



ÓBUDAI EGYETEM
ÓBUDA UNIVERSITY

**NEUMANN JÁNOS
INFORMATIKAI KAR**



DIPLOMAMUNKA

OE-NIK Hallgató neve:
2022.05.15 Hallgató törzskönyvi száma:

1

Sárközi Zoltán Attila
T/007187/FI12904/N

DIPLOMAMUNKA FELADATLAP

Hallgató neve: **Sárközi Zoltán Attila**
Törzskönyvi száma: T/007187/FI12904/N
Neptun kódja: 

A diplomamunka címe:

Adattárház IT rendszer sztochasztikus elemzése
Stochastic analysis of a Data warehouse IT system

Intézményi konzulens: Dr. Kárász Péter
Külső konzulens:

Beadási határidő: 2022. május 15.

A záróvizsga tárgyai: Mérnöki számítási módszerek
Rendszer és irányításelmélet
Sztochasztika
Optimumszámítás

A feladat

Vizsgálja meg, hogy milyen típusú sorbanállási rendszerek alkalmasak az adattárházak modellezésére. Ismertesse a megfelelőnek ítélt rendszerek elméleti felépítését, jellemzőit. Készítsen esettanulmányt.

A diplomamunkának tartalmaznia kell:

- a sorbanállási elmélet vonatkozó részeinek rövid ismertetése,
- fázis típusú eloszlások, fázis típusú sorbanállási rendszerek,
- a feldolgozó egységek állapotainak vizsgálata, leírása,
- esettanulmány.




.....
Dr. Kárász Péter
intézetigazgató

A diplomamunka elévülésének határideje: **2023. május 15.**
(OE TVSz 55.§ szerint)

A diplomamunkát beadásra alkalmasnak tartom:

.....
külső konzulens


.....
intézményi konzulens



HALLGATÓI NYILATKOZAT

Alulírott hallgató kijelentem, hogy a szakdolgozat / diplomamunka saját munkám eredménye, a felhasznált szakirodalmat és eszközöket azonosíthatóan közöltem. Az elkészült szakdolgozatomban / diplomamunkámban található eredményeket az egyetem és a feladatot kiíró intézmény saját céljára térítés nélkül felhasználhatja.

Budapest, 20 22.05.13

Szűcs Zoltán
hallgató aláírása

KONZULTÁCIÓS NAPLÓ

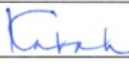
Hallgató neve: Neptun Kód: Tagozat:
 Sárközi Zoltán Attila esti.....

Telefon: Levelezési cím (pl.: lakcím):

Szakedolgozat / Diplomamunka¹ címe magyarul:
 Adattárház IT rendszer sztochasztikus elemzése.....

Szakedolgozat / Diplomamunka² címe angolul:
 Stochastic analysis of a Data warehouse IT system

Intézményi konzulens: Külső konzulens:
 Dr. Kárász Péter.....

Alk.	Dátum	Tartalom	Aláírás
1.	2020-10-15	A hálózatba szervezett feladatokhoz való modell megállapítása: PERT vagy sorbanállási elmélet	
2.	2020-10-29	Az adattárház lefutási hálózathoz való modell (sorbanállási rendszer) kitárgyalása	
3.	2020-11-20	Az adattárház lefutási hálózathoz való modell (Sztochasztikus programozás feladat) kitárgyalása	
4.	2020-12-17	A fázis típusú eloszlások hasznossága a modellben	

A Konzultációs naplót összesen 4 alkalommal, az egyes konzultációk alkalmával kell láttamoztatni bármelyik konzulenssel.

A hallgató a Szakedolgozat I. / Szakedolgozat II. (BSc) vagy Diplomamunka 1 / Diplomamunka 2 / Diplomamunka 3 / Diplomamunka 4³ tantárgy követelményét teljesítette, beszámolóra / védésre⁴bocsátható.



Intézményi konzulens

Budapest, 20. december 17.

¹ Megfelelő aláhúzendő!

² Megfelelő aláhúzendő!

³ Megfelelő aláhúzendő!

⁴ Megfelelő aláhúzendő!

KONZULTÁCIÓS NAPLÓ

Hallgató neve: Neptun Kód: Tagozat:
Sárközi Zoltán Attila [REDACTED] esti.....

Telefon: Levelezési cím (pl.: lakcím):
[REDACTED]

Szakdolgozat / Diplomamunka¹ címe magyarul:
Adattárház IT rendszer sztochasztikus elemzése.....

Szakdolgozat / Diplomamunka² címe angolul:
Stochastic analysis of a Data warehouse IT system

Intézményi konzulens: Külső konzulens:
Dr. Kárász Péter..... ..

Alk.	Dátum	Tartalom	Aláírás
1.	2022-02-24	Mátrixgeometrikus eloszlások	Kárász
2.	2022-03-17	Esettanulmány megtervezése	Kárász
3.	2022-04-14	A rendszer várakozási idő-, vagy kiszolgálóegységek eloszlással történő korlátozása, az idő függő kiszolgálóegységek	Kárász
4.	2022-05-05	Az irodalomban felsorolt fázis típusú sorelméletek végigtárgyalása. A dolgozat formai követelményeinek pontosítása	Kárász

A Konzultációs naplót összesen 4 alkalommal, az egyes konzultációk alkalmával kell láttamoztatni bármelyik konzulenssel.

A hallgató a Szakdolgozat I. / Szakdolgozat II. (BSc) vagy Diplomamunka 1 / Diplomamunka 2 / Diplomamunka 3 / Diplomamunka 4³ tantárgy követelményét teljesítette, beszámolóra / védésre⁴bocsátható.

Kárász Péter

Intézményi konzulens

Budapest, 20. május 13.

¹ Megfelelő aláhúzendő!

² Megfelelő aláhúzendő!

³ Megfelelő aláhúzendő!

⁴ Megfelelő aláhúzendő!

Adattárház IT rendszer sztochasztikus elemzése

Sárközi Zoltán Attila

Május 2022

Tartalomjegyzék

1. Bevezetés	10
1.1. A vizsgált adattárház rendszer	10
1.2. A vizsgált modell	11
2. Valószínűségszámítási alapok	14
2.1. Emlékezés nélküli eloszlások	14
2.2. Valószínűségszámítási alaptevések	15
2.2.1. Differenciaegyenletek	16
2.2.2. Differenciálegyenletek	16
2.2.3. Egyensúlyi egyenletek	17
2.3. Sorbanállási rendszerek	17
2.3.1. Kendall-féle osztályozás	17
2.3.2. Minőségi mutatók	18
2.3.3. Little-formula	19
2.3.4. Poisson folyamat	20
3. Markov-láncok	21
3.1. Diszkrét idejű Markov-láncok	21
3.2. Folytonos idejű Markov-lánc	21
3.3. Stacionárius eloszlás	22
3.4. Átmenetvalószínűség	22
3.5. Átmenetek osztályozása	23
3.6. Generátor mátrix	23
3.7. Születési-halálozási folyamat	24
3.8. Felújítási folyamat	25
4. Fázis típusú eloszlások	26
4.1. Elnyelő Markov-lánc	26
4.2. PH eloszlások jellemzői	27
4.3. Néhány nevezetes PH eloszlás	28
4.4. MAP (Markovian Arrival Process) folyamat	29
5. Fázis típusú eloszlás közelítése	31
5.1. Mintavételezés karakterisztikája	31
5.2. EM algoritmus (Expectation Maximization) PH eloszláshoz	33
5.3. Nyomokra épülő illeszkedésvizsgálat MAP-ra	35

6. Struktúrált Markov-láncú sorbanállási rendszerek	37
6.0.1. A PH eloszlások további tulajdonságai	37
6.1. Kvázi születési-halálozási folyamat	37
6.1.1. QBD folyamat stabilitása	38
6.2. Mátrix geometrikus eloszlás	39
6.2.1. Alapok	39
6.2.2. M/G/1 és GI/M/1	40
6.3. PH/M/1	41
6.3.1. Teljesítménymutatók, feldolgozóegységek állapotai	43
6.4. PH/M/N	44
6.5. PH/PH/1	46
6.6. PH/PH/N	47
6.6.1. Teljesítménymutatók, feldolgozóegységek állapotai	49
6.7. MAP/PH/1	50
6.8. MAP/PH/N	50
6.8.1. Teljesítménymutatók, feldolgozóegységek állapotai	53
7. Esettanulmány	55
7.1. Vizsgálat PH/PH/N rendszerben	56
7.1.1. Rögzített várakozási idő eloszlás mellett a feldolgozóegységek vizsgálata .	57
7.1.2. Rögzített feldolgozóegységek melletti vizsgálat	58
7.2. Vizsgálat MAP/PH/N rendszerben	60
7.2.1. Rögzített várakozási idő eloszlás mellett a feldolgozóegységek vizsgálata .	60
7.2.2. Rögzített feldolgozóegységek melletti vizsgálat	61
7.3. További vizsgálati lehetőségek	62
8. Összefoglalás	63
9. Summary	64

1. Bevezetés

1.1. A vizsgált adattárház rendszer

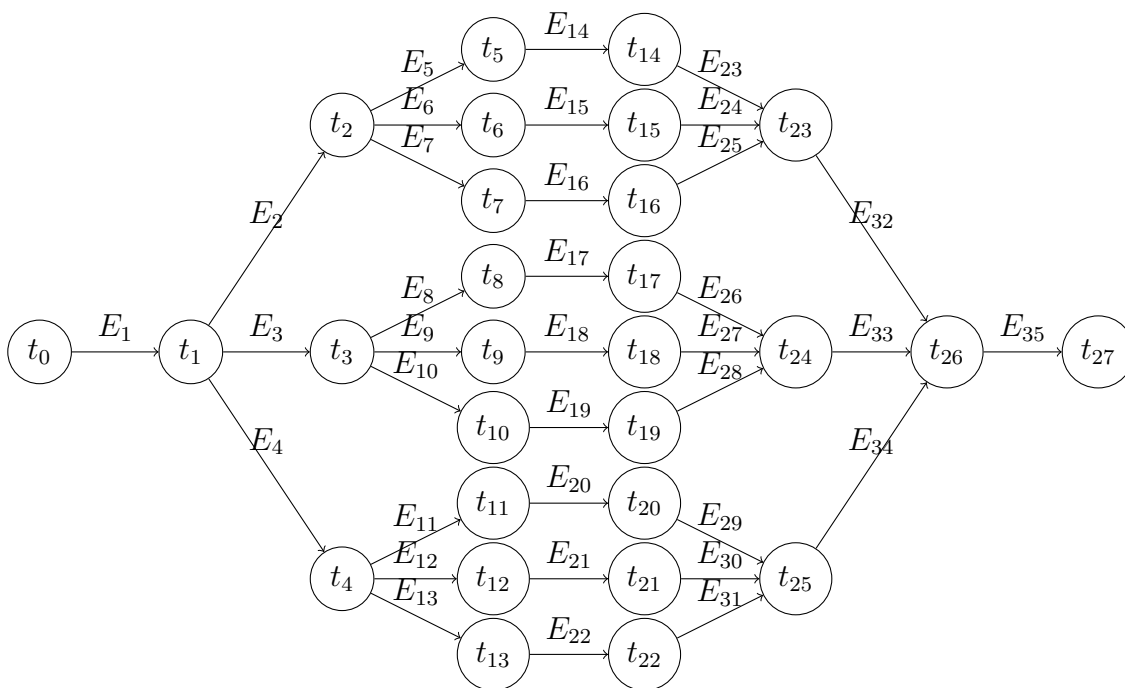
Az Adattárház egy adatösszesítő informatikai rendszer. Egy vállalatnak sok-sok különböző IT rendszere van az üzleti folyamatai mentén, és ezek mindegyike naponta átadja az adatait az Adattárház számára. Az Adattárházba érkezett adatokat kis programok, a *job*-ok dolgozzák fel. A *job* $N^n \rightarrow N$ alakú, vagyis minden *job*-ban van n darab bemeneti adat és egy kimeneti adat. A *job*-on belül az n darab bemenő halmazt kötünk össze tetszőleges logika mentén, majd ezeken tetszőleges műveleteket hajtunk végre, a *job* végén pedig az eredmény adatokat egy célhalmazba tesszük. Ezek a *job*-ok egy fába vannak szervezve. Egy *job* kimenő adata egy másik *job* bemenő adata lehet.

Az n . *job*, azaz feladat, vagy igény jelölése: J_n

A J_n mely másik *job*-októl függ: $J_n = J(J_{n-k}, J_{n-(k+l)}, \dots), k = 1, 2, \dots, l = 1, 2, \dots$, vagy $J_n = J(J_{n-k}), k = 1, 2, \dots, N$

Az Adattárház informatikai rendszernek fizikai korlátai vannak. Ilyenek a CPU, a memória, a disk elérés. Így minden futáskor egy-egy *job* más és más lefutási időt produkál, illetve az adott *job*-tól függő további *job*-ok más és más ideig várakoznak a kiszolgálásaikra. A *job*-ok lefutási idejei ezért valószínűségi változók. Az n . *job*, azaz feladat lefutási ideje: $E_n = (e_{n_1}, e_{n_2}, \dots, e_{n_m})$ Egy másik fizikai korlát, hogy az Adattárház rendszer egyszerre csak $n \leq N$ darab feladatot képes egyidejűleg végrehajtani. Az N érték egy felső korlát, az n értékét befolyásolja az éppen elérhető CPU, memória és disk kapacitás, azaz a rendszer fizikai korlátai. Így ha t_m időpillanatban $k < n$ darab feladat végrehajtása folyik, de ezek a feladatok elviszik a fizikai erőforrásokat, akkor nem lehetséges egy új feladat végrehajtása mindaddig, amíg nincs elegendő rendelkezésre álló erőforrás. Vagyis a valószínűségi változóval jellemzett *job* lefutási idők miatt az igénybe vett, illetve az igénybe vehető kiszolgálási egységek száma is véletlen folyamattal írható le.

A feladatok alkotta ütemezett hálózat lefutását bemutató modellt a következő példán keresztül magyarázom:



Ahol

- a t_0 a rendszer indulásának pillanata, a t_{27} (t_n) a rendszer futásának befejező pillanata.
- a J_1 feladat (*job*): $\overrightarrow{t_0 t_1}$ lefutási idejét jelentő valószínűségi változó: E_1 (J_n, E_n)
- t_1 pillanatban került végrehajtásra a J_1 feladat, X_1 valószínűségi változó idő alatt. (Hasonlóan, a J_n feladat végrehajtási ideje az X_n vv., a feladat befejező pillanata pedig a t_n pillanat)

1.2. A vizsgált modell

A rendszer vizsgálat során a rendelkezésemre állt az Adattárház *job*-jainak lefutási adatai:

- *job* indulásának ideje,
- *job* befejezési ideje
- és hogy az adott J_i *job* melyik (J_j , $j < i$) *job*-ok lefutása után indulhat.

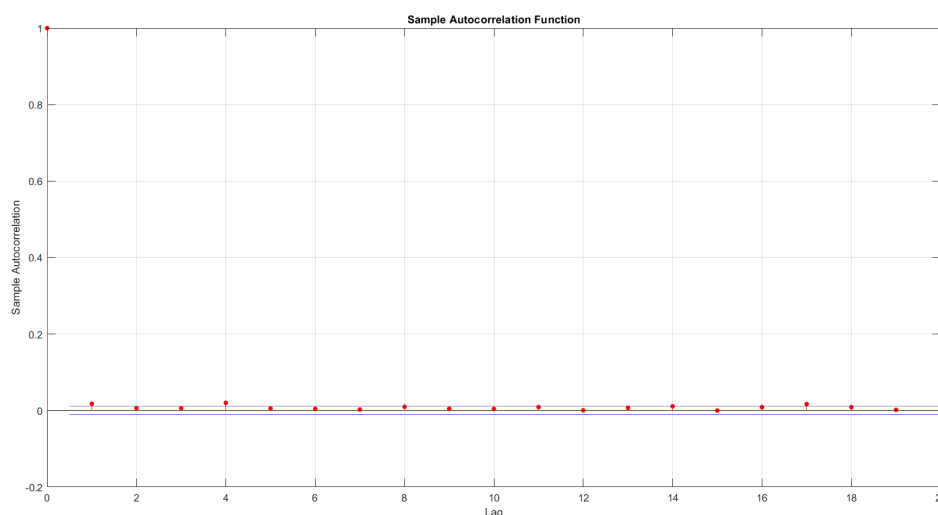
Az Adattárház IT rendszerben olyan kérdéseket kell tudnom megmagyarázni, hogy

- mennyi idő alatt fut le a rendszer,
- hogy néz ki a feldolgozóegységek állapota, esetleg ha több vagy kevesebb processzort teszünk a rendszerben, az milyen változásokat hoz?
- ha változtatok az ütemezési hálózaton (felveszek új *job*-ot, vagy törölök egy másikat), az hogyan változtathajta meg a rendszer lefutását?

Ezen kérdések megválaszolásához arra van szükségem, hogy a rendszer sztochasztikus minőségi jellemzőivel tisztában legyek. Ha ezeket ismerem, akkor van lehetőség arra, hogy a rendszeren bármilyen optimalizálást végezzek.

A dolgozat része egy esettanulmány, ahol szimulációval ellenőrzöm a feltételezett modell (PH/PH/N és MAP/PH/N) helyességét: hogy ezekkel a modellekkel választ tudok adni a fent feltett kérdésekre.

A beérkező igények kiszolgálási idejei valószínűségi változók, és az igények várakoznak a kiszolgálásukra, a rendszerben pedig véges számú kiszolgálási egység található, ezért az a feltételezésem, hogy Adattárház sorbanállási rendszerrel tudom modellezni. A kiszolgálási- és várakozási időket megvizsgálva azt tételezem fel, hogy olyan sorbanállási rendszerrel tudom jellemezni az Adattárház, amely fázis típusú érkezési és kiszolgálási folyamatokkal rendelkezik. A dolgozatomban ezért a PH/PH/N és a MAP/PH/N modelleken keresztül vizsgálom a rendszert. A mérhető várakozási- és kiszolgálási idők eloszlásfüggvényeit vizálisan ábrázoltam, és megvizsgáltam az *autokorreláció* függvényét:

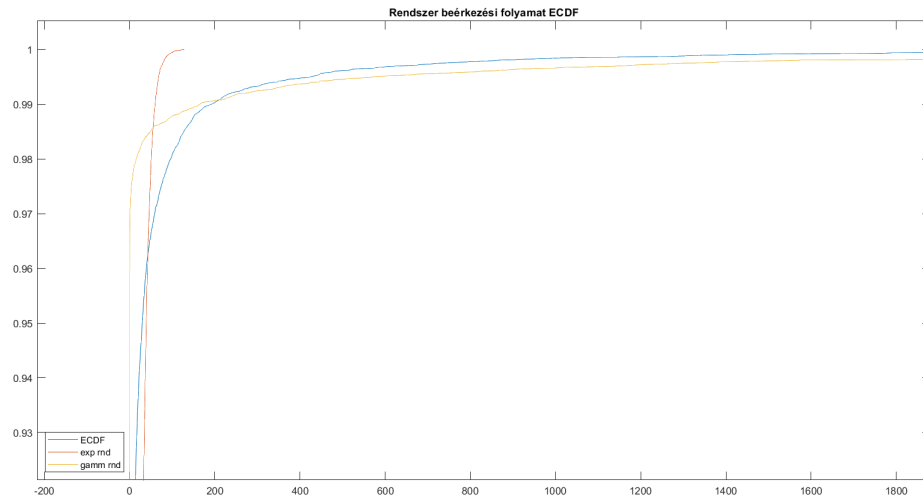


1. ábra. Érkezési folyamat autokorreláció függvénye

Arra a következtetésre jutottam, hogy emlékezés nélküli eloszlást követ az idősorom. Ezért intuitívan közelítettem, sorra

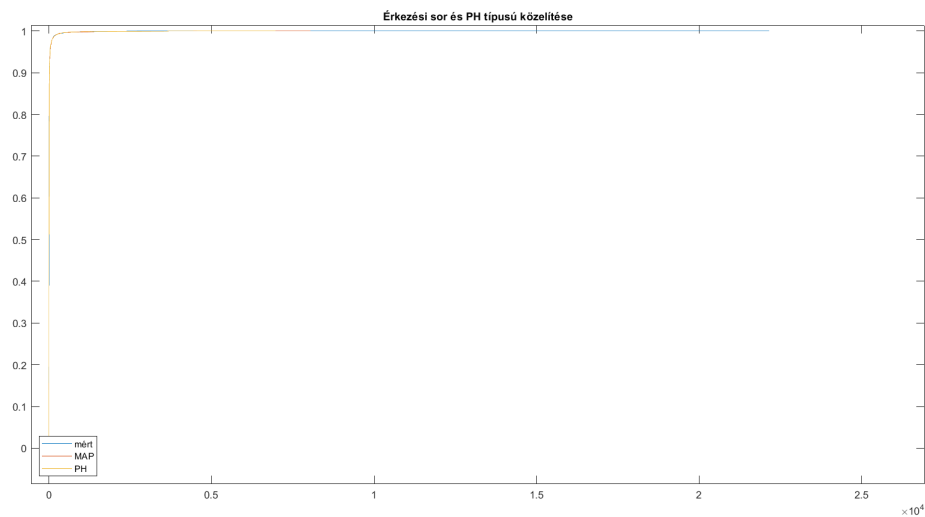
- exponenciális,
- gamma,
- valamint fázis típusú eloszlásokkal.

Ezek eredményei:



2. ábra. Kendall A paraméter CDF, és közelítései

, valamint PH eloszlással történt közelítése



3. ábra. Kendall A paraméter CDF. és közelítése (PH)

. Az érkezési sor közelítéséhez hasonló a kiszolgálási sor, amelyet szintén fázis típusú eloszlással tudtam a legpontosabban közelíteni.

2. Valószínűségyszámítási alapok

2.1. Emlékezés nélküli eloszlások

A vizsgált informatikai rendszerben a beérkező igények közötti időközök, és a kiszolgálási idők az exponenciális eloszlásokból származtatható eloszlásokkal közelíthetők. Az emlékezés nélküli eloszlások alapdefiníciói a [3, c2.8]-nak megfelelően szerepelnek:

Exponenciális eloszlás Az X valószínűségi változó exponenciális eloszlású, ha sűrűségfüggvénye:

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad (1)$$

a kumulatív eloszlásfüggvénye:

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0, \quad (2)$$

a várható értéke:

$$E[X] = \frac{1}{\lambda}, \quad (3)$$

és a momentumjai:

$$E[X^n] = \frac{n!}{\lambda^n}. \quad (4)$$

Ha x -et egy nagyon kis h -val megnöveljük, akkor annak valószínűsége, hogy x nagyobb lesz, mint h :

$$P(x > h) = e^{-\lambda h}. \quad (5)$$

Legyen $t = (t' + h)$, és akkor annak valószínűsége, hogy t nagyobb, mint $t' + h$:

$$P(t > (t' + h)) = e^{-\lambda(t' + h)}. \quad (6)$$

Tegyük fel, hogy $t' > t$ és akkor írjuk fel a következő feltételes valószínűséget:

$$P(t > t' + h \mid t > t') = \frac{e^{-\lambda(t' + h)}}{e^{-\lambda t'}} = e^{-\lambda h}. \quad (7)$$

A $P(t > t' + h \mid t > t')$ valószínűség megegyezik a $P(t > h)$ valószínűséggel, vagyis

$$P(t > t' + h \mid t > t') = P(t > h) = e^{-\lambda h} \quad (8)$$

és ezért azt mondjuk, hogy az exponenciális eloszlás az emlékezés nélküli eloszlás.

Vizsgáljuk meg az $F(x)$ kumulatív eloszlásfüggvényt:

$$\begin{aligned}
F(x) &= P(t < x) = 1 - e^{-\lambda x} \\
&= 1 - \left[\frac{(-\lambda x)^0}{0!} + \frac{(-\lambda x)^1}{1!} + \frac{(-\lambda x)^2}{2!} + \dots \right] \\
&= 1 - \left[1 + \frac{(-\lambda x)^1}{1!} + \frac{(-\lambda x)^2}{2!} + \dots \right] \\
&= \lambda x - \left[\frac{(-\lambda x)^2}{2!} + \frac{(-\lambda x)^3}{3!} + \dots \right] \\
&= \lambda x + o(h),
\end{aligned} \tag{9}$$

ahol

$$o(h) = \frac{(-\lambda x)^2}{2!} + \frac{(-\lambda x)^3}{3!} + \dots, \tag{10}$$

és

$$\lim_{h \rightarrow \infty} \frac{o(h)}{h} = 0. \tag{11}$$

Felhasználtuk, hogy

$$e^{\alpha x} = \sum_{k \geq 0} \frac{(\alpha x)^k}{k!}. \tag{12}$$

2.2. Valószínűségszámítási alaptételek

Vegyünk egy olyan sorbanállási rendszert, ahol a beérkezési folyamat és a kiszolgálási idők is egyaránt exponenciális eloszlást követnek. A beérkezési folyamatban az igények beérkezési idejeinek átlaga legyen $\frac{1}{\lambda}$, a kiszolgálási idők átlaga pedig $\frac{1}{\mu}$. Vegyünk egy kis h értéket, és vizsgáljuk meg az esetet, hogy két igény beérkezése közti időköz kevesebb, mint h , illetve hogy az igények kiszolgálási ideje szintén kisebb, mint h .

Jelölje A azt az eseményt, hogy a beérkezések közötti időköz $t_A \leq h$ valamint jelölje D azt az eseményt, hogy az igény $t_D \leq h$ időn belül távozik is a rendszerből (kiszolgálásra került).

Az [9] alapján annak valószínűsége, hogy két igény a h -n belül érkezik meg, illetve, hogy h időn belül kerül kiszolgálásra [3, c2.10]:

$$\begin{aligned}
P(t_A < h) &= \lambda h + o(h) \\
P(t_D < h) &= \mu h + o(h).
\end{aligned} \tag{13}$$

Akkor

$$\begin{aligned}
P(A) &= \lambda h + o(h) \\
P(D) &= \mu h + o(h) \\
P(\bar{A} \cup \bar{D}) &= (1 - \lambda h - o(h))(1 - \mu h - o(h)) \\
&= 1 - \mu h - \lambda h + o(h) \\
P(n \times (A \cup / \cap D)) &= o(h).
\end{aligned} \tag{14}$$

2.2.1. Differenciaegyenletek

Legyen n a kiszolgálóegységek száma. A következő pontokban egy tetszőlegesen választott t pillanatban a szükséges feldolgozóegységek számát jelölje n , és legyen P_n annak valószínűsége, hogy az adott t időpillanatban n darab egység működik.

Az $\frac{1}{\lambda}$ és az $\frac{1}{\mu}$ sorra a beérkezések közötti időközök-, és a kiszolgálási idők átlagai, és tegyük fel, hogy végtelen a beérkező igényeket "tároló" sor. A differenciaegyenletek segítségével vizsgálhatjuk a feldolgozóegységek számának valószínűségeit, amikor az idő t -ről $t + h$ -ra nő. Az $o(h)$ definíciója a [10] egyenlet.

Akkor [3, c2.11]

$$\begin{aligned}
n = 0 \quad P_0(t + h) &= (1 - \lambda h)P_0 + \mu h P_1(t) + o(h) \\
n \geq 1 \quad P_n(t + h) &= (1 - \lambda h - \mu h)P_n(t) + \lambda h P_{n-1} + \mu h P_{n+1}(t) + o(h)
\end{aligned} \tag{15}$$

2.2.2. Differenciálegyenletek

Ha a fenti differenciaegyenletekben a h értéke a 0-hoz közelít, akkor a következő három egyenletet írhatjuk fel [3, c2.12]:

$$\begin{aligned}
\lim_{h \rightarrow 0} \frac{P_n(t + h) - P_n(t)}{h} &= P_n'(t) \\
\lim_{h \rightarrow 0} \frac{(\lambda h + \mu h)P_n(t)}{h} &= (\lambda + \mu)P_n(t) \\
\lim_{h \rightarrow 0} \frac{o(h)}{h} &= 0.
\end{aligned} \tag{16}$$

Az első egyenlet mutatja, hogy néz ki a derivált. A második mutatja a valószínűséget amikor nem növeljük t -t h -val. Az utolsó pedig az $o(h)$ függvényt mutatja. A [16] alapján, ha a h értéke nullához közelít, akkor a differencia egyenletekből a következő differenciálegyenleteket kapjuk [3, c2.12]:

$$\begin{aligned}
n = 0 \quad P_0'(t) &= -\lambda P_0(t) + \mu P_1(t) \\
n \geq 1 \quad P_n'(t) &= (-\lambda - \mu)P_n(t) + \lambda P_{n-1}(t) + \mu P_{n+1}(t).
\end{aligned} \tag{17}$$

2.2.3. Egyensúlyi egyenletek

A rendszer egyensúlyban van, hogy ha derivált értéke nulla, azaz [3, c2.13] szerint:

$$\begin{aligned}\lim_{t \rightarrow \infty} P_n(t') &= 0 \\ \lim_{t \rightarrow \infty} P_n(t) &= P_n.\end{aligned}\tag{18}$$

Ezeket alkalmazva a differenciálegyenletekre [3, c2.13]:

$$\begin{aligned}n = 0 \quad 0 &= -\lambda P_0 + \mu P_1 \\ n \geq 1 \quad 0 &= -(\lambda + \mu)P_n + \lambda P_{n-1} + \mu P_{n+1}.\end{aligned}\tag{19}$$

Az egyensúlyi egyenletekből levezetve [3, c2.14] alapján :

$$\begin{aligned}n = 0 \quad 0 &= -\lambda P_0 + \mu P_1 \Rightarrow 0 = -\lambda P_0 + \mu P_1 \\ n = 1 \quad 0 &= -(\lambda + \mu)P_1 + \lambda P_0 + \mu P_2 \Rightarrow 0 = -\lambda P_1 + \mu P_2 \\ n = 2 \quad 0 &= -(\lambda + \mu)P_2 + \lambda P_1 + \mu P_3 \Rightarrow 0 = -\lambda P_2 + \mu P_3\end{aligned}\tag{20}$$

És innen általánosságban:

$$0 = -\lambda P_{n-1} + \mu P_n, \quad n \geq 1\tag{21}$$

2.3. Sorbanállási rendszerek

A sorbanállási elmélet különböző rendszerekbe érkező igények vizsgálata során az igények sorbanállását, a kiszolgálását és ezek összefüggéseit vizsgálja. Az IT rendszer feldolgozását sorbanállási elmélettel modellezem.

A vizsgálatom során a PH/PH/N és a MAP/PH/N típusú sorbanállási rendszerekkel közelítem az Adattárház működését. Az Adattárház rendszert egy olyan modellel lehet leírni, ahol egy szerveren n darab kiszolgálási egység található, és egy (elméletileg) végtelen méretű várakozási hely található. A rendszer a *job*-okat az ütemezési hálózaton végighaladva dolgozza fel. A sorbanállási rendszereket leíró, a működésüket jellemző jelölések, egyenletek sokhelyütt megtalálhatóak, a dolgozatomban a [2] által is használt jelöléseket vettem alapul.

2.3.1. Kendall-féle osztályozás

A tömegkiszolgálási rendszerek alapvető típusait a következő jelölésekkel jellemezhetjük[2, c6.2]:

$$A/B/m/n,\tag{22}$$

illetve bővebben az

$$A/B/m/n/e - x\tag{23}$$

jelölésekkel, ahol:

- A: az igények beérkezése közötti időközök eloszlása
- B: kiszolgálási idők eloszlása
- m: a kiszolgálási egységek száma
- n: a várakozási helyek száma, ahol a rendszerbe már beérkezett, de még ki nem szolgált igények várakoznak
- e: az igényforrások száma
- x: a kiszolgálás elve

Ahol A és B értéke a következők lehetnek:

- G: általános
- M: markovian (exponenciális)
- D: determinisztikus
- E_k : k-adrendű Erlang
- PH: és a fázis típusú eloszlások

2.3.2. Minőségi mutatók

A [2, c6.3] jelöléseit használva:

- Igény elvesztésének valószínűsége.
- Az igények várakozásainak eloszlása:

$$F_n(X) = \frac{1}{n} \sum_{i=1}^n I(W_i < x), x > 0 \quad (24)$$

- Átlagos várakozási idő:

$$\frac{1}{n}(W_1 + \dots + W_n) \rightarrow W^{(1)} \quad (25)$$

sztochasztikusan.

- Sorhosszúság eloszlása t időpillanatban: $L(t), t \geq 0$ a t . időpillanatban a sorhosszúság,

$$\bar{L}_k(t) = \frac{1}{t} \int_0^t I(L(s) = k) ds \quad (26)$$

és

$$p_k = \lim_{t \rightarrow \infty} \bar{L}_k, \quad k = 0, 1, \dots \quad (27)$$

Ha $\sum_{k=0}^{\infty} p_k = 1$, akkor p_k a sorhosszúság eloszlása

- Foglaltsági intervallum: Kiszolgálási egység foglalt: $[a_n, b_n), n \geq 0$, kiszolgálási egység szabad: $[b_n, a_{n+1})$. Ha létezik sztochasztikusan a

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n I(b_i - a_i < x) = F^*(x), \quad x > 0 \quad (28)$$

akkor $F^*(x)$ a foglaltsági intervallum eloszlása.

2.3.3. Little-formula

A [2, c6.3] szerint:

1. Beérkezési folyamat: $N(t), t \geq 0$. A beérkezési folyamat egy Poisson-folyamat.
2. Távozási folyamat: $M(t), t \geq 0$
3. Sorhosszúság $L(t) = N(t) - M(t), t \geq 0$
4. Rendszerben tartózkodó igények átlagos száma: $\bar{L}_t = \frac{1}{t} \int_0^t L(s) ds, t \geq 0$
5. Átlagos beérkezési intenzitás $\lambda(t) = \frac{N(t)}{t}, t > 0$
6. n-ik igény átlagos ideje a rendszerben $\tau_n(t), t > 0$
7. Az összes időmennyiség, amit a t időpontig beérkezett igények a rendszerben töltenek $\tau(t) = \sum_n \tau_n(t), t > 0$
8. Egy igényre eső átlagos rendszerben töltött idő az összes igényre nézve $T(t) = \frac{\tau(t)}{N(t)}, t > 0$

Little-formula [2, c6.4 p196 t6.3]:

$$\begin{aligned} \lim_{t \rightarrow \infty} \lambda(t) &= \lim_{t \rightarrow \infty} \frac{N(t)}{t} = \lambda \\ \lim_{t \rightarrow \infty} \bar{L}(t) &= \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t L(s) ds = L \\ \exists \lim_{t \rightarrow \infty} T(t) &= T \text{ sztochasztikusan, és érvényes a } L = \lambda T \text{ összefüggés} \end{aligned} \quad (29)$$

2.3.4. Poisson folyamat

A Poisson-folyamat egy olyan sztochasztikus folyamat, amelyik a beérkezések számát, és a beérkezések közötti időközöket modellezi.

1. Definíció. [2, c2.7.1] Az $N(t)$ folyamat λ paraméterű Poisson folyamat, ha

- $N(0) = 0$
- $N(t)$ független növekményű folyamat
- az $N(t) - N(s)$ növekmények eloszlása $\lambda(t - s)$ paraméterű Poisson folyamat

azaz

$$P(N(t) - N(s) = k) = \frac{[\lambda(t - s)]^k}{k!} e^{-\lambda(t-s)}, k = 0, 1, \dots \quad (30)$$

3. Markov-láncok

A Markov-lánc egy olyan sztochasztikus folyamat, ahol az adott jelenbeli állapot mellett a rendszer jövőbeli állapota nem függ a múltbeliektől. Azaz

3.1. Diszkrét idejű Markov-láncok

Vegyünk egy diszkrét vagy folytonos idejű sztochasztikus folyamatot: $X = (X_t, t \in \tau)$ a $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn, megszámlálható állapotot felvéve. Ha a X_t felveszi a $\chi = \{x_0, x_1, \dots\}$ állapotokat, akkor ezt az állapot térnek nevezzük. Az állapottér véges, ha χ véges számú elemet tartalmaz.

2. Definíció. [2, c3] A Markov-láncok olyan sztochasztikus folyamatok, ahol minden lehetséges $n = 0, 1, \dots$ érték és tetszőlegesen választott $i_0, \dots, i_n \in \chi$ állapotok mellett:

$$P(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n) = P(X_{n+1} = i_{n+1} | X_n = i_n). \quad (31)$$

3. Definíció. [2, c3.1, p79] Elsőrendű markov-lánc:

$$P(X_{n+1} = i_{n+1} | X_n = i) \quad (32)$$

4. Definíció. [2, c3.1, p79] m -edrendű markov-lánc:

$$P(X_{n+m} = i_{n+m} | X_0 = i_0, \dots, X_{n+m-1} = i_{n+m-1}) = P(X_{n+m} = i_{n+m} | X_n = i_n, \dots, X_{n+m} = i_{n+m} = i_{n+m-1}) \quad (33)$$

Diszkrét idejű beágyazott Markov-lánc: az $\{M_t, t \geq 0\}$ folyamat, az $\{\bar{M}_m, m \geq 0\}$ a diszkrét idejű beágyazott Markov-lánc az állapotoknak azon sorozata, amelyeket a folyamat az ugrások pillanatában felvesz. A diszkrét idejű beágyazott Markov-lánc átmenetvalószínűség mátrixa:

$$\begin{bmatrix} q_0 & p_0 & 0 & 0 & 0 & \dots \\ q_1 & 0 & p_1 & 0 & 0 & \dots \\ 0 & q_2 & 0 & p_2 & 0 & \dots \\ 0 & 0 & q_3 & 0 & p_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (34)$$

3.2. Folytonos idejű Markov-lánc

[2, c3.3] A Markov-lánc állapotttere legyen egy véges $\chi = \{0, 1, 2, \dots, N\}$ vagy pedig egy megszámlálhatóan végtelen $\chi = \{0, 1, 2, \dots\}$.

5. Definíció. [2, c.3.3] Az $X = (X_t, t \geq 0)$ egy folytonos idejű, χ állapotterű Markov-lánc (CTMC), ha

$$\begin{aligned} P(X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1}, \dots, X_{t_0} = i_0) \\ = P(X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1}) = p_{i_{n-1}, i_n}(t_{n-1}, t_n), \end{aligned} \quad (35)$$

és $0 \leq t_0 \leq t_1 \leq \dots \leq t_n$, $i_0, \dots, i_n \in \chi$ és $n = 0, 1, 2, \dots$ tetszőleges pozitív egész szám.

3.3. Stacionárius eloszlás

6. Definíció. [2, c2.3, d2.5] *Szigorú értelemben vett stacionárius folyamat*

A $\{X_t, t \in \tau\}$ folyamatot szigorú értelemben vett stacionárius, ha az $X_{t_1}, \dots, X_{t_n}$ és az $X_{t_1+1}, \dots, X_{t_n+t}$ valószínűségi változók együttes eloszlása azonos minden t -re, ahol $n = 1, 2, \dots, n$, $t_1, \dots, t_n \in \tau$ valamint $t_i + t \in \tau$

7. Definíció. [2, c2.3, d2.6] *Gyenge értelemben stacionárius folyamat*

Egy X_t sztochasztikus folyamatot stacionáriusnak nevezünk, hogy ha végesdimenziós eloszlásai az időeltolással szemben invariánsak (nem függnek az időeltolástól).

3.4. Átmenetvalószínűség

8. Definíció. *Egylépéses átmenetvalószínűség* [2, c3.1.1 p82]:

$$p_{ij}(n, n+1) = P(X_{n+1} = j | X_n = i), i, j \in \chi, \quad (36)$$

ahol $X = X_0, X_1, \dots$

9. Definíció. [2, c2.7.1 p62] *Homogén Markov-lánc*: ha egy markov lánc egylépéses átmenetvalószínűségei függetlenek az n időtől, és ekkor $p_{ij}(n, n+1) = p_{ij}$, $i, j \in \chi$, $n = 0, 1, \dots$

Ellenkező esetben a markov-lánchoz inhomogénnek nevezzük.

10. Definíció. [2, c3.1.2] $p_{ij}(s, s+m) = p_{ij}(m)$, $s, m = 0, 1, \dots$, $i, j \in \chi$,

ahol a $p_{ij}(m)$, $m = 0, 1, \dots$, $i, j \in \chi$ mennyiségeket az X homogén Markov-lánc m -lépéses átmenetvalószínűségeinek nevezzük. Az ebből képzett m -lépéses átmenetvalószínűség mátrix: $P_t(m) = (p_{ij}(m))_{i,j \in \chi}$

A homogén Markov-lánc átmenetvalószínűség-mátrixai [2, c3.11, p82, d3.7]:

Véges esetben

$$P = \begin{bmatrix} p_{00} & p_{01} & \dots & p_{0N} \\ p_{10} & p_{11} & \dots & p_{1N} \\ p_{20} & p_{21} & \dots & p_{2N} \\ \vdots & \vdots & \dots & \vdots \\ p_{N0} & p_{N1} & \dots & p_{NN} \end{bmatrix} \quad (37)$$

Megszámlálhatóan végtelen esetben

$$\begin{bmatrix} p_{00} & p_{01} & \dots \\ p_{10} & p_{11} & \dots \\ p_{20} & p_{21} & \dots \\ \vdots & \vdots & \dots \end{bmatrix} \quad (38)$$

3.5. Átmenetek osztályozása

Az $j \in \chi$ állapot elérhető az $i \in \chi$ állapotból (jelölje $\longrightarrow$), ha létezik olyan $m \geq 1$ egész szám, hogy $p_{ij}(m) > 0$. Ha két i és j állapot kölcsönösen elérhető egymásból, akkor ezek kapcsolódó állapotok. Jelölése: $i \longleftrightarrow j$ [2, c3.1.3]

11. Definíció. [2, c3.1.3 p87 d3.13] **Nem kapcsolódó állapot:** ha i, j állapotok nem kapcsolódnak, vagyis minden $m \geq 1$ -re $p_{ij}(m) = 0$, vagy $p_{ji}(m) = 0$.

12. Definíció. [2, c3.1.3, p87] **Elnyelő állapot:** ha $p_{ii} = 1$, akkor i elnyelő állapot.

13. Definíció. [2, c3.1.3, p87 d3.14] **Irreducibilis osztályok minden eleme kapcsolódó.**

14. Definíció. [2, c3.1.3, p87 d3.15] **Irreducibilis Markov-lánc:** bármely állapot elérhető tetszőleges más állapotból. Egy irreducibilis Markov-lánc egy irreducibilis osztályt képez, azaz tetszőleges $i, j \in \chi$ állapotok esetén $\exists m, m \geq 1$, és m nem függ i, j -től, hogy $p_{ij}(m) > 0$

3.6. Generátor mátrix

Legyen $X(t)$ egy folytonos idejű Markov-lánc, $X(0) = i$. A lánc a t_1 időpontban a következő állapotába ugrik, ahol $t_1 \sim e^{\lambda_i}$. Nagyon kis δ esetén felírhatjuk [1, c2.1] alapján :

$$\begin{aligned} P(t_1 < \delta) &= 1 - e^{-\lambda_i \delta} \\ &\sim 1 - (1 - \lambda_i \delta) \\ &= \lambda_i \delta \end{aligned} \quad (39)$$

Tehát nagyon kis λ hosszúságú intervallumon annak valószínűsége, hogy elhagyjuk az i . állapotot, közelítőleg $\lambda_i \delta$.

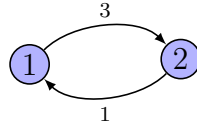
$$\lambda_i = \lim_{\delta \rightarrow \infty} \frac{P(X(\delta) \neq i | X(0) = i)}{\delta} \quad (40)$$

Annak valószínűsége, hogy i -ből j állapotba jussunk: p_{ij} , és az i -ből a j -be történő átmenet aránya: $q_{ij} = \lambda_i p_{ij}$. Vezessük be a Q generátormátrixot, amely i, j -ik eleme q_{ij} :

$$q_{ij} = \begin{cases} \lambda_i p_{ij} & i \neq j \\ -\lambda_i & i = j \end{cases} \quad (41)$$

A Q generátormátrixra igaz, hogy $\sum_j Q(i, j) = 0$.

A Q generátormátrix grafikus magyarázata:



, és

$$Q = \begin{bmatrix} -3 & 3 \\ 1 & -1 \end{bmatrix}$$

3.7. Születési-halálozási folyamat

Vegyünk egy $M(t)$, $t \geq 0$ sztochasztikus folyamatot, aminek az állapotainak halmaza: $I = 0, 1, \dots$. Egy $k \in I$ állapotban töltött tartózkodási idejének eloszlása a [2, c3.4] jelöléseit használva:

$$\alpha_k = \lambda_k + \mu_k \quad (42)$$

paraméterű exponenciális eloszlás. Ez a tartózkodási idő független az ebbe az állapotba kerülésig terjedő trajektóriától. A trajektória alatt az $\{M(t) \in \tau\}$ folyamat realizációt, $\{m(t), t \in \tau\}$ értjük.

A folyamat egy tetszőlegesen megválasztott $k \in I$, $k \geq 1$ állapotból vagy egy $k + 1$ vagy pedig

egy $k - 1$ állapotba kerülhet:

$$\begin{aligned} p_k &= \frac{\lambda_k}{\alpha_k}, \quad k + 1 \text{ állapotba jutás} \\ q_k &= \frac{\mu_k}{\alpha_k}, \quad k - 1 \text{ állapotba jutás.} \end{aligned} \tag{43}$$

A folyamat a 0 állapotból p_0 valószínűséggel kerül az 1 állapotban, és $q_0 = 1 - p_0$ annak valószínűsége, hogy ott is marad.

A folyamat kezdeti eloszlása:

$$P_k(0) = P(M_0 = k) = \rho_k, \quad k \in I. \tag{44}$$

A fenti feltételek mellett az $\{M(t), t \geq 0\}$ egy születési- és halálozási folyamat, és a λ_k a születési-, a μ_k a halálozási intenzitás, a k pedig a populáció nagysága. Ha

- $\lambda_k = 0$, akkor tiszta halálozási a folyamat,
- $\mu_k = 0$, akkor tiszta születési a folyamat.

3.8. Felújítási folyamat

[2, c4.1] Legyen $N(t)$ egy olyan sztochasztikus folyamat, amely események előfordulásainak számát tartalmazza a $[0, t)$ időintervallumon. A t idő a $t_0 = 0$ pontban kezdődik, és $0 \leq t_1 \leq t_2 \leq \dots$. Legyen T_i az események bekövetkezése közötti időszakok, és $T_i = t_i - t_{i-1}$, $i = 1, 2, 3, \dots$. Ha $T_i = t_i - t_{i-1}$, $i = 2, 3, \dots$ függetlenek és azonos eloszlásúak, ahol az eloszlásfüggvény

$$F(X) = P(T_k \leq x, \quad k = 1, 2, \dots)$$

akkor azt mondjuk, hogy $N(t)$ felújítási folyamat [2, c4.1,d4.1]. A T_n -re azt mondjuk, hogy az n . felújítási pont.

A legelső időintervallum eloszlása eltérhet a többitől. Vagyis, ha

$$\begin{aligned} P(T_k \leq x) &= F(x), \quad k = 2, 3, \dots, \text{ és} \\ P(T_1 \leq x) &\neq F(x) \end{aligned},$$

akkor azt mondjuk hogy az $N(t)$ késleltetett felújítási folyamat.

4. Fázis típusú eloszlások

A fázis típusú (Phase Type, vagy röviden PH) eloszlások több exponenciális eloszlás konvolúciójából vagy keverékéből állnak elő [2, 5.1].

15. Definíció. *PH eloszlás definíciói [1, c2.1.2 p9 d2.4] alapján:*

- *A fázis típusú eloszlás Markov-lánc az elnyelő állapotába jutási ideje által adott*
- *Ha a Markov-lánc diszkrét idejű, akkor diszkrét idejű eloszlást, ha folytonos idejű, akkor folytonos idejű eloszlást kapunk*
- *Az eloszlást a Markov-lánc kezdeti eloszlás vektora és tranziens állapotai közötti átmeneteket leíró mátrix egyértelműen definiálja*

4.1. Elnyelő Markov-lánc

[1, c2.1.2] Ha egy markov-lánc állapotai között van elnyelő, akkor elnyelő markov-lánccról beszélünk. Ha az S állapotterű folytonos idejű, az elnyelő állapotába jutó $X(t|0 \leq t < \infty)$ Markov-lánc végesállapotú, és az $S = 1, \dots, n$ állapotokból illetve az $n+1$ elnyelő állapotból áll. Akkor a Q generátor mátrixot fel tudjuk írni a következő alakban, a [1] által adott jelöléssel:

$$Q = \begin{bmatrix} D_0 & d_1 \\ \mathbf{0} & 0 \end{bmatrix} \quad (45)$$

ahol D_0 az $n \times n$ -es, a d_1 pedig $n \times 1$ méretű mátrixok, és a $\mathbf{0}$ az $n \times 1$, a 0 az 1×1 méretű csak 0 elemű mátrixok. A D_0 a kapcsolódó állapotok közötti átmeneteket írja le, a d_1 pedig az átmenetek intenzitását a kapcsolódó állapotokból az elnyelő állapotokba. A fázis típusú eloszlások irodalmában sokat idézett F. Neuts [4] ugyanezt a

$$Q = \begin{bmatrix} T & T^\circ \\ \mathbf{0} & 0 \end{bmatrix} \quad (46)$$

jelöléssel használja rendszeresen.

16. Definíció. *[1, c2.1.2] alapján a fázis típusú eloszlás egy X sztochasztikus folyamat életciklus eloszlásaként definiálható. Az életciklus az az idő, amennyi idő alatt az X folytonos idejű Markov folyamat a kapcsolódó állapotokon végighaladva eljut az elnyelő állapotába.*

Legyen S a véges Markov-lánc állapottere. Akkor $S_T = 1, \dots, n$ a kapcsolódó állapotokat tartalmazó halmaz, és $S_A = n+1$ az elnyelő állapotok halmaza. Azaz $S = S_T \cup S_A$. Az S_T állapotokat nevezzük fázisoknak. Az $X(t)$ folyamatot, aminek n darab kapcsolódó állapota van, n -edfokúnak nevezzük. A folytonos idejű Markov-láncok rendelkeznek egy $[\pi, \pi(n+1)]$ kezdeti

valószínűségi vektorral és a Q generátormátrixszal. A π egy $n \times 1$ méretű oszlopvektor és ez tartalmazza a kezdeti valószínűségeket. A $\pi(n+1)$ az elnyelő, $(n+1)$. állapotba jutás kezdeti valószínűsége [1, c2.1.2,d2.4]. A Q tulajdonságai:

- $D_0 \mathbf{1} + d_1 = \mathbf{0}$,
- $D_0(i, i) \leq 0$, $D_0(i, j) \geq 0$, ha $i \neq j$
- $d_1(i) \geq 0$, és
- $\sum_{j \in S_t} D_0(i, j) \leq 0$

. A $d_1(i)$ jelenti az S_t kapcsolódó állapotokból az $n+1$. elnyelő állapotokba lépés rátáját. A D_0 diagonálisa kifejezhető:

$$D_0(i, i) = - \left(\sum_{i \neq j} D_0(i, j) + d_1(i) \right). \quad (47)$$

És mivel $\sum Q(i, j) = 0$, ezért $D_0 \mathbf{1} + d_1 = \mathbf{0}$, innen $d_1 = -D_0 \mathbf{1}$, ezért $D_0(i, i) \leq 0$, $D_0(i, j) \geq 0$ minden $i \neq j$ esetén, $d_1(i) \geq 0$ és $\sum_{j \in S_T} D_0(i, j) \leq 0$ (S_T a kapcsolódó állapotok halmaza).

4.2. PH eloszlások jellemzői

[1, 2.1.3] megadja a következő PH jellemzőket:

Eloszlásfüggvény

$$F(x) = 1 - \pi e^{D_0 x} \mathbf{1}, \quad x \geq 0 \quad (48)$$

Sűrűségfüggvény

$$f(x) = \pi e^{D_0 x} d_1, \quad x \geq 0 \quad (49)$$

Várható érték

$$E[X] = -\pi D_0^{-1} \mathbf{1} \quad (50)$$

Momentumok

$$E[X]^n = (-1)^n n! \pi D_0^{-n} \mathbf{1} \quad (51)$$

Az $\mathbf{1}$ az 1-eseket tartalmazó oszlopvektor, és D_0 és d_1 a Q generátormátrix blokkjai [16], a π a kezdeti valószínűségi vektor. A fenti egyenletekben a mátrix exponenciális e^Q kifejezhető:

$$e^Q = \sum_{k \geq 0} \frac{1}{k!} Q^k \quad (52)$$

Az X_t markov folyamat állapotátmenet mátrixa P_t , ahol $P_t(i, j) = P(X(t) = j | X(0) = i)$, aminek a jelentése: a t időpillanatban annal valószínűsége, hogy i -ből j állapotba jut a rendszer.

$$P_t = e^{Qt}, \text{ és } e^{Qt} = \begin{bmatrix} e^{D_0 t} & \mathbf{1} - e^{D_0 t} \mathbf{1} \\ \mathbf{0} & 1 \end{bmatrix}. \quad (53)$$

Az exponenciális eloszlából is ismert λ eseményráta paraméter:

$$\lambda = \frac{1}{E[X]} = \frac{1}{\pi D_0^{-n} \mathbf{1}} \quad (54)$$

Az i . kezdőfázis mellett a j . állapotból az elnyelő állapotba jutás teljes idejének várható értéke

$$E[t_{i,j}] = -D_0^{-1}(i, j) \quad (55)$$

17. Definíció. [1, c2.3] *Aciklikus PH eloszlás: Egy (π, D) PH eloszlás aciklikus, hogy az állapotait egy felső háromszög D -be rendezhetjük.*

4.3. Néhány nevezetes PH eloszlás

A [1] és [4] megemlítik a következő, nevezetes fázis típusú eloszlásokat:

Erlang eloszlás Az X v.v. Erlang eloszlású, ha eloszlásfüggvénye:

$$F(x, k) = 1 - \sum_{n=0}^{k-1} \frac{1}{n!} e^{-\lambda x} (\lambda x)^n, \quad x \geq 0, \quad (56)$$

és sűrűségfüggvénye:

$$f(x, n) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}, \quad x \geq 0, \quad (57)$$

Az i . momentuma:

$$E[X^i] = \frac{(n+i-1)!}{(n-1)! \lambda^i} \quad (58)$$

és ebből a várható érték:

$$E[X] = \frac{n}{\lambda} \quad (59)$$

A Q generátormátrix D_0 blokkja:

$$D_0 = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & 0 \\ 0 & -\lambda & \lambda & 0 & 0 \\ 0 & 0 & -\lambda & \lambda & 0 \\ 0 & 0 & 0 & -\lambda & \lambda \end{bmatrix} \quad (60)$$

Hiperexponenciális eloszlás Az X vv. hiperexponenciál eloszlású, ha sűrűségfüggvénye

$$f(x) = \sum_{k=1}^m \pi_k \lambda_k e^{-\lambda_k x}, \quad x > 0, \quad \lambda_k \geq 0, \quad \pi(1) + \dots + \pi(m) = 1$$

Eloszlásfüggvénye:

$$F(x) = 1 - \sum_{i=1}^m \pi(i) \sum_{j=0}^{s_i-1} \frac{(\lambda(i)x)^j}{j!} e^{-\lambda(i)x}, \quad x \geq 0 \quad (61)$$

És

$$D_0 = \begin{bmatrix} -\lambda_1 & 0 & 0 & 0 & 0 \\ 0 & -\lambda_2 & 0 & 0 & 0 \\ 0 & 0 & -\lambda_3 & 0 & 0 \\ 0 & 0 & 0 & -\lambda_4 & 0 \\ 0 & 0 & 0 & 0 & -\lambda_5 \end{bmatrix} \quad (62)$$

Coxian eloszlás Általánosított Erlang-eloszlás

$$D_0 = \begin{bmatrix} -\lambda_1 & p_1 \lambda_1 & 0 & 0 & 0 \\ 0 & -\lambda_2 & p_2 \lambda_2 & 0 & 0 \\ 0 & 0 & -\lambda_3 & p_3 \lambda_3 & 0 \\ 0 & 0 & 0 & -\lambda_4 & p_4 \lambda_4 \\ 0 & 0 & 0 & 0 & -\lambda_5 \end{bmatrix} \quad (63)$$

4.4. MAP (Markovian Arrival Process) folyamat

A $MAP(\pi, D_0, D_1)$ a [1, c4] alapján egy irreducibilis, véges S állapotterű Markov-lánc, π kezdeti valószínűségi vektorral és Q generátormátrixszal, ahol

- $Q = D_0 + D_1$,
- $D_1 \geq 0$,
- $D_0 > 0$, ha $i \neq j$
- $D_0 < 0$, ha $i = j$
- A (π, D_0) PH eloszlású.

A folyamat a $\pi(i)$ valószínűséggel kezdődik az i . állapotban. Feltételezünk egy

$$\lambda(i) = \sum_{j \neq i} D_0(i, j) + \sum_j D_1(i, j)$$

paraméterű exponenciális eloszlást, ami

$$p(\lambda) = \frac{\sum_j D_1(i, j)}{\lambda(i)}$$

valószínűséggel generál egy eseményt. Ez az esemény az i -ből a j nyelő állapotba juttatja a folyamatot, aminek a valószínűsége

$$p_{ij} = \begin{cases} \frac{(D_0(i, j) + D_1(i, j))}{\lambda(i)} & , i \neq j \\ \frac{D_1(i, j)}{\lambda(i)} & , i = j. \end{cases}$$

Gyakorta a π kezdeti valószínűségi vektor nem része a MAP-folyamatnak. Az ilyen esetekben azt feltételezzük, $D_0 + D_1$ egy irreducibilis generátor mátrix, olyan módon, hogy létezik egy

$$P_S = (-D_0)^{-1} D_1$$

amely egy irreducibilis sztochasztikus mátrix. Ennek létezik egy egyedi baloldali sajátvektora:

$$\pi_s P_s = \pi_s, \text{ ahol } \pi_s \mathbf{1} = 1.$$

Az ilyen esetekben a π_s az események utáni stacionárius eloszlás, majd ezt kezdeti vektorként használja fel a *MAP* folyamat. Ilyenkor a *MAP*-ot a (D_0, D_1) formában írjuk fel és mellette implicit módon írjuk fel a π_s kezdeti vektort.

5. Fázis típusú eloszlás közelítése

A következőkben [1, c3] alapján a fázis típusú eloszlások közelítését lehetővé tevő numerikus megoldások közül mutatok néhányat. A feladat, hogy olyan PH eloszlást találjunk, amelyet a (π, D_0) alakban tudunk felírni, és a statisztikai tulajdonságai megegyezők, vagy legalábbis nagyon hasonlóak legyenek, mint egy megfigyelt P sztochasztikus folyamaté. A gyakorlatban nem kivitelezhető végtelen számú megfigyelés rögzítése, ezért véges számú eseményt rögzítünk. A P folyamatot a $\tau = (t_1, t_2, \dots, t_m)$ pillanatokban figyeljük meg. Ezt nevezzük a P folyamat nyomának (trace). A θ legyen a P folyamatból megfigyelt értékek halmaza, és $\xi \in \theta$ az egyes megfigyelések. A ξ értékét vagy a megfigyelt ξ_P (ahol az index a P folyamatot jelöli) vagy a becsült ξ_τ értékkel adhatjuk meg. Egy teljes idősor megfigyelése nagyon nagyszámú adatot eredményezhet, ezért érdemes a $\tau = \{t_i, i \geq 0\}$ értékeket alkalmasan megválasztani úgy, hogy kezelhető méretű nyomot kapjunk. Így első lépésben a mintavételezés karakterisztikájával [1, c3.1] kell foglalkoznunk, majd utána vesszük elő a megfelelő approximáció algoritmusok elméleteit a PH [1, c3.1.2] és a MAP [1, c5.2] eloszlásokhoz.

5.1. Mintavételezés karakterisztikája

A [1, c3.1.1] alapján a mintavételezés i . momentuma

$$\hat{\mu}_i = \frac{1}{m} \sum_{j=1}^m (t_j)^i \quad (64)$$

és varianciája

$$\hat{\sigma}^2 = \frac{1}{m-1} \sum_{j=1}^m (t_j - \hat{\mu}_1)^2. \quad (65)$$

Határozzuk meg az autokorrelációt és együtthatóit, hogy lássuk, a nyom adatpontjai között van-e függőség. Az együttható a k lag-paraméterrel (késleltetés-paraméter):

$$\hat{\rho}_k = \frac{1}{(m-k-1)\hat{\sigma}^2} \sum_{j=1}^{m-k} (t_j - \hat{\mu}_1)(t_{j+k} - \hat{\mu}_1) \quad (66)$$

Egymást követő két pont együttes momentumának becslése, $\mu_{i,j} = E[X_k^i X_{k+1}^j]$:

$$\hat{\mu}_{i,j} = \frac{1}{m-1} \sum_{j=1}^{m-1} (t_k)^i (t_{k+1})^j \quad (67)$$

Az empirikus eloszlásfüggvény, bevezetve az m -et, mint a lépések számát:

$$F_\tau(x) = \frac{\sum_{j=1}^m \delta(t_j \leq x)}{m} \quad (68)$$

ahol

$$\delta(b) = \begin{cases} 1 & , \text{ha } b \text{ teljesül} \\ 0 & , \text{különben} \end{cases}$$

Ha az összes megfigyelt eredmény alapján approximálnánk a PH eloszlást, akkor nagyon hatékonytalan, költséges lenne az approximációs algoritmus. Ez miatt szükséges az, hogy megfelelő méretű nyomot válasszunk az eredeti adatsorból. Ezt a szűkített nyomot aggregált-nyomnak nevezzük. Az eljárást, amivel megtaláljuk a szükséges nyomot, *nyom-aggregálásnak hívjuk*.

A nyom empírikus eloszlásfüggvényét fel tudjuk bontani kis intervallumokra:

$$\Delta_0 < \Delta_1 < \dots < \Delta_M$$

Jelentse mindegyik Δ_i intervallum az adott intervallum átlagát, illetve súlyát. Így felírhatjuk a nyomot a

$$\tau^* = \{(\hat{t}_1, w(\hat{t}_1)), \dots, (\hat{t}_m, w(\hat{t}_m))\} \quad (69)$$

formában. Jelentse tehát $\hat{t}_i$ az $[\Delta_{t-1}\Delta_t]$ intervallum átlagát, a w_i pedig azt a súlyt, hogy hány eleme esik az adott intervallumba. A τ -t osszuk fel a $\min(\tau)$ és $\max(\tau)$ intervallumot

$$[\min(\tau), \max(\tau)]$$

M darab intervallumra:

$$(\Delta_{i-1}\Delta_i], i = 1, \dots, M$$

valamint

$$0 \leq \Delta_0 < \min(\tau), \text{ és} \\ \Delta_M = \max(\tau).$$

Az intervallumok szélességét a megválasztott M paraméter adja. Vezessük be a $j \in J_i$ index változót, amely az egy intervallumba eső elemek száma:

$$t_j \in (\Delta_{t-1}\Delta_t], i = 1, \dots, M \text{ és} \\ m_i = |J_i|$$

A nyom elemeinek átlaga a $(\Delta_{t-1}\Delta_t]$ intervallumon

$$\hat{t}_i = \frac{1}{m_i} \sum_{j \in J_i} t_j$$

és az i . intervallum súlya

$$w_i = \frac{m_i}{dm}$$

ami megegyezik a

$$P(j \in J_i) = P(t_j \in (\Delta_{i-1}, \Delta_i])$$

valószínűséggel. A 0 átlagú intervallumokat nyugodtan ignorálhatjuk. Ha M értékét túl kicsire vesszük, akkor kevés intervallumra osztjuk fel a nyomot, így pontatlanul tudjuk követni az eredeti eloszlást. Az EM algoritmus használata során a jó közelítéshez (a legalkalmasabb paraméter megtalálásához) ajánlatos végrehajtani a nyom-aggregálást úgy, hogy hosszú nyomokat használjunk, abban sok hasonló elemmel.

5.2. EM algoritmus (Expectation Maximization) PH eloszláshoz

Az EM algoritmus [1, c3.1.2] egy iteratív módszer a maximum likelihood paramétereinek megtalálására.

Legyen egy $\tau = (t_1, t_2, \dots, t_m)$ megfigyelésünk, amelynek legyen egy f sűrűségfüggvénye. A sűrűségfüggvényt írjuk fel a következő alakban:

$$f = f(t_1|\theta) \quad (70)$$

ahol a θ az eloszlás szabadon választható paramétere. A mintavételezéssel kapott τ -ról azt lehet feltételezni, hogy kisebb elemszámú, mint a teljes X sokaság, amelyből τ is származik. Létezik egy u leképezés ami X -ből τ -t csinál: $\tau = u(X)$. Ha bevezetjük a $z = (\tau, X)$ struktúrát, akkor ez a teljes adatmennyiséget tartalmazza.

A t_i értékek függetlenek és egyértelműen meghatározzák az f eloszlást. Írjuk le az f -et közelítő *likelihood* függvényt:

$$\mathcal{L}(\theta|\tau) = f(\tau|\theta) = \prod_{i=1}^m f(t_i|\theta). \quad (71)$$

Gyakorta használják a log likelihood függvényt:

$$\log(\mathcal{L}(\theta|\tau)) = \log\left(\prod_{i=1}^m f(t_i|\theta)\right) = \sum_{i=1}^m \log(f(t_i|\theta)). \quad (72)$$

Jelölje Ω a θ paraméterek halmazát. Akkor θ^* legyen, ami maximalizálja a *likelihood* függvényt:

$$\theta^* = \max_{\theta \in \Omega} \mathcal{L}(\theta|\tau). \quad (73)$$

Az EM algoritmusnak két lépése van. Az első a várható érték (E), a második a maximalizálás (M). Az EM algoritmus az $(r+1)$. iterációs lépésében sorra veszi a $\theta^{(r+1)}$ paramétert, ami maximalizálja a következő kifejezést:

$$E[\log(f(X, \tau|\theta^{(r+1)})) | f(X|\tau, \theta^r)], \quad (74)$$

a θ^r az r . iterációs lépésben alapul vett paraméter. Az E lépéssel találjuk meg az adatok sokaságán az ismert- és mért adatok alapján, és szintén az E lépés adja vissza az eloszlás θ^r paraméterét. Az M lépéssel újrabecsüljük az E lépés során talált, és helyesnek elfogadott θ^r paramétert. Az EM algoritmus egy lokális maximumkereső eljárás, amellyel becslések sorozatát állítjuk elő a nemcsökkenő likelihood eljárással, de lehetséges, hogy ez a sorozat csak egy nyeregpontra, vagy csak egy lokális maximumra eredményez.

A PH eloszlások esetében a megfigyelt adatok tartalmazzák az elnyelő állapotba jutás idejét, és ezek megfelelnek a τ nyomban lévő adatoknak. A nem megfigyelt adatok megfelelnek a PH eloszlás állapotainak, amelyeket a folyamat felvesz az elnyelő állapotba jutása előtt. Az illesztési feladat alapproblémája, hogy megtaláljuk a

$$\theta = (\pi, D_0) \quad (75)$$

paramétert úgy, hogy a nyom elemeinek eloszlásfüggvényét maximalizáljuk. A (π, D_0) PH eloszlásra alkalmazott likelihood függvény:

$$\mathcal{L}((\pi, D_0)|\tau) = \prod_{i=1}^m \pi e^{D_0 t_i} d_1. \quad (76)$$

Az a célunk, hogy maximalizáljuk $\mathcal{L}$ -et:

$$(\pi, D_0)^* = \max_{(\pi, D_0)} \prod_{i=1}^m \pi e^{D_0 t_i} d_1. \quad (77)$$

Az elnyelő állapotba jutásig eltelt idő megfigyelése nem jelenti a teljes a Markov-folyamat megfigyelését. Csak részletes megfigyelésünk lehet, amit jelöljünk $\{X(t)\}_{0 \leq t \leq t_i}$ -vel. A megfigyelés során tapasztalt $t_i \in \tau$ változó egy Y Markov-folyamat része, ami az eredeti $\{X(t)\}$ folyamat elnyelő állapotba jutását tartalmazza. Csak a t_i értékeket ismerjük, ezért nincs teljes információnk arról, hogy a háttér Markov-lánc hogy jut el az elnyelő fázisba és mely állapotokat vette fel, és azokban mennyi időt töltött. A teljes adat összeállítható a beágyazott $\{X(t)\}_{0 \leq t \leq t_i}$ Markov-folyamatból a $\{X(t)\}_{0 \leq t \leq t_i}$ folyamatból, annak összes meglátogatott fázisának sorrendjét és az adott fázisban töltött időt felhasználva. Most jelöljük k -val azon lépések számát, ami alatt az $\{X(t)\}_{0 \leq t \leq t_i}$ eljut az elnyelő állapotba. Akkor az $X_k = n + 1$ az n -edfokú PH eloszlás elnyelő állapota. Ez az n adja a PH eloszlás dimenzióját, és ez az optimalizálás folyamata előtt már ismertnek tekinthető.

Az elnyelő állapotba jutásig sorra meglátogatott fázisokat jelöljük $X_0, \dots, X_{k-1}$ -el. Az ezen állapotokban eltöltött időket pedig jelöljük sorra $S_0, \dots, S_{k-1}$ -el. Akkor vegyük a következő n -eseket ($(k-1)$ -eseket):

$$z = (x_0, x_1, \dots, x_{k-1}, s_0, s_1, \dots, s_{k-1}) \quad (78)$$

amivel mindent tudunk az $\{X(t)\}_{0 \leq t \leq t_i}$ -ről a $[0, t_i]$ intervallumon. A t_i elnyelő állapotba jutási idő az egyes, odáig tartó állapotokban töltött idők összege:

$$t_i = \sum_{j=0}^k s_j. \quad (79)$$

A teljes megfigyelésből származó z és a θ paraméter alapján az f függvény eloszlása

$$\begin{aligned} f(z|\theta) &= f(z|(\pi, D_0)) \\ &= \pi(x_0) \lambda(x_0) e^{-\lambda(x_0)s_0} P(x_0, x_1) \dots \lambda(x_{k-1}) e^{-\lambda(x_{k-1})s_{k-1}} P(x_{k-1}, n+1) \\ &= \pi(x_0) e^{-\lambda(x_0)s_0} D_0(x_0, x_1) \dots e^{-\lambda(x_{k-1})s_{k-1}} d_1(x_{k-1}). \end{aligned} \quad (80)$$

A $\tau = (t_1, \dots, t_m)$ -et m -féle független módon adhat eredményt az $\{X(t)\}$ Markov-lánchról. Így a z is m számú lehetőséget ad az elnyelő állapotba jutáshoz:

$$z = (x_0^1, \dots, x_0^m, \dots, s_{k-1}^1, \dots, s_{k-1}^m). \quad (81)$$

A z sűrűségfüggvénye, amit a maximum likelihood függvényben is használunk:

$$\begin{aligned} \mathcal{L}((\pi, D_0)|\tau) &= f(z|(\pi, D_0)) \\ &= \prod_{i=1}^n \pi(i)^{B_i} \prod_{i=1}^n e^{Z_i D_0(i,j)} \prod_{i=1}^n \prod_{j=1, j \neq i}^{n+1} D_0(i, j)^{N_{ij}}, \end{aligned} \quad (82)$$

ahol B_i legyen annak száma, hogy a Markov-lánc hányszor kezdődik az i fázisban, a Z_i az i . fázisban eltöltött teljes idő, és N_{ij} az összes olyan állapotváltozás száma, amikor a fázisátmenet $i \rightarrow j$ és $i \in S_T$ és $j \in S$.

5.3. Nyomokra épülő illeszkedésvizsgálat MAP-ra

A már ismert τ alapján becsüljük a (D_0, D_1) MAP folyamat θ paraméterét [1, c5.2]. A maximum likelihood függvény:

$$\mathcal{L}((D_1, D_1)|\tau) = \pi \prod_{i(1)}^m e^{D_0 t_i} D_1 \mathbf{1}. \quad (83)$$

Az optimalizálás feladat:

$$(D_0, D_1)^* = \max_{(D_0, D_1)} \pi \prod_{i=1}^m e^{D_0 t_i} D_1 \mathbf{1}. \quad (84)$$

A likelihood függvény kifejezéséhez definiálni kell a következőket: B_i azon események száma, amikor a folyamat az i . állapotban kezdődik el. A Z_i az összes megfigyelt olyan $i \rightarrow j$

állapotátmenetek száma, amelyek nem generálnak N_{ij} eseményt. Illetve az M_{ij} az összes olyan $i \rightarrow j$ állapotátmenetek száma, amelyek viszont eseményt generálnak. Így akkor írjuk fel a likelihood függvényt:

$$\begin{aligned} \mathcal{L}((D_0, D_1)|\tau) = \\ = \prod_{i=1}^n \pi(i)^{B_i} \prod_{i=0}^n e^{Z_i D_0(i,i)} \prod_{i=1}^n \prod_{j=1, j \neq i}^n D_0(i, j)^{N_{ij}} \prod_{i=1}^n \prod_{j=1, j \neq i}^n D_1(i, j)^{M_{ij}} \mathbf{1} \end{aligned} \quad (85)$$

6. Struktúrált Markov-láncú sorbanállási rendszerek

Olyan sorbanállási rendszerekről lesz szó, ahol a beérkezési sor, és/vagy a feldolgozási idők eloszlása struktúrált markovi tulajdonságúak: PH vagy MAP. A bemutatás után a vizsgált Adattárház rendszer érkezési és kiszolgálási folyamatait a lenti modellek közül végül a MAP/PH/N és PH/PH/N típusokkal közelítettem. Az Adattárház rendszerben N kiszolgálóegység van, de az egyes modellek vizsgálata során az egy kiszolgálóegységgel rendelkező rendszerek áttekintésével kezdtem.

6.0.1. A PH eloszlások további tulajdonságai

A struktúrált Markov-láncú folyamatokhoz szükséges a PH eloszlásokhoz egy kiegészítés [4, c2.2,p59]. Az olyan alkalmazásokban, ahol az érkezési folyamat egy λ intenzitású Poisson-folyamat egy $[0, X]$ intervallumon, ahol az X a Poisson-folyamattól független valószínűségi változó. Ha F az X eloszlása, akkor

$$a_k = P(N = k) = \int_0^\infty e^{-\lambda u} \frac{(\lambda u)^k}{k!} dF(u), \quad k \geq 0. \quad (86)$$

Az F generátorfüggvénye $A(z) = \{a_k\}$:

$$A(z) = f(\lambda - \lambda z) \quad (87)$$

ahol az f az F eloszlásfüggvény Laplace-Stieltjes transzformáltja. Ha F eloszlás fázis típusú eloszlás, amit (α, T) alakban írhatunk fel, akkor az $\{a_k\}$ egy PH-sűrűségfüggvény, amit (β, S) alakban tudunk felírni. Ekkor

$$\begin{aligned} \beta &= \lambda \alpha (\lambda I - T)^{-1} \\ S &= \lambda (\lambda I - T)^{-1} \\ \beta_{m+1} &= \alpha_{m+1} + \alpha (\lambda I - T)^{-1} T^\circ \\ S^\circ &= (\lambda I - T)^{-1} T^\circ \end{aligned} \quad (88)$$

6.1. Kvázi születési-halálozási folyamat

A kvázi születési-halálozási folyamatot [2, c5.3] röviden QBD folyamatnak nevezik (Quasi Birth-Death process). Legyen az $\{N(t), J(t)\}$ egy folytonos idejű Markov-lánc, ahol a rendszer állapotát egy t időpillanatban két sztochasztikus folyamattal írjuk le. Az $\{N(t), J(t)\}$ állapottere: $\{n, j\}$ és $n \in \mathbb{N}$, $j \in \{1, 2, \dots, J\}$.

Szint: az állapotok azon részhalmaza, amihez ugyanaz a N érték tartozik. Az egy lépéses átmenetek vagy csak szinten belül, illetve a szomszédos szintek között lehetségesek. A kvázi születési-halálozási folyamat (**QBD**) struktúrált felírásához adjuk meg az átmenetek osztályait:

- felfelé történő átmenet: $(n \rightarrow n + 1)$
- lokális átmenet: $(n \rightarrow n)$
- lefelé történő átmenet: $(n \rightarrow n - 1)$

Ezek után alkalmazzuk a következő jelöléseket:

- Legyen $\mathbf{F}$ a $J \times J$ méretű mátrix, amelyik az felfelé lépéses átmenetek sebességét tartalmazza. Az $\mathbf{F}_{ij}$ elem jelentse a $\{n, i\} \rightarrow \{n + 1, j\}$, $(n \geq 0)$ átmenet intenzitását.
- Az $\mathbf{L}$ egy $J \times J$ mátrix, amelyik a lokális átmeneteket tartalmazza minden $n \geq 0$ -ra: L_{ij} jelentse a
- Az $\mathbf{L}'$ egy $J \times J$ mátrix, amelyik az olyan lokális átmenetek mértékét tartalmazza, ahol $n = 0$. A 0 szint nem szabályos, mivel a 0 állapotból nincs lehetőség visszafelé lépni.
- A $\mathbf{B}$ szintén egy $J \times J$ méretű mátrix, amely a lefelé történő átmenetek sebességét tartalmazza. Az $\mathbf{B}_{ij}$ elem jelentse a $\{n, i\} \rightarrow \{n - 1, j\}$, $(n \geq 0)$ átmenetet.

A fenti jelölések ismeretében a kvázi születési-halálozási folyamat generátormátrixa:

$$Q = \begin{bmatrix} L' & F & & & \\ B & L & F & & \\ & B & L & F & \\ & & B & L & F \\ & & & B & L & F \end{bmatrix}.$$

A kvázi születési-halálozási (QBD) folyamat reguláris része, azaz amikor $n > 1$ egy folytonos idejű markov-lánc, aminek a generátormátrixa

$$A = F + L + B. \quad (89)$$

6.1.1. QBD folyamat stabilitása

Az A legyen egy irreducibilis folyamat egy α stacionárius eloszlással, ekkor

$$\alpha A = \mathbf{0}$$

$$\alpha \mathbf{1} = \mathbf{1}.$$

Az α stacionárius eloszláshoz tartozó *sodródást* jelöljük d -vel:

$$d = \alpha F \mathbf{1} - \alpha B \mathbf{1}.$$

A d sodródás azt mutatja meg, hogy a szintfolyamat hogyan változik meg a reguláris részben (amikor $n > 1$). Ha $d < 0$, akkor a folyamat pozitív rekurrens, vagyis a folyamat stabil:

$$\begin{aligned} d &= \alpha F \mathbf{1} - \alpha B \mathbf{1} < 0, \text{ ahol} \\ \alpha(F + L + B) &= 0 \\ \alpha \mathbf{1} &= 1 \end{aligned} \tag{90}$$

[2, c5.3,p182]

6.2. Mátrix geometrikus eloszlás

6.2.1. Alapok

[2, c5.3.1]. A Q generátormátrixsal rendelkező kvázi születési-halákozási (QBD) folyamat stacionárius megoldása megegyezik a következő egyenletek megoldásával:

$$\begin{aligned} \pi Q &= 0 \\ \pi \mathbf{1} &= 1 \end{aligned} \tag{91}$$

A π felírhatjuk a következő formában: $\pi = \{\pi_0, \pi_1, \dots\}$. Ezt a formát használva felírható a generátormátrix egyenlete:

$$\begin{aligned} \pi_0 L' + \pi_1 B &= 0 \\ \pi_{n-1} F + \pi_n L + \pi_{n+1} B &= 0, n \geq 1, \end{aligned} \tag{92}$$

és $\sum_{n=0}^{\infty} \pi_n \mathbf{1} = 1$.

1. Tétel. [2, 5.3.1 t5.9] A π stacionárius valószínűség vektor felbontható m -vektorra: $\pi_k, k \geq 0$. Stabil állapotú QBD esetén a [92] egyenletek megoldása

$$\pi_n = \pi_0 R^n \tag{93}$$

ahol az R mátrix megoldása a

$$F + RL + R^2 + B = 0 \tag{94}$$

kvadratikus egyenletnek. Az egyenletrendszer megoldása a π_0 oszlopvektor:

$$\pi_0(L' + RB) = 0, \tag{95}$$

és a normalizáló feltétel:

$$\pi_0(I - R)^{-1} \mathbf{1} = 1 \tag{96}$$

18. Definíció. [8] A Normalizáló állandó egy olyan konstans, amellyel megszorozva egy sehol sem negatív függvényt, annak görbe alatti területe 1 lesz.

6.2.2. M/G/1 és GI/M/1

A PH/PH/1, PH/PH/N, MAP/PH/1, MAP/PH/N rendszerek bemutatásához az alábbi rendszereken keresztül részletezem tovább a mátrixgeometriai eljárást [4, c1.1]. Az M/G/1 típus markov-láncának blokk alakú felírása :

$$P = \begin{bmatrix} B_0 & B_1 & B_2 & B_3 & B_4 & \cdots \\ C_0 & A_1 & A_2 & A_3 & A_4 & \cdots \\ 0 & A_0 & A_1 & A_2 & A_3 & \cdots \\ 0 & 0 & A_0 & A_1 & A_2 & \cdots \\ 0 & 0 & 0 & A_0 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (97)$$

Az $A_v, v \geq 0$ egy $m \times m$ méretű mátrix. A $B_v, v \geq 1$ mérete $n \times m$, a B_0 mérete $n \times n$ és legutoljára C_0 mérete $m \times n$.

Legyen $\tilde{P}$ a GI/M/1 átmenetvalószínűségmátrixa, akkor

$$\tilde{P} = \begin{bmatrix} B_0 & A_0 & 0 & 0 & 0 & \cdots \\ B_1 & A_1 & A_0 & 0 & 0 & \cdots \\ B_2 & A_2 & A_1 & A_0 & 0 & \cdots \\ B_3 & A_3 & A_2 & A_1 & A_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Az összes blokk mérete $m \times m$. A $\tilde{P}$ sztochasztikus mátrix (azaz sorainak összege 1), akkor

$$B_k \mathbf{1} + \sum_{v=0}^k A_v \mathbf{1} = \mathbf{1}, \quad k \geq 0. \quad (98)$$

1. Tétel. [4, c1.2 t2.1.2] A $\tilde{Q}$ irreducibilis Markov folyamat pozitív rekkurrens, akkor és csakis akkor, ha a

$$\sum_{k=0}^{\infty} R^k A_k = 0$$

egyenlet minimális, nemnegatív R megoldása spektrális rádiusza kisebb mint egy: $sp(R) < 1$, és létezik egy π_0 pozitív vektor, ami kielégíti a következő feltételt:

$$\pi_0 B[R] = 0$$

és $B[R] = \sum_{k=0}^{\infty} R^k B_k$.

19. Definíció. [9] *Spektrális rádiusz: egy négyzetes mátrix legnagyobb sajátértékének abszolút értéke.*

$$sp(R) = \max\{|\lambda_1|, |\lambda_2|, \dots\}.$$

A GI/M/1 folyamat markov-lánca egy kvázi születési-halálozási folyamat, az x kezdeti valószínűségvektorát feloszthatjuk m vektorra, ezt jelöljük x_k -val, ahol $k \geq 0$, és akkor adott a

$$x_k = x_0 R^k, \quad k \geq 1, \quad (99)$$

és az R a nemnegatív megoldása a mátrixkvadrátikus egyenletnek.

Vezessük be a következő paramétert a folytonos idejű markov-folyamat esetében:

$$\tilde{P}^\circ = \begin{bmatrix} B_0^\circ & A_0^\circ & 0 & 0 & 0 & \dots \\ B_1^\circ & A_1^\circ & A_0^\circ & 0 & 0 & \dots \\ B_2^\circ & A_2^\circ & A_1^\circ & A_0^\circ & 0 & \dots \\ B_3^\circ & A_3^\circ & A_2^\circ & A_1^\circ & A_0^\circ & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Az új paraméter blokkjai:

$$\begin{aligned} B_0^\circ &= \Delta_0^{-1} B_0 + I \\ A_0^\circ &= \Delta_0^{-1} A_0 \\ B_\nu^\circ &= \Delta^{-1} B_\nu \\ A_\nu^\circ &= \Delta^{-1} A_\nu + \delta_{1\nu} I \\ \Delta &= -\text{diag}(A_1) \\ \Delta_0 &= -\text{diag}(B_0) \end{aligned} \quad (100)$$

Az $A^\circ = \sum_{\nu=0}^{\infty} A_\nu^\circ$, és $A^\circ = I + \Delta^{-1} A$. Az A° akkor és csak akkor sztochasztikus mátrix, ha $A\mathbf{1} = 0$. Az A° akkor és csak akkor reducibilis, ha A is reducibilis. A π° valószínűségi vektor pedig az alábbi által adott: $\pi^\circ = (\pi \Delta \mathbf{1})^{-1} \pi \Delta$.

6.3. PH/M/1

A PH/M/1 modellben [2, c9.1] a beérkezési folyamat PH eloszlású, amit a (τ, T) jelöléssel adunk meg. A τ a kezdeti eloszlásvektor, a T a generátormátrix. A kiszolgálási idők eloszlása μ paraméterű exponenciális. Egy feldolgozóegység található a rendszerben, és végtelen a várakozási sor.

. A rendszert le tudjuk írni úgy, mint egy folytonos idejű $\{N(t), J(t)\}$ folytonos idejű markov-láncot. Ahol az $N(t)$ a rendszerben található igények száma, a $J(t)$ pedig a Markov-lánc

fázisfolyamata. A $J(t)$ pedig a PH-eloszlású érkezéseket jellemi a t időpillanatban.

Egy t pillanatban a folyamat: $(N(t), J(t)) = (n, j)$. Ekkor a következők történhetnek:

- Ha $n \geq 1$, akkor $(n, j) \rightarrow (n, k)$ és $k \neq j$, és ezalatt a PH eloszlás "háttér" Markov-láncban nem történik érkezés. Az $(n, j) \rightarrow (n, k)$ -t a T_{jk} állapotátmenet írja le.
- A PH folyamat háttér Markov-lánca elnyelő állapotba juthat, és történhet egy igény beérkezése. Ha a folyamatban egy igény beérkezése történik, akkor a rendszerben található igények száma eggyel növekszik, és egy új, PH eloszlású intervallum kezdődik a τ kezdeti eloszlásvektor szerint. Legyen $\mathbf{t}$ egy oszlopvektor, amelyik a nyelő állapotba jutás rátáját tartalmazza.

$$\mathbf{t} = -T\mathbf{1}$$

Akkor a $(n, j) \rightarrow (n + 1, k)$ átmenet esetén az átmenet rátája $\mathbf{t}_j \tau_k$

- $n > 0$ esetében a rendszerben található lesz egy olyan igény, ami kiszolgálás alatt van, és a kiszolgálási idők eloszlása egy μ paraméterű exponenciális. Az igény kiszolgálásának befejeztével a rendszerben található igények száma eggyel csökken. Mivel az érkezési és a kiszolgálási folyamatok függetlenek, a kiszolgálás sikeressége nem befolyásolja az érkezési folyamatot. Így tehát $(n, j) \rightarrow (n - 1, j)$ átmenet rátája pontosan a μ lesz, és $(n, j) \rightarrow (n - 1, k)$ értéke pedig 0, ahol $j \neq k$.

A két fázissal rendelkező PH érkezési folyamatú markov-lánc generátormátrixa:

$$Q = \begin{bmatrix} \begin{bmatrix} x & T_{1,2} \\ T_{2,1} & x \end{bmatrix} & \begin{bmatrix} \mathbf{t}_1 \tau_1 \mathbf{t}_1 \tau_2 \\ \mathbf{t}_2 \tau_1 \mathbf{t}_2 \tau_2 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ \begin{bmatrix} \mu & 0 \\ 0 & \mu \end{bmatrix} & \begin{bmatrix} x & T_{1,2} \\ T_{2,1} & x \end{bmatrix} & \begin{bmatrix} \mathbf{t}_1 \tau_1 \mathbf{t}_1 \tau_2 \\ \mathbf{t}_2 \tau_1 \mathbf{t}_2 \tau_2 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} \mu & 0 \\ 0 & \mu \end{bmatrix} & \begin{bmatrix} x & T_{1,2} \\ T_{2,1} & x \end{bmatrix} & \begin{bmatrix} \mathbf{t}_1 \tau_1 \mathbf{t}_1 \tau_2 \\ \mathbf{t}_2 \tau_1 \mathbf{t}_2 \tau_2 \end{bmatrix} \end{bmatrix} \quad (101)$$

A diagonálisban látható x értékét az adja, hogy a Q mátrixban a sorok összege nulla. A fenti Q generátor átalánosítása:

$$Q = \begin{bmatrix} T & t\tau & & \\ \mu\mathbf{I} & T - \mu\mathbf{I} & t\tau & \\ & \mu\mathbf{I} & T - \mu\mathbf{I} & t\tau \\ & & \mu\mathbf{I} & T - \mu\mathbf{I} & t\tau \end{bmatrix} \quad (102)$$

A mátrixban a sorok összege nulla, vagyis, ha

- $n = 0$, akkor $T\mathbf{1} + t\tau\mathbf{1} = 0$. A $\tau\mathbf{1} = 1$ és $t = -T\mathbf{1}$.

- $n \geq 1$, akkor $\mu \mathbf{I} \mathbf{1} + (T - \mu \mathbf{I}) \mathbf{1} + t\tau \mathbf{1} = 0$

A QBD folyamat generátormátrixa: $A = F + L + B$, és

- $F = t\tau$
- $L = T - \mu \mathbf{I}$
- $B = \mu \mathbf{I}$
- és $L' = T$.

6.3.1. Teljesítménymutatók, feldolgozóegységek állapotai

Stabilitás Akkor tekintjük a rendszert stabilnak [2, c9.1.2], hogy ha több igényt szolgál ki, mint amennyi beérkezik. Azaz a beérkezések közötti időközök átlaga $(\tau(-T^{-1})\mathbf{1})$ nagyobb, mint a kiszolgálási idő:

$$\tau(-T)^{-1}\mathbf{1} > \frac{1}{\mu}$$

A kvázi születési-halálozási stabilitási feltétele mentén vizsgáljuk a PH/M/1 stabilitását. A QBD folyamat generátormátrixa:

$$A = B + L + F = T + t\tau,$$

Ha az α a fázis folyamat stacionárius eloszlása, és $\alpha(F + L + B) = 0$, $\alpha \mathbf{1} = 1$ akkor

$$\alpha = \frac{\tau(-T)^{-1}}{\tau(-T)^{-1}\mathbf{1}} \quad (103)$$

Az PH/M/1 rendszer stabil, hogy ha

$$\alpha B \mathbf{1} > \alpha F \mathbf{1}, \quad (104)$$

ahol

$$\begin{aligned} \alpha B \mathbf{1} &= \alpha \mu \mathbf{I} \mathbf{1}, \text{ és} \\ \alpha F \mathbf{1} &= \frac{1}{\tau(-T)^{-1}\mathbf{1}} (-T \mathbf{1} \tau) \mathbf{1} = \frac{1}{\tau(-T)^{-1}\mathbf{1}} \end{aligned} \quad (105)$$

Teljesítménymutatók, feldolgozóegységek állapotai A PH/M/1 rendszer teljesítményének [2, c9.1.3] mértékét a $\{N(t), J(t)\}$ QBD folyamat stacionárius eloszlása jellemzi. A mátrix geometrikus eloszlásból [6.2.2] ismert módon a stacionárius eloszlás az R mátrixszal számolható: $\pi_n = \pi_0 R^n$, ahol n a rendszerben található igények száma. Az R pedig a $F + RL + R^2B = 0$ megoldása.

Kihasznátság A PH/M/1 modellben egy szerver van. A Little-formulában [2.3.3] megadott $N(t)$ a rendszerben található igények száma. A rendszer akkor foglalt, ha legalább egy igény van benne. Akkor

$$\rho = \lim_{t \rightarrow \infty} P(N(t) \geq 1) = \sum_{n=1}^{\infty} \pi_n \mathbf{1} = 1 - \pi_0 \mathbf{1} \quad (106)$$

Igények száma Az igények számának stacionárius eloszlása

$$p_n = \lim_{t \rightarrow \infty} P(N(t) = n) = \pi_n \mathbf{1} = \pi_0 R^n \mathbf{1} \quad (107)$$

Az igények számának várható értéke

$$\begin{aligned} E(N) &= \lim_{t \rightarrow \infty} E(N(t)) \\ &= \sum_{n=0}^{\infty} n p_n \\ &= \sum_{n=0}^{\infty} n \pi_0 R^n \mathbf{1} \\ &= \sum_{n=0}^{\infty} \sum_{k=0}^n \pi_0 R^n \mathbf{1} \\ &= \pi_0 \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} R^n \mathbf{1} \\ &= \pi_0 \sum_{k=0}^{\infty} R^k (I - R)^{-1} \mathbf{1} \\ &= \pi_0 (I - R)^{-2} \mathbf{1} \end{aligned} \quad (108)$$

6.4. PH/M/N

A PH/M/N típusú modellben [4, c3.2,p88] az érkezési folyamat PH eloszlású (τ, T) , és a $\lambda'_1 = -\alpha T^{-1} \mathbf{1}$. A kiszolgálási idők eloszlása μ paraméterű exponenciális (szerverenként), és N darab kiszolgálóegységünk van. Azt feltételezhetjük, hogy a rendszerben egyidejűleg maximum $K + N$, $K \geq 0$ igény tartózkodhat.

A sorunk egy QBD folyamat az $E = \{(i, j), i \geq 0, 1 \leq j \leq m\}$ állapottérrel végtelen, és $E = \{(i, j), 0 \leq i \leq K + N, 1 \leq j \leq m\}$ állapottérrel véges esetben, ahol jelölje i a rendszerben található igények számát, és legyen j a beérkezés folyamat fázisa, Q pedig a generátormátrix.

Végtelen esetben:

$$Q = \begin{bmatrix} T & T^\circ A^\circ & \dots & \dots \\ \mu I & T - \mu I & T^\circ A^\circ & \dots \\ & 2\mu I & T - 2\mu I & T^\circ A^\circ & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & c\mu I & T - c\mu I & T^\circ A^\circ & \dots \\ & & & c\mu I & T - c\mu I & \dots \end{bmatrix} \quad (109)$$

A [45] és [46] alapján a Q generátormátrixot több módon is jelölik az irodalomban. A D_0 vagy T blokkok, illetve a d_1 és T° megegyeznek. Véges esetben:

$$Q = \begin{bmatrix} T & T^\circ A^\circ & \dots & \dots \\ \mu I & T - \mu I & T^\circ A^\circ & \dots \\ & 2\mu I & T - 2\mu I & T^\circ A^\circ & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & c\mu I & T - c\mu I & T^\circ A^\circ \\ & & & c\mu I & T - c\mu I & \end{bmatrix}, \quad (110)$$

ahol a Q az $K + N \times m$ méretű.

A kvázi születési folyamat akkor és csak akkor pozitív visszatérő (rekurrens), ha

$$N\mu \geq \pi(T^\circ A^\circ)\mathbf{1} = \pi T^\circ = \lambda_1'^{-1}$$

és

$$\begin{aligned} \pi(T + T^\circ A^\circ)\mathbf{1} &= 0 \\ \pi\mathbf{1} &= 1 \end{aligned} \quad (111)$$

Ez egyébként a G/PH/N sorbanállási rendszerek egyesnűyi feltételével egyezik meg. A véges állapotterű PH/M/N rendszer állandósult állapotainak egyenletei (amikor a különböző állapotok ismétlődésének valószínűsége állandó marad):

$$\begin{aligned} \pi_0 T + \mu \pi_1 &= 0 \\ \pi_{i-1} T^\circ A^\circ + \pi_i (T - i\mu I) + (i+1)\mu \pi_{i+1} &= 0, \quad 1 \leq i \leq (N-1) \\ \pi_{i-1} T^\circ A^\circ + \pi_i (T - N\mu I) + n\mu \pi_{i+1} &= 0, \quad N \leq i \leq K+N-1 \\ \pi_{K+N-1} T^\circ A^\circ + \pi_{K+N} (T + T^\circ A^\circ - N\mu I) &= 0, \text{ és} \\ \sum_{i=0}^{K+N} \pi_i \mathbf{1} &= 1 \end{aligned} \quad (112)$$

Az $A^\circ = (1 - \alpha_{m+1})^{-1} \text{diag}(\alpha_0, \dots, \alpha_m)$.

6.5. PH/PH/1

A PH/PH/1 [4, c3.7] rendszerben a diszkrét idejű PH, $(m + 1)$ állapotú Markov-lánc felírható a következő formában:

$$P = \begin{bmatrix} T & T^\circ \\ \mathbf{0} & 0. \end{bmatrix}$$

Legyen a beérkezési folyamat eloszlása F , a kiszolgálási idők eloszlása pedig K . Az F -et írjuk fel az (α, T) formában, a K -t pedig a (β, S) formában. Az érkezési folyamat átlag jelölje λ'_1 , a kiszolgálási idők átlagát pedig μ'_1 . A PH/PH/1-et QBD-folyamatként vizsgáljuk:

$$E = \{(0, j), 1 \leq j \leq m\} \cup \{(i, j, k), i \geq 1, 1 \leq j \leq m, 1 \leq k \leq \nu\}$$

Az i a rendszerben lévő igények száma, a j az érkezési fázisfolyamat fázisa, a k a szerverek száma. A generátormátrix:

$$Q = \begin{bmatrix} T & T^\circ A^\circ \otimes \beta & & \dots \\ I \otimes S^0 & A_1 & A_0 & \dots \\ & A_2 & A_1 & A_0 & \dots \\ & & A_2 & A_1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (113)$$

és

$$\begin{aligned} A_0 &= T^\circ A^\circ \otimes I \\ A_1 &= T \otimes I + I \otimes S \\ A_2 &= I \otimes S^0 B^0. \end{aligned}$$

Az $A = A_0 + A_1 + A_2$ -alapján ($A = \sum_{n=0}^{\infty} A_n$):

$$A = (T + T^\circ A^\circ) \otimes I + I \otimes (S + S^0 B^0),$$

aminek van egy $\pi \otimes \phi$ stacionárius kezdeti valószínűségvektora (A π a $(T + T^\circ A^\circ)$, a ϕ pedig a $(S + S^0 B^0)$ kezdeti valószínűségvektora).

A rendszer akkor stabil hogy ha a beérkezési időközök átlaga nagyobb, mint a kiszolgálási idők

átlaga, azaz $\lambda'_1 > \mu'_1$. Ebből:

$$\begin{aligned}\frac{1}{\lambda'_1} &= \pi T^\circ = (\pi \otimes \phi) A_0(I \otimes I) \\ \frac{1}{\mu'_1} &= \phi S^0 = (\pi \otimes \phi) A_2(I \otimes I).\end{aligned}\tag{114}$$

A [94] alapján az R mátrix a

$$R^2(I \otimes S^0 B^0) + R(T \otimes I + I \otimes S) + T^\circ A^\circ \otimes I = 0\tag{115}$$

megoldása. A Q generátormátrixhoz tartozó z valószínűségvektor felírható

$$z = [z_0, z_1, z_1 R, z_1 R^2, \dots]\tag{116}$$

A z_1 mérete m , a z_2 mérete $m\nu$. A z_1 -et és z_2 -öt az alábbi egyenletekből kapjuk meg (kihasználjuk, hogy a Q oszlopainak összege 0):

$$\begin{aligned}z_0 T + z_1(i \otimes S^\circ) &= 0 \\ z_0(T^\circ A^\circ \otimes \beta) + z_1(T \otimes I + I \otimes S + R(I \otimes S^\circ B^\circ)) &= 0 \\ z_0 e + z_1(I - R)^{-1} e &= 1\end{aligned}\tag{117}$$

Ha $z_0 T^\circ = z_1(e \otimes S^\circ)$, akkor

$$z_0(T^\circ A^\circ \otimes \beta) = (z_0 T^\circ)(\alpha \otimes \beta) = z_1(A^{\circ\circ} \otimes S^\circ B^\circ)\tag{118}$$

ahol $T^\circ \alpha = T^\circ A^\circ$, és $\mathbf{1}\alpha = A^{\circ\circ}$.

6.6. PH/PH/N

A PH/PH/N [7, c3] rendszerben legyen az érkezési folyamat a $T = (\alpha, T)$, a kiszolgálási idők folyamata pedig $S = (\beta, S)$. A kezdeti eloszlásvektorok sorra: $\alpha = [0|\alpha^T]$ és $\beta = [0|\beta^T]$. Tegyük fel, hogy mindkét folyamat irreducibilis. Az érkezési folyamat generátormátrixa a [16] alapján:

$$Q^{0T} = \begin{bmatrix} r^T & Q^T \\ 0 & \mathbf{0} \end{bmatrix},\tag{119}$$

és ugyanígy a kiszolgálási folyamat generátormátrixa pedig

$$Q^{0S} = \begin{bmatrix} r^S & Q^S \\ 0 & \mathbf{0} \end{bmatrix}.\tag{120}$$

Minden pillanatban, amikor beérkezik egy új igény a rendszerbe, tegyük fel, hogy elindul egy újabb és újabb Markov-lánc, ami végigfut az összes állapotán, amíg el nem jut a saját elnyelő állapotába. Az érkezési folyamatot tekintsük egy ilyen Markov-láncnak. Ezt jelöljük Q^T -vel. A feldolgozási folyamatot képzeljük el hasonlóan és ezt a markov láncot pedig jelöljük Q^S -sel, és ha éppen nulla igényt dolgozunk fel, akkor a feldolgozási folyamat fázisa legyen 0. A két Markov láncot a (a^T, Q^T) illetve a (a^S, Q^S) alakban írjuk fel. A rendszerállapotot $(N+2)$ -essel jellemezzük:

$$(n, j, i_1, \dots, i_N), \quad (121)$$

ahol n igény van a rendszerben, amelyek vagy várakoznak, vagy éppen kiszolgálás alatt vannak, az érkezési folyamat a j fázisban van, és végül az l . kiszolgálóegység ($l = 1, 2, \dots, N$) az i_l fázisban van. Az S állapotterű PH/PH/N folyamat generátormátrixa legyen

$$Q = \begin{bmatrix} B_0 & C_0 & & & & \\ A_1 & B_1 & C_1 & & & \\ \vdots & \vdots & \vdots & & & \\ & & A_{c-1} & B_{c-1} & C_{c-1} & \\ & & & A_c & B_c & C_c \\ & & & & A & B & C \dots \end{bmatrix} \quad (122)$$

Az A_n, B_n, C_n blokkmátrixok az az átmenetek aránya az S_n, S_{n-1}, S_{n+1} állapotokban. Az A, B, C részmátrixok is az átmenetek arányai, A az $n \geq (N+1)$, B, C az $n \geq N$ esetben (az N a kiszolgálóegységek száma a PH/PH/N formulából).

$$\begin{aligned} A &= 0^T \oplus (r^S a^S) \oplus \dots \oplus (r^S a^S) \\ B &= Q^T \oplus Q^S \oplus \dots \oplus Q^S \\ C &= (r^T a^T) \oplus 0^S \oplus \dots \oplus 0^S \end{aligned} \quad (123)$$

Az A, B, C egyenleteiben a kronecker összegeket ($\oplus$) $(N+1)$ -szer kell elvégezni. A 0^T és a 0^S nullákat tartalmazó (zéró) négyzetes mátrixok, a rendjük a Q^{0T} és a Q^{0S} -nek megfelelő. A [91] alapján a PH/PH/N folyamat kezdeti valószínűségvektora, π , akkor a már ismert

$$\begin{aligned} \pi Q &= 0 \\ \pi \mathbf{1} &= 1 \end{aligned} \quad (124)$$

A $\pi_i = (n, j, i_1, \dots, i_N)$ az egy oszlopvektor az S_n állapotokhoz társítva. Ha a Q pozitív rekurrens, a $\pi_n, n \geq N$ esetén

$$\pi_n = \pi_N R^{n-N}, n \geq N \quad (125)$$

akkor az R a minimális nemnegatív megoldása a

$$R^2 A + RB + C = 0 \quad (126)$$

egyenletnek.

6.6.1. Teljesítménymutatók, feldolgozóegységek állapotai

[5, c2,Page198] A stabilitás feltétele, a szokásos: a beérkezéseknél $(\frac{1}{\lambda})$ nagyobb intenzitással történjen a kiszolgálás $(\frac{1}{\mu})$, azaz

$$\rho = \frac{\lambda}{N\mu} < 1. \quad (127)$$

A rendszerállapotot egy $(N + 2)$ méretű egységekbe lehet leírni:

$$S = (k, i_0, i_1, \dots, i_N) \quad (128)$$

amikor k darab igény van a rendszerben, az érkezési folyamat az i_0 állapotban van, és az N -ből a j -ik kiszolgáló egység a i_j fázisban van. Az egyszerűség kedvéért ezt jelöljük (k, i) -vel. Az i állapotokat a k igények mellett írjuk fel az L_k alakban. A legkevesebb, mint k igény esetében az állapotok halmaza $\mathcal{N}_k = \bigcup_{n=k}^{\infty} L_n$. Definiáljuk a következő átmenetet (k, i) -ből (l, j) -be:

$$p_{(k,i)(l,j)}^m = P(X(t) = (l, j), X(u) \in \mathcal{N}_m, 0 \leq u \leq t | X(0) = (k, i)). \quad (129)$$

és $(k, i), (l, j) \in \mathcal{N}_m$.

Észrevétel: ha $(k, i) = (l, j)$ -vel, akkor $p_{(k,i)(l,j)}^m(0) = 1$, különben 0.

Ha $(k, i), (l, j) \in \mathcal{N}_1$, akkor

$$q_{(k,i)(l,j)}^1(t) = P(X(t) = (l, j) | X(0) = (k, i); X(u) \in \mathcal{N}_1, 0 \leq u \leq t) \quad (130)$$

és akkor annak valószínűsége, hogy a folyamat (k, i) állapotból t idő alatt (l, j) -be jut úgy, hogy legalább egy kiszolgálóegység foglalt:

$$q_{(k,i)(l,j)}^1(t) = \frac{p_{(k,i)(l,j)}^1(t)}{\sum_{(a,b) \in \mathcal{N}_1} p_{(k,i)(a,b)}^1(t)} \quad (131)$$

Ha pedig azt nézzük hogy (k, i) -ből t idő alatt (l, j) -be jutunk úgy, hogy minden (N) darab kiszolgálóegység foglalt:

$$q_{(l,j)}^N = \lim_{t \rightarrow \infty} \frac{p_{(k,i)(l,j)}^N(t)}{\sum_{(a,b) \in \mathcal{N}_N} p_{(k,i)(a,b)}^N(t)} \quad (132)$$

6.7. MAP/PH/1

A MAP/PH/1 rendszerben [2, c9.1.1] az érkezési- és a kiszolgálási folyamat háttér markov-folyamtként jellemezhető, a $(N(t), J(t))$ QBD folyamaton keresztül nézünk rá a rendszerre. A kvázi születési-halálozási folyamat $J(t)$ folyamata egyszerre jelenti az érkezési és a kiszolgálási folyamatok fázisfolyamatát. A QBD folyamat fázisfolyamata a két folyamat fázisainak descartes-szorzata. Az érkezési- és a kiszolgálási folyamatok független működését leíró markov-láncot a kronecker-operátorral tudjuk kifejezni.

Az érkezési folyamat legyen MAP (D_0, D_1) folyamat, a kiszolgálás pedig PH (τ, T) folyamat, ahol a $t = -T\mathbf{1}$.

$$Q = \begin{bmatrix} D_0 & D_1 \otimes \tau & & & \\ I \otimes t & D_0 \oplus T & D_1 \otimes I & & \\ & I \otimes t\tau & D_0 \oplus T & D_1 \otimes I & \\ & & I \otimes t\tau & D_0 \oplus T & D_1 \otimes I \end{bmatrix} \quad (133)$$

Az ismert $A = F + B + L$ formulából

$$\begin{aligned} F &= D_0 \otimes I \\ L &= D_0 \otimes I + I \otimes T = D_0 \otimes T \\ B &= I \otimes t\tau \end{aligned}$$

és az irreguláris rész:

$$\begin{aligned} F' &= D_1 \otimes \tau \\ L' &= D_0 \\ B' &= I \otimes t. \end{aligned}$$

6.8. MAP/PH/N

A MAP/PH/N rendszerre [6] tekintsünk úgy, mint egy $\{r(t), t \geq 0\}$ sztochasztikus folyamatra, amelyről azt gondoljuk, hogy egy irreducibilis folytonos idejű markov-lánc Q generátormátrixszal és $\{1, \dots, R\}$ állapottérrel. A [6] alapvetően egy "retrial" típusú modelltől értekezik, ahol ha foglalt kiszolgálóval találkozunk az igény, akkor egy úgynevezett "orbit" várakozási sorba kerül. A modellt megfelelő átalakítottam, és így le lehet írni a MAP/PH/N "sima" sort. Az érkezési folyamat egy MAP folyamat legyen $D^r = D_0^r + D_1^r$, a kezdeti valószínűségvektora pedig α^r , $[1, \dots, W]$ állapottérrel. A kiszolgálási idők PH eloszlásúak $[1, \dots, M]$ állapottérrel, S^r generátormátrixszal és β kezdeti valószínűségvektorral, a kiszolgálási idők eloszlása

$$B^r(x) = 1 - \beta^r e^{S^r x} \mathbf{1}. \quad (134)$$

A rendszert írjuk le egy Q_A kis generátormátrixszal:

$$Q_A = \begin{bmatrix} A_{0,0} & A_{0,1} & \cdots & & & \\ A_{1,0} & A_{1,1} & A_{1,2} & \cdots & & \\ & A_{2,1} & A_{2,2} & A_{2,3} & \cdots & \\ & & \vdots & & & \\ & & & & A_{N-1,N_2} & A_{N-1,N-1} & A_{N-1,N} \\ & & & & & A_{N,N-1} & A_{N,N} \end{bmatrix} \quad (135)$$

A Q_A generátormátrix blokkjait a következőkben érthetjük meg:

A Q_A diagonálisa "alatti" blokkok:

$$A_{i,i-1} = \text{diag}(a_1, a_2, \dots, a_n) \quad (136)$$

az a_n blokkok számítási módja:

$$\begin{aligned} a_1 &= (\mathcal{G}_0^{(L-i,i)})\mathcal{U}^{(L-i,i)} \\ a_2 &= (\mathcal{G}_0^{(L-i-1,i)})\mathcal{U}^{(L-i-1,i)} \\ &\dots = \dots \\ a_n &= (\mathcal{G}_0^{(N-i,i)})\mathcal{U}^{(N-i,i)} \end{aligned} \quad (137)$$

ahol i végigfut az egész számokon N -ig: $i = 1, \dots, N$. Az a_n blokkokkal kifejezve a Q diagonálisa "alatti" blokkokat:

$$A_{i,i-1} = \text{diag}(a_1, a_2, \dots, a_n). \quad (138)$$

Az $A_{i,i-1}$ blokkjait tartalmazó a_n -ek kiszámolásához definiálnunk kell néhány további blokkot. A kiszolgálási idők (S^r, β) eloszlásában a mátrix geometrikus eloszlásokból ismert módon írjuk fel S_0 -át:

$$S_0^{(r)} = -S^{(r)}\mathbf{1}. \quad (139)$$

Akkor a [137] ben leírt G_0 meghatározása:

$$\mathcal{G}_0^{(L-i-j,i)} = \text{diag}(I_{W^{L-i-j}V} \otimes (S_0^{(r)})^{\oplus i}) \quad (140)$$

és $i = \{1, \dots, N\}, j = \{1, \dots, L - N - 1, L - N\}, j + i > 0$. Az I az egyesekből álló mátrix, a dimenziója pedig az indexeiben van: W a MAP-folyamat állapotainak száma, a $V = 1$. Az N a kiszolgálóegységek száma, az L az igényeket szolgáltató folyamatok (véges) száma. Az $\mathcal{U}$ blokk a beérkezési MAP folyamat kezdeti eloszlásvektorából származik az alábbi módon:

$$\mathcal{U}^{L-i-j,i} = \text{diag}(I_W^{L-i-j} \otimes \alpha^{(r)} \otimes I_{V^{M^i}}), \quad (141)$$

ahol az r végigmegy R -ig, ami a kiszolgálási folyamat β kezdeti valószínűségvektorának dimenziója: $r = \{1, \dots, R\}$, az i a szerverek számán fut végig: $i = \{1, \dots, N\}$, a j pedig a rendszert tápláló és a feldolgozó folyamatok különbségén: $j = \{0, 1, \dots, L - N - 1, L - N\}$, illetve $i + j > 0$. A Q_A generátormátrix $A_{i,i+1}$ blokkjait a következőképpen kell megadni:

$$A_{i,i+1} = \begin{bmatrix} A_{a,i}(0) & & & & & \\ A_{b,i}(1) & A_{a,i}(1) & & & & \\ & A_{b,i}(2) & A_{a,i}(2) & & & \\ & & \vdots & \ddots & & \\ & & & A_{b,i}(L - N - 1) & A_{a,i}(L - N - 1) & \\ & & & & A_{b,i}(L - N) & A_{a,i}(L - N) \end{bmatrix} \quad (142)$$

Az $A_{a,i}$ és $A_{b,i}$ blokkok definíciói:

$$\begin{aligned} A_{a,i} &= \mathcal{D}_1^{(L-i-j,i)} \mathcal{B}^{(L-i-j-1,i)}, j = \{0, 1, \dots, L - N - 1, L - N\} \\ A_{b,i} &= \Lambda_j^{(L-i-j,i)} \mathcal{B}^{(L-i-j,i)}, j = \{1, \dots, L - N - 1, L - N\} \end{aligned} \quad (143)$$

Az i végigfut a kiszolgáló egységeken: $i = \{0, 1, \dots, N - 1\}$. A fenti, további blokkok definíciói: A MAP érkezési folyamat (D_0, D_1) blokkjai alapján

$$\mathcal{D}_1^{(L-i-j,i)} = \text{diag}((D_1^{(r)} e_W)^{\oplus(L-j-i)} \otimes I_{VM^i}) \quad (144)$$

A β kiszolgálási idő kezdeti valószínűségvektor alapján

$$\mathcal{B}^{(L-i-j)} = \text{diag}(I_{W^{(L-i-j)}VM^i} \otimes \beta^{(r)}), r = \{1, \dots, R\} \quad (145)$$

és a α MAP folyamat kezdeti eloszlásvektor alapján a

$$\Lambda^{(L-i,j)} = \text{diag}(\alpha_j^{(r)} I_{W^{L-i-j}VM^i}), r = \{1, \dots, R\} \quad (146)$$

és $i = \{0, 1, \dots, N - 1\}$ valamint $j = \{0, 1, \dots, L - N - 1, L - N\}$.

Az $A_{0,0}$ blokk:

$$\begin{aligned} A_{0,0} &= \text{diag}(\mathcal{D}_0^{(L,0)} + \mathcal{Q}^{(L,0)} + \mathcal{D}_0^{(L,0)} + \mathcal{Q}^{(N,0)} + \mathcal{D}_0^{(L-1,0)} \\ &\quad + \mathcal{Q}^{(L-1,0)} - \Lambda_1^{(L-1,0)} + \dots + \mathcal{D}_0^{(N,0)} + \mathcal{Q}^{(N,0)} - \Lambda_{L-N}^{(N,0)}) \end{aligned} \quad (147)$$

A $\mathcal{D}^{(L-i-j,i)} = Q \otimes I_{W^{L-i-j}VM^i}$, ahol $i = \{0, 1, \dots, N\}$ és $j = \{0, 1, \dots, L-N-1, L-N\}$.
Már csak a Q_A diagonálisa hiányzik:

$$A_{N,N} = \begin{bmatrix} S(N,0) & T(N,0) & & & \\ & S(N,1) & T(N,1) & & \\ & & \vdots & \ddots & \\ & & & S(N,L-N-1) & T(N,L-N-1) \\ & & & & S(N,L-N) \end{bmatrix} \quad (148)$$

A diagonálishoz definiáljuk az alábbi blokkokat:

$$S(N, J) = \mathcal{G}^{(L-N-j,M)} + \mathcal{D}_0^{(L-N-j,N)} + \mathcal{Q}^{(L-N-j,N)} \quad (149)$$

ahol a $j = \{0, 1, \dots, L-N-1, L-N\}$ és

$$T(N, J) = \mathcal{D}_1^{(L-N-j,N)} \quad (150)$$

és a $j = \{0, 1, \dots, L-N-1\}$.

6.8.1. Teljesítménymutatók, feldolgozóegységek állapotai

A [6, c5] alapján Q_A generátormátrix alapján felírható a stacionárius eloszlás:

$$\begin{aligned} \Pi_r &= \Pi_{r+1} A_{r+1,r} (-C_r^{-1}), \quad 0 \leq r \leq (N-1) \\ \Pi_N C_N &= 0 \end{aligned} \quad (151)$$

A képletben szereplő C a Q_A blokkjai alapján:

$$\begin{aligned} C_0 &= A_{0,0} \\ C_n &= A_{n,n} + A_{n,n-1} (-C_{n-1}^{-1}) A_{n-1,n}, \quad 1 \leq n \leq N \end{aligned} \quad (152)$$

Legyen $P(i, j)$ annak valószínűsége, hogy i darab kiszolgálóegység foglalt, és j igény várakozik. Akkor

$$P(i, j) = \Pi_i e_{[RW^{(VM^i j+1, RW^{(L-i-j)} VM^i (j+1))]} \quad (153)$$

Ebből következnek:

Annak valószínűsége, hogy i foglalt kiszolgálóegység van:

$$P(i) = \sum_{j=0}^{L-N} P(i, j) = \Pi_i e \quad (154)$$

Annak valószínűsége, hogy j igény várakozik:

$$P(j) = \sum_{i=0}^N P(i, j) = \sum_{i=0}^N \Pi_i e_{[RW(VM^i j+1, RW^{(L-i-j)} VM^i(j+1))]} \quad (155)$$

ahol $j = \{0, 1, \dots, L - N - 1, L - N\}$.

Az éppen foglalt szerverek átlagos száma:

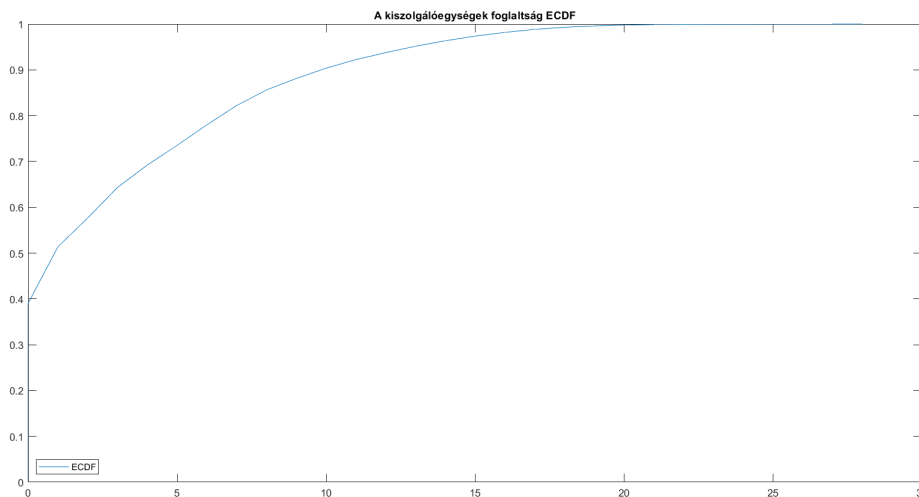
$$L_b = \sum_{r=1}^N r \Pi_r e. \quad (156)$$

7. Esettanulmány

A bevezetőben leírtak szerint a PH/PH/N és MAP/PH/N modellekkel hasonlítom össze az eredeti IT rendszer adatait. Mindkettő modellben rögzített, N -számú feldolgozó egység feltételezt. Az IT rendszerben ténylegesen $N = 40$ kiszolgálóegység van, de a mért adatokban az látszik, hogy sosem éri el ezt a számot. Ennek több oka is van:

- Akkor kerülhet a (sorbanállási) rendszerbe egy igény, ha minden előzménye megvan (azaz minden előzmény *job* ki lett szolgálva a rendszerben),
- És ha az IT rendszernek van elegendő CPU, memória és disk erőforrása, hogy új igény feldolgozásába kezdjen.

Vagyis a PH/PH/M modell esetében PH/PH/M(t)-ről beszélhetünk, fizikailag ugyanakkor létezik az N feldolgozóegység fizikai korlát. Ennek tükrében nézzük meg a feldolgozó egységek igénybevételének eloszlását, pontosabban az empirikus kumulált eloszlásfüggvényét (ECDF):



4. ábra. Kiszolgáló egységek foglaltsága, ECDF

. Az $..../N$ modellek rögzített számú kiszolgálót feltételez, az IT rendszerben ez dinamikusán változik. Azt a megközelítést választom, hogy

- a modelljeimben $N := E[J(t)]$, vagyis a kiszolgáló egységek számának várható értékén rögzítem az N -t, így készítek szimulációkat,
- vagy pedig nem az N -et rögzítem, hanem vagy a várakozási eloszlását és a $J(t)$ -t vizsgálom.

A fenti modelleket összevetve kaphatok képet arról, hogy az adott (PH/PH/N, vagy MAP/PH/N) modellemmel sikeresen tudom-e modellezni az eredeti IT-rendszer lefutását. Az egyes esetek vizsgálatához felhasználok a szabadon elérhető <http://webspn.hit.bme.hu/~telek/tools/butools/> csomagot. A csomag lehetővé teszi egyebek között PH és MAP eloszlások vizsgálatát, mind *Python*, mind *MATLAB* környezetben. A fejezetben a következőkben felsorolt modelleket vizsgálom úgy, hogy

- veszem az Adattárház érkezési- és feldolgozási sorait,
- majd az akutálisan vizsgált modell érkezési- és várakozási sorainak eloszlásai alapján generálok egy hasonló méretű mintát,
- a vizsgált modell típusa szerint szimulálok egy új folyamatot, és
- a szimulált folyamatot összevetem a mért Adattárház adatokkal, így validálok az éppen vizsgált modell alkalmasságát az Adattárház rendszer modellezésére.

A vizsgálat során a következő, a bevezetőben már megfogalmazott kérdéseket teszem fel:

- Milyen a feldolgozóegységek állapota (hogyan kell méreteznem a fizikai rendszert)
- Mennyi idő alatt fut le az IT-rendszer.

7.1. Vizsgálat PH/PH/N rendszerben

A beérkezési sort először PH eloszlással közelítem. A közelítés során 6 Erlang-eloszlás konvolúciójával közelítem a mért érkezési adatokat. A butools-nevű szoftvereszközzel vizsgálva, a beérkezési folyamat generátormátrixa:

$$Q_A = \begin{bmatrix} -0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & -0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & -0.07 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & -0.07 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & -1.16 & 1.16 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -1.16 \end{bmatrix},$$

ami a [62] alapján hiperexponenciális, és a hozzá tartozó kezdeti eloszlásvektor:

$$\pi_A = [0.01 \quad 0.06 \quad 0.04 \quad 0.07 \quad 0.82 \quad 0.00].$$

A kiszolgálási idők eloszlásának generátormátrixa, hasonló paraméterekkel:

$$Q_B = \begin{bmatrix} -0.50 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & -2.50 & 2.50 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & -2.50 & 2.50 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & -2.50 & 2.50 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & -2.50 & 2.50 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -2.50 \end{bmatrix}, \quad (157)$$

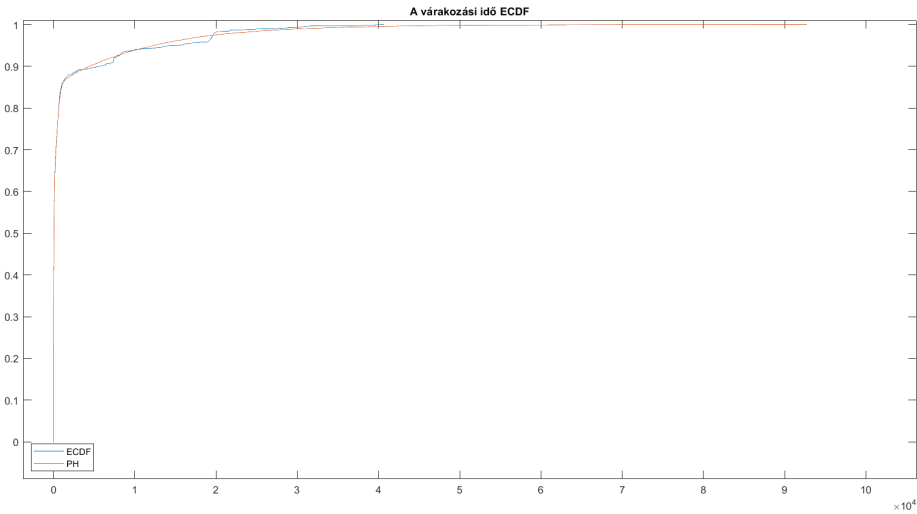
és a kiszolgálási PH-folyamathoz tartozó kezdeti eloszlásvektor, ami a [63] alapján a Coxian eloszlás.

$$\pi_B = \begin{bmatrix} 0.01 & 0.06 & 0.04 & 0.07 & 0.82 & 0.00 \end{bmatrix}.$$

7.1.1. Rögzített várakozási idő eloszlás mellett a feldolgozóegységek vizsgálata

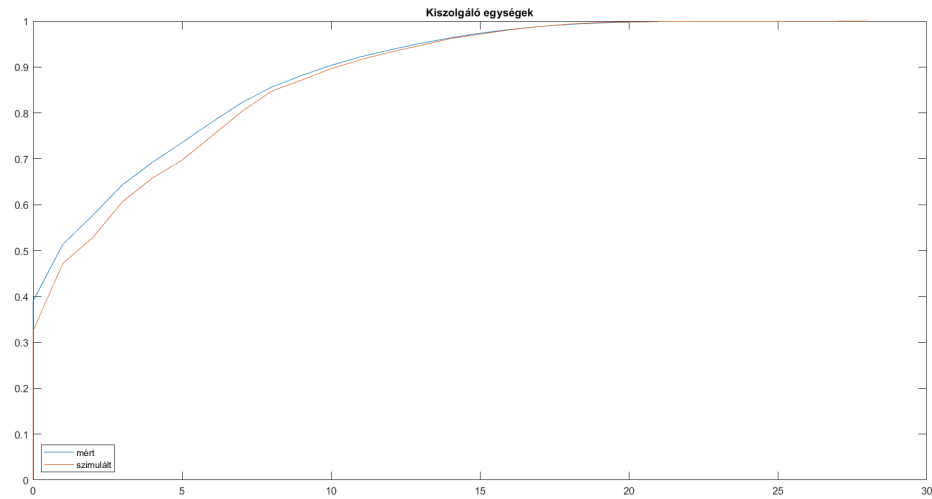
A fenti $F_A = (\pi_A, Q)$ és $F_B = (\pi_B, Q_B)$ PH eloszlások alapján a *butools*-al generálok egy új érkezési és kiszolgálási sort. A várakozási idők a mért adatok eloszlása szerint szintén generáltak. A IT rendszerben Az Adattárházban az egyes igények a [1.1] struktúrában találhatók.

Az Adattárházban a várakozási idők eloszlása, és annak közelítése PH eloszlással:



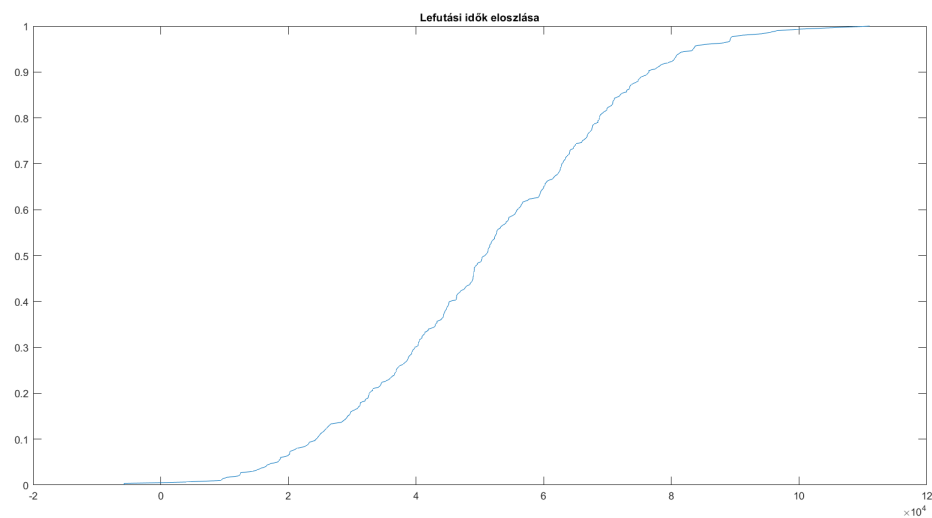
5. ábra. Várakozás eloszlásfüggvénye

A rendszerben a kiszolgáló egységek foglaltságainak, $J(t)$ -nek az eloszlása összehasonlítva egy kisebb méretű, szimulált mintával:



6. ábra. Kiszolgáló egységek foglaltsága, ECDF

. A lefutások eloszlása

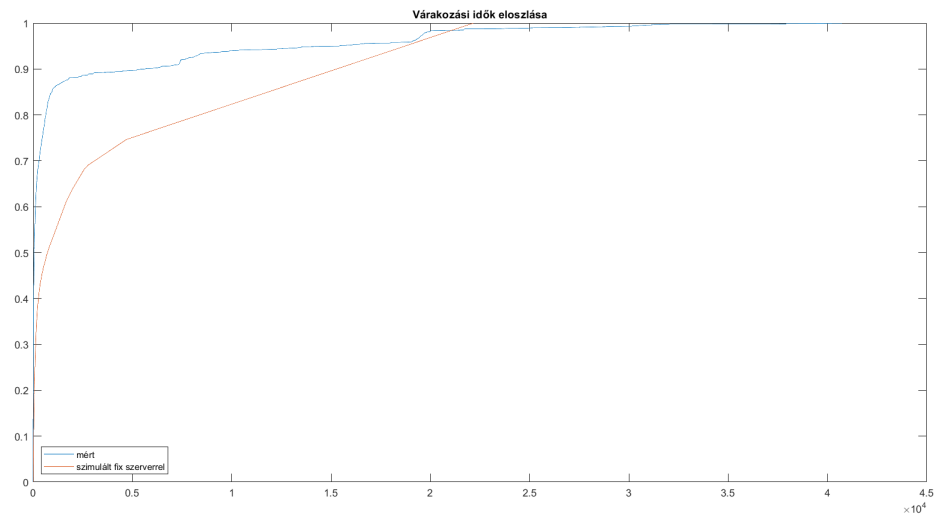


7. ábra. Rendszer lefutási idők eloszlása

ami $E[T] = 49901$ várható értékű, $D(x) = 20691$ szórású normális eloszlás. Ez az érték nem számít jónak, csupán 75% pontossággal követi az eredeti rendszeradatokat.

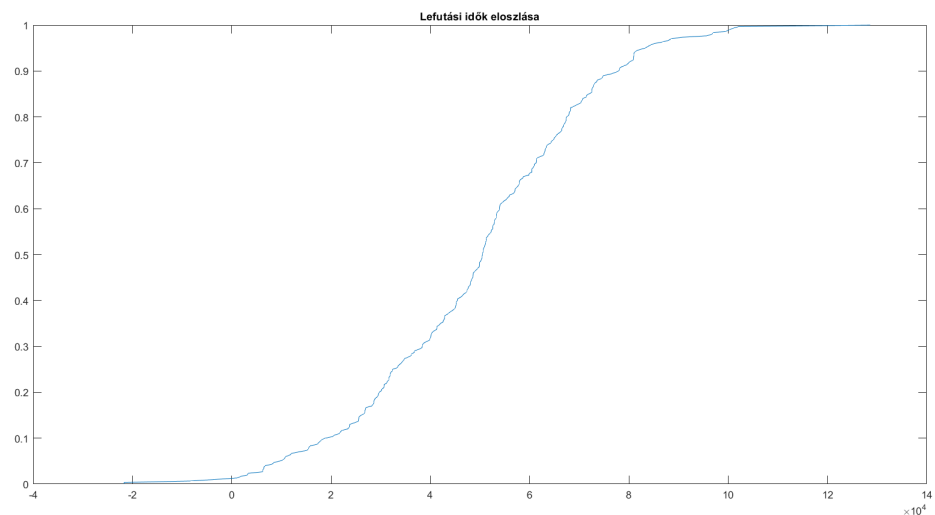
7.1.2. Rögzített feldolgozóegységek melletti vizsgálat

Az $N = E(J(t))$, ahol legyen $J(t)$ a t időpillanatban használt feldolgozók száma, és akkor N darab, rögzített számú kiszolgálóegységgel végeztem ezt a szimulációt. Így a tényleges PH/PH/N modellt tudom vizsgálni. A szimuláció lefuttatása után a kapott várakozási idők eloszlása:



8. ábra. Várakozási idők eloszlása

. A rögzített számú szerveres modell futtatása után az látszik, hogy a várakozási idők eloszlása nem követi jól az eredeti, mért adatokat. A szimulált lefutási idők eloszlása



9. ábra. Lefutási idők eloszlása

. ami egy $E[X] = 49709.375$ várható értékű, $D(X) = 22239.479$ szórású normélis eloszlású v.v.. Ez szintén nem pontos közelítése az eredeti adatoknak.

7.2. Vizsgálat MAP/PH/N rendszerben

A *butools*-al a beérkezési folyamatot a következő paraméterű (D_0, D_1) MAP folyamattal lehet közelíteni:

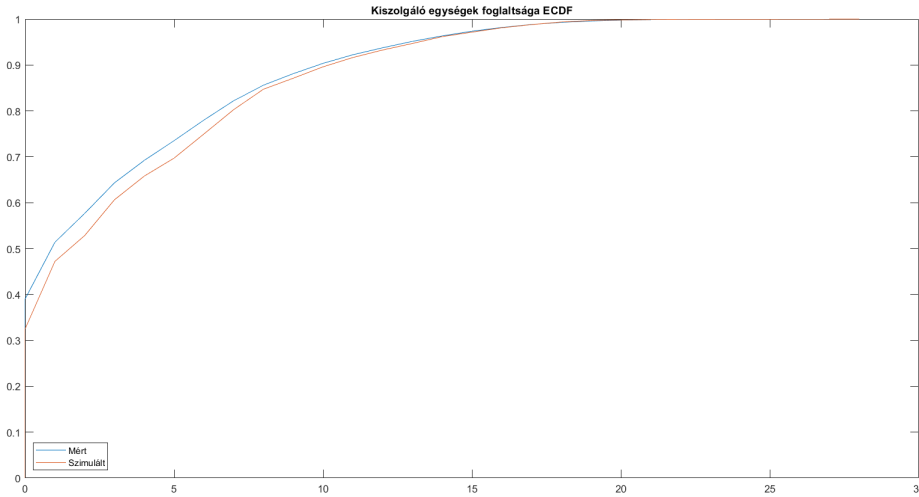
$$D_0 = \begin{bmatrix} -0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & -0.03 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & -0.39 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & -0.63 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & -0.63 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -0.63 \end{bmatrix}$$

, és

$$D_1 = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.02 & 0.01 & 0.00 & 0.00 & 0.00 \\ 0.01 & 0.06 & 0.32 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.01 & 0.16 & 0.21 & 0.25 \\ 0.00 & 0.00 & 0.01 & 0.16 & 0.21 & 0.25 \\ 0.00 & 0.00 & 0.01 & 0.16 & 0.21 & 0.25 \end{bmatrix}$$

. A kiszolgálási idők megegyeznek a [157]-el. A vizsgálat egyik tárgya, a rendszer átlagos lefutási ideje $T_{teljes} = 36600$ másodperc.

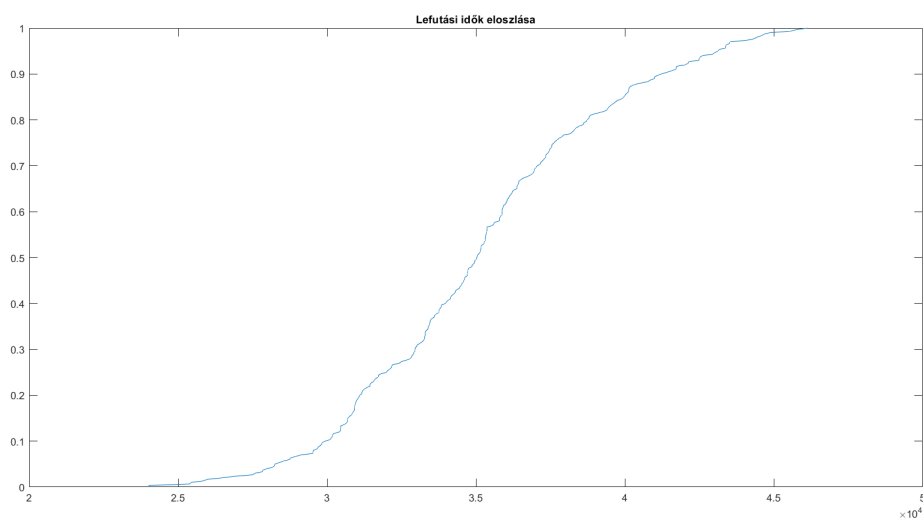
7.2.1. Rögzített várakozási idő eloszlás mellett a feldolgozóegységek vizsgálata



10. ábra. Feldolgozóegysége kihasználtságának eloszlásfüggvénye

A szimulációval generált rendszer lefutási idők várható értéke $E[T_{szimult}] = 35396$, azaz a szimuláció $P(T_{szim} < T_{teljes}) = 0.9671$ valószínűséggel követi az éles IT rendszer lefutásait. A

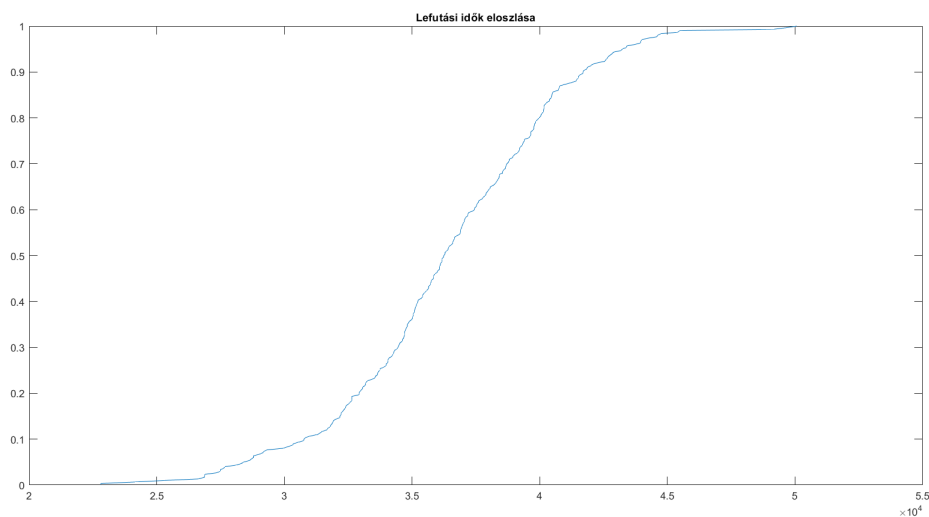
szimulált lefutások eloszlása



11. ábra. Rendszer lefutási idők eloszlása

ami egy $E[X] = 35396$ várható értékű, $D(X) = 4447.7957$ szórású normál eloszlás. Ez az adat 96%-os pontosságot jelent.

7.2.2. Rögzített feldolgozóegységek melletti vizsgálat



12. ábra. Rendszer lefutási idők eloszlása

ami egy $E[X] = 36369$ várható értékű, $D(X) = 4382.11$ szórású normál eloszlás, ami ez esetben 99%-os (!!!) pontosságot jelent.

7.3. További vizsgálati lehetőségek

A PH/PH/M(t) / MAP/PH/M(t) sorbanállási rendszerek további vizsgálatában a "server can take vacation" [10] modell irányában indulnék el. A modell szerint egy kiszolgáló egység "vakációra mehet", azaz átmenetileg kiesik, és helyette, átmeneti időre egy tartalék szerver próbál "beállni". Az ilyen, a példában egyszerveres rendszer állapotát a következő idősorokkal tudjuk leírni:

- $N(t)$ a t időpillanatban a rendszerben lévő igények száma,
- $J_1(t)$ a szerver státusza,
- $J_2(t)$ a kiszolgálás fázisa,
- $J_3(t)$ pedig a beérkezési sor fázisa.

A $J_1(t)$ értékei a következők lehetnek:

- $J_1(t) = 0$, amennyiben a szerver foglalt,
- $J_1(t) = 1$ abban az esetben, hogy ha a szerver "vakációra ment"
- $J_1(t) = 2$ pedig akkor, ha a szerver javítás alatt áll.

Ezt a modellt úgy lehet használni az IT rendszer modellezésére, hogy a szerver kiesésre nem elromlásként tekintek, hogy egy szerver elromlik, hanem hogy költséget rendelek az egyes igényekhez, és azzal számolok, hogy az N szerver egy időben csak bizonyos összköltséget tud kiszolgálni.

Ehhez a be kell vezetni egy $C(J(t), N(t))$ (az $J(t)$ legyen a kiszolgálás fázisa a t időpillanatban) költségfüggvényt, ami megmutatja, hogy a t időpillanatban, a rendszerben kiszolgálás alatt álló $N(t)$ igénynek milyen költsége van, és ez hogy viszonyul a $C_{\max}$ egyidőben maximálisan kiszolgálható költséghez.

Numerikusan ezt a fenti [7.1.1] rögzített várakozási idők mellett vizsgált modellel is el tudom végezni. A rendszer pontos vizsgálatához azonban le kell írni a rendszert leíró, folytonos idejű $\{N(t), J_1(t), J_2(t), J_3(t)\}$ modell Q generátormátrixát és π kezdeti valószínűségvektorát, illetve ugyanezt a többszerveres modellre.

8. Összefoglalás

A dolgozatomban egy létező banki informatikai rendszer, az Adattárház működését vizsgáltam sztochasztikus módszerekkel. A Banki üzemidő indulásáig, azaz a fiókhálózat nyitásáig, elő kell állítani az aznapi értékesítéshez szükséges adatokat, illetve a Magyar Nemzeti Bank irányába határidőre el kell küldeni az előző napi kötelező jelentéseket, ezért kritikus és folyamatosan jelenlévő kérdés az, hogy mennyi idő alatt fut le a rendszerben lévő feldolgozási folyamat. Ezzel együtt az is állandó kérdés - és ez minden banki program kapcsán igaz -, hogy hogyan érdemes méreteznünk a rendszerhez szükséges erőforrásokat (CPU-kat).

A rendszer vizsgálata során ezért azokra a kérdésekre kerestem a választ, hogy milyen modellel tudom megmondani azt, hogy az mennyi idő alatt fut le, és a rendszer alatt lévő infrastruktúrát (feldolgozó egységek) hogyan méretezzük. A rendszer lefutásának vizsgálatához a sorbanállás elméletet használtam, azon belül a PH/PH/N és MAP/PH/N típusú sorbanállási modelleket. A dolgozatban bemutattam a modelleket, majd az Adattárház rendszerből kinyert, mért tényadatok alapján szimulációt végeztem. A kezdeti kérdésekre, miszerint megbecsülhető-e, hogy mennyi idő alatt fut le az Adattárház rendszer, milyen feldolgozóegység igényei vannak, azt a választ kaptam, hogy a kérdésekre a modelljeimen alapuló szimulációk kielégítő választ adnak, és a MAP/PH/N modell pontosabb. Összességében, a dolgozatban a sorelmélet egy modelljét egy gyakorlati alkalmazásban próbáltam ki, és ez a kísérlet kedvező eredményt adott.

A vizsgálat során felmerült, hogy a jövőben a [PH/MAP]/PH/N modellemet ki lehet egészíteni [PH/MAP]/PH/N(t) modellre, ahol az $N(t)$ a t időben változó a feldolgozóegységek száma, amely függ a rendszerben éppen kiszolgálás alatt lévő igények "súlyától".

9. Summary

I examined the operation of an existing banking IT system, the Data Warehouse, with stochastic methods in my diploma thesis. Until the start of the Bank's operating hours, i.e. the branch network is opened, the data required for the same day sales must be generated, and the previous day's mandatory reports must be sent to the Magyar Nemzeti Bank by deadline, therefore a critical and ongoing question is how long the processing process in the system takes. However, the constant question - and this is true of all banking programs - is how to scale the resources (CPUs) needed for the system. When examining the system, I therefore looked for answers to the questions of what model I can use to tell how long it will run and how to scale the infrastructure (processing units) under the system. I used the queuing theory to examine the operation of the system, including the PH/PH /N and MAP/PH/N queuing models. In the thesis I presented the models, then I performed a simulation based on the measured factual data extracted from the Data Warehouse system. To the initial questions about how long it takes to estimate how long the Data Warehouse system runs, what processing unit needs it was, the simulations based on my models provide satisfactory answers, and the MAP / PH / N model is more accurate. Overall in the thesis I tested a model of row theory in a practical application, and this experiment yielded favorable results. The study suggested that in the future my [PH/MAP]/PH/N model could be supplemented with a [PH/MAP]/PH/N(t) model, where $N(t)$ is t the number of processing units varies over time, depending on the "weight" of needs currently being served in the system.

Hivatkozások

- [1] Peter Buchholz, Jan Kriege, Iryna Felko (2014) *Input modelling with Phase-Type Distributions and Markov Models, Theory and applications*, Springer.
- [2] Lakatos Szeidl Telek (2010) *Introduction to Queueing Systems with Telecommunication Applications*, Springer.
- [3] Nick T. Thomopoulos (2012) *Fundamentals of Queueing Systems, Statistical Methods for Analyzing Queueing Models*, Springer
- [4] Marcel F. Neuts (1981) *Matrix-Geometric Solutions in Stochastic Models, An Algorithmic Approach*, The Johns Hopkins University Press Baltimore and Londo
- [5] Naoki Makimoto (1993) *Quasi-stationary distributions in a PH/PH/c queue*, Communications in Statistics, Stochastic Models, 9:2, 195-212, DOI: 10.1080/15326349308807262
- [6] Jinbiao Wu, Zaiming Liu, Gang Yang (2011) *Analysis of the finite source MAP/PH/N retrial G-queue operating in a random environment*, Applied Mathematical Modelling 35 (2011) 1184–1193
- [7] Yukio Takahashi (1981) *Asymptotic Exponentiality of the Tail of the Waiting-Time Distribution in a PH/PH/c Queue*, Advances in Applied Probability, Vol. 13, No. 3 (Sep., 1981), pp. 619-630
- [8] Feller, William (1968) *An Introduction to Probability Theory and its Applications (volume I).*, John Wiley & Sons.
- [9] <https://math.jhu.edu/~brown/courses/s18/Lectureess/421Week5S15.pdf>, last seen at 2022.05.10
- [10] William J. Gray, Pu Patrik Wang, Meckinley Scott *A vacation queueing model with service breakdowns*, Applied Mathematical Modelling 24 (2000) 391