačně, pak $i$ a $j$ jsou stejného typu (stejná nulovost a periodicita ??).
2. Je-li $j$ stav trvalý a $k$ z něj dosažitelný, pak $k$ je trvalý a zároveň $j$ je dosažitelný z $k$ (protože se do $j$ a z něj zas do $k$ určitě v konečném čase vrátím).

**Důsledek**

V konečném homogenním Markovově řetězci nemůžou všechny stavy být přechodné a neexistují nulové stavy. Po určité době pohyb v homogenním konečném M. řetězci dospěje do něčeho, čemu říkáme *stacionární rozdělení.*

**Definice** *(Stacionární pravděpodobnostní rozdělení)*

Mějme homogenní Markovův řetězec $\{X_n\}$, množinu stavů $S$ a matici pravděpodobností přechodů $\mathcal{P}$. Pak je-li $\Pi = \{\pi_, j \in S\}$ nějaké pravděpodobnostní rozdělení na $S$ (tj. $\pi_j \geq 0 \; \forall j \in S, \sum_{j \in S} \pi_j = 1$), nazveme ho *stacionárním rozdělením*, platí-li $\Pi^T = \Pi^T P$ (tj. zachovává pro jednotku času rozložení „částic").

**Poznámka**

Pro homogenní M. ř. se stacionárním (počátečním) rozdělením platí $P_1(X_0 = i_0, \ldots, X_n = i_n) = P_k(X_0 = i_0, \ldots x_n = i_n) \; \forall k, \forall i_j, j = 1 \ldots k$

**Věta** *(Vlastnosti nerozložitelného homogenního řetězce)*

Nerozložitelný homogenní M. řetězec:

1. má-li všechny stavy přechodné nebo trvale nulové, pak v něm stacionární rozdělení neexistuje
2. má-li všechny stavy tvale nenulové, pak v něm existuje *právě jedno* stacionární rozdělení.
   - Jsou-li stavy neperiodické, potom navíc platí $\lim_{n\to\infty} p_{ij}^{(n)} = \Pi_j(>0), i, j \in S$
     a $\lim_{n\to\infty} p_{jj}^{(n)} = \Pi_j(>0), j \in S$
   - jsou-li stavy periodické, pak platí (Cauchyovsky) $\lim_{k\to\infty} \frac{1}{n}\sum_{i=1}^{n} p_{ij}^{(k)} = \Pi_j(>0), i, j \in S$
     a $\lim_{k\to\infty} \frac{1}{n}\sum_{i=1}^{n} p_{jj}^{(k)} = \Pi_j(>0), j \in S$

   tj. stacionární rozdělení je limitní, existuje-li.

Pro konečný počet stavů nastává určitě varianta 2.

**Příklady**

- Markovův řetězec daný maticí

$$P = \begin{pmatrix} q & p & 0 & 0 & \ldots \\ q & 0 & p & 0 & \ldots \\ 0 & q & 0 & p & \ldots \\ 0 & 0 & q & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

  kde $p + q = 1, 0 < p < 1$. Je nerozložitelný, hledám stacionární rozdělení, najdu z $(\Pi_1, \Pi_2 \ldots) = (\Pi_1, \Pi_2 \ldots) \cdot P$ pomocí rozepsání na nekonečno rovnic, vyjde (z rekurentního vztahu $\Pi_j = q\Pi_{j-1} + p\Pi_{j+1}$) vzorec $\Pi_j = (\frac{p}{q})^j \Pi_0 (j \in \mathbb{N}_0)$, potom z nerozložitelnosti jsou stavy nenulové a trvalé; pokud $\sum_{j=1}^{\infty} \Pi_j$ konverguje, stacionární řešení existuje.

- Nějaký automat s posouváním konečné pásky délky $n$ s pravidly $1 \to 0, L$ a $0 \to 1, R$ a $* \to 1, R$ (kde $*$ je jen na začátku a na konci), na počátku mám na pásce nějaké náh. číslo, kde prvních $m$ ($0 \leq m \leq n$) jsou jedničky – mám „useknutě“ geometricky rozdělenou náh. veličinu. Hledání první jedničky se dá popsat M. ř. s konečně mnoha stavy.
- Mám $N$ kuliček ve dvou pikslách $A$ a $B$. V každém kroku náhodně vyberu jednu kuličku a z té piksly, ve které je, jí přesunu do druhé. Tohle se dá simulovat Markovovým řetězcem – stav systému popíšu jako počet kuliček v piksle $A$.
  Potom ze stavu $E_n$, kdy v $A$ je $n$ kuliček, jdu s pravděpodobností $\frac{n}{N}$ do stavu $E_{n-1}$ a s $\frac{N-n}{N}$ do stavu $E_{n+1}$. Z toho si můžu udělat matici $P$ a pak hledat $\Pi$ tak, že $\Pi^T = \Pi^T P$ (nebo $P^T \Pi = \Pi$).
  To se dělá tak, že pro každý stav $E_n$ si napíšu rovnici $\pi_n = \frac{N-(n-1)}{N}\pi_{n-1} + \frac{n+1}{N}\pi_{n+1}$. (V té matici se podívám na příslušný sloupec – jde o součet toho, s jakou pravděpodobností se z jakého stavu dostanu do $E_i$. Když se na to podíváte, tady to platí i pro krajní stavy, kdy jeden z těch členů je prostě nulový.) Z toho můžu vyjádřit rekurentní vztah (protože $\pi_0 = \frac{1}{N}\pi_1$ a takhle se dá dosazovat dál a dál a dál), vyjde $\pi_n = \frac{n}{N}\pi_n + \frac{n+1}{N}\pi_{n+1}$ a z toho dostanu $\pi_{n+1} = \frac{N-n}{n+1}\pi_n$.
  Z rekurentního vztahu jde pomocí $\pi_0$ vyjádřit úplně všechny členy, díky kouzlům s násobením pak vychází, že $\pi_n = \binom{N}{n}\pi_0$. Abych dostal platné pravděpodobnostní rozdělení, součet všech pravděpodobností všech stavů musí dávat jedničku, tj. $\sum_{i=1}^{N}\binom{N}{i}\pi_0 = 1$. Z binomické věty vyjde, že $2^N\pi_0 = 1$ a tedy $\pi_0 = \frac{1}{2^N}$, a tak dostanu celé rozdělení.

# 4 Fronty

## 4.1 Exponenciální rozdělení

**Poznámka** *(Vztah Exp a Ge)*

Mezi exponenciálním a geometrickým rozdělením je oboustranný vztah:

- $Exp(\lambda) \sim Ge(1 - e^{-\lambda})$
- $Ge(p) \sim Exp(-\log(1-p))$

Dá se to použít na generování náh. čísel z geom. rozdělení.

**Důkaz**

Pro $Exp(\lambda)$ je hustota pravd. $f(x) = \lambda e^{-\lambda x}$, z toho $F(i) = 1 - \lambda e^{-\lambda i}$ a potom $F(i+1) - F(i) = (1-e^{-\lambda})(e^{-\lambda})^i = pq^i$.

**Poznámka** *(Vlastnosti Exp)*

„Nezáleží na tom, kolik už výrobek přežil, ptám-li se na jeho další přežití – *Exp* nemá paměť.“

$$P(X > x + y | X > x) = e^{-\lambda(x+y)}/e^{-\lambda x} = P(X > y)$$

V realitě je to trochu moc pesimistické, ale dobře se s tím počítá.

**Definice** *(Intenzita poruch)*

Jde o pravděpodobnost okamžitého selhání:

$$P(x < X < x + \delta | X > x) = \frac{P(x \leq X \leq x + \delta | X > x)}{P(X > x)} = \frac{\delta \cdot \frac{F(x+\delta)-F(x)}{\delta}}{1 - F(x)} \approx \delta \frac{f(x)}{1 - F(x)}$$

a právě $f(x)/1 - F(x)$ je *intenzita poruch.*

**Poznámka**

Intenzita poruch exponenciálního rozdělení je konstantní – $\lambda$ a exponenciální je jediné spojité rozdělení, které má tuto vlastnost. Podobně geometrické rozdělení je jediné diskrétní rozdělení s konstatní intenzitou poruch.

## 4.2 Modely zrodu a zániku

**Definice** *(Model zrodu a zániku)*

Mějme nějaký systém s konečně nebo spočetně mnoha stavy $E_i$, kde je možné se v 1 kroku (v intervalu délky $h$) buď posunout k sousedním, nebo zůstat ve stejném:

$$E_i \to \begin{cases} E_{i+1} & \lambda_i h + o(h) \\ E_i & 1-(\lambda_i+\mu_i)h + o(h) \\ E_{i-1} & \mu_i h + o(h) \\ E_{i+k}\ (k \in \mathbb{Z}) & o(h) \end{cases}$$

kde $o(h)$ je libovolně malý člen vzhledem k $h$ ($\frac{o(h)}{h} \to 0$ pro $h \to 0$). $\lambda_i h, \mu_i h$ ukazuje, že doby mezi událostmi se řídí exp. rozdělením.

Označme $P_n(t)$ pravděpodobnost, že v čase $t$ je systém ve stavu $E_n$. Zajímá nás $\lim_{t\to\infty} P_n(t) := p_n$, tj. pravděpodobnosti obsazení systému po té, co se stabilizoval.

$X_i, Y_i$ označíme nějaké náh. veličiny – doby mezi přechody do vyššího, resp. nižšího stavu.

**Algoritmus** *(Výpočet obsazení modelu)*

Chci zjistit $P_n(t+h)$ pro nějaké malé $h$ (tj. pravd., že v čase $t+h$ budu ve stavu $n$). Potom

$$P_n(t+h) = \underbrace{P_n(t)(1-\lambda_n h - \mu_n h + o(h))}_{\text{v čase } t \text{ jsem byl v } P_n} \underbrace{+P_{n+1}(t)(\mu_{n+1}h + o(h))}_{\text{příchod z } P_{n+1}} \underbrace{+P_{n-1}(t)(\lambda_{n-1}h + o(h))}_{\text{příchod z } P_{n-1}} \underbrace{+o(h)}_{\text{přechod o víc}}$$

Z toho získám pro $n \geq 1$ (vydělím $h$ a zahodím $\frac{o(h)}{h}$)

$$P_n'(t) = -(\lambda_n + \mu_n)P_n(t) + \mu_{n+1}P_{n+1}(t) + \lambda_{n-1}P_{n-1}(t)$$

Pro $E_0$ dostanu rovnici

$$P_0'(t) = -\lambda_0 P_0(t) + \mu_1 P_1(t)$$

Z toho mám soustavu dif. rovnic, vezmu nějaké poč. podmínky ($P_i(0) = 1, P_j(0) = 0, j \neq i$, tj. začátek ve stavu $E_i$) a spočítám. Limity v $\infty$ – $p_n$ – dostanu položením levé strany rovnice (derivací) rovné nule a nahrazením $P_n(t)$ přímo $p_n$.

**Poznámka**

1. Model lze charakterizovat jak počty událostí za daný časový úsek (očekávané $\lambda T$ pro délku $T$), tak dobami mezi příchody požadavků – tj. je vidět, že mezi těmito věcmi ($Po(\lambda)$ a $Exp(\lambda)$) je nějaký vztah, že jsou převoditelná (viz dále).
2. Více strojů se stejnou intenzitou lze „sesypat" do jednoho, nezáleží na začátku práce strojů (exp. rozdělení nemá paměť).

**Definice** *(Lineární model zrodu a zániku)*

Mám $n$ jedinců, v čase $[t, t+h)$ každý z nich může umřít (s pravd. $\mu$) nebo se rozdělit na 2 (s pravd. $\lambda$).

- Zvýšení počtu: binomické rozdělení – vyberu 1 dělícího se, ostatní se nedělí

$$\binom{n}{1}(\lambda h + o(h))(1 - \lambda h - \mu h + o(h))^{n-1} = \cdots \approx n\lambda h + o(h)$$

- Snížení: podobně, vyjde $n\mu h + o(h)$.

Označím $F(t,s) := \sum_{n=0}^{\infty} P_n(t)s^n$ – vytvořující funkce $P_n(t)$ pro nějaký pevný čas $t$.

**Věta** *(O vytvořující funkci lineárního modelu)*

Pro případ lineárního modelu zrodu a zániku má $F(t,s)$ tvar

$$F(t,s) = \frac{\mu(1-e^{(\lambda-\mu)t}) - (\lambda - \mu e^{\lambda-\mu)t})s}{\mu - \lambda e^{(\lambda-\mu)t} - \lambda(1-e^{(\lambda-\mu)t})s}$$

protože řešení diferenciálních rovnic dává

$$\begin{aligned} P_n(t) &= (1-A(t))(1-B(t))(B(t))^{n-1} \quad n \geq 1 \\ P_0(t) &= A(t) \end{aligned}$$

kde

$$A(t) = \frac{\mu(e^{\lambda-\mu}-1)}{\lambda e^{(\lambda-\mu)t}-\mu} \quad \text{a} \quad B(t) = \frac{\lambda(e^{(\lambda-\mu)t}-1)}{\lambda e^{(\lambda-\mu)t}-\mu}$$

**Model** *(M1 – Neomezená ústředna)*

Mějme tel. ústřednu s neomezeně mnoho linkami, potom $\mu_n = n\mu$ (intenzita ukončení hovoru) a $\lambda_n = \lambda$ (intenzita navázání hovoru) – z exp. rozdělení. Narozdíl od předch. případu nezávisí příchozí hovory na počtu probíhajících. Dosadím do obecných rovnic moje $\mu_n$ a $\lambda_n$ a dostanu limitní pravděpodobnosti $p_1 = \frac{\lambda}{\mu}p_0$ atd., obecně

$$p_n = \frac{\lambda^n}{n!\mu^n}p_0 = (\frac{1}{n!})(\frac{\lambda}{\mu})^n p_0$$

což odpovídá Poissonovu rozdělení. Najdu $p_0$, tak aby $\sum_{i=1}^{\infty} p_i = 1$; z Taylora $p_0 \sum_{i=0}^{\infty} \frac{1}{i!}(\frac{\lambda}{\mu})^i = p_0 e^{\lambda}\mu$, takže $p_0 = e^{-\frac{\lambda}{\mu}}$ a hledané řešení je $Po(\frac{\lambda}{\mu})$.

**Model** *(M2 – Dvojbudka s neomezenou frontou)*

Příchod zákazníka – intenzita $\lambda$, obsluha – intenzita $\mu$, stav – počet obsluhovaných + počet ve frontě. Potom $\lambda_n = \lambda$ a $\mu_n = \begin{cases} 0 & n=0 \\ \mu & n=1 \\ 2\mu & n \geq 2 \end{cases}$.

$$p_1 = \frac{\lambda}{\mu}p_0 \quad p_2 = \frac{1}{2}\frac{\lambda^2}{\mu^2}p_0 \quad p_3 = \frac{1}{4}(\frac{\lambda}{\mu})^3 p_0 \quad p_n = \frac{1}{2^{n-1}}(\frac{\lambda}{\mu})^n p_0$$

Dostanu nějaké $Ge$. Aby se netvořila nekonečně dlouhá fronta, je nutné zachovat podmínku $\mu > \lambda$.

**Model** *(M3 – Parkoviště bez fronty / konečná ústředna)*

$N$ míst, 1 vjezd a 1 výjezd, nesmí se tvořit fronta. Potom

- $\lambda_n = \lambda$ (v krátké době projede jen 1 auto),
- $\mu_n = n\mu h + o(h)$ (vyberu 1 auto a nechám ho odjet $\sim Bi(n, \mu h + o(h))$)

Rovnice pak vyjdou takhle, konkrétní řešení pro $P_n(t)$ máme dosazením do obecného vztahu pro dif. rovnice:

$$\begin{aligned} P_0'(t) &= -\lambda P_0(t) + \mu P_1(t) \\ P_n'(t) &= -(\lambda + n\mu)P_n(t) + (n+1)\mu P_{n+1}(t) + \lambda P_{n-1}(t) && 1 \leq n < N \\ P_N'(t) &= -N\mu P_n(t) + \lambda P_{N-1}(t) && \text{nemůže vzniknout odjezdem auta} \end{aligned}$$

Limitní rozdělení dostaneme položením derivací nule a vyřešením soustavy za podmínky $p_i \geq 0$, $\sum_{i=1}^{N} p_i = 1$ (aby šlo o platné rozdělení pravděpodobnosti), z rekurzivních vztahů vyjde $UPo(\frac{\lambda}{\mu})$ (useknuté Poissonovo rozdělení – Poissonovo s normalizační konstantou „přerozdělením chvostu“, useknutí se ale v praxi často zanedbává):

$$p_1 = \frac{\lambda}{\mu}p_0 \qquad p_n = \frac{1}{n!}(\frac{\lambda}{\mu})^n p_0 \;\; n = 1, \dots, N \qquad p_n = \frac{e^{-\frac{\lambda}{\mu}}(\frac{\lambda}{\mu})^n}{n!} \cdot \frac{1}{1 - \sum_{j=N+1}^{\infty} \frac{e^{-\frac{\lambda}{\mu}}(\frac{\lambda}{\mu})^j}{j!}}$$

**Model** *(M2a – Budka s omezenou frontou)*

Intenzita příchodů a odchodů je konstantní ($\lambda_n = \lambda$, $\mu_n = m$), tohle platí pro 1 budku, bylo-li by jich víc, měl bych podobný, ale složitější výpočet. Dostávám pak rekurentní rovnici $p_n = (\frac{\lambda}{\mu})^n p_0$, podle které vidím, že jde o $UGe(\frac{\lambda}{\mu})$ ($p_n = \frac{\lambda}{\mu} \cdot$ „něco“). Z té rovnice pak (podle vzorce pro součet prvních $N+1$ členů geometrické posloupnosti) vyjde:

$$\sum_{i=0}^{N} p_i = \sum_{i=0}^{N} \left(\frac{\lambda}{\mu}\right)^i p_0 = \frac{\left(\frac{\lambda}{\mu}\right)^{N+1} - 1}{\frac{\lambda}{\mu} - 1}$$

A tento součet musí být roven jedné, aby šlo o korektní rozdělení pravděpodobnosti, takže celkově dostávám:

$$p_0 = \frac{\mu^N(\lambda - \mu)}{\lambda^{N+1} - \mu^{N+1}}$$

**Model** *(M4 – 1 opravář na M strojů)*

Jde v podstatě o inverzi parkoviště. Mám $M$ strojů, intenzita poruch $\lambda_n = (M-n) \cdot \lambda$ (protože pokud se už $n$ strojů porouchalo, ještě se může porouchat $M-n$) a doba obsluhy (opravy) je $\mu_n = \mu$ (konstantní). Dostávám také $UPo$, jen s prohozenými parametry $UPo(\frac{\mu}{\lambda})$ a opačným počítáním indexů. Vychází $p_{M-k} = \frac{1}{k!}(\frac{\mu}{\lambda})^k p_M$ a $p_M$ slouží jako normalizační konstanta.

**Model** *(M5 – Líní svářeči)*

Mám $a$ svářečů, pokud svářeč nepracuje, začne pracovat s intenzitou $\lambda$, při práci s intenzitou $\mu$ přestane.

$$\lambda_n = (a-n)\lambda \qquad \mu_n = n\mu \qquad p_n = \binom{a}{n}(\frac{\mu}{\lambda+\mu})^{a-n}(1 - \frac{\mu}{\lambda+\mu})^n$$

Intenzity jdou „proti sobě“, ovlivňují se navzájem. Výsledné rozdělení je binomické.

!Tady chybí ještě model M/M/1 s hodnotami $W$ a $L$, probíraný na přednášce 18.12!

## 4.3 Poissonovo rozdělení

**Definice** *(Poissonovo rozdělení („rozdělení řídkých jevů“)*

Náh. veličina $X$ má Poissonovo rozdělení s parametrem $\kappa$, jestliže $P(X = n) = e^{-\kappa}\frac{\kappa^n}{n!}$ pro $n = 0, 1 \ldots$. Vytvořující funkce (z Taylora) $A(x) = e^{-\kappa}e^{\kappa x}$, stř. hodnota $\kappa$ a rozptyl taky $\kappa$ (z derivací vytv. funkce).

**Poznámka** *(Součet Poissonových rozdělení)*

Mám-li 2 veličiny $X$, $Y$ a $X \sim Po(\kappa_1)$, $Y \sim Po(\kappa_2)$ a jsou nezávislé, pak platí $X + Y \sim Po(\kappa_1 + \kappa_2)$ (ze součinu vytvořujících funkcí).

**Poznámka** *(Aproximace binomického rozdělení normálním)*

Binomické rozdělení lze podle centrální limitní věty aproximovat rozdělením $No(np, np(1-p))$, nejde ale dělat vždy (potřebuji pevné $p$).

**Poznámka** *(Aproximace binomického rozdělení Poissonovým)*

Mám $Bi(n,p)$ jako součet $S_n$ nezávislých $X_1 + \cdots + X_n$ z $Alt(p)$. Pokud $p \to 0$ a $n \to \infty$ (hodnota $p$ není pevná, záleží na $n$, platí $np_n \to \kappa$), lze aproximovat (kam se podělo $(n-k)!$) ??):

$$p_k = \frac{n!}{(n-k)!k!}p^k(1-p)^{n-k} = np(n-1)p \ldots (n-k+1)p \cdot \frac{1}{k!}(1-p)^{n-k} \approx \kappa^k(1 - \frac{\kappa}{n})^{n-k} \approx \kappa^k \frac{e^{-\kappa}}{k!}$$

## 4.4 Poissonův proces

**Definice** *(Poissonův proces)*

Pro nějaké události (proces příchodu) a náh. veličiny $X_i$ (doby mezi příchody) definuji *čítací proces* $N(t)$, označující počet událostí v intervalu $[0, t]$. pak tento čítací proces nazvu (homogenní) *Poissonův proces*, pokud platí:

1. $N(0) = 0$
2. počet událostí v disjunktních čas. úsecích jsou nezávislé náh. veličiny
3. rozdělení počtu událostí v intervalu $I$ záleží pouze na délce tohoto intervalu, nikoliv na hodnotě procesu v čase, kdy $I$ začíná (stacionarita)
4. $P(N(h) = 1) = \lambda h + o(h)$, $P(N(h) = 2) = o(h)$ – za krátký čas. úsek se stane povětšinou jen 1 událost (podobně jako v předch. modelech)

**Věta** *(Exponenciální rozdělení pravděpodobnosti, že nenastane žádná událost)*

Platí $P(N(t) = 0) = e^{-\lambda t}$, tj. pravděpodobnost, že se do času $t$ nic nestane, klesá exponenciálně.

**Důkaz**

Označme $P_0(t) := P(N(t) = 0)$. Pak platí (ze vzorců) $P_0(t+h) = P_0(t)(1-\lambda h+o(h))$, převedením $P_0(t)$ a vydělením $h$ dostávám dif. rovnici $\P'_0(t) = -\lambda P_0(t)$ a její řešení odpovídá (protože $P_0(0) = 1$).

**Věta** *(Doby mezi událostmi $\sim Exp(\lambda)$)*

Doby mezi událostmi v homogenním Poissonově procesu se řídí exp. rozdělením s parametrem $\lambda$.

**Důkaz**

Vím, že $P(X_1 > t) = P(N(t) = 0) = e^{-\lambda t}$, takže $P(X_1 \leq t) = 1 - e^{-\lambda t}$ – distrib. fce odpovídá exp. rozdělení. Potom pro $P(X_2 > s + t | X_1 = s)$ vychází stejně. S odvoláním na nezávislost dostávám indukcí všechny $X_i$.

**Věta** *(Doba čekání na n-tou událost $\sim \Gamma(n, \lambda)$)*

Doba čekání na n-tou událost se řídí gamma-rozdělením s parametry $n, \lambda$.

**Důkaz**

Definujme čas do $n$-té události jako $S_n = X_1 + \cdots + X_n$, z konvolucí vyjde, že hustota pravděpodobnosti náh. veličiny $S_n$ má tvar $g(x, n, \lambda) = \frac{\lambda^n}{\Gamma(n)} e^{-\lambda x} x^{n-1}$, kde $\Gamma(n) = \int_0^\infty e^{-x} x^{n-1} \mathbf{d}x$, což pro přirozená $n$ dává $\Gamma(n) = (n-1)!$. Výsledná hustota pravděpodobnosti odpovídá gamma-rozdělení:

$$g(x, n, \lambda) = \lambda \cdot \frac{e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!}$$

**Věta** *(Poissonovo rozdělení pravděpodobnosti, že nastane n událostí)*

Platí $P(N(t) = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$, tedy počet událostí v intervalu délky $t$ se řídí $Po(\lambda t)$.

**Důkaz**

Plyne z $N(t) \geq n \iff S_n \leq t$: $P(N(t) = n) = P(S_n \leq t) - P(S_{n+1} \leq t) = I_n - I_{n+1}$ pro $I_k = \int_0^t \frac{\lambda^k}{\Gamma(k)} e^{-\lambda x} x^{k-1} \mathbf{d}t$. Dokážu, že $I_n = I_{n+1} + \frac{(\lambda t)^n}{n!} e^{-\lambda t} - I_{n+1}$ (rozdělím $I_n$ metodou per partes, kde $u = e^{-\lambda x}$ a $v' = x^{n-1}$).

### 4.4.1 Hledání $\lambda$ pro naměřené hodnoty

**Algoritmus** *(Nalezení $\lambda$)*

Mám nějaké naměřené hodnoty $X_1, X_2 \ldots X_n$, hledám $\lambda$ pro exponenciální rozdělení, jímž by se měly řídit (takovou, aby byly hodnoty nejpravděpodobnější).

- Spočtu $\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i$.
- Pro $Exp(\lambda)$ platí $EX = \frac{1}{\lambda}$, takže z Čebyševovy nerovnosti ($\lim_{n\to\infty} P(|\overline{X_n} - \mu| \geq \varepsilon) = 0$) dostanu $\lambda \approx \frac{1}{\overline{X_n}}$.

Nalezené $\hat{\lambda}$ dále testuji pomocí testů dobré shody.

**Algoritmus** *(Testování $\hat{\lambda}$ Kolmogorov-Smirnovovým testem)*
Situace:

- Hypotéza: mám náhodný výběr z nějaké distr. funkce $F(x)$.
- Alternativa: rozdělení výběru je jiné.

Postup:

1. Odhadnu $F(x)$ pomocí *empirické distr. funkce* $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x)$, kde $I$ je indikátorová funkce (pro každé $X_i$ dá 1, pokud je menší než $x$ a 0 jinak). Tato funkce má všechny vlastnosti distribuční funkce a platí věta, že pro $n \to \infty$ jde $\hat{F}_n(x) \to F(x)$.
2. Spočtu $KS = \sqrt{n}\sup_{X\in\mathbb{R}} |\hat{F}_n(x) - F(x)|$ a podle toho se rozhoduju.

Problém: rozdělení této naší statistiky je zadané jen tabulkou, nelze parametr odhadovat a hned testovat vůči odhadu (vždy vyjde dobře)! Je nutné náhodně rozdělit data na 2 části, z 1 odhadovat parametr a druhou testovat. Výběr jedné části lze provést např. postupným náh. generováním pořadí prvku, který vyjmu (a meze generování klesají s počtem kandidátů na vybrání).

**Algoritmus** *(Testování $\hat{\lambda}$ pomocí $\chi^2$-testu (Pearsonova))*
Postup:

Máme naměřená data $U_1, \ldots U_n$, a $\mathbb{R}$ rozdělíme na $J$ intervalů pevné délky. Potom $n_i$ skutečný je počet hodnot v $i$-tém intervalu a $np_i$ je očekávaný počet (podle testovaného rozdělení pravděpodobnosti). Sledujeme statistiku, která by se měla v případě, že máme správné rozdělení, řídit $\chi^2$ rozdělením s $J-1$ stupni volnosti:

$$X^2 = \sum_{i=1}^{J} \frac{(n_i - np_i)^2}{np_i} \sim \chi^2_{J-1}$$

Pro nějakou hodnotu tolerance $\alpha$ správnost rozdělení odmítáme, když $X^2 \geq \chi^2_{J-1}(\alpha)$. Neznáme-li parametr $\lambda$ z homogenního Poissonova procesu, musíme jej odhadnout a ve srovnávání použít *modifikovanou metodu* $\chi^2$, kde se snižuje počet stupňů volnosti.

### 4.4.2 Testování změny $\lambda$

**Situace**

Mám nějaký Poissonův proces, kde se doby příchodů požadavků $X_1, \ldots X_n$ řídí stále $Exp(\lambda)$. Chci zjistit, jestli se v nějakém čas. okamžiku – po nějakém $X_m$ (např. po úpravě stroje) hodnota $\lambda$ nezměnila (tj. jestli je $\lambda$ konstantní – *homogenní*). Takže testuji:

- Hypotéza: $X_1, \ldots X_n \sim Exp(\lambda)$, tj. $\lambda$ je konstantní
- Alternativa: existuje $m$ tak, že $X_1, \ldots X_m \sim Exp(\lambda_1)$ a $X_{m+1}, \ldots X_n \sim Exp(\lambda_2)$, tj. $\lambda$ se změnila

Máme 2 možnosti, jak změnu testovat. Buď OFF-LINE – tj. na konci urč. období zkontroluji, zda se $\lambda$ nemění – jde o klasické testování hypotéz a odhad (viz výše), nebo ON-LINE – s každým novým pozorováním (nebo každých $n$ pozorování) zkontroluji, zda se něco nemění – sekvenční metody. Výsledky obou přístupů se mohou lišit.

**Algoritmus** *(On-line testování)*
Cíle:

1. Běží-li proces požadavků stále s $Exp(\lambda)$, tak nechat běžet co nejdéle
2. Dojde-li ke změně $\lambda$, co nejdříve ji odhalit

Použijeme *Stopping rule* (zde podle Shewarta):

$$\tau_s = \inf\{n | \log \frac{f_1(X_n)}{f_0(X_n)} > h\}$$

kde $\tau_s$ značí čas změny a funkce $f_0$ a $f_1$ určují podobnost původnímu, resp. změněnému rozdělení. $h$ je potom zvolená hladina, kterou poměr nesmí překročit. Cíle jsou protichůdné, takže zvolím $h$, abych posloužil prvnímu a testuji citlivost na druhý. Hodnoty $\tau_s$ záleží jen na posledním pozorování, takže vlastně pro každé pozorování zjišťuji, zda $\log \frac{f_1(X_n)}{f_0(X_n)} > h$, což dává booleanovské výsledky, tj. mám novou náhodnou veličinu s alternativním rozdělením. Čekám pak na její první zdar – celkově dostávám geometricky rozdělenou náh. veličinu (s parametrem $p$ – pravděpodobností, že překročím hladinu $h$).

**Důsledek** *(Shewartovo pravidlo)*

1. Zafixuj $f_0$, $f_1$. (podobnost tomu co chci a čemu se bráním)
2. Stanov střední dobu mezi „falešnými poplachy“. (střední hodnota mého geom. rozdělení, volím sám)
3. K této hodnotě najdi $h$.
4. Pro dané $h$ si spočítej, jak dlouho pro různá $\mu$ (parametry odpovídající $f_1$) bude trvat odhalení změny (citlivost).

**Příklad**

Vezmeme $f_0 : X_i \sim Exp(\lambda)$ a $f_1 : X_i \sim Exp(\mu)$, předpokládejme, že $\mu = k \cdot \lambda$. Potom

$$\log \frac{f_1(X_n)}{f_0(X_n)} = \log \frac{k\lambda e^{-k\lambda X_n}}{\lambda e^{-\lambda X_n}} = \log k + \lambda X_n(1-k)$$

Ze střední doby mezi poplachy $\Upsilon$ (pro Geom. rozdělení odpovídá $1/p$) vypočítám $h$:

$$\frac{1}{\Upsilon} = p = P(\log k + \lambda X_n(1-k) > h) = P(X_n > \frac{h - \log k}{\lambda(1-k)}) = P(Exp(\lambda) \leq \frac{h - \log k}{\lambda(1-k)}) = 1 - e^{\frac{h-\log k}{\lambda(1-k)}}$$

Odhalení změny – jde o to, s jakou pravděpodobností $X_n \sim Exp(\mu)$ překročí moje $h$:

$$P(X_n \sim Exp(\mu) > h) = \int_h^\infty \mu e^{-\mu x} \mathbf{d}x$$

**Poznámka** *(Simulace Markovovým řetězcem)*

Celý Shewartův proces je vlastně Markovův řetězec se 2 stavy – $E_0$ (v pořádku) a $E_1$ (chyba, pohlcující stav), kdy s pravd. $p$ se dostanu z $E_0$ do $E_1$. $L_0$ – očekávaná délka procházky z $E_0$ do $E_1$ mi dává střední dobu mezi poplachy.

**Poznámka** *(Zesložitění)*

Chci-li brát v úvahu víc pozorování, zavedu „warningový“ stav – např. 2 pozorování z „warning“ za sebou nebo 1 pozorování z „halt“ zastaví stroj. Mám tedy 3 stavy Markovova řetězce a jeden speciální, kam se dostanu po 2. pozorování hodnoty odpovídající „warning“. Z toho pak mohu počítat stř. doby mezi poplachy, Markovovy řetězce jsou nejjednodušší způsob modelování.

Další zesložitění je „stop-lo“, „stop-hi“, „warning-lo“, „warning-hi“ – dostanu ještě víc stavů.