3. Little-formula, kiszolgálási idő elemzése

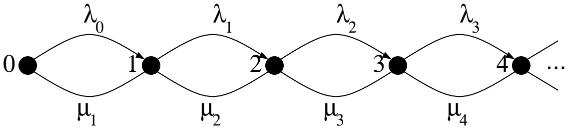
Kommunikációs hálózatok teljesítményének elemzése

Horváth Illés

2024/09/18

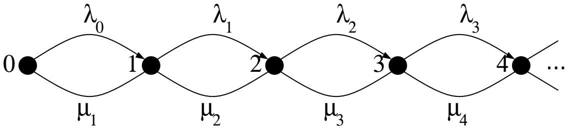
Markov-sorok, M/M/1 és M/M/1/K sor

Emlékeztető: Markov-sor.

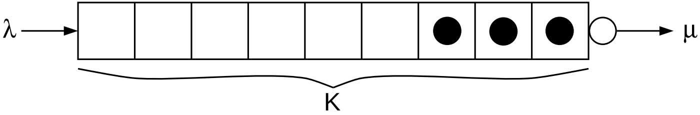


Markov-sorok, M/M/1 és M/M/1/K sor

Emlékeztető: Markov-sor.



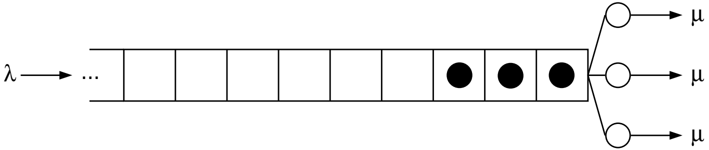
Azon belül M/M/1 és M/M/1/K sorok: 1 szerver, FIFO (érkezési sorrend szerinti) kiszolgálás.



M/M/c és M/M/c/K sor

További érdekes sorok.

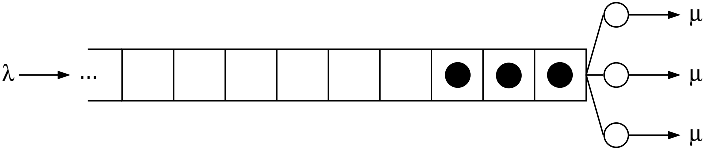
Az M/M/c sorban c szerver van és végtelen buffer:



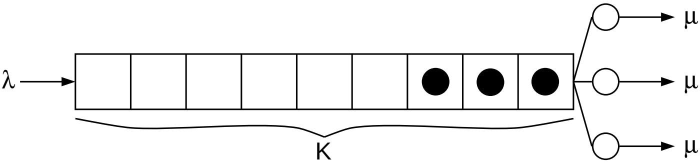
M/M/c és M/M/c/K sor

További érdekes sorok.

Az M/M/c sorban c szerver van és végtelen buffer:



Az M/M/c/K sorban pedig véges, K hosszú buffer:



Várakozási idő

Egy kiszolgáló sor teljesítményét általában az ügyfelek/igények szemszögéből mérjük.

Várakozási idő

Egy kiszolgáló sor teljesítményét általában az ügyfelek/igények szemszögéből mérjük.

Ennek megfelelően sorbanállási problémáknál a tipikusan használt teljesítményjellemző a *kiszolgálási idő* vagy *rendszerben töltött idő*, ami egy igénynek a sorba való beérkezés pillanatától a kiszolgálás végéig eltelt idő. Ez általában véletlenszerű; érdekes a teljes eloszlása is, de ha egyetlen számmal kell jellemezni, akkor az általában a várható értéke.

Várakozási idő

Egy kiszolgáló sor teljesítményét általában az ügyfelek/igények szemszögéből mérjük.

Ennek megfelelően sorbanállási problémáknál a tipikusan használt teljesítményjellemző a *kiszolgálási idő* vagy *rendszerben töltött idő*, ami egy igénynek a sorba való beérkezés pillanatától a kiszolgálás végéig eltelt idő. Ez általában véletlenszerű; érdekes a teljes eloszlása is, de ha egyetlen számmal kell jellemezni, akkor az általában a várható értéke.

Ezt általában stacionárius rendszerekre vizsgáljuk (azaz feltesszük, hogy a rendszer hosszú ideje fut).

Várakozási idő

Egy kiszolgáló sor teljesítményét általában az ügyfelek/igények szemszögéből mérjük.

Ennek megfelelően sorbanállási problémáknál a tipikusan használt teljesítményjellemző a *kiszolgálási idő* vagy *rendszerben töltött idő*, ami egy igénynek a sorba való beérkezés pillanatától a kiszolgálás végéig eltelt idő. Ez általában véletlenszerű; érdekes a teljes eloszlása is, de ha egyetlen számmal kell jellemezni, akkor az általában a várható értéke.

Ezt általában stacionárius rendszerekre vizsgáljuk (azaz feltesszük, hogy a rendszer hosszú ideje fut).

A korábban tárgyaltak (pl. a sorhossz stacionárius eloszlása) az átlagos várakozási időről közvetlenül nem adnak információt. Szerencsére az *átlagos* várakozási időre van egy egyszerű képlet.

Little-formula

Tétel. (Little-formula)

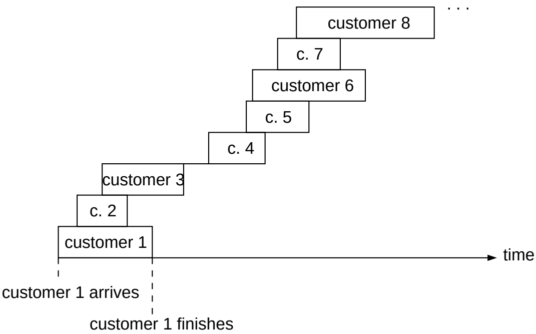
$$L = \lambda_e W,$$

ahol

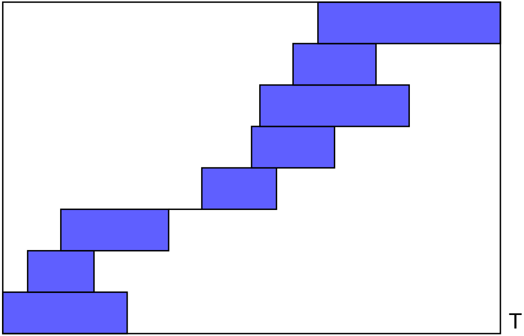
- *W az átlagos várakozási idő értéke.*
- *λ_e az effektív érkezési ráta (véges buffer esetén a teli buffer miatt elveszett érkezések nem számítanak bele),*
- *L az átlagos sorhossz.*

Little-formula

Biz. (vázlat) Ábrázoljuk az igényeket a következő módon (az érkezés ideje szerint rendezve):

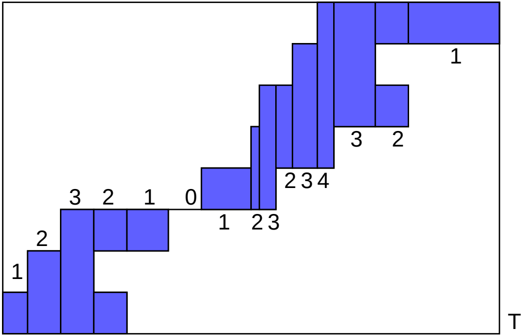


Little-formula



Nagy T esetén a téglalapok száma T-ig $\lambda_e T + o(T)$ a NSzT alapján, egy téglalap területe pedig átlagosan W, tehát a kék terület összesen $B = \lambda_e TW + o(T)$.

Little-formula



Másrészt a kék területet T-vel osztva éppen az átlagos sorhosszt kapjuk meg:

$$L = \frac{B}{T} = \frac{\lambda_e T W}{T} = \lambda_e W.$$

Little-formula

Megjegyzések. Véges állapottér esetén a teli buffer miatt elveszett érkezések nem számítanak bele λ_e-be.

Little-formula

Megjegyzések. Véges állapottér esetén a teli buffer miatt elveszett érkezések nem számítanak bele λ_e-be.

λ_e kiszámítható a $v_{\mathrm{st}} = (x_0\, x_1\, \ldots)$ stacionárius eloszlásból:

$$\lambda_e = \lambda_0 x_0 + \lambda_1 x_1 + \cdots + \lambda_K x_K,$$

speciálisan ha az érkezési ráta konstans λ, akkor ez a következő alakban is írható:

$$\lambda_e = \lambda(1 - x_K).$$

Little-formula

Megjegyzések. Véges állapottér esetén a teli buffer miatt elveszett érkezések nem számítanak bele λ_e-be.

λ_e kiszámítható a $v_{\mathrm{st}} = (x_0\, x_1\, \dots)$ stacionárius eloszlásból:

$$\lambda_e = \lambda_0 x_0 + \lambda_1 x_1 + \dots + \lambda_K x_K,$$

speciálisan ha az érkezési ráta konstans λ, akkor ez a következő alakban is írható:

$$\lambda_e = \lambda(1 - x_K).$$

Végtelen, stabil sorra pedig, ha az érkezési ráta konstans λ, akkor

$$\lambda_e = \lambda.$$

Little-formula

Megjegyzések. Véges állapottér esetén a teli buffer miatt elveszett érkezések nem számítanak bele λ_e-be.

λ_e kiszámítható a $v_{st} = (x_0\ x_1\ \dots)$ stacionárius eloszlásból:

$$\lambda_e = \lambda_0 x_0 + \lambda_1 x_1 + \dots + \lambda_K x_K,$$

speciálisan ha az érkezési ráta konstans λ, akkor ez a következő alakban is írható:

$$\lambda_e = \lambda(1 - x_K).$$

Végtelen, stabil sorra pedig, ha az érkezési ráta konstans λ, akkor

$$\lambda_e = \lambda.$$

Az átlagos sorhossz L is megkapható a stacionárius eloszlásból az ergodtétel révén:

$$L = 0 \cdot x_0 + 1 \cdot x_1 + 2 \cdot x_2 + \dots$$

Little-formula

Mi most Markov-sorokra használjuk, de a Little-formula gyakorlatilag tetszőleges sorbanállási rendszerre teljesül, például a következő esetekben:

- ha például az érkezési időköz vagy a kiszolgálási idő nem exponenciális eloszlású (ilyenkor a sor nem is Markov-lánc);
- ha a rendszer nem beérkezési sorrendben szolgálja ki az igényeket;
- ha több sorból áll a rendszer (sőt, ilyenkor igaz külön-külön az egyes sorokra és globálisan az egész rendszerre is).

Little-formula

Mi most Markov-sorokra használjuk, de a Little-formula gyakorlatilag tetszőleges sorbanállási rendszerre teljesül, például a következő esetekben:

- ha például az érkezési időköz vagy a kiszolgálási idő nem exponenciális eloszlású (ilyenkor a sor nem is Markov-lánc);
- ha a rendszer nem beérkezési sorrendben szolgálja ki az igényeket;
- ha több sorból áll a rendszer (sőt, ilyenkor igaz külön-külön az egyes sorokra és globálisan az egész rendszerre is).

Viszont a Little-formula csak az összes beérkező igényre kiátlagolt várakozási időt számítja ki. Következőnek megnézzük, hogy mi a helyzet, ha ennél részletesebb információra van szükség.

Konvolúció

Valószínűségszámítás előkészületekkel kezdünk.

Legyenek X és Y nemnegatív folytonos eloszlású valószínűségi változók, sűrűségfüggvényeik $f_X, f_Y : [0, \infty) \to [0, \infty)$, eloszlásfüggvényeik $F_X, F_Y : [0, \infty) \to [0, 1]$.

Tétel.

Legyen $Z = X + Y$. Ha X és Y függetlenek, akkor Z sűrűségfüggvénye

$$f_Z(x) = \int_0^x f_Y(y) f_X(x - y) \mathrm{d}y,$$

eloszlásfüggvénye

$$F_Z(x) = \int_0^x F_Y(y) f_X(x - y) \mathrm{d}y = \int_0^x f_Y(y) F_X(x - y) \mathrm{d}y.$$

Z eloszlását úgy hívjuk, hogy X és Y eloszlásának a konvolúciója.

Exponenciális eloszlás

Az $X \sim \mathrm{EXP}(\lambda)$ (λ paraméterű exponenciális) eloszlás sűrűségfüggvénye

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

eloszlásfüggvénye

$$F(x) = 1 - e^{-\lambda x} \qquad (x \geq 0).$$

A várható értéke $\mathbb{E}(X) = \frac{1}{\lambda}$.

Erlang eloszlás

Ha $X_1, \ldots, X_n \sim \mathrm{EXP}(\lambda)$ függetlenek, akkor

$$Y = X_1 + \cdots + X_n$$

eloszlását Erlang eloszlásnak nevezzük n és λ paraméterekkel (röviden $Y \sim \mathrm{Erlang}(n, \lambda)$); a sűrűségfüggvénye

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

eloszlásfüggvénye

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

várható értéke $\mathbb{E}(Y) = \frac{n}{\lambda}$.

Erlang eloszlás

Ha $X_1, \ldots, X_n \sim \mathrm{EXP}(\lambda)$ függetlenek, akkor

$$Y = X_1 + \cdots + X_n$$

eloszlását Erlang eloszlásnak nevezzük n és λ paraméterekkel (röviden $Y \sim \mathrm{Erlang}(n, \lambda)$); a sűrűségfüggvénye

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

eloszlásfüggvénye

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

várható értéke $\mathbb{E}(Y) = \frac{n}{\lambda}$.

Speciálisan $\mathrm{Erlang}(1, \lambda) = \mathrm{EXP}(\lambda)$.

Geometriai eloszlás és pesszimista geometriai eloszlás

Az, hogy hányadik független érmedobásra jön ki először fej egy olyan érmével, aminél p a fej valószínűsége, geometriai eloszlású: $N \sim \mathrm{GEO}(p)$ $(0 < p < 1)$:

$$P(N = n) = p(1-p)^{n-1} \qquad n = 1, 2, \ldots$$

Geometriai eloszlás és pesszimista geometriai eloszlás

Az, hogy hányadik független érmedobásra jön ki először fej egy olyan érmével, aminél p a fej valószínűsége, geometriai eloszlású: $N \sim \mathrm{GEO}(p)$ $(0 < p < 1)$:

$$P(N = n) = p(1 - p)^{n-1} \qquad n = 1, 2, \ldots$$

A pesszimista geometriai eloszlás ugyanez, csak a fej dobást nem számoljuk. $M \sim \mathrm{PGEO}(p)$:

$$P(M = n) = p(1 - p)^{n} \qquad n = 0, 1, \ldots$$

Geometriai eloszlás és pesszimista geometriai eloszlás

Az, hogy hányadik független érmedobásra jön ki először fej egy olyan érmével, aminél p a fej valószínűsége, geometriai eloszlású: $N \sim \mathrm{GEO}(p)$ $(0 < p < 1)$:

$$P(N = n) = p(1-p)^{n-1} \qquad n = 1, 2, \dots$$

A pesszimista geometriai eloszlás ugyanez, csak a fej dobást nem számoljuk. $M \sim \mathrm{PGEO}(p)$:

$$P(M = n) = p(1-p)^{n} \qquad n = 0, 1, \dots$$

A pesszimista geometriai eloszlás "eggyel kevesebb", mint a geometriai eloszlás:

$$N \sim \mathrm{GEO}(p) \quad \Longleftrightarrow \quad N - 1 \sim \mathrm{PGEO}(p)$$

Geometriai eloszlás és pesszimista geometriai eloszlás

A geometriai és pesszimista geometriai eloszlások memóriamentesek.

Lemma

(a) *Ha* $N \sim GEO(p)$, *akkor*

$$P(N > m + n | N > m) = P(N > n) \qquad \forall m, n \in \mathbb{Z}^+.$$

(b) *Ha* N *eloszlására teljesül az előbbi tulajdonság, akkor* $N \sim GEO(p)$ *valamilyen* p*-re.*

(c) *Ha* $M \sim PGEO(p)$, *akkor*

$$P(M \geq m + n | M \geq m) = P(M \geq n) \qquad \forall m, n \in \mathbb{Z}^+.$$

(d) *Ha* M *eloszlására teljesül az előbbi tulajdonság, akkor* $M \sim PGEO(p)$ *valamilyen* p*-re.*

Csonkolt geometriai eloszlás

A csonkolt (truncated) geometriai eloszlás a geometriai eloszlás azon feltétel mellett, hogy az értéke legfeljebb K.
$N \sim \mathrm{TGEO}(K, p)$:

$$P(N = n) = \frac{p(1-p)^{n-1}}{1-(1-p)^K} \qquad n = 1, 2, \ldots, K.$$

Csonkolt geometriai eloszlás

A csonkolt (truncated) geometriai eloszlás a geometriai eloszlás azon feltétel mellett, hogy az értéke legfeljebb K.
$N \sim \mathrm{TGEO}(K, p)$:

$$P(N = n) = \frac{p(1-p)^{n-1}}{1-(1-p)^K} \qquad n = 1, 2, \ldots, K.$$

A csonkolt pesszimista geometriai eloszlás a pesszimista geometriai eloszlás azon feltétel mellett, hogy az értéke legfeljebb K.
$N \sim \mathrm{TPGEO}(K, p)$:

$$P(N = n) = \frac{p(1-p)^{n}}{1-(1-p)^{K+1}} \qquad n = 0, 1, \ldots, K.$$

Kapcsolat a geometriai és exponenciális eloszlás között

Lemma

Ha $X_1, \ldots, X_n \sim EXP(\lambda)$ *függetlenek, és* $N \sim GEO(p)$ *tőlük független, akkor*

$$Y = X_1 + \cdots + X_N$$

eloszlása

$$Y \sim EXP(p\lambda).$$

Kapcsolat a geometriai és exponenciális eloszlás között

Lemma

Ha $X_1, \dots, X_n \sim EXP(\lambda)$ függetlenek, és $N \sim GEO(p)$ tőlük független, akkor

$$Y = X_1 + \cdots + X_N$$

eloszlása

$$Y \sim EXP(p\lambda).$$

Biz. Azon feltétel mellett, hogy $N = n$, Y feltételes eloszlása Erlang(k, λ). Teljes valószínűség tételét használjuk Y sűrűségfüggvényére:

$$f_Y(x) = \sum_{n=1}^{\infty} p(1-p)^{n-1} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} = p\lambda e^{-\lambda x} \sum_{n=1}^{\infty} \frac{(\lambda(1-p)x)^{n-1}}{(n-1)!} =$$
$$= p\lambda e^{-\lambda x} e^{\lambda(1-p)x} = p\lambda e^{-p\lambda x},$$

ami éppen EXP$(p\lambda)$ sűrűségfüggvénye.

M/M/1 sor kiszolgálási idő eloszlása

Tekintsünk egy stacionárius M/M/1 sort. Az érkezési ráta λ, a kiszolgálási ráta μ, $\lambda < \mu$. A sor terhelése

$$\rho = \lambda/\mu < 1.$$

Tétel.

(a) *Az M/M/1 sor stacionárius sorhossz eloszlása PGEO$(1-\rho)$.*

(b) *Feltéve, hogy egy igény a sorban a k-adik helyre érkezik (vagyis beérkezéskor $k-1$ igény van előtte), a kiszolgálási idő eloszlása Erlang(k, λ).*

(c) *Egy véletlen igény várakozási ideje EXP$((1-\rho)\mu)$.*

M/M/1 sor kiszolgálási idő eloszlása

Biz.

(a) A dinamikus egyensúly egyenletekből

$$\mu x_{i+1} = \lambda x_i \quad \Longrightarrow \quad x_{i+1} = \frac{\lambda}{\mu} x_i = \rho x_i,$$

és

$$x_n = \rho^n \cdot (1 - \rho),$$

ami éppen a PGEO$(1 - \rho)$ eloszlás.

M/M/1 sor kiszolgálási idő eloszlása

Biz.

(a) A dinamikus egyensúly egyenletekből

$$\mu x_{i+1} = \lambda x_i \quad \Longrightarrow \quad x_{i+1} = \frac{\lambda}{\mu} x_i = \rho x_i,$$

és

$$x_n = \rho^n \cdot (1 - \rho),$$

ami éppen a PGEO$(1 - \rho)$ eloszlás.

(b) Az aktuális igény távozásához ki kell szolgálni az előtte lévő $k - 1$ igényt és az aktuális igényt is; mindegyiknek a kiszolgálási ideje EXP(λ) eloszlású, és k darab EXP(λ) eloszlás konvolúciója éppen az Erlang(k, λ) eloszlás.

M/M/1 sor kiszolgálási idő eloszlása

(c) Itt annyi különbség van a (b) részhez képest, hogy a beérkező igény egy véletlen hosszú sorba érkezik.

Kérdés: milyen sorhossz-eloszlást lát a beérkező igény?

M/M/1 sor kiszolgálási idő eloszlása

(c) Itt annyi különbség van a (b) részhez képest, hogy a beérkező igény egy véletlen hosszú sorba érkezik.

Kérdés: milyen sorhossz-eloszlást lát a beérkező igény?

A sorhossz stacionárius eloszlását látja a beérkező igény is.

M/M/1 sor kiszolgálási idő eloszlása

(c) Itt annyi különbség van a (b) részhez képest, hogy a beérkező igény egy véletlen hosszú sorba érkezik.

Kérdés: milyen sorhossz-eloszlást lát a beérkező igény?

A sorhossz stacionárius eloszlását látja a beérkező igény is.

A sorhossz stacionárius eloszlása PGEO$(1-\rho)$, a beérkező igénnyel együtt az igények száma eggyel nagyobb, azaz GEO$(1-\rho)$ eloszlású, és ennyi darab EXP(μ) eloszlás konvolúciója az előző lemma szerint éppen EXP$((1-\rho)\mu)$ eloszlású.

M/M/1/K sor kiszolgálási idő eloszlása

Tétel.

(a) *Az M/M/1/K sor stacionárius sorhossz eloszlása* $PTGEO(1-\rho)$.

(b) *Feltéve, hogy egy igény a sorban a k-adik helyre érkezik (vagyis beérkezéskor* $k-1$ *igény van előtte), a kiszolgálási idő eloszlása* $\mathrm{Erlang}(k,\lambda)$.

(c) *Egy, a rendszerbe bekerülő véletlen igény várakozási ideje* $\mathrm{Erlang}(k,\lambda)$ *eloszlások keveréke* $PGEO(1-\rho)$ *eloszlás szerinti súlyokkal.*

Nem biz., de ugyanaz, mint az M/M/1, csak a véges buffer miatt mindenhol a geometriai eloszlás csonkolt változata szerepel.

M/M/1/K sor kiszolgálási idő eloszlása

Tétel.

(a) *Az M/M/1/K sor stacionárius sorhossz eloszlása* $PTGEO(1-\rho)$.

(b) *Feltéve, hogy egy igény a sorban a k-adik helyre érkezik (vagyis beérkezéskor* $k-1$ *igény van előtte), a kiszolgálási idő eloszlása* $Erlang(k,\lambda)$.

(c) *Egy, a rendszerbe bekerülő véletlen igény várakozási ideje* $Erlang(k,\lambda)$ *eloszlások keveréke* $PGEO(1-\rho)$ *eloszlás szerinti súlyokkal.*

Nem biz., de ugyanaz, mint az M/M/1, csak a véges buffer miatt mindenhol a geometriai eloszlás csonkolt változata szerepel.

Fontos: véges buffer esetén az igények egy része elvész.

M/M/c és M/M/c/K sorok kiszolgálási idő eloszlása

Az M/M/c és M/M/c/K sorokra (ahol 1 helyett c szerver szolgálja ki az igényeket) hasonlóan számolható a kiszolgálási idő, mint eddig, és Erlang eloszlások megfelelő keveréke lesz. Az M/M/c és M/M/c/K sorok stacionárius eloszlása sajnos nem pontosan geometriai eloszlás, és a képletek valamivel rondábbak, de ugyanúgy expliciten számolható minden. Ezt most nem tesszük meg.

M/M/c és M/M/c/K sorok kiszolgálási idő eloszlása

Az M/M/c és M/M/c/K sorokra (ahol 1 helyett c szerver szolgálja ki az igényeket) hasonlóan számolható a kiszolgálási idő, mint eddig, és Erlang eloszlások megfelelő keveréke lesz. Az M/M/c és M/M/c/K sorok stacionárius eloszlása sajnos nem pontosan geometriai eloszlás, és a képletek valamivel rondábbak, de ugyanúgy expliciten számolható minden. Ezt most nem tesszük meg.

Egy fontos közös tulajdonság ezekben a sorokban, hogy ha egy igény bekerült a rendszerbe, akkor a kiszolgálási idejét már nem befolyásolja az, ami mögötte történik a sorban. Ez más Markov-sorokra nem teljesül, és nem Erlang eloszlások keveréke lesz a kiszolgálási idő.

M/M/c és M/M/c/K sorok kiszolgálási idő eloszlása

Ami szintén közös az M/M/1, M/M/1/K, M/M/c, M/M/c/K sorokban, hogy egy véletlenszerű beérkező igény a stacionárius sorhossz-eloszlást látja.

M/M/c és M/M/c/K sorok kiszolgálási idő eloszlása

Ami szintén közös az M/M/1, M/M/1/K, M/M/c, M/M/c/K sorokban, hogy egy véletlenszerű beérkező igény a stacionárius sorhossz-eloszlást látja.

Ez más Markov-sorokra is teljesül, egészen pontosan abban az esetben, ha az érkezési ráta konstans λ minden állapotra. Ez PASTA-elv (Poisson Arrival Sees Time Average) vagy Érkezési tétel (Arrival theorem) néven is ismert.

M/M/c és M/M/c/K sorok kiszolgálási idő eloszlása

Ami szintén közös az M/M/1, M/M/1/K, M/M/c, M/M/c/K sorokban, hogy egy véletlenszerű beérkező igény a stacionárius sorhossz-eloszlást látja.

Ez más Markov-sorokra is teljesül, egészen pontosan abban az esetben, ha az érkezési ráta konstans λ minden állapotra. Ez PASTA-elv (Poisson Arrival Sees Time Average) vagy Érkezési tétel (Arrival theorem) néven is ismert.

A "Poisson Arrival Sees Time Average" megfogalmazás mögött az van, hogy még olyan (nem-markovi) sorokban is teljesül, ahol a kiszolgálási idő nem exponenciális eloszlású. Ilyenkor a rendszer nem folytonos idejű Markov-lánc, és nincs stacionárius eloszlás, hanem helyette hosszú távú időátlag (time average) szerepel - de az teljesül, hogy a véletlenszerű beérkező igény pont ezt az eloszlást látja.

M/M/c és M/M/c/K sorok kiszolgálási idő eloszlása

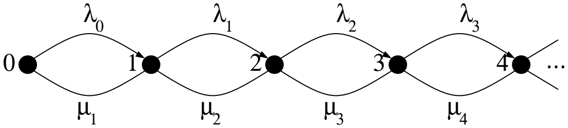
Ami szintén közös az M/M/1, M/M/1/K, M/M/c, M/M/c/K sorokban, hogy egy véletlenszerű beérkező igény a stacionárius sorhossz-eloszlást látja.

Ez más Markov-sorokra is teljesül, egészen pontosan abban az esetben, ha az érkezési ráta konstans λ minden állapotra. Ez PASTA-elv (Poisson Arrival Sees Time Average) vagy Érkezési tétel (Arrival theorem) néven is ismert.

A "Poisson Arrival Sees Time Average" megfogalmazás mögött az van, hogy még olyan (nem-markovi) sorokban is teljesül, ahol a kiszolgálási idő nem exponenciális eloszlású. Ilyenkor a rendszer nem folytonos idejű Markov-lánc, és nincs stacionárius eloszlás, hanem helyette hosszú távú időátlag (time average) szerepel - de az teljesül, hogy a véletlenszerű beérkező igény pont ezt az eloszlást látja.

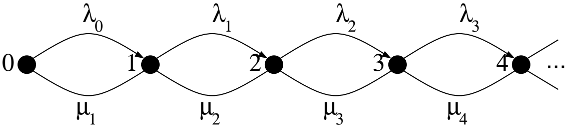
Nem biz., de lényegében azon múlik, hogy a beérkező igények "függetlenek" a sor hosszától.

Általános Markov-sorok



Egy általános Markov-sorban minden állapotban eltérő lehet az érkezési és a kiszolgálási ráta.

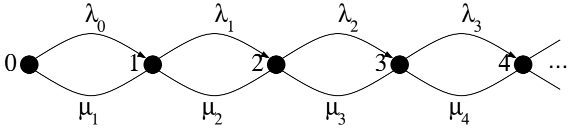
Általános Markov-sorok



Egy általános Markov-sorban minden állapotban eltérő lehet az érkezési és a kiszolgálási ráta.

Eltérő kiszolgálási ráta lehet például amiatt, hogy a szerver (részben) párhuzamosítani tudja az igényeket: több igény esetén a teljes kiszolgálási ráta valamivel nagyobb. Ilyenkor általában a kiszolgálás sem FIFO, hanem valamilyen megosztás történik több igény között. Ezt meg fogjuk vizsgálni.

Általános Markov-sorok



Egy általános Markov-sorban minden állapotban eltérő lehet az érkezési és a kiszolgálási ráta.

Eltérő kiszolgálási ráta lehet például amiatt, hogy a szerver (részben) párhuzamosítani tudja az igényeket: több igény esetén a teljes kiszolgálási ráta valamivel nagyobb. Ilyenkor általában a kiszolgálás sem FIFO, hanem valamilyen megosztás történik több igény között. Ezt meg fogjuk vizsgálni.

Eltérő érkezési ráta lehet például amiatt, hogy a vizsgált sor egy nagyobb rendszer része, ahol egy diszpécser terheléselosztást végez - ha hosszabb a sor, az érkezéseket máshová irányítja. Ezt is meg fogjuk nézni (de nem ma).

Laplace-transzformált

Előkészület: Laplace-transzformáció.

Laplace-transzformált

Előkészület: Laplace-transzformáció.

Egy $f : [0, \infty) \to \mathbb{R}$ függvény Laplace-transzformáltja

$$f^*(s) = \int_0^\infty e^{-st} f(t) \mathrm{d}t.$$

Laplace-transzformált

Előkészület: Laplace-transzformáció.

Egy $f : [0, \infty) \to \mathbb{R}$ függvény Laplace-transzformáltja

$$f^*(s) = \int_0^\infty e^{-st} f(t) \mathrm{d}t.$$

Egy $X \geq 0$ folytonos valószínűségi változó Laplace-transzformáltja

$$X^*(s) = E\left(e^{-sX}\right) = \int_0^\infty e^{-st} f(t) \mathrm{d}t,$$

ahol $f(t)$ az X sűrűségfüggvénye.

Laplace-transzformált

Előkészület: Laplace-transzformáció.

Egy $f : [0, \infty) \to \mathbb{R}$ függvény Laplace-transzformáltja

$$f^*(s) = \int_0^\infty e^{-st} f(t) \mathrm{d}t.$$

Egy $X \geq 0$ folytonos valószínűségi változó Laplace-transzformáltja

$$X^*(s) = E\left(e^{-sX}\right) = \int_0^\infty e^{-st} f(t) \mathrm{d}t,$$

ahol $f(t)$ az X sűrűségfüggvénye.

Ha f sűrűségfüggvény, akkor az integrál konvergens minden $\mathrm{Re}(z) \geq 0$ komplex számra.

Tulajdonságok

Példa. Az $X \sim \mathrm{EXP}(\lambda)$ eloszlás Laplace-transzformáltja

$$X^*(s) = \int_0^\infty e^{-st}\lambda e^{-\lambda t}\mathrm{d}t = \lambda \int_0^\infty e^{-(\lambda+s)t}\mathrm{d}s = \frac{\lambda}{\lambda+s}.$$

Tulajdonságok

Példa. Az $X \sim \mathrm{EXP}(\lambda)$ eloszlás Laplace-transzformáltja

$$X^*(s) = \int_0^\infty e^{-st}\lambda e^{-\lambda t}\mathrm{d}t = \lambda \int_0^\infty e^{-(\lambda+s)t}\mathrm{d}s = \frac{\lambda}{\lambda+s}.$$

Lemma (Teljes várható érték tétel Laplace-transzformáltra)

Ha X, Y, Z valószínűségi változók f_X, f_Y, f_Z sűrűségfüggvénnyel, és

$$\mathbb{P}(Z = X) = p, \qquad P(Z = Y) = 1 - p,$$

akkor

$$f_Z^*(s) = pf_X^*(s) + (1-p)f_Y^*(s).$$

Tulajdonságok

Példa. Az $X \sim \mathrm{EXP}(\lambda)$ eloszlás Laplace-transzformáltja

$$X^*(s) = \int_0^\infty e^{-st}\lambda e^{-\lambda t}\mathrm{d}t = \lambda \int_0^\infty e^{-(\lambda+s)t}\mathrm{d}s = \frac{\lambda}{\lambda+s}.$$

Lemma (Teljes várható érték tétel Laplace-transzformáltra)

Ha X, Y, Z valószínűségi változók f_X, f_Y, f_Z sűrűségfüggvénnyel, és

$$\mathbb{P}(Z = X) = p, \qquad P(Z = Y) = 1 - p,$$

akkor

$$f_Z^*(s) = pf_X^*(s) + (1-p)f_Y^*(s).$$

Biz. (vázlat) A Laplace-transzformált lineáris: ha a, b valós konstansok, akkor

$$(af + bg)^*(s) = a \cdot f^*(s) + b \cdot g^*(s).$$

Tulajdonságok

További fontos tulajdonsága, hogy konvolúcióból szorzatot csinál.

Lemma

Ha X, Y valószínűségi változók függetlenek f_X, f_Y sűrűségfüggvénnyel, és

$$Z = X + Y,$$

akkor

$$f_Z^*(s) = f_X^*(s) \cdot f_Y^*(s).$$

Tulajdonságok

További fontos tulajdonsága, hogy konvolúcióból szorzatot csinál.

Lemma

Ha X, Y valószínűségi változók függetlenek f_X, f_Y sűrűségfüggvénnyel, és

$$Z = X + Y,$$

akkor

$$f_Z^*(s) = f_X^*(s) \cdot f_Y^*(s).$$

Tulajdonságok

Biz. Ha $Z = X + Y$, akkor $f_Z = f_X \otimes f_Y$, és

$$f_Z^*(s) = \int_0^\infty f_Z(t) e^{-st} \mathrm{d}t = \int_0^\infty \left(\int_0^t f_Y(u) f_X(t-u) \mathrm{d}u \right) e^{-st} \mathrm{d}t =$$

$$= \int_0^\infty \int_u^\infty f_Y(u) f_X(t-u) e^{-st} \mathrm{d}t \mathrm{d}u.$$

Tulajdonságok

Biz. Ha $Z = X + Y$, akkor $f_Z = f_X \otimes f_Y$, és

$$f_Z^*(s) = \int_0^\infty f_Z(t)e^{-st}\mathrm{d}t = \int_0^\infty \left(\int_0^t f_Y(u)f_X(t-u)\mathrm{d}u\right) e^{-st}\mathrm{d}t =$$
$$= \int_0^\infty \int_u^\infty f_Y(u)f_X(t-u)e^{-st}\mathrm{d}t\mathrm{d}u.$$

$v = t - u$ változócserével

$$\int_0^\infty \int_u^\infty f_Y(u)f_X(t-u)e^{-st}\mathrm{d}t\mathrm{d}u =$$
$$= \int_0^\infty \int_0^\infty f_Y(u)f_X(v)e^{-s(v+u)}\mathrm{d}v\mathrm{d}u =$$
$$= \int_0^\infty f_Y(u)e^{-su}\mathrm{d}u \cdot \int_0^\infty f_X(v)e^{-sv}\mathrm{d}v =$$
$$= f_Y^*(S) \cdot f_X^*(S).$$

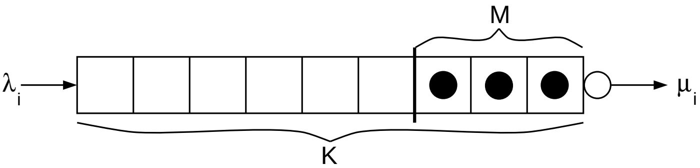
Általános Markov-sor

Tekintsünk egy általános véges bufferű Markov-sort. A buffer mérete K. Az i állapotban az érkezési ráta λ_i, a kiszolgálási ráta μ_i.

Általános Markov-sor

Tekintsünk egy általános véges bufferű Markov-sort. A buffer mérete K. Az i állapotban az érkezési ráta λ_i, a kiszolgálási ráta μ_i.

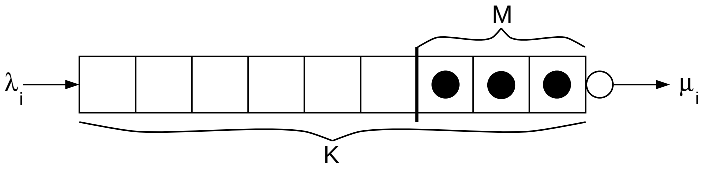
A kiszolgálási elv FIFO helyett Limited Processor Sharing (LPS): megengedjük, hogy egyszerre több igényt szolgáljon ki a szerver párhuzamosan, maximum M-et.



Általános Markov-sor

Tekintsünk egy általános véges bufferű Markov-sort. A buffer mérete K. Az i állapotban az érkezési ráta λ_i, a kiszolgálási ráta μ_i.

A kiszolgálási elv FIFO helyett Limited Processor Sharing (LPS): megengedjük, hogy egyszerre több igényt szolgáljon ki a szerver párhuzamosan, maximum M-et.

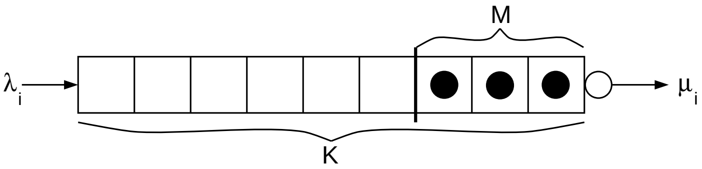


Ilyenkor, ha legfeljebb M igény van a sorban, akkor a teljes μ_i kiszolgálási ráta egyenletesen oszlik meg a sorban lévő igények között. Ha M-nél több igény van a sorban, akkor az első M igény között oszlik meg a kiszolgálási ráta, a későbbi igények várakoznak.

Általános Markov-sor

Tekintsünk egy általános véges bufferű Markov-sort. A buffer mérete K. Az i állapotban az érkezési ráta λ_i, a kiszolgálási ráta μ_i.

A kiszolgálási elv FIFO helyett Limited Processor Sharing (LPS): megengedjük, hogy egyszerre több igényt szolgáljon ki a szerver párhuzamosan, maximum M-et.



Ilyenkor, ha legfeljebb M igény van a sorban, akkor a teljes μ_i kiszolgálási ráta egyenletesen oszlik meg a sorban lévő igények között. Ha M-nél több igény van a sorban, akkor az első M igény között oszlik meg a kiszolgálási ráta, a későbbi igények várakoznak.

Speciálisan az $M = 1$ esetben visszakapjuk a FIFO kiszolgálást.

Kiszolgálási idő eloszlás általános Markov-sorra

Tekintsünk egy kitüntetett igényt egy j hosszú sorban az i-edik pozícióban ($1 \leq i \leq j \leq K$). Jelölje $X_{i,j}$ a kiszolgálási idejét.

Kiszolgálási idő eloszlás általános Markov-sorra

Tekintsünk egy kitüntetett igényt egy j hosszú sorban az i-edik pozícióban ($1 \leq i \leq j \leq K$). Jelölje $X_{i,j}$ a kiszolgálási idejét.

$X_{i,j}$ sűrűségfüggvénye $f_{i,j}(t)$, Laplace-transzformáltja $f^*_{i,j}(s)$.

Kiszolgálási idő eloszlás általános Markov-sorra

Tekintsünk egy kitüntetett igényt egy j hosszú sorban az i-edik pozícióban ($1 \leq i \leq j \leq K$). Jelölje $X_{i,j}$ a kiszolgálási idejét.

$X_{i,j}$ sűrűségfüggvénye $f_{i,j}(t)$, Laplace-transzformáltja $f^*_{i,j}(s)$.

A következő változások történhetnek a sorral:

Kiszolgálási idő eloszlás általános Markov-sorra

Tekintsünk egy kitüntetett igényt egy j hosszú sorban az i-edik pozícióban ($1 \leq i \leq j \leq K$). Jelölje $X_{i,j}$ a kiszolgálási idejét.

$X_{i,j}$ sűrűségfüggvénye $f_{i,j}(t)$, Laplace-transzformáltja $f^*_{i,j}(s)$.

A következő változások történhetnek a sorral:

- Érkezik egy igény a sor végére. Ennek rátája λ_j. Ekkor a kitüntetett igény továbbra is az i-edik pozícióban van, de a sor hossza $(j+1)$-re változik.

Kiszolgálási idő eloszlás általános Markov-sorra

Tekintsünk egy kitüntetett igényt egy j hosszú sorban az i-edik pozícióban ($1 \leq i \leq j \leq K$). Jelölje $X_{i,j}$ a kiszolgálási idejét.

$X_{i,j}$ sűrűségfüggvénye $f_{i,j}(t)$, Laplace-transzformáltja $f^*_{i,j}(s)$.

A következő változások történhetnek a sorral:

- Érkezik egy igény a sor végére. Ennek rátája λ_j. Ekkor a kitüntetett igény továbbra is az i-edik pozícióban van, de a sor hossza $(j+1)$-re változik.
- Történik egy kiszolgálás a sorban. Ennek rátája μ_j. Ha az a kitüntetett igény, akkor a kiszolgálása befejeződött. Ha egy másik igényt szolgált ki a szerver, akkor a kitüntetett igény $i > M$ esetén előrébb lép az $(i-1)$-edik pozícióba. $i \leq M$ esetén tulajdonképpen mindegy, előre lép-e, mivel az első M igény egyforma helyzetben van.

Kiszolgálási idő eloszlás általános Markov-sorra

A kétféle változás együttes rátája $\lambda_j + \mu_j$, így a következő állapotváltozásig eltelt idő $\mathrm{EXP}(\lambda_j + \mu_j)$ eloszlású (ennek Laplace-transzformáltja $\frac{\lambda_j+\mu_j}{\lambda_j+\mu_j+s}$), és a változáskor $\frac{\lambda_j}{\lambda_j+\mu_j}$ az érkezés valószínűsége és $\frac{\mu_j}{\lambda_j+\mu_j}$ a kiszolgálás valószínűsége.

Kiszolgálási idő eloszlás általános Markov-sorra

A kétféle változás együttes rátája $\lambda_j + \mu_j$, így a következő állapotváltozásig eltelt idő EXP$(\lambda_j + \mu_j)$ eloszlású (ennek Laplace-transzformáltja $\frac{\lambda_j+\mu_j}{\lambda_j+\mu_j+s}$), és a változáskor $\frac{\lambda_j}{\lambda_j+\mu_j}$ az érkezés valószínűsége és $\frac{\mu_j}{\lambda_j+\mu_j}$ a kiszolgálás valószínűsége.

Legyen először $M < i \leq j < K$. A fentiek alapján a Laplace-transzformáltra teljesül a következő egyenlet:

$$f^*_{i,j}(s) = \frac{\lambda_j + \mu_j}{\lambda_j + \mu_j + s}\left(\frac{\lambda_j}{\lambda_j + \mu_j} f^*_{i,j+1}(s) + \frac{\mu_j}{\lambda_j + \mu_j} f^*_{i-1,j-1}(s)\right)$$

Kiszolgálási idő eloszlás általános Markov-sorra

A kétféle változás együttes rátája $\lambda_j + \mu_j$, így a következő állapotváltozásig eltelt idő EXP$(\lambda_j + \mu_j)$ eloszlású (ennek Laplace-transzformáltja $\frac{\lambda_j+\mu_j}{\lambda_j+\mu_j+s}$), és a változáskor $\frac{\lambda_j}{\lambda_j+\mu_j}$ az érkezés valószínűsége és $\frac{\mu_j}{\lambda_j+\mu_j}$ a kiszolgálás valószínűsége.

Legyen először $M < i \leq j < K$. A fentiek alapján a Laplace-transzformáltra teljesül a következő egyenlet:

$$f^*_{i,j}(s) = \frac{\lambda_j + \mu_j}{\lambda_j + \mu_j + s}\left(\frac{\lambda_j}{\lambda_j + \mu_j} f^*_{i,j+1}(s) + \frac{\mu_j}{\lambda_j + \mu_j} f^*_{i-1,j-1}(s)\right)$$

A cél: felírni hasonló egyenleteket az összes $f^*_{i,j}(s)$-re, majd a kapott egyenletrendszert megoldani.

Kiszolgálási idő eloszlás általános Markov-sorra

További esetek.

Ha $i \leq j \leq M$ és $j < K$, akkor egyrészt

$$f^*_{1,j}(s) = \dots f^*_{j,j}(s),$$

hiszen az első j igény helyzete egyforma, másrészt

$$f^*_{i,j}(s) = \frac{\lambda_j + \mu_j}{\lambda_j + \mu_j + s}\left(\frac{\lambda_j}{\lambda_j + \mu_j} f^*_{i,j+1}(s) + \frac{\mu_j}{\lambda_j + \mu_j}\frac{j-1}{j} f^*_{i-1,j-1}(s) + \right.$$
$$\left. + \frac{\mu_j}{\lambda_j + \mu_j}\frac{1}{j}\cdot 1\right),$$

mivel az első j igényből $1/j$ eséllyel a kitüntetett igényt szolgálja ki a szerver, és olyankor az elhagyja a rendszert (0 további időt tölt a rendszerben, aminek a Laplace-transzformáltja 1).

Kiszolgálási idő eloszlás általános Markov-sorra

Ha $i \leq M < j < K$, akkor is

$$f_{1,j}^*(s) = \ldots f_{M,j}^*(s),$$

hiszen az első M igény helyzete egyforma, másrészt

$$f_{i,j}^*(s) = \frac{\lambda_j + \mu_j}{\lambda_j + \mu_j + s}\Big(\frac{\lambda_j}{\lambda_j + \mu_j} f_{i,j+1}^*(s) + \frac{\mu_j}{\lambda_j + \mu_j}\frac{M-1}{M} f_{i-1,j-1}^*(s) + \\ + \frac{\mu_j}{\lambda_j + \mu_j}\frac{1}{M} \cdot 1\Big),$$

mivel ilyenkor az első M igény között oszlik meg a kiszolgálási ráta, és azon belül $1/M$ eséllyel a kitüntetett igényt szolgálja ki a szerver.

Kiszolgálási idő eloszlás általános Markov-sorra

$j = K$, azaz tele buffer esetén nem történhet érkezés, csak kiszolgálás. Ha $i \leq M$, akkor is teljesül

$$f^*_{1,K}(s) = \ldots f^*_{M,K}(s),$$

továbbá

$$f^*_{i,K}(s) = \frac{\mu_j}{\mu_j + s}\Big(\frac{M-1}{M} f^*_{i-1,M-1}(s) + \frac{1}{M} \cdot 1\Big),$$

Kiszolgálási idő eloszlás általános Markov-sorra

$j = K$, azaz tele buffer esetén nem történhet érkezés, csak kiszolgálás. Ha $i \leq M$, akkor is teljesül

$$f^*_{1,K}(s) = \dots f^*_{M,K}(s),$$

továbbá

$$f^*_{i,K}(s) = \frac{\mu_j}{\mu_j + s}\left(\frac{M-1}{M} f^*_{i-1,M-1}(s) + \frac{1}{M} \cdot 1\right),$$

Hasonlóan tele buffer, csak $i > M$ esetén a kitüntetett igény biztosan nem kap kiszolgálást, így

$$f^*_{i,K}(s) = \frac{\mu_j}{\mu_j + s} f^*_{i-1,M-1}(s).$$

Inverz Laplace-transzformáció

Minden (i, j) párra $(1 \leq j \leq j \leq K)$ felírva az előző egyenleteket, kapunk egy egyenletrendszert, ami lineáris az $f_{i,j}^*(s)$ változókban $\rightarrow$ meg tudjuk oldani, ezzel a $f_{i,j}^*(s)$ függvények megvannak.

Inverz Laplace-transzformáció

Minden (i, j) párra $(1 \leq j \leq j \leq K)$ felírva az előző egyenleteket, kapunk egy egyenletrendszert, ami lineáris az $f^*_{i,j}(s)$ változókban $\rightarrow$ meg tudjuk oldani, ezzel a $f^*_{i,j}(s)$ függvények megvannak.

Ezután a dinamikus egyenletekből kiszámoljuk a sorhossz stacionárius eloszlását: $v_{\mathrm{st}} = (x_0, x_1, \ldots x_K)$. Figyelem! Ha λ_i nem konstans, akkor egy véletlen beérkező igény nem a sorhossz stacionárius eloszlását látja! Ehelyett v_{st}-t újra kell súlyozni a λ_i érkezési rátákkal:

$$y_i = \frac{\lambda_i x_i}{\sum_{i=0}^{K-1} \lambda_i x_i} \qquad i = 0, \ldots, K-1.$$

Inverz Laplace-transzformáció

Minden (i, j) párra $(1 \leq j \leq j \leq K)$ felírva az előző egyenleteket, kapunk egy egyenletrendszert, ami lineáris az $f^*_{i,j}(s)$ változókban $\rightarrow$ meg tudjuk oldani, ezzel a $f^*_{i,j}(s)$ függvények megvannak.

Ezután a dinamikus egyenletekből kiszámoljuk a sorhossz stacionárius eloszlását: $v_{\mathrm{st}} = (x_0, x_1, \ldots x_K)$. Figyelem! Ha λ_i nem konstans, akkor egy véletlen beérkező igény nem a sorhossz stacionárius eloszlását látja! Ehelyett v_{st}-t újra kell súlyozni a λ_i érkezési rátákkal:

$$y_i = \frac{\lambda_i x_i}{\sum_{i=0}^{K-1} \lambda_i x_i} \qquad i = 0, \ldots, K-1.$$

Amikor egy igény megérkezik egy i hosszú sorba, akkor ő az utolsó lesz és a sor hossza $i+1$ lesz, így egy véletlen igény kiszolgálási idejének Laplace-transzformáltja

$$f^*(s) = \sum_{i=0}^{K-1} y_i f^*_{i+1,i+1}(s).$$

Inverz Laplace-transzformáció

Ha $f^*(s)$ ismert, akkor egyszerűbb esetekben $f(t)$ visszakapható analitikusan. A gyakorlatban azért többnyire numerikus inverz Laplace transzformációt végzünk. Általában $f^*(s)$ bizonyos (esetleg komplex) pontbeli értékeinek lineáris kombinációval közelítjük $f(t)$-t:

$$f(t) \approx \sum_{k=1}^{N} \frac{\eta_k}{t} f^*\left(\frac{\beta_k}{t}\right),$$

ahol a β_k alappontokra és η_k súlyokra többféle választás is lehetséges.

Inverz Laplace-transzformáció

Ha $f^*(s)$ ismert, akkor egyszerűbb esetekben $f(t)$ visszakapható analitikusan. A gyakorlatban azért többnyire numerikus inverz Laplace transzformációt végzünk. Általában $f^*(s)$ bizonyos (esetleg komplex) pontbeli értékeinek lineáris kombinációval közelítjük $f(t)$-t:

$$f(t) \approx \sum_{k=1}^{N} \frac{\eta_k}{t} f^*\left(\frac{\beta_k}{t}\right),$$

ahol a β_k alappontokra és η_k súlyokra többféle választás is lehetséges.

Bővebben ajánlom a

http://inverselaplace.org/

honlapot, ahol többek között python implementáció is található.